Auditory spatial tuning to moving speech in quiet or background babble measured in cortex

Katherine Noel Palandrani, and Erol James Ozmeral

Citation: Proc. Mtgs. Acoust. **36**, 050001 (2019); doi: 10.1121/2.0001009
View online: https://doi.org/10.1121/2.0001009
View Table of Contents: https://asa.scitation.org/toc/pma/36/1
Published by the Acoustical Society of America

POMA Proceedings of Meetings on Acoustics

Volume 36 http://acousticalsociety.org/

177th Meeting of the Acoustical Society of America

Louisville, Kentucky

13-17 May 2019

Psychological and Physiological Acoustics: Paper 1aPP2

Auditory spatial tuning to moving speech in quiet or background babble measured in cortex

Katherine Noel Palandrani and Erol James Ozmeral
Department of Communication Sciences and Disorders, University of South Florida, Tampa, FL, 33612; katpaland@gmail.com; eozmeral@usf.edu

The ability to understand speech in complex backgrounds often relies on spatial factors that contribute to forming discernible auditory objects. From stimulus-evoked onset responses in normal hearing listeners using electroencephalography (EEG), we have shown measurable spatial tuning to moving noise bursts in quiet, revealing a potential window into cortical object formation. However, it is still unknown whether comparable effects are observed with speech stimuli, and whether and how much the presence of noise disrupts EEG responses to moving speech. To test whether the presence of noise has deleterious effects on object formation and potential selective auditory attention, we measured cortical responses to moving speech 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 when it occurred at a new location, while indicating yes or no to whether the stimulus occurred at a block-specific location. Results clearly show evoked responses with speech stimuli comparable to those obtained with moving noise-bursts. Measurable spatial tuning and subsequent sharpening was evident in both speech-in-noise and speech-in-quiet conditions, showing that background noise did not have deleterious effects on observed responses.

Published by the Acoustical Society of America

1. INTRODUCTION

The “cocktail party” scenario is often used to illustrate the difficulties associated with understanding speech in a complex acoustic scene. In order to successfully follow speech signals amid competing background stimuli, it is thought that the brain is able to attend to relevant signals by “filtering out” the irrelevant or extraneous signals (Cherry, 1953). Much like visual attention, auditory attention relies on the ability to perceptually group objects of which we can then decide to emphasize or ignore (e.g., Shinn-Cunningham, 2008). Many complex, sequential cognitive processes are required in order to achieve this object-based auditory attention. The brain must be able to first form distinct auditory objects on the basis of common spectro-temporal properties of stimuli before segregating between foreground and background through competition. Over time, this process is called “streaming,” where these complex processes are subject to top-down modulatory effects of attention following behavioral and decision making processes. This phenomenon of increasing neural representation for sensory stimuli features is referred to as sensory gain control (Hillyard et al., 1998) and is the focus of the following investigation.

One of the key features used for the formation of auditory objects in competing backgrounds is space. It is well established in the literature that the magnitude of neural responses are dependent on the extent of shift in spatial location (Briley et al., 2013) or lateral position in the case of binaural cues (Magezi and Krumbholz, 2010; Salminen et al., 2010; Ozmeral et al., 2016). Eddins et al. (2018) demonstrated that a large spatial change in the horizontal location of a stimulus evokes a larger neural response than a small spatial change. Neural activity tends to be largest when a stimulus changes in location far to the left or right in relation to its starting point, even when the listener is not actively attending to the specific spatial location of the stimulus. According to principles of sensory gain control, it is reasonable to hypothesize that active attention to spatial location will yield sharper spatial tuning and greater response magnitudes relative to passive listening under the same conditions.

Eddins et al. (2018) investigated the effects of spatial attention on neural response magnitude and sharpness of spatial tuning. In this study, 10 young listeners and 8 older listeners attended to a left or right target location, while being presented with stimuli consisting of narrowband noise bursts in the front spatial field from either ±60°, ±30°, or 0° azimuth. The participants were instructed to push a button when the noise stimulus moved to the attended target location, while evoked responses were triggered and recorded for the given location at the onset of the stimulus. A passive condition was included where the participants did not attend or respond to a spatial location. While results showed prominent N1 and P2 components, an attention-related P3 component emerged when the stimuli were located in the target location for the young listeners. The older listeners, however, did not show an obvious P3 component in addition to the N1 and N2 observed, leading the authors to suggest an overall reduction to evoked responses, possibly on the basis of reduced temporal synchrony. Both groups showed a more prominent N1 and P2 response for attentive, rather than passive conditions. The results of this previous study established the potential to observe sensory gain control when attending to the perceptual features of auditory objects (i.e., spatial location). However, the stimulus construct was far from what would be considered a “cocktail party.” 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 observable, 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 with electrophysiological measures in order to investigate the dynamic modulation of sensory-evoked brain activity by attention (i.e., sensory gain control). If objects are successfully formed, attention to a perceptual feature like pitch or in this case, 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 stimuli that change in location in an unpredictable manner. These experiments were designed to determine how response patterns differ between attended versus unattended locations, with interest in the effects of responses at both attended and unattended locations. Due to the inherent challenges evident when acoustic scenes are complex (i.e., “cocktail party scenario”), an additional goal was to explore stimulus conditions including speech and speech-in-babble to better reflect real-world listening situations. The difference in effects between arbitrary signals, like noise, versus a speech stream were investigated while evaluating the ways in which a poor signal-to-noise ratio affects neural responses to moving speech.

2. METHOD

Participants included 15 adults (13 females) between the ages of 21 and 25 years of age with audiometrically normal hearing (≤ 20 dB HL at octave frequencies between 250 and 8000 Hz; ANSI, 2010). Data collection was completed over the course of 3-4 visits lasting approximately 2 hours in duration, with a total of 6-8 hours of total testing completed for each participant. All participants provided written consent for study participation, and all procedures were approved by the university Institutional Review Board. Participants were compensated for their time at an hourly rate.

In each experiment, participants were seated in a sound attenuated chamber surrounded by a 360° 24-channel horizontal speaker array at ear level. Electroencephalography (EEG) measures were obtained with the use of an ANT (Advanced Neuro-Technology eego^{TM}) high-speed amplifier and an active shield, Waveguard^{TM} cap with 64 sintered Ag/AgCl electrodes (International 10–20 electrode system). Target stimuli consisted of monosyllabic English words recorded from a male speaker with 100-ms interstimulus intervals at 76 dB SPL. In Experiment I, stimuli were presented in quiet (no background noise). 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 (+6 dB SNR). Target stimuli could be presented from only one location at a time, at either ±60°, ±30°, or 0° azimuth (checkered boxes in Figure 1) and moved to a new, random location with replacement every 2 seconds. In Experiment II, background babble was presented from locations at ±165°, ±105°, ±75°, and ±15° azimuth (black boxes in Figure 1) simultaneously.

In each experiment, listeners were instructed to either attend left (-30°), attend right (+30°), or passively listen while maintaining a forward head position and watching a silent nature-related video (i.e., Planet Earth) without subtitles. The user interface generated in MATLAB consisted of two buttons reading “yes” and “no” in a horizontal row on the southeast location (all participants were right-handed) of a touchscreen in front of the participant. In the active (attend left or attend right) conditions, the corresponding speakers were marked visually by either a blue x (attend -30°), or a red o (attend +30°). The participants were asked 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, or “no” when the target moved to any other location. In the passive conditions, stimuli remained the same; however, the participants were not asked to attend or respond.

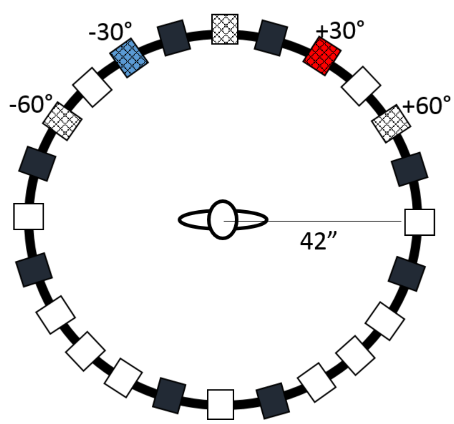

Figure 1. Schematic of the laboratory spatial array. The target speech stream came from alternate locations labeled, -60, -30, 0, +30, and +60 degrees. For the directed-attention conditions, listeners were instructed to either attend to the speaker at -30 (blue; Attend Left) or attend to the speaker at +30 (red; Attend Right). 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).

3. RESULTS

Experiment I presented speech in quiet, whereas Experiment II presented speech in noise. Analyses centered on the auditory event-related potentials associated with the change in location of the target stream. To capture the overall activity across the 64 scalp electrodes, the rms activity, or global field power (GFP; Skrandies et al., 1990) was computed for subsequent analyses. Figure 2 shows the average GFP in Experiment I for all correct trials in which the stimulus moved to one of five possible loudspeaker locations. Figure 3 shows the same from Experiment II. Correct trials were determined by the behavioral response (i.e., 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. Colored lines in Figures 2 and 3 indicate the instructions given to the participant – either to passively listen (black lines), attend to the left (blue), or attend to the right (red).

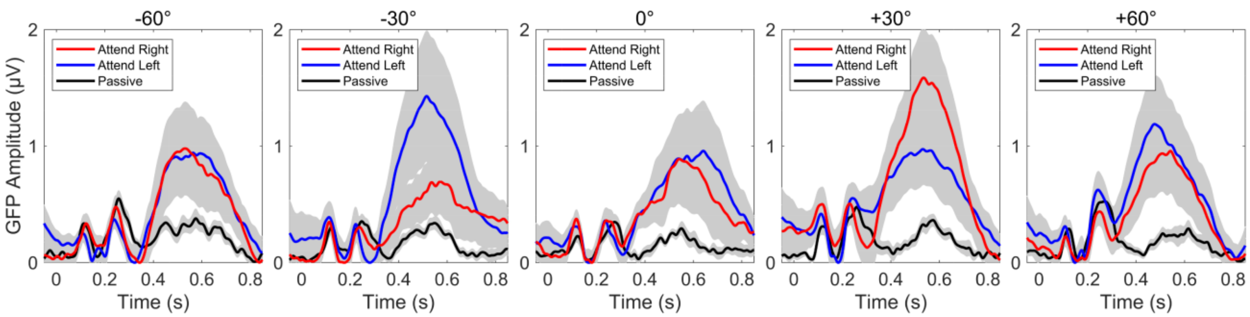

Figure 2. Grand average global field power (GFP) from Experiment I in which the speech stream was presented in quiet. Panels represent trials in which the target arrived at each of the five speaker locations (from left to right: -60 to +60). The average waveforms are separated by attention condition (Attend Right [red]; Attend Left [blue]; and Passive [black]).

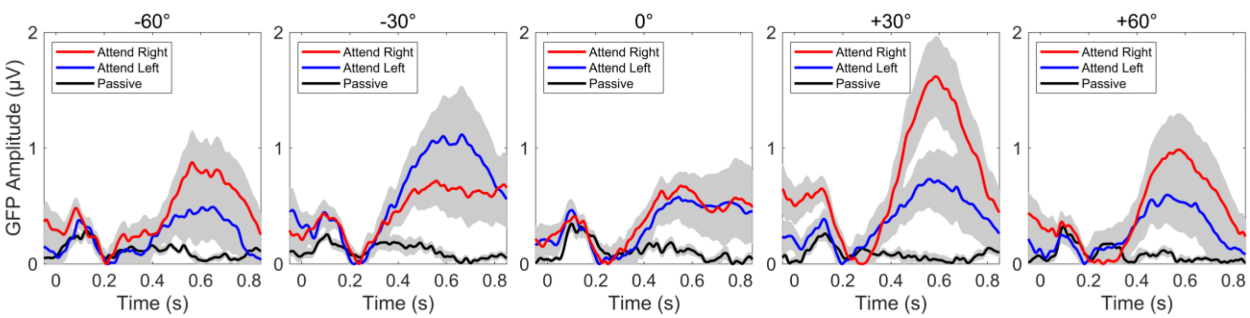

Figure 3. Grand average global field power (GFP) from Experiment II in which the speech stream was presented in background speech babble (+6-dB SNR). Panels represent trials in which the target arrived at each of the five speaker locations (from left to right: -60 to +60). The average waveforms are separated by attention condition (Attend Right [red]; Attend Left [blue]; and Passive [black]).

Results indicate that much like responses to narrowband noise burst stimuli (e.g., Eddins et al., 2018), the morphology of the GFP has prominent deflections from baseline at latencies near 150 ms, 250 ms, and around 500 ms that correspond to the N1, P2, and possibly a P3 component, respectively (Picton, 1992; see Figure 2). Because the stimuli are presented in the free field, these latencies tend to be later than the commonly held assumption regarding the latency of peak deflections in auditory event-related potentials (e.g., Briley et al., 2013). Of key interest was the effect on the late-latency component, P3, which is believed to be associated with higher-order processing following the change in spatial location. If listeners are indeed able to latch on to the spatial features of the auditory object or stream, then attention to that feature should modulate its neural representation. In Figure 2, the late-latency deflection was modest for the passive condition at all locations, which would be expected if listeners were not motivated behaviorally to attend to the stimuli. On the other hand, prominent late-latency deflections were observed for active attention conditions (red and blue). This was consistent with Eddins et al. (2018), who showed that evoked responses during a passive condition were less robust than active conditions. Moreover, it is clear from the present results that *where* the listener was attending played a key role in the activity level of the late-latency component. At -30 degrees to the left, the largest peak was observed for the Attend Left condition, and at +30 degrees to the right, the largest peak was observed for the Attend Right condition. At non-target locations (-60, 0, and +60 degrees), no discernible difference in amplitude was observed between the two attention conditions.

The addition of background babble in Experiment II posed the threat of reducing or eliminating observable sensory gain control. However, the responses measured under the +6 dB SNR babble condition were consistent with effects shown in the prior experiment (compare Figures 2 and 3). At all locations, there was a robust P3 component with overall greater potentials for active rather than passive conditions, which was mostly driven by the late-latency component. For conditions in which listeners attended to a specific target location (i.e., ± 30°), responses were demonstrably larger, whereas at non-target locations, attention conditions were less robust. Together, the results described by Figures 2 and 3 demonstrate that overall neural activity is strengthened when attending to the auditory stream and its spatial location, and modulatory effects are present when there is a congruence between attended location and the source location.

Finally, to explore the effect of attention on spatial tuning, data were re-analyzed with respect to not only the location of the speech stream after any given location change, but also with respect to where the stimulus was previous to the change. In Figure 4, colored lines represent GFP at N1 for the five possible stimulus locations (-60 to + 60 degrees) as a function of the change in location in degrees (a negative sign indicates to the left and positive sign indicates to the right). Figure 4A shows the cortical tuning curves for Experiment I (speech in quiet), and Figure 4B shows the cortical tuning curves for Experiment II (speech in babble). The left panels are for passive conditions, and the right panels are for the Attend-Left conditions (the Attend-Right conditions were left out for the sake of brevity). By overlapping these curves, the general shape of a funnel emerges, in which the lowest activity occurs for no change in space (0 degrees), and an increase in activity is observed when the stimulus changes location by 30 degrees or more. In previous studies in the free field, noise bursts that changed location evoked similar responses (e.g., Briley et al., 2013) and attention to noise sharpened spatial tuning and increased overall magnitude (unpublished). In the present study, it is clear that such tuning and subsequent sharpening with attention was possible both for a speech stream in quiet (Figure 4A) and speech stream in noise (Figure 4B). Importantly, the present results focus on the earliest component, N1, but do not necessarily rule out such effects at later latency components. Rather, it is quite remarkable that the effects of attention are seen for cortical responses believed to be driven by peripheral coding.

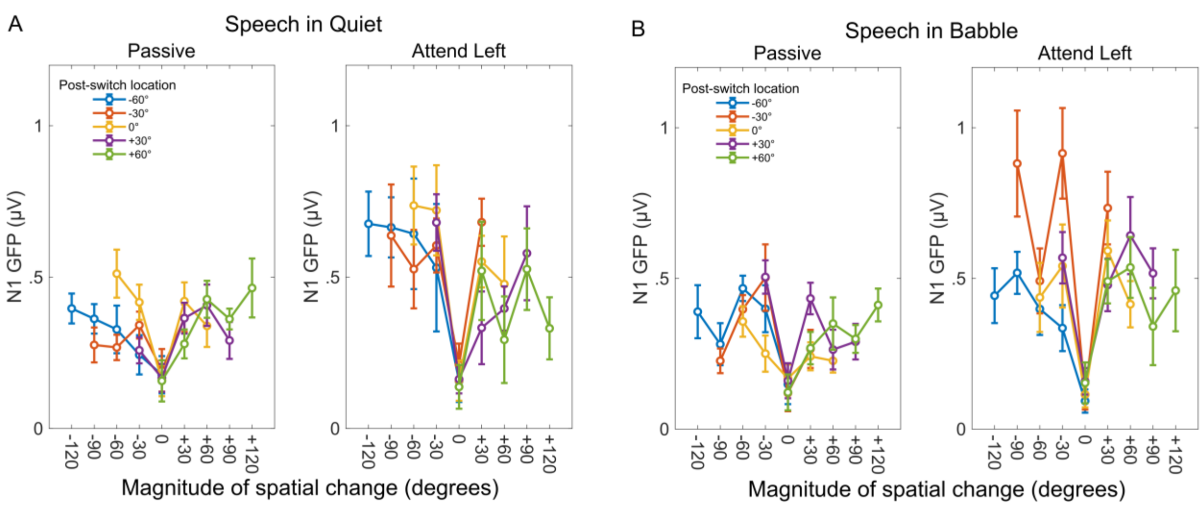

Figure 4: Cortical spatial tuning curves derived from the global field power (GFP) at the N1 latency as a function of the magnitude of spatial change in degrees. Colored lines represent the five possible post-switch locations. Figure 4A shows the tuning curve for the Passive condition (left panel) and Attend Left (right panel) for Experiment I (Speech in Quiet), and Figure 4B shows the same for Experiment II (Speech in Babble).

4. DISCUSSION AND CONCLUSION

The goal of this study was to demonstrate sensory gain control when the brain is attending to the spatial location of an auditory object. Earlier work demonstrated specific neural correlates for noise stimuli, however, it was unknown if more real-world stimuli would show comparable evoked responses. Results clearly demonstrate an evoked P3 response that was similar to the neural responses established for noise bursts (Eddins et al., 2018). Active attention to specific target locations modulated the overall responses, and background babble (+6 dB SNR) did not have deleterious effects on observed measures. Spatial tuning and subsequent sharpening was evident in this group of young normal-hearing listeners when attending to a single location, in quiet or in competing background babble at a sufficient signal-to-noise ratio.

While young normal hearing listeners do not show difficulty localizing speech with competing background signals, there is a challenge in individuals with hearing loss (Best et al., 2011). 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 and/or their success with potential interventions. It is unclear, for example, whether a better signal-to-noise ratio resulting from directional processing in hearing aids could help to mediate spatial hearing challenges. Future work will focus on the consequences of aging and hearing impairment on object-based auditory attention, while evaluating efficacy of spatial hearing enhancement (i.e., directional microphones) in hearing aids.

ACKNOWLEDGMENTS

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. These data were presented at the 177th Meeting of the Acoustical Society of America in Louisville, KY. This work was supported by an internal grant from the College of Behavioral and Community Sciences at USF.

REFERENCES

American National Standards Institute. (2010). Specifications for audiometers (ANSI S3. 6-2010).

Best, V., Mason, C. R., & Kidd, G. Jr. (2011). Spatial release from masking in normally hearing and hearing-impaired listeners as a function of the temporal overlap of competing talkers. *The Journal of the Acoustical Society of America*, *129*(3), 1616-1625.

Briley, P. M., Kitterick, P. T., & Summerfield, A. Q. (2013). Evidence for opponent process analysis of sound source location in humans. *Journal of the Association for Research in Otolaryngology*, *14*(1), 83-101.

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.

Eddins, A. C., Ozmeral, E. J., & Eddins, D. A. (2018). How aging impacts the encoding of binaural cues and the perception of auditory space. *Hearing Research*, *369*, 79-89.

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.

Magezi, D. A., & Krumbholz, K. (2010). Evidence for opponent-channel coding of interaural time differences in human auditory cortex. *Journal of Neurophysiology*, *104*(4), 1997-2007.

Ozmeral, E. J., Eddins, D. A., & Eddins, A. C. (2016). Reduced temporal processing in older, normal-hearing listeners evident from electrophysiological responses to shifts in interaural time difference. *Journal of Neurophysiology*, *116*(6), 2720-2729.

Picton, T. W. (1992). The P300 Wave of the human event-related potential. *Journal of Clinical Neurophysiology*, 9(4), 456-479.

Salminen, N. H., Tiitinen, H., Miettinen, I., Alku, P., & May, P. J. (2010). Asymmetrical representation of auditory space in human cortex. *Brain Research*, *1306*, 93-99.

Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences*, *12*(5), 182-186.

Skrandies, W. (1990). Global field power and topographic similarity. *Brain Topography*, *3*(1), 137-141.