Strength of Attentional Modulation on Cortical Auditory Evoked Responses Correlates with Speech-in-Noise Performance in Bimodal Cochlear Implant Users

Jae-Hee Lee[1,2], Hwan Shim[3], Bruce Gantz[2] and Inyong Choi[1,2] (iD)

Abstract

Auditory selective attention is a crucial top-down cognitive mechanism for understanding speech in noise. Cochlear implant (CI) users display great variability in speech-in-noise performance that is not easily explained by peripheral auditory profile or demographic factors. Thus, it is imperative to understand if auditory cognitive processes such as selective attention explain such variability. The presented study directly addressed this question by quantifying attentional modulation of cortical auditory responses during an attention task and comparing its individual differences with speech-in-noise performance. In our attention experiment, participants with CI were given a pre-stimulus visual cue that directed their attention to either of two speech streams and were asked to select a deviant syllable in the target stream. The two speech streams consisted of the female voice saying "Up" five times every 800 ms and the male voice saying "Down" four times every 1 s. The onset of each syllable elicited distinct event-related potentials (ERPs). At each syllable onset, the difference in the amplitudes of ERPs between the two attentional conditions (attended - ignored) was computed. This ERP amplitude difference served as a proxy for attentional modulation strength. Our group-level analysis showed that the amplitude of ERPs was greater when the syllable was attended than ignored, exhibiting that attention modulated cortical auditory responses. Moreover, the strength of attentional modulation showed a significant correlation with speech-in-noise performance. These results suggest that the attentional modulation of cortical auditory responses may provide a neural marker for predicting CI users' success in clinical tests of speech-in-noise listening.

Keywords

auditory selective attention, attentional modulation, event-related potential, cortical auditory evoked responses, speech-in-noise, cochlear implants

1. Introduction

Because the world poses complex living environments with many sources of sound, it can be challenging to comprehend speech in noise. The term "cocktail party problem (Cherry, 1953)" can explain this difficulty. To resolve the cocktail party problem, a person needs to be able to separate the mixed sounds into auditory objects or streams (McDermott, 2009) and focus on the target object.

Hearing loss can hinder the process of sound segregation, and consequentially, have a negative impact on the performance of speech in noise (Mackersie et al., 2001; Oxenham, 2008; Paredes-Gallardo et al., 2018), even in

those who wear hearing aids that can amplify the sound signals (Nelson et al., 2003). Cochlear implants (CI) may

[1]Dept. Communication Sciences and Disorders, University of Iowa, Iowa City, IA, 52242, USA
[2]Dept. Otolaryngology – Head and Neck Surgery, University of Iowa Hospitals and Clinics, Iowa City, IA, 52242, USA
[3]Dept. Electrical and Computer Engineering Technology, Rochester Institute of Technology, Rochester, NY, 14623, USA

Corresponding Author:
Inyong Choi, 250 Hawkins Dr., Iowa City, IA 52242, USA.
Email: inyong-choi@uiowa.edu

even further exacerbate sound segregation due to the lack of critical acoustic cues such as pitch (Galvin et al., 2009; Gfeller et al., 2007).

Although hearing remediation through CIs has evolved in the recent decade to combine amplified acoustic hearing (Gantz & Turner, 2004; von Ilberg et al., 1999), such "bimodal" CI users still exhibit large variability in speech-in-noise performance [e.g., (Berger et al., 2021)]. Understanding the factors that contribute to this variability has clinical value. Such information can suggest potential targets for diagnostic tests and rehabilitative strategies.

The ability to segregate sound is correlated with speech-in-noise performance both in normal hearing (Holmes & Griffiths, 2019) and in bimodal CI users (Choi et al., 2022). However, it is not the only contributing factor to speech-in-noise performance. Selective attention (Posner & Driver, 1992; Treisman, 1969), which can function as a neural attentional filter (Bester et al., 2016; da Costa et al., 2013; Lakatos et al., 2013) or as a form of sensory gain control (Hillyard et al., 1998), must be considered as well because it occurs simultaneously with auditory scene analysis (Alain & Arnott, 2000; Shinn-Cunningham et al., 2020). Selective attention results in suppressed neural representations of undesired auditory objects and enhanced neural responses to the attended sound (Mesgarani & Chang, 2012; Shinn-Cunningham et al., 2020; Woldorff & Hillyard, 1991). Such attentional modulation of neural responses can be measured using event-related potentials (ERPs) [(Choi et al., 2013; Hillyard et al., 1973)]. The strength of attentional modulation has been shown to have a relationship with how an individual performs in speech-in-noise or similar tasks (Choi et al., 2014; Dai & Shinn-Cunningham, 2016; Kim, Schwalje, et al., 2021b). It has been also reported that speech-in-noise performance can be improved by perceptual training that enhances the attentional modulation of ERPs (Kim, Emory, et al., 2021a), implying that stronger auditory selective attention predicts better speech-in-noise performance.

A few studies have explored whether attention modulates ERPs in CI users (Bester et al., 2016; Nogueira & Dolhopiatenko, 2022; Paredes-Gallardo et al., 2018; Paul et al., 2020). Among those studies, Paul et al. (2020) showed that attentional modulation observed in bilateral CI users' neural responses to a mixture of two competing continuous speech streams correlated with their behavioral attention performance. Recently, Nogueira and Dolhopiatenko (2022) reported that the attentional modulation of neural responses to continuous speech directly presented to the CI processor was associated with speech-in-noise performance.

The present study was designed to show whether 1) CI users exhibit attentional modulation and 2) the strength of attentional modulation is correlated with clinical outcomes in the emerging population of bimodal CI users when they perform speech-in-noise tasks in the sound field using their usual combination of hearing devices. We used relatively shorter (i.e., 4 s) and repetitive speech streams that have been used in our previous study about perceptual training of selective attention (Kim, Emory, et al., 2021a), expecting that findings from this study can evolve to post-CI rehabilitation using the similar training paradigm in (Kim, Emory, et al., 2021a).

2. Method

2.1 Participant Information

Thirteen CI users, between 37 and 77 years of age (mean = 62.2 years, $SD = 13.5$ years; median = 70.0 years; 7 (53.8%) female), were recruited from the University of Iowa Cochlear Implant Research Center. They gave written consent according to a written protocol that followed the guidelines provided by the Institutional Review Board at the University of Iowa, and they were provided $50 U.S. dollars for their 2-h participation in the EEG session. Table 1 shows the list of subjects with their demographic information, hearing profiles, and the type of CI devices.

2.2 Auditory Selective Attention Experiment: Stimulus

Both selective attention and speech-in-noise experiments were implemented in custom-written Matlab scripts (R2016b, Mathworks) using the Psychtoolbox 3 toolbox (Brainard, 1997). Tests were conducted in an acoustically-treated, electrically-shielded booth with a single loudspeaker (model LOFT40, JBL) positioned at a 0° azimuth angle at a distance of 1.2 m. Sound levels were the same across subjects at 70 dB SPL.

The auditory stimuli contained two clearly separated speech streams: the "Up" and the "Down" streams. The "Up" stream consisted of a female voice repeating the word "up" every 800 ms five times while the "Down" stream had the word "down" spoken by a male voice every 1 s for four times. Thus, by design, the timing of the words was misaligned except for the very first utterance of each stream so that there is no physical overlap between the competing streams.

In each stream, one of the utterances had a 26% higher pitch than the others so that an oddball detection task can be given. The high pitch utterance's duration was maintained the same. The same-duration higher pitch deviant was made by down-sampling recorded audio waveform by 26% and time-stretching it at the same rate. The 26% pitch increase was equal to four semitones, which have been reported to be distinguishable even by CI users without any acoustic hearing (Gfeller et al., 2007). All the subjects easily distinguished deviant from standard utterances when a single-stream voice (either "Up"s or "Down"s) was given. See Figure 1 for the stimulus (single utterance) spectrograms.

Table 1. List of Subjects with Their Demographic & Audiometric Information.

Subject #	Gender	Age (Years)	Device experience (Montds)	Age of Implantation (Years)	Deafness Duration (Years)
I	F	77	3	77	20
2	F	52	12	53	7
3	F	64	38	61	3
4	F	71	49	67	unsure
5	M	54	61	49	13
6	M	50	25	48	3
7	M	72	3	72	4
8	F	70	3	69	15
9	F	42	6	42	40
10	M	37	12	36	6
11	M	73	6	73	14
12	F	74	12	73	24
13	M	73	12	72	17

Subject #	Clear	External Device 01	Internal Device 01	External Device 02	PTA low-freq better ear (dB HL)
I	Left	Naida CI Q90	Advanced Bionics High Resolution 3D Ultra Slim	HA	58
2	Right	Naida CI Q91	Advanced Bionics Slim J	HA	68
3	Left	Sonnet EAS	Med El Syncrhony Flex 24		50
4	Right	Sonnet EAS	Med El Syncrhony	HA	67
5	Left	Nuclues 6 CP910	Nucleus L24 Hybrid		78
6	Left	Nucleus 7 CP1000	Nucleus CI 532	HA	92
7	Left	Naida CI Q90	Advanced Bionics High Resolution 3D Ultra Slim	HA	62
8	Left	Naida CI Q90	Advanced Bionics High Resolution 3D Ultra Slim	HA	78
9	Right	Naida CI Q90	Advanced Bionics High Resolution 3D Ultra Slim	HA	83
10	Right	Nucleus 7	Nuclus Hybrid S12	HA	68
11	Left	Naida CI Q90	Advanced Bionics High Resolution 3D with High Focus Slim	HA	65
12	Right	Naida CI Q90	Advanced Bionics High Resolution 3D with High Focus Slim		93
13	Right	Naida CI Q90	Advanced Bionics High Resolution 3D with High Focus Slim	HA	102

Frequency (Hz) (Right Ear) Subject number	0125	0250	0500	1000	2000	4000	8000
I	45	45	50	55	70	80	90
2	60	60	70	90	115	115	90
3	30	25	40	50	60	60	70
4	70	85	85	100	115	115	90
5	25	30	50	75	110	115	95
6	80	90	95	95	85	85	90
7	30	35	55	60	70	80	105
8	40	50	65	75	95	105	90
9	90	100	110	115	115	115	100
10	45	45	45	65	95	105	90
11	30	30	35	50	110	115	90
12	50	50	65	100	115	115	90
13	N/A	N/A	N/A	N/A	N/A	N/A	N/A

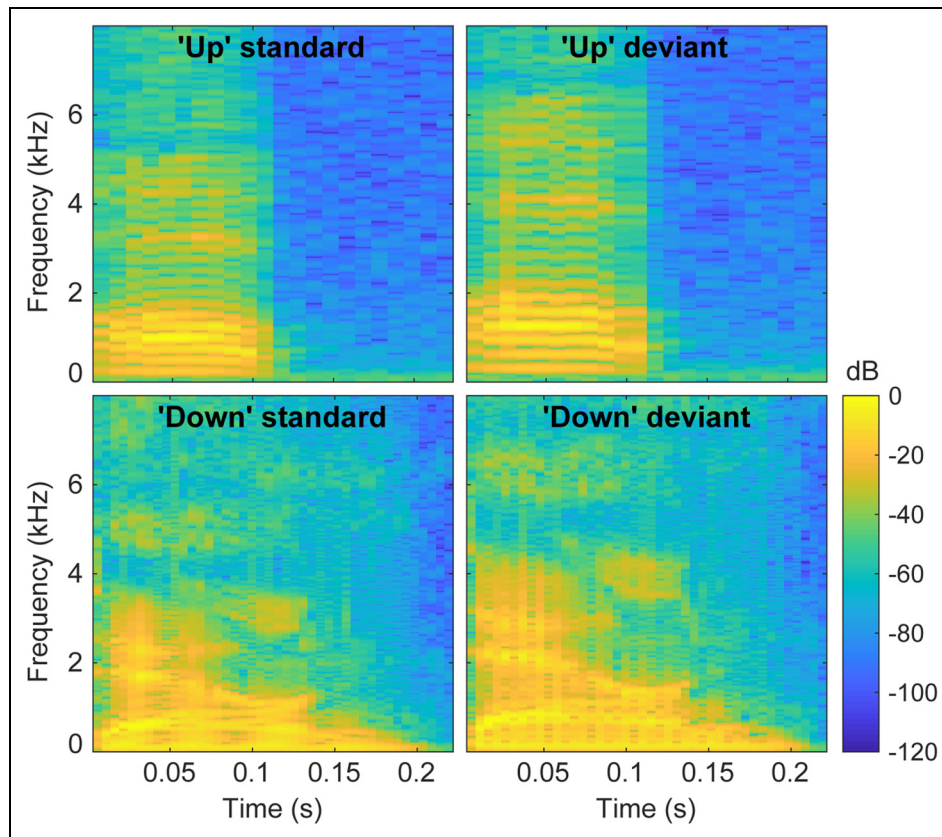| Frequency (Hz) (Left Ear) Subject Number	0125	0250	0500	1000	2000	4000	8000
I	55	65	65	80	80	100	95
2	50	55	60	70	75	65	75
3	20	40	85	100	95	85	85
4	30	30	30	65	105	115	90
5	70	70	85	115	115	115	95
6	95	100	110	115	115	115	90
7	60	75	85	85	85	115	105
8	50	55	80	85	110	115	90
9	45	60	70	90	90	90	75
10	35	35	50	65	95	105	90
11	50	65	75	90	110	115	90
12	35	45	75	115	115	115	90
13	50	55	85	105	115	115	90

Figure 1. Stimulus (single utterance) spectrograms.

The top part of Figure 2 showed example waveforms of the two competing streams of "Up"s and "Down"s.

2.3 Auditory Selective Attention Experiment: Task Design

At the beginning of each trial, the subjects fixated on the stationary sign (+) at the center of the blank screen of a computer for a second. Then, a written instruction (i.e., a visual cue) was provided on the screen to direct subjects to attend to either the "Up" or "Down" streams. While a deviant utterance (i.e., the utterance with a higher pitch) existed in each of the "Up" and "Down" streams, subjects were instructed to choose a deviant occurrence from the attended stream only.

The participants were asked to fixate on the screen again at the sign (+) before and while the auditory stimulus was provided. After the auditory stimulus, the participants could report their answer (the deviant occurrence number: 3, 4, or 5 for the attend "up" condition and 3 or 4 for the attend "down" condition) by pressing a number into the keypad. 60 attend-up and 60 attend-down trials were randomly intermixed in a single block. Subjects were given a break after every 30 trials.

2.4 Speech-in-Noise Test: California Consonant Test (CCT) Stimulus and Design

In every trial, the participants were shown a white cross sign on the black screen and were instructed to fix their gaze upon it so that eye movement artifacts were reduced. The auditory cue phrase "Check the word" (spoken by the same person whose voice was used to introduce the target words) was presented to the listeners for 800 ms so that the participants could be primed for the beginning of the babble. Then, there was a 700 ms silence, which was followed by the eight-talker babble that lasted for 2 s. Target words, which were 100 monosyllabic words chosen from the California Consonant Test (CCT) (Owens & Schubert, 1977), were presented 1 s after the onset of the babble at 70 dB SPL. The signal-to-noise ratio was fixed to +7 dB for all the trials. The sound levels were selected using the same process as described by Berger et al. (2021) and Kim et al. (2021b). Every participant was provided with 100 words in random order. After 100 ms following the offset of the target word stimulus, participants were given four options, which consisted of four words, displayed on the center of the screen. The participants then used the keypad to choose the word they thought they heard. They were not informed about the
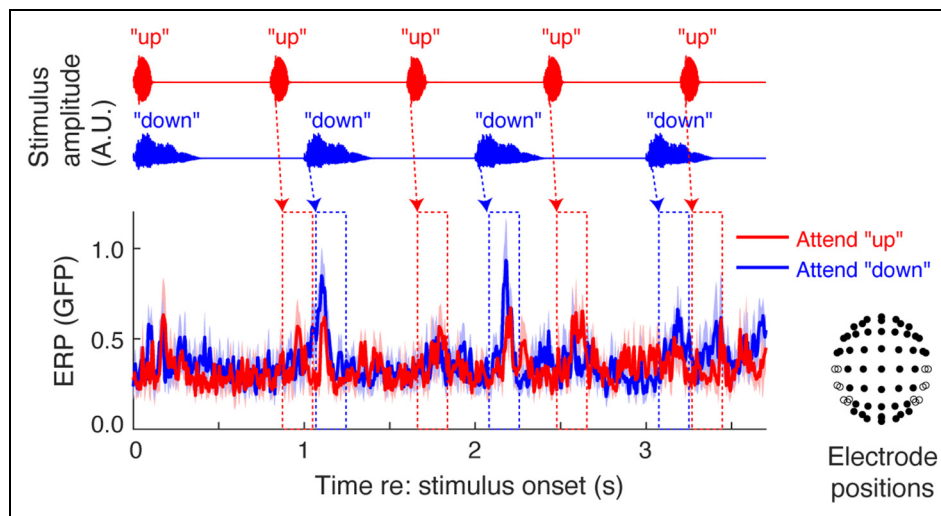
Figure 2. Grand average ERPs calculated in global field power (GFP), along with stimulus waveforms, is graphed against time in seconds. The red curve is from the "Attend-up" condition while the blue is from the "Attend-down." Shaded ranges around the ERPs represent standard error across subjects. Red and blue dashed boxes represent the expected timings of peaks of ERPs for each syllable with red corresponding to the utterance "up" and blue corresponding to the utterance "down" (i.e., 50 – 250ms after each utterance onset). For the attentional modulation index (AMI) calculation, the peak amplitude of ERP was found within those boxes. The positions of EEG electrodes that were used for the GFP computation are represented as filled circles at the rightmost corner (the top view).

correctness of their answer. The task did not move on to the next trial until 1 s after the participants gave their responses. Although the participants were not given time to be trained for CCT, verbally explained instructions were provided by the audiologists (Berger et al., 2021).

2.5 Speech-in-Noise Test: AzBio Stimulus and Design

The same participants were also tested with a sentence-based speech-in-noise test, AzBio (Spahr et al., 2012). The sentences were presented in 10-talker babble at 5 dB SNR, at 70 dB SPL presentation level. The total number of words repeated correctly was scored.

2.6 EEG Data Acquisition and Analysis

The EEG data were recorded during the selective attention task using a BioSemi ActiveTwo system with a 64-electrode cap with a standard 10–20 position at a 2048 Hz sampling rate. For each subject, 4–5 EEG electrodes were excluded (i.e., not inserted into the cap) because they were situated on top of the CI coil and magnet. For the EEG data analysis, first, the raw EEG data were band-pass filtered offline between 1 to 50 Hz with a 1024-point zero-phase non-causal FIR filter. Then we epoched data from 0.5 s before the sound onset to 0.5 s after the sound offset. Baseline correction was applied by subtracting the mean voltage between −200 and 0 ms before stimulus onset for each channel. Then epoched data were down-sampled to 256 Hz.

Next, electric field-evoked artifacts introduced by CIs and eye-blink artifacts were removed using independent component analysis based on the principles suggested by (Gilley et al.,

2006) implemented in the Matlab EEGLab toolbox (Delorme & Makeig, 2004). The artifact-related components were determined by visual inspection that observed temporal and spatial patterns. Out of ∼60 independent components in each subject, the number of removed independent components (because contaminated by CI and eye-blink artifacts) ranged from 3 to 19. The mean number of removed components was 8.8.

Then epochs with maximum absolute amplitude exceeding 100 µV were discarded from the following analyses. After this preprocessing, the number of remaining artifact-free epochs ranged from 51 to 59 in each condition in each subject. The mean numbers of epochs were 53.6 and 54.0 for Attend-Up and Attend-Down conditions, respectively. There was no significant difference in the number of epochs between conditions.

3. Results

3.1 Behavioral Results

Participants' performance in the selective attention test ranged from 91.67% to 100% correct (mean accuracy 96.67%, standard deviation 3.07%). This accuracy from the selective attention task was ceiling as expected and not used for any further analyses.

The same participants' performance at the CCT speech-in-noise test ranged from 30% to 80% (mean accuracy 56.47%, standard deviation 16.82%).

The performance at the AzBio test ranged from 3% to 87% (mean 52.15, standard deviation 30.51%).

There was a significant correlation between the two speech-in-noise tests (i.e., CCT and AzBio); a Pearson correlation coefficient $R = 0.74$, $p = 0.0039$.

3.2   EEG Results: Quantification of Attentional Modulation

For each subject, at least 50 epochs in each attentional condition (i.e., Attend-Up or Attend-Down) were averaged in the time domain at each EEG channel to yield evoked responses. Then, global field power (GFP) was computed for each condition at each time sample in each subject by taking the standard deviation of the evoked responses across 50 electrodes (out of 64) that were commonly used for all the 13 subjects. The filled circles in the rightmost panel of Figure 2 show the positions of the 50 electrodes that were commonly used for each subject's GFP computation. Open circles represent the positions of excluded electrodes as they were situated near CI coils and magnets.

The expected timings of the amplitude peaks of evoked potentials in GFP were hypothesized to occur between 50–250 ms after the onset of each utterance, which are represented by dashed boxes in Figure 2. The red dashed boxes correspond to the utterance "up" while the blue dashed boxes correspond to the utterance "down." As seen in Figure 2, the evoked peaks following "up" line up within the red dashed boxes show greater amplitude in the "Attend-Up" condition while the evoked peaks following "down" located within the blue dashed boxes show greater amplitude in the "Attend-Down," which implies graphically that attentional modulation exists at the group level.

Each participant's attentional modulation was quantified through the following procedures.

(i)   For each of the "Attend-Up" and "Attend-Down" conditions, the peak evoked amplitude at each utterance was measured by finding the maximum GFP value in the time window of 50–250 ms after the onset of each utterance (i.e., the dashed boxes in Figure 2).

(ii) The "attended" evoked amplitude was computed by averaging peak evoked amplitudes across "Up" windows (i.e., red dashed boxes in Figure 2) for the "Attend-Up" condition and "Down" windows (i.e., blue dashed boxes) for the "Attend-Down" condition.

(iii) The "ignored" evoked amplitude was computed by averaging peak evoked amplitudes across "Up" windows (i.e., red dashed boxes in Figure 2) for the "Attend-Down" condition and "Down" windows (i.e., blue dashed boxes) for the "Attend-Up" condition.

(iv) The attentional modulation index (AMI) was calculated by dividing the difference between the "attended" and the "ignored" evoked amplitudes with the sum of the "attended" and the "ignored" evoked amplitudes [i.e., (attended – ignored) / (attended + ignored)] (Dai & Shinn-Cunningham, 2016; O'Sullivan et al., 2019).

When computed this way, a positive AMI indicates that evoked response is stronger when the sound is attended, which implies a successful sensory gain control (Kim et al., 2021b). A one-sample t-test against the mean of zero proved that attentional modulation was positive at the group

level: $T = 3.26$, $p = 0.0069$. To note, this main AMI result is from the combination of both evoked time windows following the "Up" and "Down" streams (i.e., "All" in Figure 3).

3.3   EEG Results: Stimulus Effect on the Strength of Attentional Modulation

In this section, to observe the stimulus effect on attentional modulation, we computed AMI for the "Up" and "Down" streams separately (i.e., "Up" and "Down" columns in Figure 3, respectively).

When computed in the "Up" time windows only, AMIs did not show an "above-zero" distribution (two-sided one-sample t-test, $p = 0.118$). However, when computed in the "Down" time windows only, AMI distribution was above zero ($p < 0.01$ after Bonferroni correction), indicating significant attentional modulation of evoked responses at the group level (Figure 3).

3.4   EEG Results: Relationship Between the Strength of Attentional Modulation, Speech-in-Noise Performance, and Other Factors

Then, we observed the relationship between AMI and speech-in-noise performance. As Figure 4 shows, significant
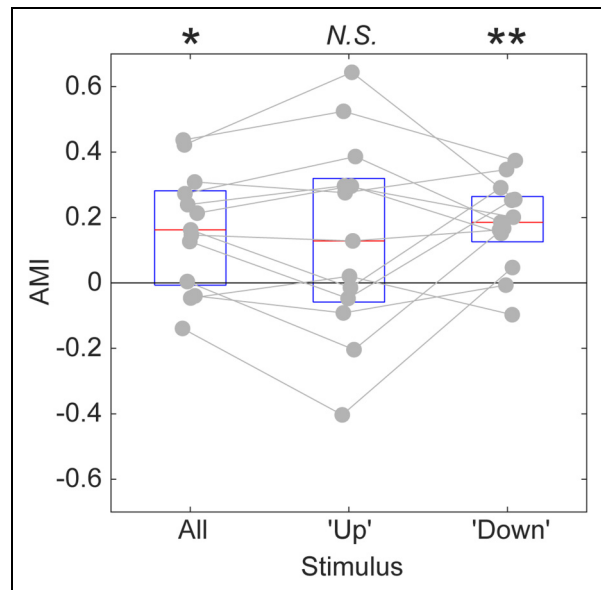

Figure 3. Comparisons of stimulus-specific attentional modulation indices (AMI). All: Computed over the time windows following both "up" and "down" utterances (i.e., both red and blue dashed boxes in Figure 2). "Up": AMI computed over the time windows following "up" utterances (i.e., red dashed boxes in Figure 2) only. "Down": AMI computed over the time windows following "down" utterances (i.e., blue dashed boxes in Figure 2) only. Gray dots represent the AMIs of individual subjects. The top and bottom edges of blue boxes represent the 75 and 25 percentile, respectively. Red bars depict the median. *: $p < 0.05$, **: $p < 0.01$, N.S.: Non-significant (from two-sided one-sample t-tests, after Bonferroni corrections of multiple comparisons).
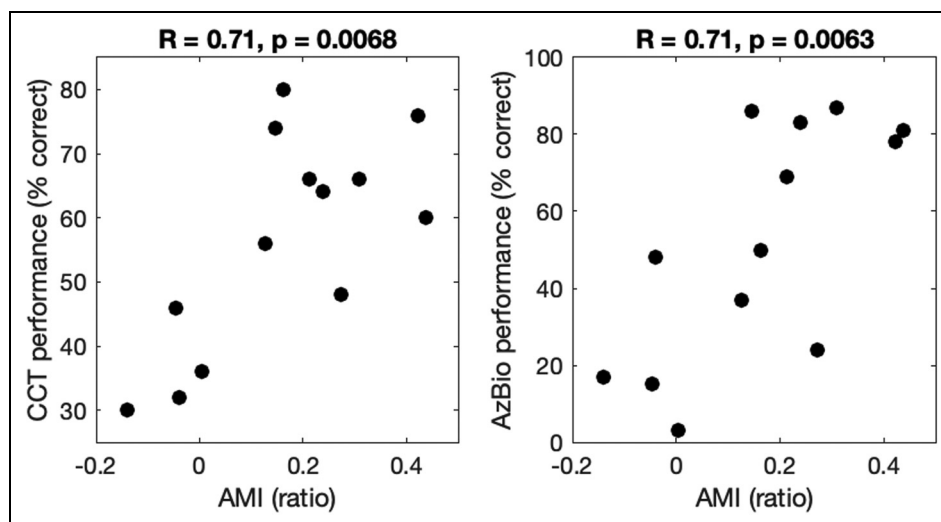
Figure 4. Correlation between AMI and accuracy in the speech-in-noise tasks.

correlations were found between AMI and participants' accuracy in the speech-in-noise tasks. Pearson correlation coefficients (R) were 0.71 (uncorrected $p = 0.0068$) against CCT and 0.71 (uncorrected $p = 0.0063$) against AzBio scores. These results indicate that AMI and CI users' performance in speech-in-noise tasks are associated.

None of the observed demographic and audiometric factors (i.e., age, device experience, and unaided hearing threshold) showed a significant correlation with any speech-in-noise performance or with the AMIs. See Figure 5 for the scatter plots and Pearson correlation coefficients. In Figure 5, "Speech-in-noise performance" represented mean percent-correct values averaged across CCT and AzBio scores. All the P values were uncorrected.

4. Discussion

The present study has shown that attentional modulation of cortical auditory evoked responses exists at a group level in CI listeners. This result is consistent with the study by Paredes-Gallardo et al. (2018). The current study also showed that an individual's performance at a speech-in-noise task was correlated with her/his strength of attentional modulation, also consistent with previous studies that used different stimulus presentation methods and EEG analyses (Nogueira & Dolhopiatenko, 2022; Paul et al., 2020).

Previous studies found that the variance of speech-in-noise performance in CI users is not well predicted by demographic factors (Berger et al., 2021). Rather, it has been found that auditory encoding fidelity (Friesen et al., 2001; Won et al., 2007) or auditory streaming (Choi et al., 2022; Hong & Turner, 2006) showed correlations with speech-in-noise performance in CI users. Although it is unclear whether our suggested measure of selective attention is redundant with the measures of auditory encoding fidelity

and streaming, our results still imply that cognitive factors such as selective attention can work as a neural marker of speech-in-noise performance. The results of this study suggest including an analysis of auditory cognitive functions in the assessment of CI benefits and outcomes. Potential strategies for compensating for hearing loss can be devised through rehabilitation of auditory cognitive functions in the future.

There are certain limitations to the study. First, since all the subjects had some degree of amplified acoustic hearing through contralateral hearing aids, it is unclear to what extent the finding of this study is related to CI-only listening. This is because we focused on the emerging population of bimodal CI users and their outcomes measured using their usual everyday hearing devices in the sound field. A future within-subject study may investigate the relative contributions of acoustic and electric hearing to the neural processing of selective attention and speech-in-noise performance.

Second, although a multiple linear regression analysis could be a good option for this study to investigate the relative contributions of various predictors, we did not conduct it because the sample size (i.e., 13 subjects) was too small. Instead, we showed a series of bivariate relationships of demographic and audiometric factors with AMI and speech-in-noise performance (see Figure 5). Future studies with larger sample sizes will provide a better interpretation of the preliminary result of the current study by conducting a multiple linear regression analysis.

Third, behavioral performance in the selective attention task exhibited a strong ceiling effect, which prevented observing correlations between the attentional modulation of neural responses and attention performance. We intentionally made the attention task easy so that we can obtain a similar number of correctly-answered trials across subjects for a consistent number of EEG epochs. A future study
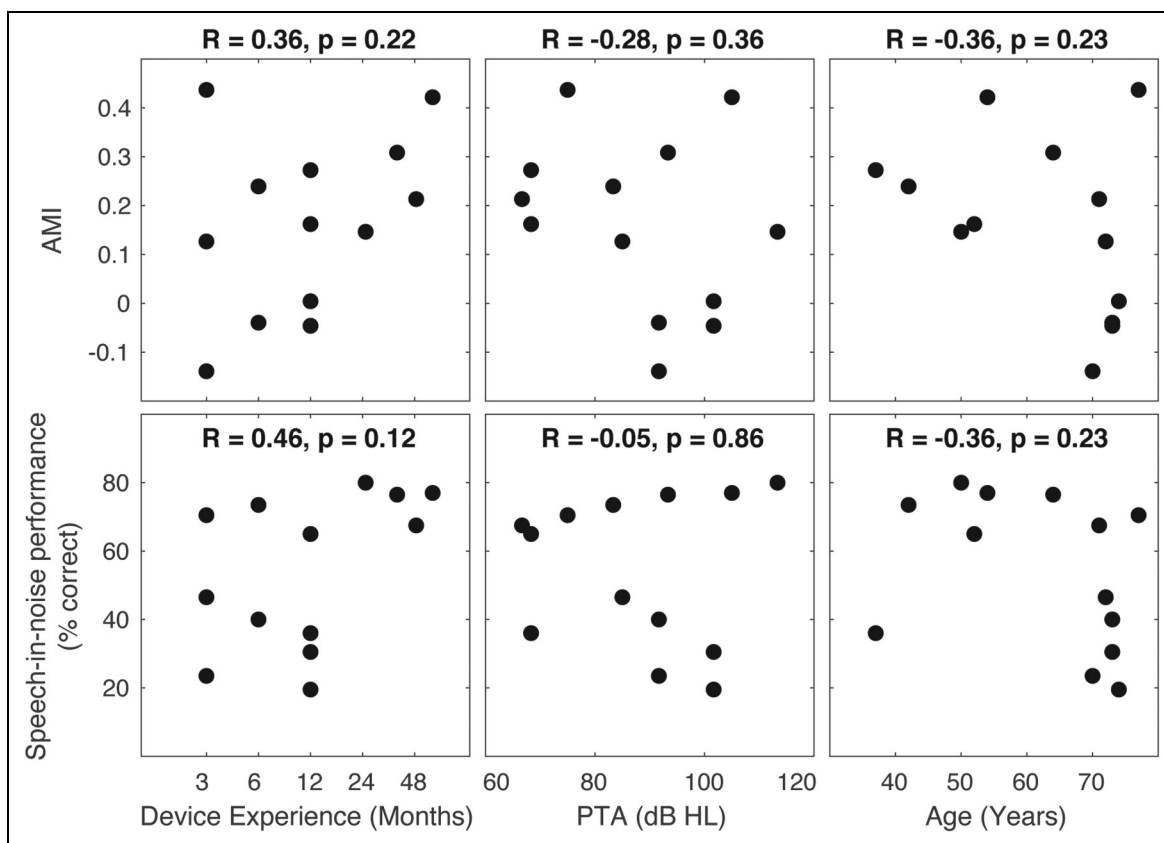
Figure 5. Relationship between device experience (i.e., post-implantation time in months), unaided pure-tone average (PTA) hearing thresholds (in dB HL), age (years) and AMI (the top row) and speech-in-noise performance (the bottom row). R: Pearson correlation coefficient. P-values were uncorrected.

may add a more difficult attentional condition to investigate the neural vs. behavioral relationship of attention.

Forth, by our stimulus design, there were two distinct features that differentiated competing streams: the voice identity (i.e., male vs. female) and the word content ("up" vs. "down"). Either or both features could be used to engage attention, meaning that we cannot disentangle relative contributions of different acoustic and/or linguistic features for engaging attention. This can be controlled in a future study.

Lastly, the deviations by pitch modulations in the speech streams were somewhat predictable (the deviations were always fourth or fifth syllables). It is possible that the participants may have been less engaged or may have been paying attention to both speech streams simultaneously. Thus, it must be acknowledged that the task may not have measured the participants' capability of employing selective attention to the full desired effect.

5. Conclusion

The current study has demonstrated that CI listeners exhibit attentional modulation of cortical evoked responses at the group level and that the strength of attentional modulation correlates with performance in SiN tasks.

ORCID iD

Inyong Choi ⓘD https://orcid.org/0000-0002-6663-9152

References

Alain, C., & Arnott, S. (2000). Selectively attending to auditory objects. *Frontiers in Bioscience*, 5(1), d202. http://www.bioscience.org/2000/v5/d/alain/fulltext.htm. https://doi.org/10.2741/Alain

Berger, J. I., Gander, P. E., Kim, S., Schwalje, A. T., Woo, J., Na, Y., Holmes, A., Hong, J., Dunn, C., Hansen, M., Gantz, B., McMurray, B., Griffiths, T. D., & Choi, I. (2021). Neural correlates of individual differences in speech-in-noise performance in a large cohort of cochlear implant users. *BioRxiv*, 2021.04.22.440998. https://doi.org/10.1101/2021.04.22.440998

Bester, C., Jayakody, D. M. P., Hammond, G., & Robertson, D. (2016). Selective attentional processes in cochlear implant recipients: Measurements of the attentional filter. *The Journal of the Acoustical Society of America*, 140(6), 4091–4100. https://doi.org/10.1121/1.4968783

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10(4), 433–436. https://doi.org/10.1163/156856897X00357

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America*, 25(5), 975–979. https://doi.org/10.1121/1.1907229

Choi, I., Gander, P. E., Berger, J. I., Hong, J., Colby, S., McMurray, B., & Griffiths, T. D. (2022). Auditory grouping ability predicts speech-in-noise performance in cochlear implants. *MedRxiv*, 2022.05.30.22275790. https://doi.org/10.1101/2022.05.30.22275790

Choi, I., Rajaram, S., Varghese, L. A., & Shinn-Cunningham, B. G. (2013). Quantifying attentional modulation of auditory-evoked cortical responses from single-trial electroencephalography. *Frontiers in Human Neuroscience*, 7(April), 1–19. https://doi.org/10.3389/fnhum.2013.00115

Choi, I., Wang, L., Bharadwaj, H., & Shinn-Cunningham, B. (2014). Individual differences in attentional modulation of cortical responses correlate with selective attention performance. *Hearing Research*, 314, 10–19. https://doi.org/10.1016/j.heares.2014.04.008

da Costa, S., van der Zwaag, W., Miller, L. M., Clarke, S., & Saenz, M. (2013). Tuning in to sound: Frequency-selective attentional filter in human primary auditory cortex. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 33(5), 1858–1863. https://doi.org/10.1523/JNEUROSCI.4405-12.2013

Dai, L., & Shinn-Cunningham, B. G. (2016). Contributions of sensory coding and attentional control to individual differences in performance in spatial auditory selective attention tasks. *Frontiers in Human Neuroscience*, 10, 530. https://doi.org/10.3389/FNHUM.2016.00530/BIBTEX

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1), 9–21. https://doi.org/10.1016/J.JNEUMETH.2003.10.009

Friesen, L. M., Shannon, R. V., Baskent, D., & Wang, X. (2001). Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. *The Journal of the Acoustical Society of America*, 110(2), 1150–1163. https://doi.org/10.1121/1.1381538

Galvin, J. J., Fu, Q. J., & Shannon, R. v. (2009). Melodic contour identification and music perception by cochlear implant users. *Annals of the New York Academy of Sciences*, 1169(1), 518–533. https://doi.org/10.1111/J.1749-6632.2009.04551.X

Gantz, B. J., & Turner, C. (2004). Combining acoustic and electrical speech processing: Iowa/Nucleus hybrid implant. *Acta Oto-Laryngologica*, 124(4), 344–347. https://doi.org/10.1080/00016480410016423

Gfeller, K., Turner, C., Oleson, J., Zhang, X., Gantz, B., Froman, R., & Olszewski, C. (2007). Accuracy of cochlear implant recipients on pitch perception, melody recognition, and speech reception in noise. *Ear and Hearing*, 28(3), 412–423. https://doi.org/10.1097/AUD.0B013E3180479318

Gilley, P. M., Sharma, A., Dorman, M., Finley, C. C., Panch, A. S., & Martin, K. (2006). Minimization of cochlear implant stimulus artifact in cortical auditory evoked potentials. *Clinical Neurophysiology*, 117(8), 1772–1782. https://doi.org/10.1016/J.CLINPH.2006.04.018

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science*, 182(4108), 177–180. https://doi.org/10.1126/science.182.4108.177

Hillyard, S. A., Vogel, E. K., & Luck, S. J. (1998). Sensory gain control (amplification) as a mechanism of selective attention: Electrophysiological and neuroimaging evidence. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 353(1373), 1257–1270. https://doi.org/10.1098/rstb.1998.0281

Holmes, E., & Griffiths, T. D. (2019). 'Normal' hearing thresholds and fundamental auditory grouping processes predict difficulties with speech-in-noise perception. *Scientific Reports*, 9(1), 1–11. https://doi.org/10.1038/s41598-019-53353-5

Hong, R. S., & Turner, C. W. (2006). Pure-tone auditory stream segregation and speech perception in noise in cochlear implant recipients. *The Journal of the Acoustical Society of America*, 120(1), 360–374. https://doi.org/10.1121/1.2204450

Kim, S., Emory, C., & Choi, I. (2021a). Neurofeedback training of auditory selective attention enhances speech-in-noise perception. *Frontiers in Human Neuroscience*, 15, 337. https://doi.org/10.3389/fnhum.2021.676992

Kim, S., Schwalje, A. T., Liu, A. S., Gander, P. E., McMurray, B., Griffiths, T. D., & Choi, I. (2021b). Pre- and post-target cortical processes predict speech-in-noise performance. *NeuroImage*, 228, 117699. https://doi.org/10.1016/j.neuroimage.2020.117699

Lakatos, P., Musacchia, G., O'Connel, M. N., Falchier, A. Y., Javitt, D. C., & Schroeder, C. E. (2013). The spectrotemporal filter mechanism of auditory selective attention. *Neuron*, 77(4), 750–761. https://doi.org/10.1016/j.neuron.2012.11.034

Mackersie, C. L., Prida, T. L., & Stiles, D. (2001). The role of sequential stream segregation and frequency selectivity in the perception of simultaneous sentences by listeners with sensorineural hearing loss. *Journal of Speech, Language, and Hearing Research*, 44(1), 19–28. https://doi.org/10.1044/1092-4388(2001/002)

McDermott, J. H. (2009). The cocktail party problem. *Current Biology*, 19(22), R1024–7. https://doi.org/10.1016/j.cub.2009.09.005

Mesgarani, N., & Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature*, 485(7397), 233–236. https://doi.org/10.1038/nature11020

Nelson, P. B., Jin, S.-H., Carney, A. E., & Nelson, D. A. (2003). Understanding speech in modulated interference: Cochlear

implant users and normal-hearing listeners. *The Journal of the Acoustical Society of America*, *113*(2), 961–968. https://doi.org/10.1121/1.1531983

Nogueira, W., & Dolhopiatenko, H. (2022). Predicting speech intelligibility from a selective attention decoding paradigm in cochlear implant users. *Journal of Neural Engineering*, *19*(2), 026037. https://doi.org/10.1088/1741-2552/AC599F

O'Sullivan, J., Herrero, J., Smith, E., Schevon, C., McKhann, G. M., Sheth, S. A., Mehta, A. D., & Mesgarani, N. (2019). Hierarchical encoding of attended auditory objects in multi-talker speech perception. *Neuron*, *104*(6), 1195–1209. https://doi.org/10.1016/j.neuron.2019.09.007

Owens, E., & Schubert, E. D. (1977). Development of the California consonant test. *Journal of Speech, Language, and Hearing Research*, *20*(5), 463–474. https://doi.org/10.3102/0002831 2040003655

Oxenham, A. J. (2008). Pitch perception and auditory stream segregation: Implications for hearing loss and cochlear implants. *Trends in Amplification*, *12*(4), 316–331. https://doi.org/10.1177/1084713808325881

Paredes-Gallardo, A., Innes-Brown, H., Madsen, S. M. K., Dau, T., & Marozeau, J. (2018). Auditory stream segregation and selective attention for cochlear implant listeners: Evidence from behavioral measures and event-related potentials. *Frontiers in Neuroscience*, *12*, 581. https://doi.org/10.3389/fnins.2018.00581

Paul, B. T., Uzelac, M., Chan, E., & Dimitrijevic, A. (2020). Poor early cortical differentiation of speech predicts perceptual difficulties of severely hearing-impaired listeners in multi-talker environments. *Scientific Reports*, *10*(1), 1–12. https://doi.org/10.1038/s41598-020-63103-7

Posner, M. I., & Driver, J. (1992). The neurobiology of selective attention. *Current Opinion in Neurobiology*, *2*(2), 165–169. https://doi.org/10.1016/0959-4388(92)90006-7

Shinn-Cunningham, B. G., Gazzaniga, M. S., Mangun, G. R., & Poeppel, D. (2020). Brain mechanisms of auditory scene analysis. In *The cognitive neurosciences VI* (pp. 159–166). MIT Press.

Spahr, A. J., Dorman, M. F., Litvak, L. M., van Wie, S., Gifford, R. H., Loizou, P. C., Loiselle, L. M., Oakes, T., & Cook, S. (2012). Development and validation of the AzBio sentence lists. *Ear and Hearing*, *33*(1), 112–117. https://doi.org/10.1097/AUD.0B013E31822C2549

Treisman, A. M. (1969). Strategies and models of selective attention. *Psychological Review*, *76*(3), 282–299. https://doi.org/10.1037/H0027242

von Ilberg, C., Kiefer, J., Tillein, J., Pfenningdorff, T., Hartmann, R., Stürzebecher, E., & Klinke, R. (1999). Electric-acoustic stimulation of the auditory system. New technology for severe hearing loss. *ORL; Journal for Oto-Rhino-Laryngology and Its Related Specialties*, *61*(6), 334–340. https://doi.org/10.1159/000027695

Woldorff, M. G., & Hillyard, S. A. (1991). Modulation of early auditory processing during selective listening to rapidly presented tones. *Electroencephalography and Clinical Neurophysiology*, *79*(3), 170–191. http://www.ncbi.nlm.nih.gov/pubmed/1714809. https://doi.org/10.1016/0013-4694(91)90136-R

Won, J. H., Drennan, W. R., & Rubinstein, J. T. (2007). Spectral-ripple resolution correlates with speech reception in noise in cochlear implant users. *Journal of the Association for Research in Otolaryngology*, *8*(3), 384–392. https://doi.org/10.1007/s10162-007-0085-8