Selective auditory attention modulates cortical responses to sound location change for speech in quiet and in babble

Erol J. Ozmeral[1] & Katherine N. Palandrani[2]

[1] *Department of Communication Sciences and Disorders, University of South Florida, Tampa, FL 33612, USA*

[2] *Department of Hearing and Speech Sciences, University of Maryland, College Park, MD 20742, USA*

Corresponding author: Erol Ozmeral (eozmeral@usf.edu)

Abstract

Listeners use the spatial location or change in spatial location of coherent acoustic cues to aid in auditory object formation. From stimulus-evoked onset responses in normal-hearing listeners using electroencephalography (EEG), we have previously shown measurable tuning to stimuli changing location in quiet, revealing a potential window into cortical object formation. These earlier studies used non-fluctuating, spectrally narrow stimuli, so it was still unknown whether previous observations would translate to speech stimuli and whether responses would be preserved for stimuli in the presence of background maskers. To examine the effects that selective auditory attention and interferers have on object formation, we measured cortical responses to speech changing location in the free field with and without background babble (+6 dB SNR) during both passive and active conditions. Active conditions required listeners to respond to the onset of the speech stream when it occurred at a new location, explicitly indicating yes or no to whether the stimulus occurred at a block-specific location either 30 degrees to the left or right of midline. In the aggregate, results show similar evoked responses to speech stimuli changing location in quiet compared to babble background. However, the effect of the two background environments diverges when considering the magnitude and direction of the location change, in which there was a clear influence of change vector in quiet but not in babble. Therefore, consistent with challenges associated with cocktail party listening, directed spatial attention can be shunted in the presence of stimulus noise and likely leads to poorer use of spatial cues in auditory streaming.

Introduction

Our ability to perceive auditory motion has been studied for over a century (1). Its relevance to virtual and augmented auditory spaces and hearing device technologies has led to a resurgence in interest in the last couple of decades. Psychophysically, listeners are able to detect a change in the spatial location of a stimulus with as little as 1°-resolution at the front azimuth but perform worse for stimuli that change location off-of-center (between 1.5° and 11.3° when the reference stimulus is at 60° and depending on the center frequency and bandwidth; i.e., minimum audible angle [MAA]) (2, 3). For stimuli that move at a fixed velocity, the smallest average arc to detect the movement (i.e., minimum audible movement angle [MAMA]) (3, 4) tends to be greater than the MAA with poorer performance associated with faster velocities. Neural imaging has also shown that responses to location changes are dependent on the reference location and extent of a shift in the free field (5-7) or lateral position in the case of binaural stimuli over headphones (8-11). Understanding the neural mechanisms associated with spatial hearing, auditory motion perception, and selective attention offers unique opportunities to address fundamental challenges that listeners with hearing loss, mainly older, face daily.

Natural acoustic environments often include dynamic sources in their relative location to the listener. In the classic "cocktail party" (12) and other substantiations (13, 14), most normal-hearing listeners are adept at following speech and other auditory objects to maximize communication goals and awareness of potential threats. Much like visual attention, auditory attention relies on the ability to perceptually group objects of which we can then decide to emphasize or ignore (15, 16). Many complex, sequential cognitive processes are required to achieve this object-based auditory attention. The brain must first be able to form distinct auditory objects based on common spectro-temporal properties of stimuli before segregating between foreground and background (17). Over time, this streaming process is subject to top-down modulatory effects of attention, which manifests as increased neural representation for sensory stimuli features (referred to as sensory gain control) (18, 19). To successfully follow speech signals amid competing background stimuli, therefore, it is thought that the brain can attend to relevant signal channels both by adding gain to encoding of relevant signals and "filtering out" the extraneous signals (20, 21).

In an earlier study, we investigated the effects of selective attention on an auditory event-related potential (AERP) (22) evoked by narrowband noise stimuli that changed locations in the front horizontal plane (7). Active attention to a distinct spatial location was predicted to yield sharper tuning to spatial changes at or near the target location. The results of this previous study were consistent with sensory gain control; however, the stimulus construct was far from what would be considered a "cocktail party," and therefore, it was unknown whether similar effects could be observed with more ecologically relevant stimuli. It may be that such neural responses would be quite different for speech stimuli, and the addition of energetic noise could potentially eliminate any evoked responses to spatial change. On the other hand, if similar responses are indeed observable with and without background interferers, there is the potential to use this objective measure as a conduit for measuring successful auditory object formation in complex acoustic scenes.

The approach of the present study combines behavioral and electrophysiological measures to investigate the dynamic modulation of sensory-evoked brain activity by spatial attention. If objects are successfully formed, attention to a perceptual feature like location is presumed to modulate the neural activity related to changes in that attended perceptual feature. A primary goal of the present study was to evaluate the effect of attention on neural responses to a speech stream that changes azimuth in an unpredictable manner. These experiments were designed to determine how response patterns differ between attended versus unattended locations, with an interest in the effects of responses at both attended and unattended locations. Due to the inherent challenges evident when acoustic scenes are complex, an additional goal was to explore stimulus conditions, including speech-in-quiet (Experiment I) and speech-in-babble (Experiment II), to better reflect real-world listening situations. The difference in effects between an arbitrary signal (like noise), versus a speech stream was investigated while evaluating the ways in which a favorable signal-to-noise ratio (SNR) affects neural responses to speech changing locations.

Materials and Methods

Participants

Participants included 18 adults (16 females) between the ages of 18 and 25 years of age (mean: 21.6) with audiometrically normal hearing (≤ 20 dB HL at octave frequencies between 250 and 8000 Hz). Data collection was completed over the course of 3-4 visits lasting approximately 2 hours each in duration. Exclusion from participating included any reported history of neurological dysfunction, middle ear disease, or ear surgery. The Montreal Cognitive Assessment (MoCA) was administered to all participants to screen for cognitive impairment, and all listeners had a passing score of at least 26 (23). All participants provided written consent for study participation prior to testing, and all procedures were approved by the University of South Florida Institutional Review Board. Participants were compensated for their time at an hourly rate.

Stimulus presentation

Target stimuli consisted of monosyllabic English words (2535 tokens) recorded from a male speaker with 100-ms inter-stimulus intervals at 76 dB SPL. In Experiment I, stimuli were presented in quiet. In Experiment II, stimuli were presented in multi-talker babble, consisting of eight turn-taking conversations spoken in eight foreign languages at 70 dB SPL overall (+6 dB SNR; for more details on target and background stimuli, see 24). Stimulus files (.wav) were loaded in MATLAB (MathWorks, Natick, MA) at a 44.1 kHz sampling rate and presented in the free field. Digital-to-analog conversion was performed by a 24ao (MOTU, Cambridge, MA) soundcard routed to three ne8250 amplifiers (Ashly Audio, Webster, NY) to 24 possible Q100 loudspeakers (KEF, Maidstone, England) in the azimuthal plane (Figure 1a). Target stimuli could be presented from only one location at a time, at either ±60°, ±30°, or 0° azimuth (checkered boxes in Figure 1a) and switched to a random location with replacement every 2 seconds. Catch trials were included at equal probability in which no change in location occurred. In Experiment II, background babble was presented simultaneously from locations at ±165°, ±105°, ±75°, and ±15° azimuth (black boxes in Figure 1a). Figure 1b shows a 10-second example stimulus presentation for Experiment II with time on the x-axis and stimulus waveforms centered at their respective speaker location angle on the y-axis.

[ENTER FIGURE 1 ABOUT HERE]

Procedure

Testing took place in a double-walled, sound-treated booth. Listeners sat in a height-adjustable chair facing the arc of loudspeakers approximately 1 m away from the listener at ear level (Figure 1a). Throughout testing, listeners were instructed to remain as still as possible and to maintain their head position such that their nose was pointed towards the center speaker (0°), with monitoring conducted by the experimenter through the sound booth window. To measure the effects of spatial attention, listeners were either instructed to actively attend to one of two possible speaker locations, or they were told to watch a silent video on a display in front of them during the audio presentation. In the active (Attend Left or Attend Right) conditions, the corresponding speakers were marked visually by either a blue ‘x’ (Attend Left) or a red ‘o’ (Attend Right). A touchscreen display was positioned near the participant’s right hand with a user interface (all participants were right-handed). The user interface was generated in MATLAB and consisted of two buttons reading “Yes” and “No” in a horizontal row on the southeast location of the display. During active attention conditions, the participants were told to respond every time that the target stimulus changed positions with either a tap of a “Yes” button when the target moved to the location that they were attending to (16% of trials) or “No” when the target moved to any other location (64% of trials). Only trials in which the signal changed locations required a response from the participant. Catch trials (20% of trials) – those in which the stimulus location had no change – were not meant to elicit a button press for target locations. Except for the attention component (including the button press), active and passive conditions were identical in stimulus presentation and environment. Blocks of listening, with self-paced breaks in between, were 150 2-second trials.

Electroencephalography (EEG)

Continuous EEG was recorded using a Waveguard™ (ANT Neuro BV, Enschede, Netherlands) elastic cap with 64 sintered AG/AgCl electrodes (International 10-20 electrode system). All electrode impedances were below 10 kΩ, with digitization at 512 Hz and 24-bit precision. Ground was located at the central forehead (AFz). All electrodes were initially referenced to the Cpz electrode. Recordings were made through asalab™ acquisition software (ANT Neuro, Enschede, Netherlands), and triggering was controlled using custom MATLAB scripts via the digital input/output stage of an RZ6 (System III, Tucker-Davis Technologies, Alachua, FL).

Recordings were processed offline using the software suite Brainstorm (25), running within MATLAB. Pre-processing of raw recordings consisted of band-pass filtering between 0.1 and 100 Hz (slope of 48 dB/octave), notch filtering at 60 Hz and harmonics, automatic detection of eye blinks based on the frontal electrodes (25), re-referencing to the average, and artifact removal based on spatial-source projections (SSP) (26, 27). The SSP approach is very similar to an independent component analysis that performs a spatial decomposition of the signals to identify the topographies of an idiosyncratic event, such as an eye blink. Because these events are reproducible and occur at the same location, this analysis can use their spatial topographies to remove their contribution from the recording while preserving contributions from other generators.

Following pre-processing stages, each continuous recording was epoched by trigger type, corresponding to shift pairs (i.e., the combination of where the stimulus occurred and where it was immediately before the location change). Trigger labels were applied at the time of measurement in which both the pre- and post-shift location were coded in the label. There were five possible target locations, so 25 possible pre- and post-shift combinations, including the catch trials. Epochs were further separated for attention conditions based on whether they were followed by either a “Yes” or “No” button press within 2 seconds of the stimulus location change. All epochs of a given pair combination (and button response) were then averaged following the removal of DC offset and linear drift. To capture the overall activity across the

64 scalp electrodes, the RMS activity, or global field power (GFP; 28), was computed for subsequent analyses.

Results

Cortical response to speech stream changing location in the free field

Experiment I: Quiet background. Initial analyses centered on the AERPs associated with the effect of attention on a change in location of the target speech stream. Only trials with accurate responses were considered for Attend Left and Attend Right conditions. Correct trials were determined by the button press indicating whether the target moved to an attended location or not (all trials were "correct" for the Passive condition). Accuracy was above 95% for all participants. Each panel in Figure 2 shows the average scalp response for the 2-second period following a change in the speech stream location in Experiment I for the three attention conditions (Passive, Attend Left, and Attend Right). Colored lines in Figure 2 panels represent the baseline-corrected global field power (GFP). Deflections near 125 ms, 250 ms, 550 ms, and around 1.4 s characterize the morphology in each panel at latencies. The two early deflections likely correspond to N1 and P2 activity in response to the change in location of the target stimulus (29, 30). In contrast, the third peak could reflect a P3 response indicating a higher-level of awareness and uncertainty in the environmental change (31-33). The only substantive difference between attention conditions was whether listeners responded to target location changes via button press. This likely explains the greater amplitudes at and sustained activity between the latter two latencies in the Attend Left and Attend Right versus Passive conditions. However, the presence of the P3 component in the Passive condition also indicates that this response was driven by stimulus factors and was separate from the behavioral response.

[ENTER FIGURE 2 ABOUT HERE]

To explore some of the spatial factors contributing to the scalp responses, the GFP was extracted for each condition as a function of where the target stream was presented (Figure 3). Colored lines in Figure 3 indicate the attention condition (Passive, black; Attend Left, blue; Attend Right, red). As in Figure 2, we see morphological similarities and differences among the Passive, Attend Left, and Attend Right conditions at each azimuth. The differences in the two late deflections and overall sustained activity between the Attend Left and Attend Right conditions are most notable. At locations in which the listener responded ‘no’ for all correct trials in both conditions (-60°, 0°, and 60°), the average responses are aligned, indicated by seldom significant differences in a paired-sample t-test (alpha = 0.5; black horizontal markers). However, for the two possible target locations (-30° and 30°), there is a clear increase in activity in which the listener responded ‘yes’ for correct trials (compare larger blue than red curve at -30° and larger red than blue curve at 30°), indicated by consistent significant differences starting around 500 ms. Black horizontal markers indicate the time points that yielded significant differences in amplitude between the Attend Left and Attend Right conditions in a paired-samples t-test (alpha = 0.05).

[ENTER FIGURE 3 ABOUT HERE]

Experiment II: Babble background. The addition of background babble in Experiment II posed the threat of reducing or eliminating observable cortical activity to the target stream. In Figure 4, the average responses for the 2-second period after a target location change are plotted. The morphologies of these responses to a target stream in +6 dB-SNR babble are like those in quiet (Experiment I; Figure 2), with comparable deflections at equivalent latencies. However, there were slight differences worth noting; specifically, the N1 deflection was less evident, and the responses overall were smaller compared to those in Experiment I. Figure 5 shows the activity when separated by spatial location (like Figure 3 for Experiment I). At all locations, there was a robust component around 550 ms and 1.4 s with overall greater potentials for active rather than passive attention conditions. For conditions in which listeners attended to a specific target location (i.e., ± 30°), responses were demonstrably larger in the corresponding attention condition, whereas at non-target locations, the two attention conditions did not elicit remarkably different responses from one another.

[ENTER FIGURE 4 ABOUT HERE]

[ENTER FIGURE 5 ABOUT HERE]

Late-latency P3 modulated by spatial attention

The late-latency P3 responses were also analyzed with respect to both the pre- and post-switch location of the speech stream to test the question of whether spatial attention would modulate activity dependent on the magnitude and direction of the change in speech location. In Figure 6, results per attention and background condition are reported as matrices of GFP activity at around 550 ms with rows indicating the pre-switch location and columns indicating the post-switch location. Matrix cells along the diagonal from top left to bottom right represent catch trials in which stimuli had no location change. Panels on the left are for Passive conditions and have a different range of values than the two attention conditions (middle: Attend Left; right: Attend Right), as indicated by the color bars. The top panels are for quiet conditions, and the bottom panels are for the babble conditions. Overall, as was seen in earlier analyses, there was less activity in the Passive condition than the two active attention conditions. There were generally larger responses to speech moving in quiet than in babble background. In both the quiet and babble backgrounds, for Attend Left and Attend Right, there was a tendency for larger responses in the column associated with the target location, -30° and +30° respectively.

[ENTER FIGURE 6 ABOUT HERE]

To measure the effect of magnitude and direction of the location change, trials with the same magnitudes and directions were grand averaged. Figure 7 shows the average GFP at the late-latency P3 near 550 ms. The resulting mean spatial-change tuning curve is v-shaped with a minimum response for the average of the no-change catch trials and maximum responses typically near ±120°. In separate two-way repeated-measures ANOVAs with nine levels of spatial change and three levels of attention, there was a main effect of attention condition in each background (quiet: $F[2,34] = 24.2$, $p < .001$; babble: $F[2,34] =$

9.1, p < .001) and angle (quiet: F[8,136] = 8.3, p < .001 ; babble: F[8,136] = 3.1, p = .003). Bonferroni-corrected *posthoc* tests confirm that the amplitudes in the attention conditions were greater that the Passive condition (all p < .01). In addition, Bonferroni-corrected *posthoc* analyses indicated there was a significant difference between Attend Left and Attend Right for two spatial-change vectors in the rightward direction in the quiet background (indicated by asterisks in Figure 7) but no differences between the two for speech in babble.

[ENTER FIGURE 7 ABOUT HERE]

Discussion

In a previous study, we explored the neural modulatory effects of spatial attention on a narrowband noise burst periodically changing between five locations in the front horizontal plane (7). We observed a clear effect of attending to one spatial hemifield versus the other, such that location-change responses were largest at the attended location, and all active attention conditions yielded stronger cortical responses than passive conditions overall in both younger and older normal-hearing adults. The present study similarly tested the effects of spatial attention on evoked scalp responses to a speech stream in younger, normal-hearing listeners because it was unknown whether more spectro-temporally complex stimuli with greater ecological relevance would elicit comparable location-change responses. Also, of interest was whether a babble background would disrupt modulatory effects of attention on the speech stream, which has the potential to explain mechanisms associated with cocktail party listening.

The morphology of the electroencephalogram for speech changing location in quiet and babble background included sensory potentials, N1 and P2, that were not modulated by attention. The N1 component was less prominent when background babble was present, which is consistent with previous research on the effects of signal-to-noise ratio on cortical responses to speech in noise (34). In all conditions, the N1 and P2 latencies were delayed relative to classic sensory potentials to the onset of stimulus energy (35, 36). Previous studies that have measured EEG responses to location changes have also shown delayed latencies relative to those elicited by stimulus onsets and have labeled these components as "change"-N1 or "change"-P2 that make up the motion-onset-response (MOR) (37, 38). Though the MOR is influenced by stimulus-dependent factors (39-41), it is also thought to reflect higher-level auditory areas and can change with task-relevant attentional processes (42). We did not observe significant differences between the attention and passive conditions in the present study at these earlier latencies. Still, there are key differences between this study and those that specifically explore auditory motion. The first major methodological difference is that the present stimulus construct included an instantaneous randomized location change rather than a successive change in location to each loudspeaker along a motion path. This "jump" in location may have precluded any attentional effects on auditory motion. A second key difference is that our study required listeners to actively respond to changes in the stimulus location only in the attention conditions and therefore lacked a motor control in the Passive conditions. Previous studies included a non-spatial or irrelevant task in the baseline condition (e.g., 42), and therefore, any differences at N1 or P2 could be attributed to differences in attentional processes. Thus, the present study focused more on the varied responses between the Attend Left and Attend Right conditions, which differed by the location listeners were required to attend.

For the passive and both attention conditions, P3 amplitudes were small for catch trials, in which no location change occurred 20% of the time, and they were large for location changes. This was consistent with our earlier study using narrowband noise stimuli (7). The fact that a P3 response was elicited in the Passive condition also suggests that despite instruction to ignore the speech during these blocks, the discrete changes in location of the speech stream were salient, and the P3 reflects an automatic process (43, 44). Early studies of the P3 often describe it as being elicited by surprising or unexpected stimuli or stimulus changes (e.g., 45), and it has been shown to be larger for task-relevant stimuli requiring a response rather than unattended or irrelevant stimuli (46, 47). Here, the late-latency P3 response is believed to be associated with higher-order processing related to its uncertain change in location (80% of the trials had equal probability to change to one of four possible locations).

Directing spatial attention in both studies led to even greater P3 responses when the stimuli changed to the attended location (see Figures 3 and 5 in this study, and see Figure 1 in 7). At -30° to the left, the largest P3 response was observed for the Attend Left condition, and at +30° to the right, the largest peak was observed for the Attend Right condition. At non-target locations (-60°, 0°, and +60°), no meaningful differences in amplitudes were observed between the two attention conditions. Variations on this paradigm have previously been used in the visual domain and have found similar results (48-50). In the present study, cortical responses also appeared higher for active attention conditions in quiet than in babble backgrounds. This is consistent with the view that sensory gain control is dependent on signal-to-noise-ratio (51), in which adding gain to the system is predicated on the saliency of the target stimulus. Together, the results demonstrate that overall neural activity is modulated by the occurrence of a speech stream at an attended location in both quiet and babble backgrounds, but stimulus noise can interfere with the salience of location changes in target speech.

Conclusions

This EEG study aimed to measure the effect of active attention to a spatial location of speech in the free field, either in quiet or babble background. Earlier work demonstrated that for narrowband noise stimuli, younger and even older listeners show significant modulatory effects of attention depending on the magnitude and direction of a spatial change (7); however, it was unknown if attention to more real-world stimuli, like speech, would show comparable modulation to evoked responses. Results demonstrated a late-latency P3 indicator of engaged attention to changes in the environment that was like the neural responses established for noise bursts. Active attention to specific target locations modulated the overall responses, and background babble at +6 dB SNR diminished responses somewhat. Still, the babble background did not eliminate the overall effect in these younger, normal-hearing listeners.

Young normal-hearing listeners are not known to have difficulty localizing speech with competing background signals at positive SNRs, but there are known challenges in individuals with hearing loss (52). By understanding the consequences of attention on auditory evoked neural measures, it is possible that objective tasks can be designed that directly assess hearing-impaired listeners' perceptual limitations or their success with potential interventions. It is unclear, for example, whether a better SNR resulting from directional processing in hearing aids could help mediate spatial hearing challenges at the neural level. Future work will focus on the consequences of aging and hearing impairment on object-based auditory attention while evaluating the efficacy of spatial hearing enhancement (i.e., directional microphones) in hearing aids.

Acknowledgments

This work was supported by an internal grant from the College of Behavioral and Community Sciences at the University of South Florida and by the NIH NIDCD to E.J.O. (R21 DC017832). The authors would like to thank Abigail Biberdorf and the staff of the USF Auditory & Speech Sciences Laboratory for assistance with subject recruitment and data collection. A portion of these data were presented at the 177th Meeting of the Acoustical Society of America in Louisville, KY, and were published as proceedings of that meeting (53).

Author Contributions

E.J.O. developed the study concept and design. K.N.P. performed testing and data collection. Both authors contributed to data processing, analysis, and interpretation. E.J.O. drafted the manuscript, and K.N.P. provided critical revisions.

References

1. Carlile S, Leung J. The Perception of Auditory Motion. Trends Hear. 2016;20.

2. Mills AW. On the Minimum Audible Angle. J Acoust Soc Am. 1958;30(4):237-46.

3. Chandler DW, Grantham DW. Minimum Audible Movement Angle in the Horizontal Plane as a Function of Stimulus Frequency and Bandwidth, Source Azimuth, and Velocity. J Acoust Soc Am. 1992;91(3):1624-36.

4. Harris JD, Sergeant RL. Monaural-binaural minimum audible angles for a moving sound source. J Speech Hear Res. 1971;14(3):618-29.

5. Briley PM, Kitterick PT, Summerfield AQ. Evidence for Opponent Process Analysis of Sound Source Location in Humans. J Assoc Res Oto. 2013;14(1):83-101.

6. Eddins AC, Ozmeral EJ, Eddins DA. How aging impacts the encoding of binaural cues and the perception of auditory space. Hear Res. 2018;369:79-89.

7. Ozmeral EJ, Eddins DA, Eddins AC. Selective auditory attention modulates cortical responses to sound location change in younger and older adults. J Neurophys. 2021;126(1):803-15.

8. Magezi DA, Krumbholz K. Evidence for Opponent-Channel Coding of Interaural Time Differences in Human Auditory Cortex. J of Neurophys. 2010;104(4):1997-2007.

9. Salminen NH, Tiitinen H, Yrttiaho S, May PJC. The neural code for interaural time difference in human auditory cortex. J Acoust Soc Am. 2010;127(2):EL60-EL5.

10. Ozmeral EJ, Eddins DA, Eddins AC. Reduced temporal processing in older, normal-hearing listeners evident from electrophysiological responses to shifts in interaural time difference. J of Neurophys. 2016;116(6):2720-9.

11. Altman JA, Vaitulevich SF. Auditory Image Movement in Evoked-Potentials. Electroen Clin Neuro. 1990;75(4):323-33.

12. Cherry EC. Some Experiments on the Recognition of Speech, with One and with 2 Ears. J Acoust Soc Am. 1953;25(5):975-9.

13. Darwin CJ. Listening to speech in the presence of other sounds. Phil Trans Royal Soc B-Bio Sci. 2008;363(1493):1011-21.

14. Kidd G, Colburn HS. Informational Masking in Speech Recognition. Springer Handbook Audit. 2017;60:75-109.

15. Shinn-Cunningham BG. Object-based auditory and visual attention. Trends Cog Sci. 2008;12(5):182-6.

16. Bronkhorst AW. The cocktail-party problem revisited: early processing and selection of multi-talker speech. Atten Perc & Psychophysics. 2015;77(5):1465-87.

17. Shamma SA, Elhilali M, Micheyl C. Temporal coherence and attention in auditory scene analysis. Trends Neurosci. 2011;34(3):114-23.

18. Hillyard SA, Vogel EK, Luck SJ. Sensory gain control (amplification) as a mechanism of selective attention: electrophysiological and neuroimaging evidence. Phil Trans Royal Soc B-Biol Sci. 1998;353(1373):1257-70.

19. Bharadwaj HM, Lee AKC, Shinn-Cunningham BG. Measuring auditory selective attention using frequency tagging. Front Integr Neurosc. 2014;8.

20. Shinn-Cunningham B, Best V. Auditory selective attention. The handbook of attention. Cambridge, MA, US: Boston Review; 2015. p. 99-117.

21. Kerlin JR, Shahin AJ, Miller LM. Attentional Gain Control of Ongoing Cortical Speech Representations in a "Cocktail Party". J Neurosci. 2010;30(2):620-8.

22. Luck SJ. An introduction to the event-related potential technique. London, England: The MIT Press; 2014.

23. Nasreddine ZS, Phillips NA, Bedirian V, Charbonneau S, Whitehead V, Collin I, et al. The montreal cognitive assessment, MoCA: A brief screening tool for mild cognitive impairment. J Am Ger Soc. 2005;53(4):695-9.

24. Ozmeral EJ, Hoover EC, Gabbidon P, Eddins DA. Development of the Continuous Number Identification Test (CNIT): feasibility of dynamic assessment of speech intelligibility. Int J Audiol. 2020;59(6):434-42.

25. Tadel F, Baillet S, Mosher JC, Pantazis D, Leahy RM. Brainstorm: a user-friendly application for MEG/EEG analysis. Comput Intell Neurosci. 2011;2011:879716.

26. Gramfort A, Luessi M, Larson E, Engemann DA, Strohmeier D, Brodbeck C, et al. MNE software for processing MEG and EEG data. Neuroimage. 2014;86:446-60.

27. Uusitalo MA, Ilmoniemi RJ. Signal-space projection method for separating MEG or EEG into components. Med Biol Eng Comput. 1997;35(2):135-40.

28. Skrandies W. Global field power and topographic similarity. Brain Topo. 1990;3(1):137-41.

29. Picton TW. The P300 Wave of the Human Event-Related Potential. J Clin Neurophysiol. 1992;9(4):456-79.

30. Zhang FW, McGuire K, Firestone G, Dalrymple K, Greinwald J, Fu QJ. Cortical processing of location and frequency changes of sounds in normal hearing listeners. Hear Res. 2021;400.

31. Mccarthy G, Donchin E. Effects of Temporal and Event Uncertainty in Determining Waveforms of Auditory Event Related Potential (Erp). Psychophys. 1976;13(6):581-90.

32. Michalewski HJ, Prasher DK, Starr A. Latency Variability and Temporal Interrelationships of the Auditory Event-Related Potentials (N1, P2, N2, and P3) in Normal Subjects. Electroen Clin Neuro. 1986;65(1):59-71.

33. Niemczak CE, Vander Werff KR. Informational Masking Effects of Similarity and Uncertainty on Early and Late Stages of Auditory Cortical Processing. Ear Hear. 2021;42(4):1006-23.

34. Billings CJ, McMillan GP, Penman TM, Gille SM. Predicting Perception in Noise Using Cortical Auditory Evoked Potentials. J Assoc Res Oto. 2013;14(6):891-903.

35. Naatanen R, Picton T. The N1 Wave of the Human Electric and Magnetic Response to Sound - a Review and an Analysis of the Component Structure. Psychophys. 1987;24(4):375-425.

36. Lutkenhoner B, Steinstrater O. High-precision neuromagnetic study of the functional organization of the human auditory cortex. Audiol Neuro-Otol. 1998;3(2-3):191-213.

37. Getzmann S. Auditory motion perception: onset position and motion direction are encoded in discrete processing stages. Eur J Neurosci. 2011;33(7):1339-50.

38. Krumbholz K, Hewson-Stoate N, Schonwiesner M. Cortical response to auditory motion suggests an asymmetry in the reliance on inter-hemispheric connections between the left and right auditory cortices. J Neurophys. 2007;97(2):1649-55.

39. Getzmann S, Lewald J. Shared Cortical Systems for Processing of Horizontal and Vertical Sound Motion. J Neurophys. 2010;103(4):1896-904.

40. Getzmann S, Lewald J. Effects of natural versus artificial spatial cues on electrophysiological correlates of auditory motion. Hear Res. 2010;259(1-2):44-54.

41. Getzmann S. Effect of auditory motion velocity on reaction time and cortical processes. Neuropsychologia. 2009;47(12):2625-33.

42. Kreitewolf J, Lewald J, Getzmann S. Effect of attention on cortical processing of sound motion: An EEG study. Neuroimage. 2011;54(3):2340-9.

43. Polich J. Updating p300: An integrative theory of P3a and P3b. Clin Neurophysiol. 2007;118(10):2128-48.

44. Polich J. Overview of P3a and P3b. In: Polich J, editor. Detection of Change: Event-Related Potential and fMRI Findings. Boston: Kluwer Academic Press; 2003. p. 83-98.

45. Donchin E. Society for Psychophysiological Research - Presidential-Address, 1980 - Surprise ... Surprise. Psychophys. 1981;18(5):493-513.

46. Duncan-Johnson CC, Donchin E. Quantifying Surprise - Variation of Event-Related Potentials with Subjective-Probability. Psychophys. 1977;14(5):456-67.

47. Lang S, Kotchoubey B. Brain responses to number sequences with and without active task requirement. Clin Neurophysiol. 2002;113(11):1734-41.

48. Mangun GR, Hillyard SA. Event-Related Brain Potentials and Perceptual Sensitivity during Visual-Spatial Selective Attention. Psychophys. 1988;25(4):467-.

49. Hillyard SA, Munte TF. Selective Attention to Color and Location - an Analysis with Event-Related Brain Potentials. Percept Psychophys. 1984;36(2):185-98.

50. Eimer M. Sensory Gating as a Mechanism for Visuospatial Orienting - Electrophysiological Evidence from Trial-by-Trial Cueing Experiments. Percept Psychophys. 1994;55(6):667-75.

51. Hawkins HL, Hillyard SA, Luck SJ, Downing CJ, Mouloua M, Woodward DP. Visual-Attention Modulates Signal Detectability. J Exp Psychol Human. 1990;16(4):802-11.

52. Best V, Carlile S, Kopco N, van Schaik A. Localization in speech mixtures by listeners with hearing loss. J Acoust Soc Am. 2011;129(5):E210-E5.

53. Palandrani KN, Ozmeral EJ. Auditory spatial tuning to moving speech in quiet or background babble measured in cortex. Proc Meet on Acoust. 2019;36(1):050001.

Figure Captions

Figure 1: Panel A shows a schematic of the laboratory spatial array. The target speech stream came from alternate locations labeled, -60, -30, 0, +30, and +60°. For the directed-attention conditions, listeners were instructed to either attend to the speaker at -30° (blue; Attend-30°) or attend to the speaker at +30° (red; Attend+30°). In Experiment I, the speech stream was presented in quiet. In Experiment II, the speech stream was presented in background speech babble at +6-dB SNR from 8 speaker locations (black boxes). Loudspeaker positions indicated by white boxes were not used in this study. Panel B shows a 10-second example presentation (Experiment II). Target (grey; 76 dB SPL) and babble (black; 70 dB SPL overall) waveforms are positioned at their respective speaker locations over time. The location of the target could change every 2 seconds. In Experiment I, only speech targets were presented.

Figure 2: Average potentials measured at the scalp for three attention conditions (Passive, Attend Left, and Attend Right) in quiet background. Dark grey lines represent 64 recording sites, and colored lines are the baseline-corrected global field power (GFP) across all electrodes.

Figure 3: Grand average global field power (GFP) for stimuli presented in quiet background for each attention condition (Passive [green]; Attend left [blue]; Attend Right [red]). The five panels represent trials in which the target arrived at each of the five horizontal speaker locations (from left to right: -60° to +60°). Grey shaded regions represent standard error of the mean, and black horizontal markers represent time latencies in which a significant difference was found between Attend Left and Attend Right conditions in a paired-sample t-test (alpha = 0.5; n = 18).

Figure 4: Average potentials measured at the scalp for three attention conditions (Passive, Attend Left, and Attend Right) in babble background. Dark grey lines represent 64 recording sites, and colored lines are the baseline-corrected global field power (GFP) across all electrodes.

Figure 5: Grand average global field power (GFP) for stimuli presented in babble background for each attention condition (Passive [green]; Attend left [blue]; Attend Right [red]). The five panels represent

trials in which the target arrived at each of the five horizontal speaker locations (from left to right: -60° to +60°). Grey shaded regions represent standard error of the mean, and black horizontal lines represent time latencies in which a significant difference was found between Attend Left and Attend Right conditions in a paired-samples t-test (alpha = 0.5; n = 18).

Figure 6: Results per attention and background condition are reported as matrices of GFP activity at around 550 ms with rows indicating the pre-switch location and columns indicating the post-switch location. Matrix cells along the diagonal from top left to bottom right represent catch trials in which stimuli had no location change. Panels on the left are for Passive conditions and have a different range of values compared to the two attention conditions (middle: Attend Left; right: Attend Right) as indicated by the color bars. The top panels are for quiet conditions and the bottom panels are for the babble conditions.

Figure 7: The average GFP activity around 550 ms for spatial-change vectors spanning ±120°.

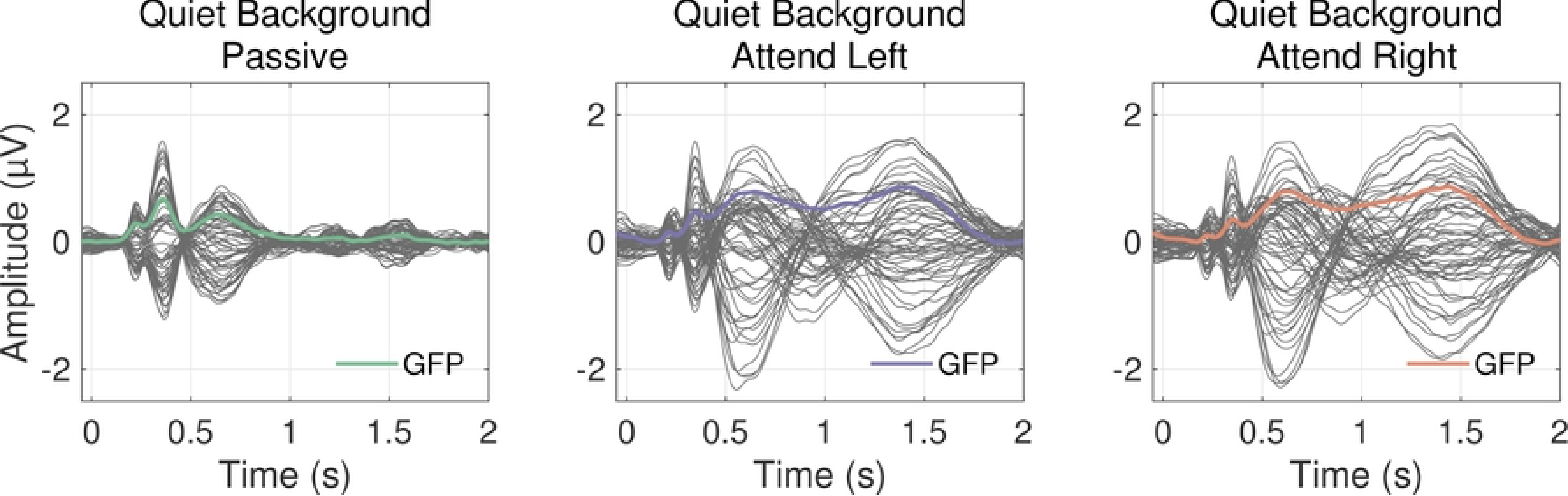
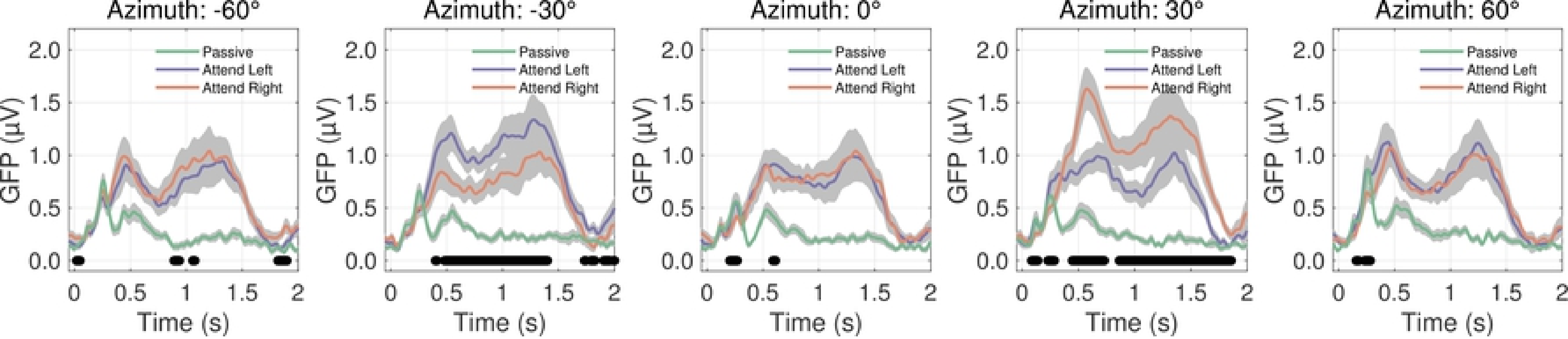

Quiet Background
Passive
Quiet Background
Attend Left
Quiet Background
Attend Right
Amplitude (μV)
2
0
-2
GFP
0
0.5
1
1.5
2
Time (s)

Figure 2



Azimuth: -60°
Azimuth: -30°
Azimuth: 0°
Azimuth: 30°
Azimuth: 60°
Passive
Attend Left
Attend Right
GFP (μV)
2.0
1.5
1.0
0.5
0.0
Time (s)
0
0.5
1
1.5
2

Figure 3

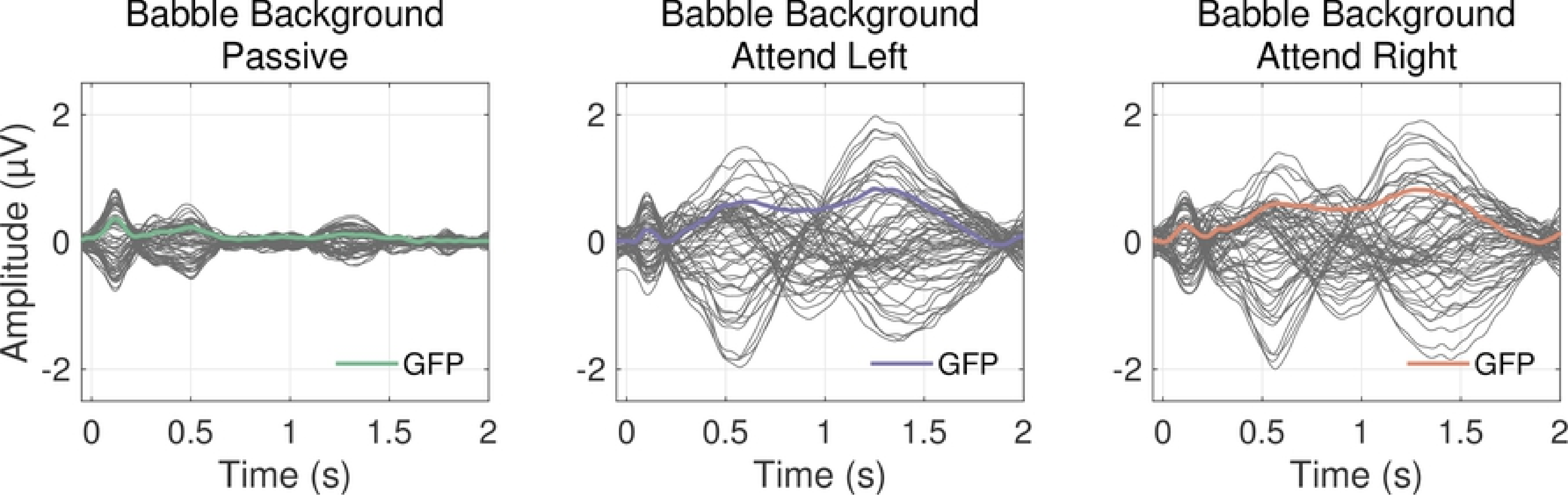
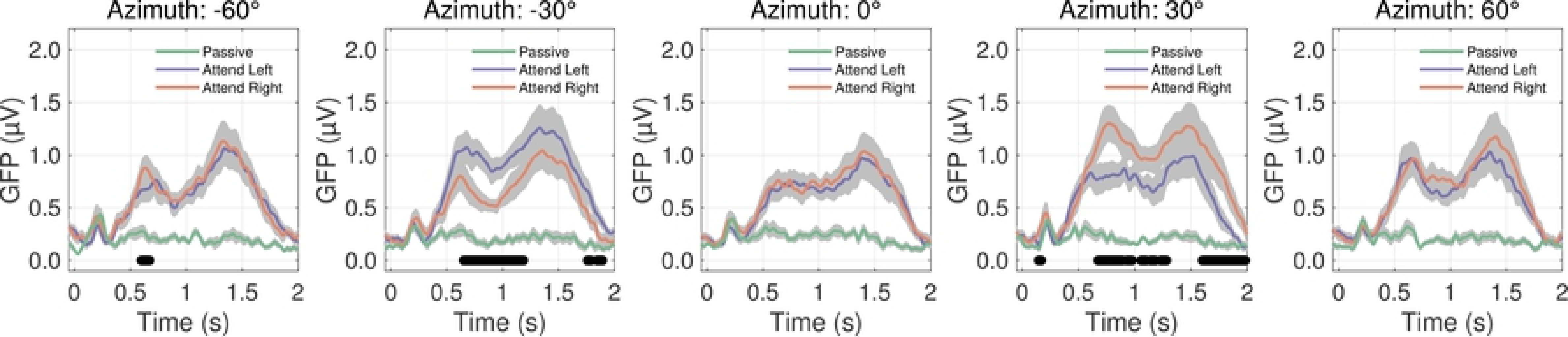

Babble Background
Passive
Babble Background
Attend Left
Babble Background
Attend Right
Amplitude (μV)
2
0
-2
GFP
0
0.5
1
1.5
2
Time (s)

Figure 4



Azimuth: -60°
Azimuth: -30°
Azimuth: 0°
Azimuth: 30°
Azimuth: 60°
Passive
Attend Left
Attend Right
GFP (µV)
2.0
1.5
1.0
0.5
0.0
0
0.5
1
1.5
2
Time (s)

Figure 5

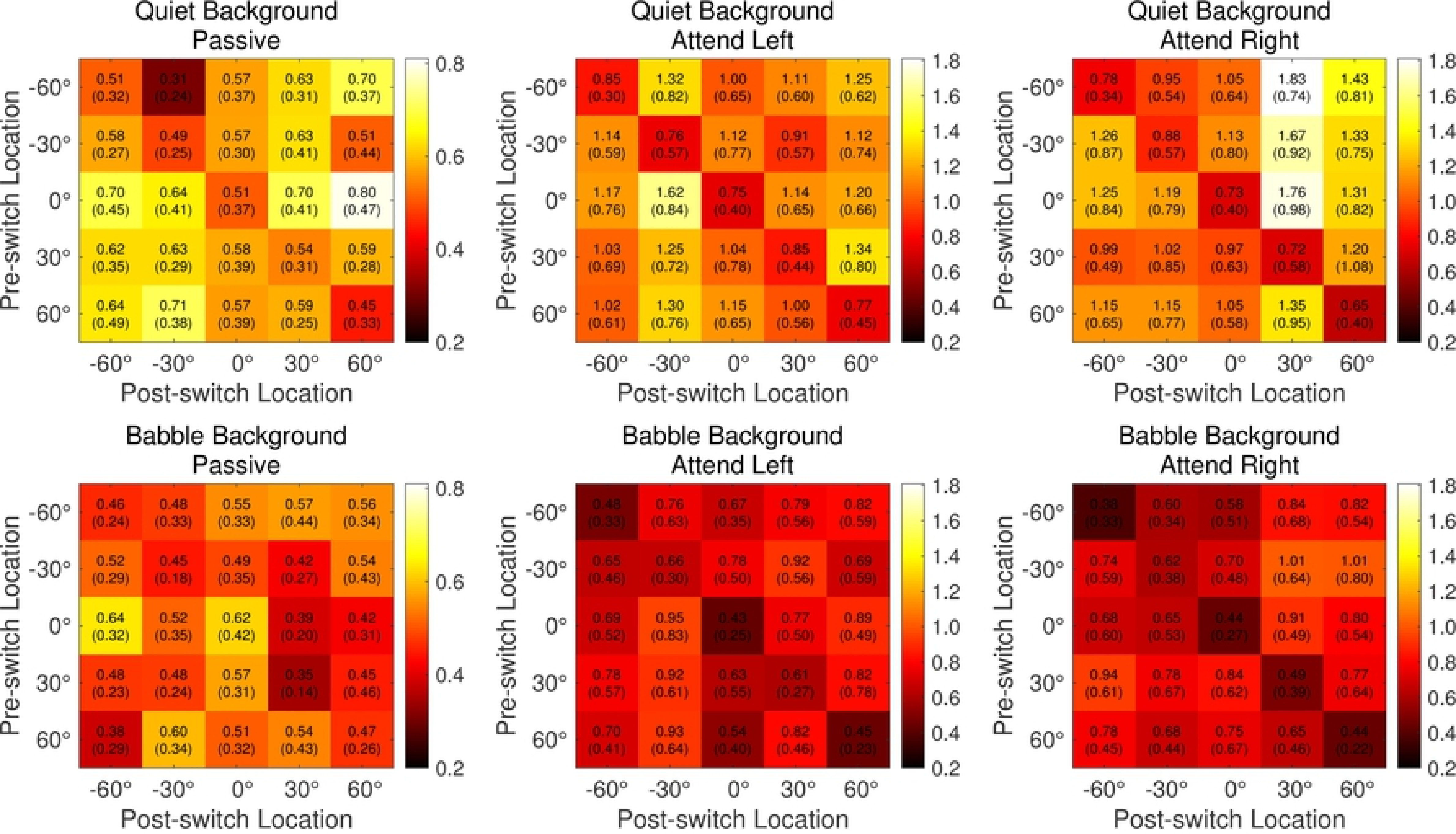

Quiet Background
Passive
Pre-switch Location
Post-switch Location
-60°
-30°
0°
30°
60°
0.51 (0.32) 0.31 (0.24) 0.57 (0.37) 0.63 (0.31) 0.70 (0.37)
0.58 (0.27) 0.49 (0.25) 0.57 (0.30) 0.63 (0.41) 0.51 (0.44)
0.70 (0.45) 0.64 (0.41) 0.51 (0.37) 0.70 (0.41) 0.80 (0.47)
0.62 (0.35) 0.63 (0.29) 0.58 (0.39) 0.54 (0.31) 0.59 (0.28)
0.64 (0.49) 0.71 (0.38) 0.57 (0.39) 0.59 (0.25) 0.45 (0.33)
0.8
0.6
0.4
0.2
Quiet Background
Attend Left
0.85 (0.30) 1.32 (0.82) 1.00 (0.65) 1.11 (0.60) 1.25 (0.62)
1.14 (0.59) 0.76 (0.57) 1.12 (0.77) 0.91 (0.57) 1.12 (0.74)
1.17 (0.76) 1.62 (0.84) 0.75 (0.40) 1.14 (0.65) 1.20 (0.66)
1.03 (0.69) 1.25 (0.72) 1.04 (0.78) 0.85 (0.44) 1.34 (0.80)
1.02 (0.61) 1.30 (0.76) 1.15 (0.65) 1.00 (0.56) 0.77 (0.45)
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
Quiet Background
Attend Right
0.78 (0.34) 0.95 (0.54) 1.05 (0.64) 1.83 (0.74) 1.43 (0.81)
1.26 (0.87) 0.88 (0.57) 1.13 (0.80) 1.67 (0.92) 1.33 (0.75)
1.25 (0.84) 1.19 (0.79) 0.73 (0.40) 1.76 (0.98) 1.31 (0.82)
0.99 (0.49) 1.02 (0.85) 0.97 (0.63) 0.72 (0.58) 1.20 (1.08)
1.15 (0.65) 1.15 (0.77) 1.05 (0.58) 1.35 (0.95) 0.65 (0.40)
Babble Background
Passive
0.46 (0.24) 0.48 (0.33) 0.55 (0.33) 0.57 (0.44) 0.56 (0.34)
0.52 (0.29) 0.45 (0.18) 0.49 (0.35) 0.42 (0.27) 0.54 (0.43)
0.64 (0.32) 0.52 (0.35) 0.62 (0.42) 0.39 (0.20) 0.42 (0.31)
0.48 (0.23) 0.48 (0.24) 0.57 (0.31) 0.35 (0.14) 0.45 (0.46)
0.38 (0.29) 0.60 (0.34) 0.51 (0.32) 0.54 (0.43) 0.47 (0.26)
Babble Background
Attend Left
0.48 (0.33) 0.76 (0.63) 0.67 (0.35) 0.79 (0.56) 0.82 (0.59)
0.65 (0.46) 0.66 (0.30) 0.78 (0.50) 0.92 (0.56) 0.69 (0.59)
0.69 (0.52) 0.95 (0.83) 0.43 (0.25) 0.77 (0.50) 0.89 (0.49)
0.78 (0.57) 0.92 (0.61) 0.63 (0.55) 0.61 (0.27) 0.82 (0.78)
0.70 (0.41) 0.93 (0.64) 0.54 (0.40) 0.82 (0.46) 0.45 (0.23)
Babble Background
Attend Right
0.38 (0.33) 0.60 (0.34) 0.58 (0.51) 0.84 (0.68) 0.82 (0.54)
0.74 (0.59) 0.62 (0.38) 0.70 (0.48) 1.01 (0.64) 1.01 (0.80)
0.68 (0.60) 0.65 (0.53) 0.44 (0.27) 0.91 (0.49) 0.80 (0.54)
0.94 (0.61) 0.78 (0.67) 0.84 (0.62) 0.49 (0.39) 0.77 (0.64)
0.78 (0.45) 0.68 (0.44) 0.75 (0.67) 0.65 (0.46) 0.44 (0.22)

Figure 6

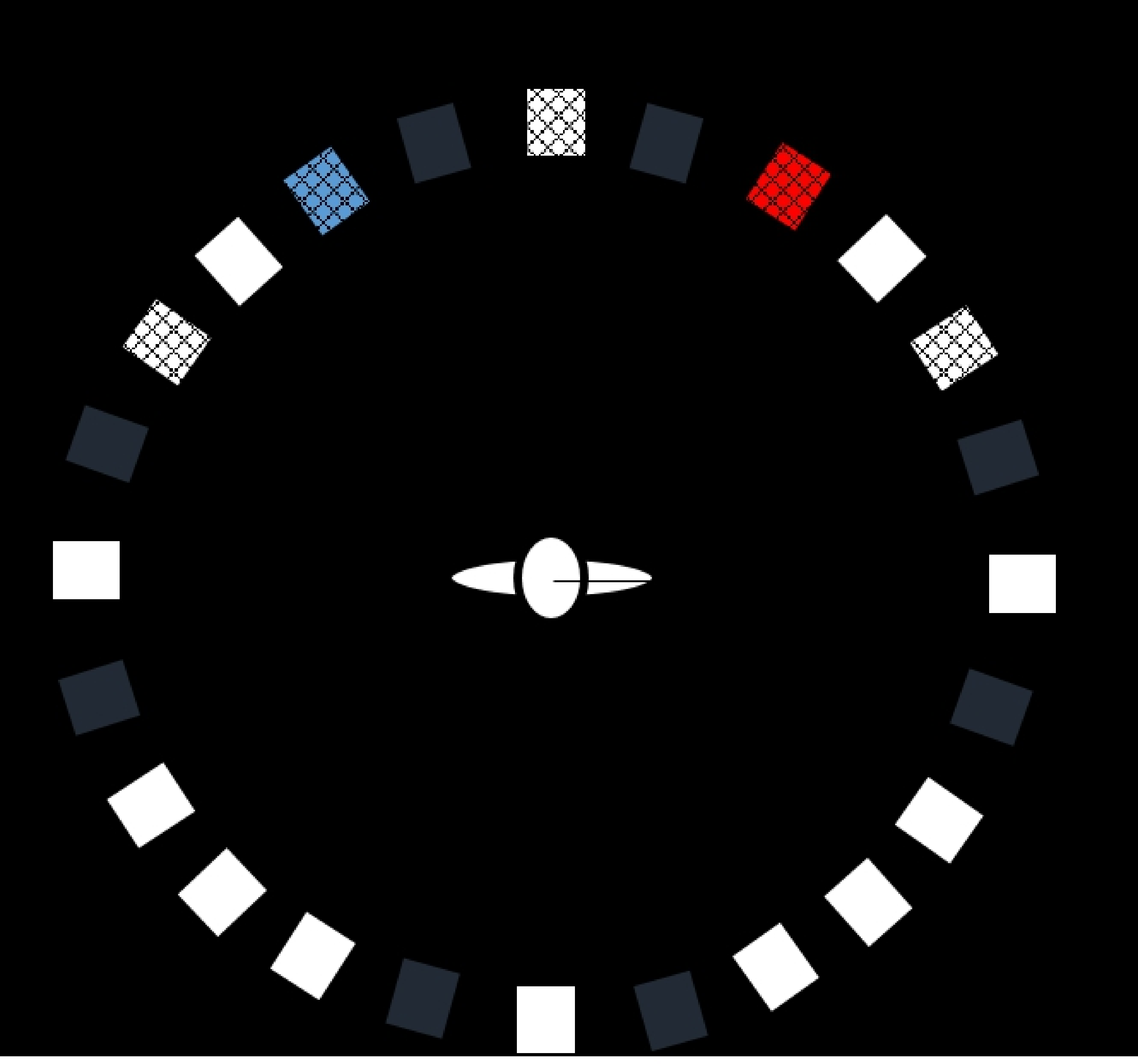


Figure 1a

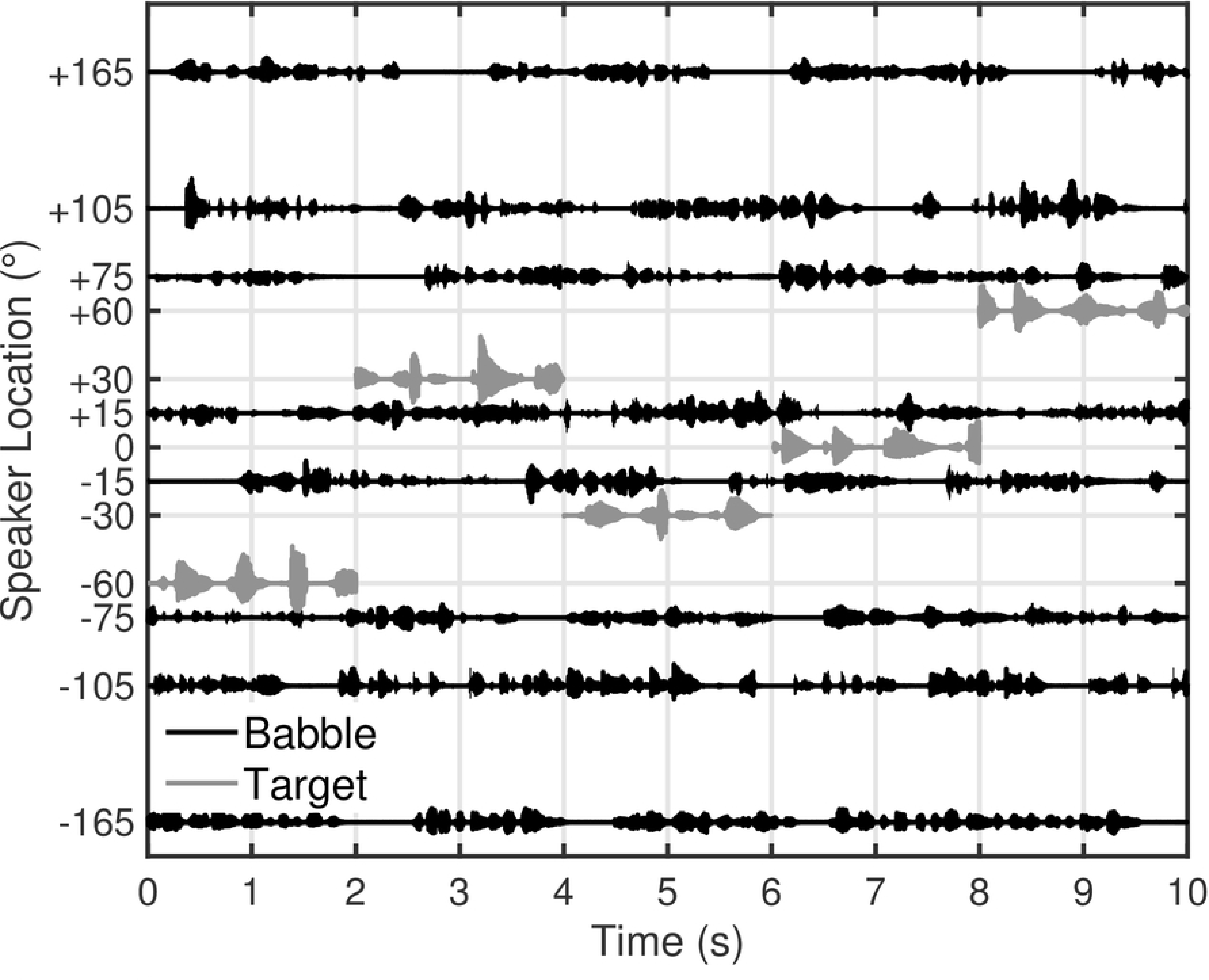

+165
+105
+75
+60
+30
+15
0
-15
-30
-60
-75
-105
-165
Speaker Location (°)
Babble
Target
0
1
2
3
4
5
6
7
8
9
10
Time (s)

Figure 1b

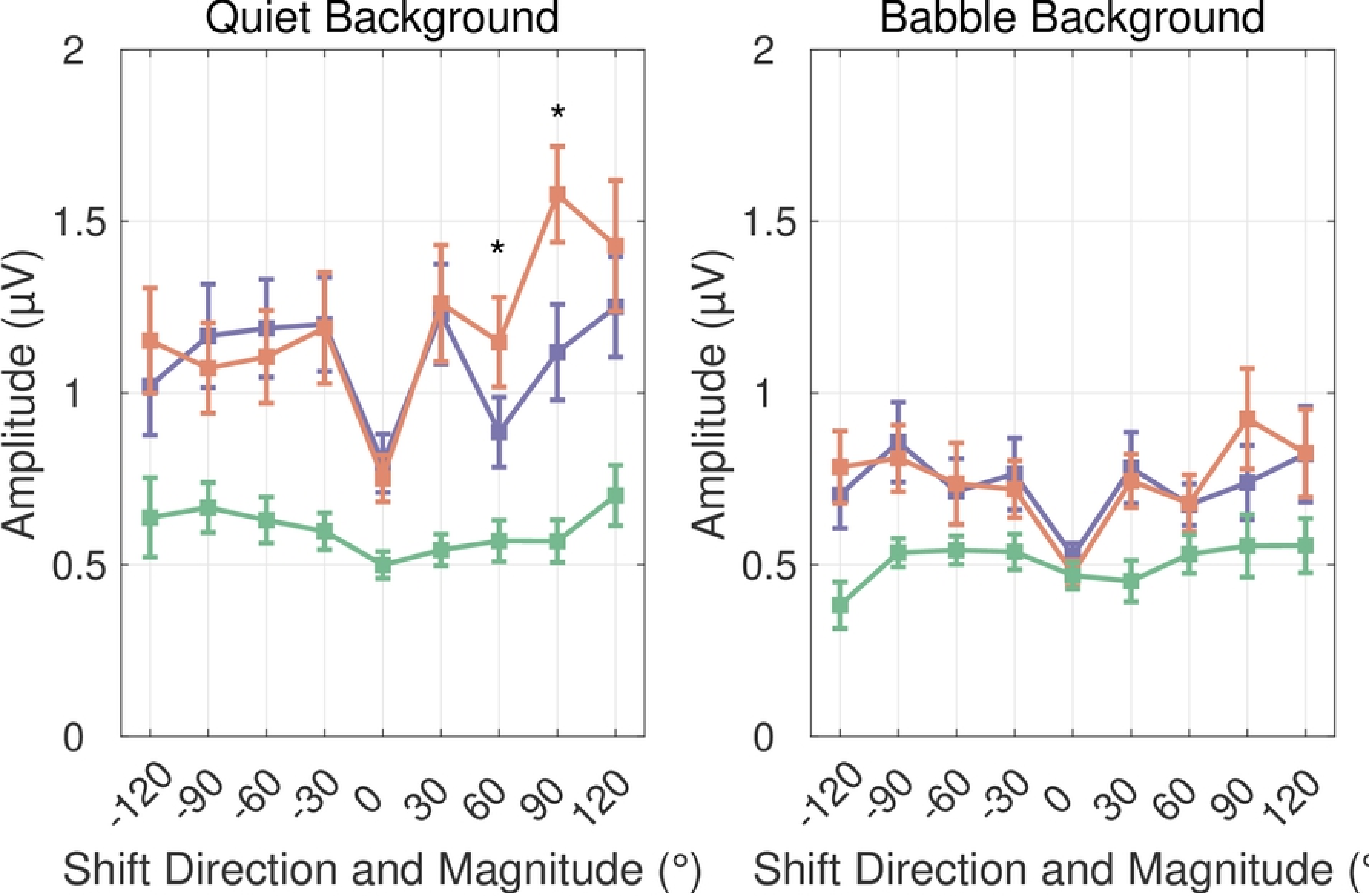

Quiet Background
Babble Background
Amplitude (µV)
2
1.5
1
0.5
0
*
*
-120
-90
-60
-30
0
30
60
90
120
Shift Direction and Magnitude (°)
Shift Direction and Magnitude (°

Figure 7