Parametric information about eye movements is sent to the ears

Stephanie N Lovich[1-4*] Cynthia D King[1-4*], David LK Murphy[1,3,4*], Rachel Landrum[1-4], Christopher A Shera[7], Jennifer M Groh[1-6]

*These authors contributed equally to this work and are listed in reverse chronological order of their period of most intense involvement.

[1]Department of Psychology and Neuroscience
[2]Department of Neurobiology
[3]Center for Cognitive Neuroscience
[4]Duke Institute for Brain Sciences
[5]Department of Computer Science
[6]Department of Biomedical Engineering
Duke University

[7]Dept. of Otolaryngology
University of Southern California

Acknowledgments

We are grateful to Dr. Matthew Cooper, Dr. Kurtis Gruters, Dr. David Kaylie, Dr. Jeff Mohl, Dr. Shawn Willett, Meredith Schmehl, Dr. David Smith, Chloe Weiser and Justine Shih for discussions and assistance concerning this project. This work was supported by NIH (NIDCD) grant DC017532 to JMG.

Abstract

Eye movements alter the relationship between the visual and auditory spatial scenes. Signals related to eye movements affect neural pathways from the ear through auditory cortex and beyond, but how these signals contribute to computing the locations of sounds with respect to the visual scene is poorly understood. Here, we evaluated the information contained in eye movement-related eardrum oscillations (EMREOs), pressure changes recorded in the ear canal that occur in conjunction with simultaneous eye movements. We show that EMREOs contain parametric information about horizontal and vertical eye displacement as well as initial/final eye position with respect to the head. The parametric information in the horizontal and vertical directions combines linearly, allowing accurate prediction of the EMREOs associated with oblique eye movements from their respective horizontal and vertical components. Target location can also be inferred from the EMREO signals recorded during eye movements to those targets. We hypothesize that the thus-far unknown mechanism underlying EMREOs could impose a two-dimensional eye-movement related transfer function on any incoming sound, permitting subsequent processing stages to compute the positions of sounds in relation to the visual scene.

Introduction

Every time we move our eyes to localize multisensory stimuli, our retinae move in relation to our ears. These movements shift the alignment of the visual scene (as detected by the retinal surface) with respect to the auditory scene (as detected based on timing, loudness, and frequency differences in relation to the head and ears). Precise information about each eye movement is therefore needed to connect the brain's views of visual and auditory space to one another (e.g. e.g. Groh and Sparks, 1992; Boucher et al., 2001; Metzger et al., 2004). Most previous work about how eye movement information is incorporated into auditory processing has focused on cortical and subcortical brain structures (Jay and Sparks, 1984, 1987a, b; Russo and Bruce, 1994; Hartline et al., 1995; Stricanne et al., 1996; Cohen and Andersen, 2000; Groh et al., 2001; Zella et al., 2001; Werner-Reiss et al., 2003; Fu et al., 2004; Populin et al., 2004; Zwiers et al., 2004; Mullette-Gillman et al., 2005; Porter et al., 2006; Mullette-Gillman et al., 2009; Maier and Groh, 2010; Bulkin and Groh, 2012a, b; Lee and Groh,

2012; Caruso et al., 2021), but the recent discovery of eye-movement related eardrum oscillations (EMREOs) (Gruters et al., 2018) suggests that the process might be manifest much earlier in the auditory periphery. EMREOs can be thought of as a biomarker of underlying efferent information impacting the internal structures of the ear in association with eye movements. What information this efferent signal contains is currently unknown.

We reasoned that if this efferent signal is to play a role in linking auditory and visual space across eye movements, EMREOs should be *parametrically* related to the associated eye movement. Specifically, EMREOs should vary in a regular and predictable fashion with both horizontal and vertical displacements of the eyes, and some form of information regarding the initial position of the eyes should also be present. These properties are required if the efferent signal underlying EMREOs is to play a role in linking hearing and vision. Notably, this parametric relationship is not required of alternative possible roles, such as synchronizing visual and auditory processing in time or enhanced attentional processing of sounds regardless of their spatial location (Barczak et al., 2019; O'Connell et al., 2020).

Accordingly, we evaluated the parametric spatial properties of EMREOs in human participants by varying the starting and ending positions of visually-guided saccades in two dimensions. We find that EMREOs do in fact vary parametrically depending on the saccade parameters in both horizontal and vertical dimensions and as a function of both initial eye position in the orbits and the change in eye position relative to that initial position. EMREOs associated with oblique saccades can be predicted by the linear combination of the EMREOs associated with strictly horizontal and vertical saccades. Furthermore, an estimate of target location can be decoded from EMREOs alone – i.e. where subjects looked in space can be roughly determined from their observed EMREOs.

These findings suggest that the eye-movement information needed to accomplish a coordinate transformation of incoming sounds into a visual reference frame is fully available in the most peripheral part of the auditory system. While the precise mechanism that creates EMREOs remains unknown, we propose that the underlying mechanisms might introduce a transfer function to the sound transduction process that serves to adjust

the gain, latency, and/or spectral dependence of responses in the cochlea. In principle, this could provide later stages of auditory processing access to an eye-centered signal of sound location for registration with the eye-centered visual scene (Groh and Sparks, 1992), Indeed, recent work has shown that changes in muscular tension on the ossicular chain would be expected to affect gain and latency of sound transmission through the middle ear, thus supporting the plausibility of this hypothesis (Gallagher et al., 2021; Cho et al., in revision).

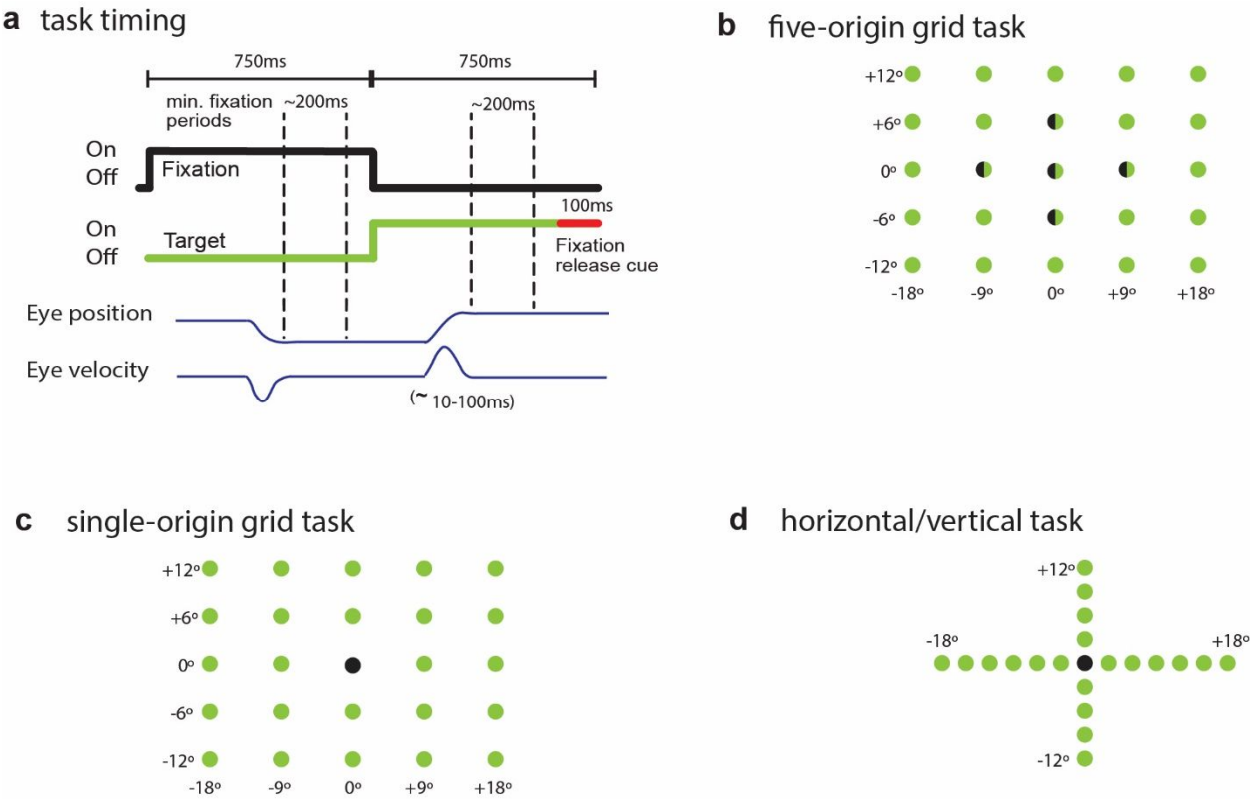


Figure 1. Events of the tasks in time and space. a. Task events across time. Each trial began with the onset of an initial "fixation" cue (black trace). Participants made saccades to the fixation point, then maintained fixation for a minimum of 200 ms. The fixation point was then turned off and a new "target" was turned on (green trace). Participants saccaded to this target and fixated for another 200 ms, at which point the target turned red indicating that the trial was over. The ear-canal recordings were analyzed in conjunction with the fixation-point-to-target saccade. b-d. Spatial layouts of fixation points and targets for the three task designs used in this study. Points in space that were used as both a fixation and a target across different trials are half-green, half-black.

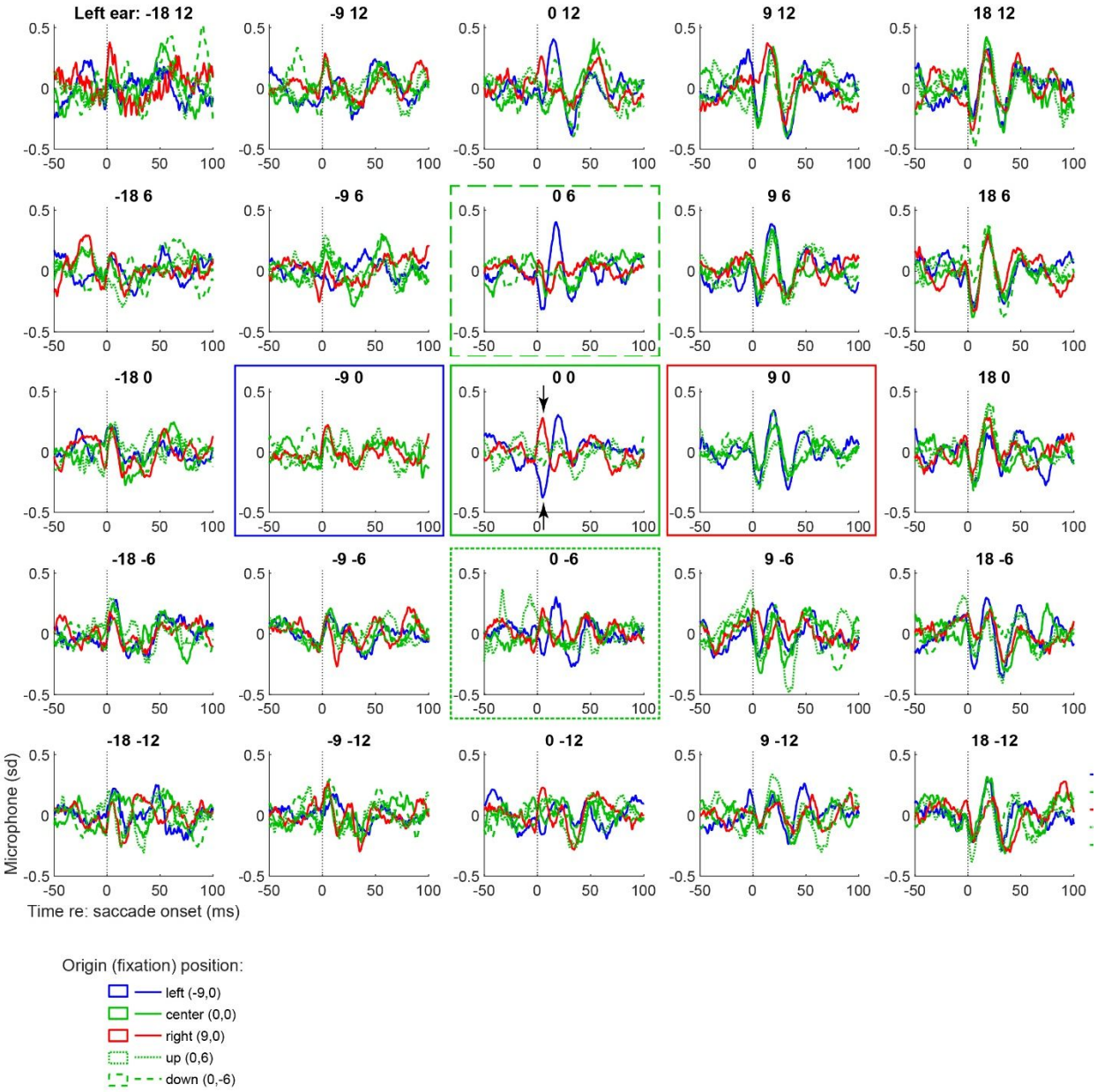


Figure 2. EMREOs recorded during the five-origin grid task. Each panel shows the grand average EMREO signal generated when saccades were made to that location on the screen (average of N=10 subjects' individual left ear averages). For example, the top right panel shows microphone recordings during saccades to the top right (contralateral) target location, and the color and line styles of each trace in that panel correspond to saccades from different initial fixation points. e.g. the red traces originated from the rightward fixation, the blue from the leftward fixation etc as indicated by the legend and boxes of the same color and line style. Both magnitude and phase vary as a function of initial eye position and target location, with contralateral responses being larger than ipsilateral. Phase reversal occurs based on the location of the target with respect to the initial fixation position, as can be seen for the central target location (central panel), where the EMREOs evoked for saccades from the rightward fixation (ed

traces) show an opposite phase relationship as those evoked for saccades from the leftward fixation (blue traces). Corresponding grand averages for right ear data are shown in Supplementary Figure 1.

Results

We used earbud microphones to record internally-generated oscillations in the ear canals of human subjects (n=10 for each task) while they performed eye-movement tasks involving various visual fixation and target configurations (Figure 1). No external sounds were presented in any task. The events of the tasks in time are shown in Figure 1a. At the beginning of each trial, subjects fixated a visual fixation point for 750 ms and then made a saccade to a second target, which they then fixated for another 200 ms. Any trials with micro- or corrective saccades during the 200 ms prior to or following main fixation-point-to-target saccade were discarded, to ensure a stable baseline ear-canal recording could be established without intrusions by other eye movements. Ten subjects were tested in the single origin and horizontal/vertical tasks, and ten were tested in the five-origin grid task. Four subjects participated in both groups, so that 16 subjects (8 female, 8 male) were tested overall. Female-male ratios were equal in both subgroups.

We first tested subjects (N=10) on a task involving variation in both initial fixation position and target locations varying along both horizontal and vertical dimensions – the "five-origin grid task". Subjects fixated on an initial fixation light located either straight ahead, 9º left or right, or 6º up or down, and then made a saccade to a target located within the array of possible target locations spanning +/- 18º horizontally and +/- 12º vertically as shown in Figure 1B. Results of this task are shown in Figure 2. Each panel shows the average microphone signal recorded in the left ear canal (averaged across all subjects) associated with saccades to a target at that location – e.g. the top right panel shows all saccades to the top right target location. The color and line styles of the waveforms correspond to the five initial fixation positions from which the saccades could originate in space.

The first overall observation from this figure is that the *magnitude* of the waveform of the EMREO depends on both the *horizontal* and *vertical* dimensions. In the horizontal dimension, EMREOs are larger for more contralateral target locations: compare the column on the right (contralateral) to the column on the left (ipsilateral).

The pattern is reversed for right ear canal recordings (Supplementary Figure 1). In the vertical dimension, EMREOs are larger for higher vs lower targets in both left and right ears (compare top row to bottom row in Figure 1/Supplementary Figure 1).

The second overall observation from this figure is that the *phase* of the EMREO waveform depends on the horizontal location of the target *with respect to* the fixation position. Specifically, the first deflection after saccade onset is a peak for the most ipsilateral targets (left-most column) and trough for the most contralateral targets (right-most column). But *where* this pattern reverses depends on the initial fixation position. Specifically, consider the red vs blue traces in the middle column of the figure, which correspond to targets along the vertical meridian. Red traces involve saccades to these targets from the fixation position on the right, and thus involve leftward (ipsiversive) saccades. The red traces in this column begin with a peak followed by a trough. In contrast, the blue traces involve saccades to these targets from the fixation position on the left, i.e. rightward or contraversive saccades. The blue traces begin with a trough followed by a peak. The pattern is particularly evident in the central panel (see arrows).

The phase reversal as a function of the combination of target location and initial eye position suggests that the EMREO waveforms might align *better* when plotted in an eye-centered frame of reference. Figure 3 demonstrates that this is indeed the case: the data from Figure 2 is re-plotted as a function of target location *relative* to the fixation position. The eight panels around the center represent the traces for the subset of targets that can be fully analyzed in an eye-centered frame, i.e. the targets immediately left, right, up, down, and diagonal relative to the five fixation locations. By plotting the data based on the relative location of the targets to the origins, the waveforms are better aligned, showing no obvious phase reversals.

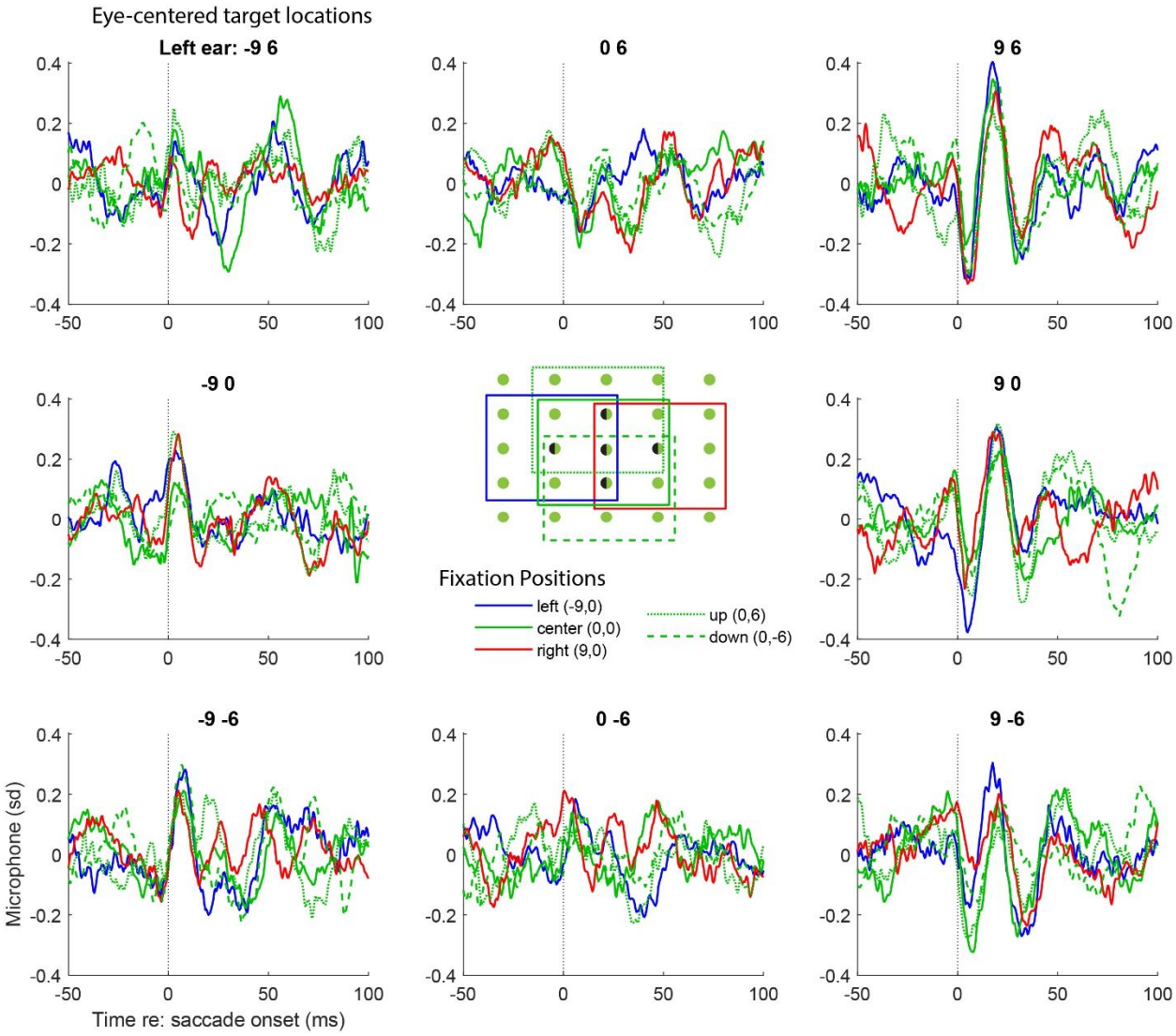


Figure 3. Replotting the grand average EMREOs as a function of relative *target location shows better, but not perfect, correspondence of the EMREOs across different fixation positions. The data shown are a subset of those shown in Figure 2, but here each panel location corresponds to a particular target location defined relative to the associated fixation position. The color/linestyle indicates the associated relative fixation position. For example, the waveforms in the upper right panel all involved 9º rightward and 6º upward saccades; the red trace in that panel indicates those that originated from the 9º right fixation; the blue those from the 9º left fixation etc. Only relative target locations that existed for all 5 fixation positions are plotted, as indicated by the inset. Corresponding right ear data are shown in Supplementary Figure 2.*

Although the waveforms are better aligned when plotted based on target location *relative* to initial eye position, some variation related to fixation position is still evident in the traces. That is, in each panel, the EMREO waveforms with different colors/line styles (corresponding to different fixation positions) do not necessarily superimpose perfectly. This suggests that a model that incorporates both relative target position and

original fixation position, in both horizontal and vertical dimensions, is needed to account for the findings. Furthermore, a statistical accounting of these effects is needed. Accordingly, we fit the data to the following regression equation:

$$Mic(t) = B_H(t)\,H + B_{\Delta H}(t)\,\Delta H + B_V(t)\,V + B_{\Delta V}(t)\,\Delta V + C(t) \quad \text{Eq 1}$$

where H and V correspond to the initial horizontal and vertical eye position and ΔH and ΔV correspond to the respective changes in position associated with that trial. The slope coefficients B_H, $B_{\Delta H}$, B_V, $B_{\Delta V}$ are time-varying and reflect the dependence of the microphone signal on the respective eye position/movement parameters. The term $C(t)$ contributes a time-varying "constant" independent of eye movement metrics. It can be thought of as the best fitting average oscillation across all initial eye positions and changes in eye position. We used the measured values of eye position/change in eye position for this analysis rather than the associated fixation and target locations so as to incorporate trial-by-trial variability in fixation and saccade accuracy.

Figure 4 shows the average time-varying values of the slope coefficients across subjects (blue = left ear; red = right ear) and provides information about the contribution of these various eye movement parameters to the EMREO signal ~~ear~~. A strong, consistent dependence on horizontal eye displacement is observed, consistent with our previous report (Figure 4A) (Gruters et al., 2018). This component is oscillatory and begins slightly before the onset of the eye movement, inverting in phase for left vs right ears. The thickened parts of the line indicate periods of time when this coefficient differed significantly from 0 with 95% confidence (Shaded areas are +/-SEM). There is also an oscillatory and binaurally phase-inverting signal related to the initial position of the eyes in the horizontal dimension (Figure 4B). This signal is smaller and more variable across subjects.

In the vertical dimension, the effect of vertical saccade amplitude is in phase for both the left and right ears; it exhibits an oscillatory pattern, although not obviously sinusoidal like the one observed for the horizontal saccade amplitude. Initial position of the eyes in the vertical dimension exerts a variable effect across participants such that it is not particularly evident in this grand average analysis; this may be related to poorer

abilities to localize sounds in the vertical vs. horizontal dimensions (Hebrank and Wright, 1974a, b; Middlebrooks and Green, 1991; Macpherson and Sabin, 2013).

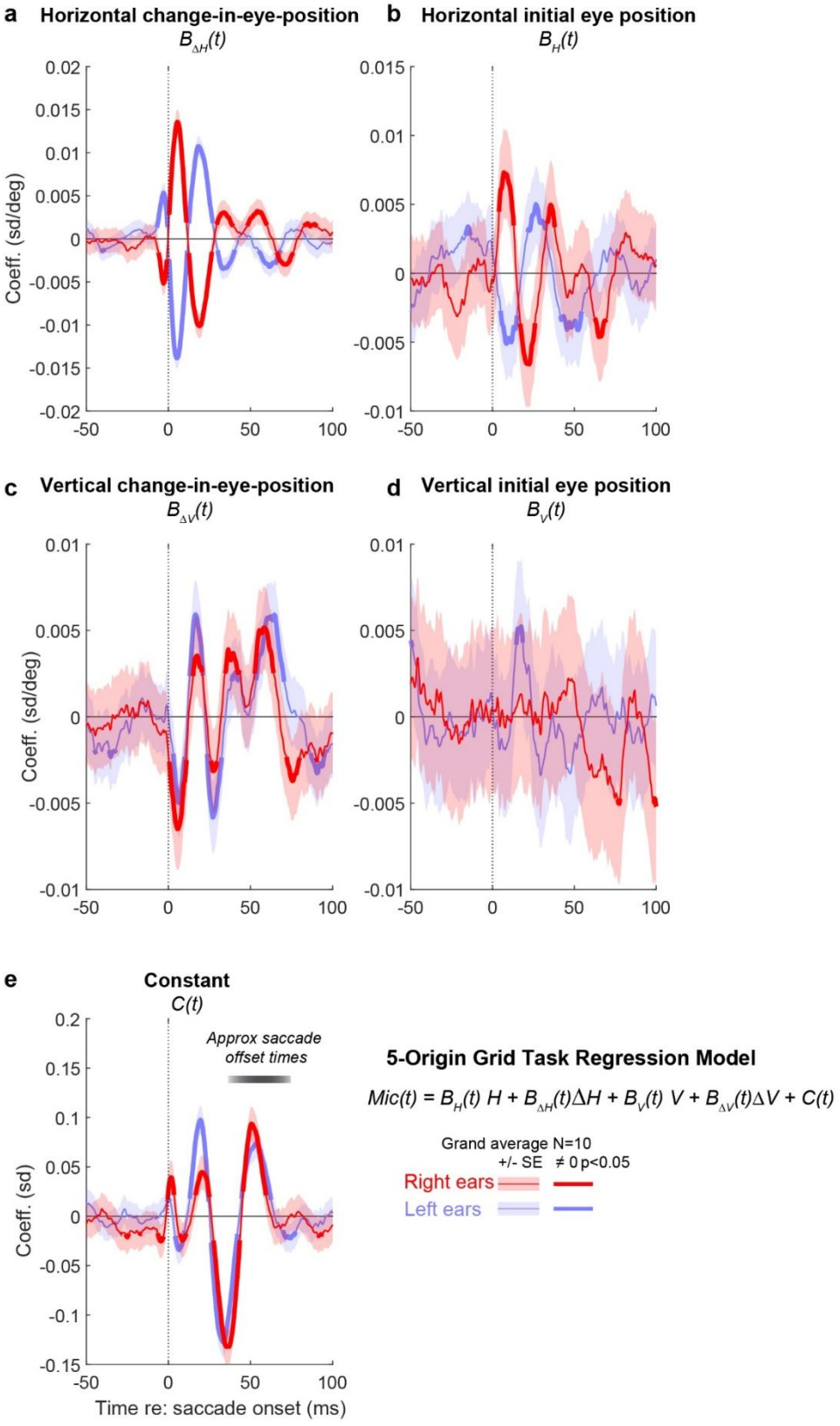


Figure 4. Regression analysis of EMREOs shows contributions from multiple aspects of eye movement: horizontal and vertical change-in-eye-position (A, C), horizontal initial eye position (B), as well as a constant component that was consistent across saccades (E). The contribution of vertical initial eye position was weaker (D). The regression involved modeling the microphone signal at each time point, and each panel shows the time varying values of the coefficients associated with the different aspects of the eye movement (horizontal vs. vertical, change-in-position and initial position). The regressions were fit to individual subjects' microphone recordings, and plotted here as grand averages of these regression coefficients across the N=10 subjects tested in the 5-origin grid task. Microphone signals were z-scored in reference to baseline variability during a period -150 to 120 ms prior to saccade onset. Results are presented in units of standard deviation (panel e) or standard deviation per degree (panels a-d). Shaded areas represent +/-SEM.

Finally, there is a constant term that is similar in the two ears and is larger later in the saccade than early in the saccade (Figure 4E). As noted above, this constant term can be thought of as encapsulating the average EMREO waveform that occurs when pooling across all the eye movements in the dataset, regardless of their initial positions or horizontal or vertical components.

The basic regression model assumes a roughly linear relationship between the contributions of the horizontal and vertical dimensions of the eye movements – but is this appropriate? To test this, we collected data using two simplified tasks, the single-origin-grid task (with a single initial fixation in the center, Figure 1C) and the horizontal/vertical task (with fixation-target pairs on the horizontal and vertical meridians, generating purely horizontal or vertical saccades, Figure 1D). We sought to determine if we could predict an EMREO associated with an oblique eye displacement using the corresponding components of purely horizontal and purely vertical eye movements. Ten subjects (four of whom also completed the 5-origin grid task) completed both the single-origin grid task and the horizontal/vertical saccade. We fit the results from these tasks using the same regression procedure but omitting the initial fixation position terms, i.e.:

$$Mic(t) = B_{\Delta H}(t)\,\Delta H + B_{\Delta V}(t)\,\Delta V + C(t) \qquad Eq\ 2$$

As shown in Figure 5, both tasks yield similar values of the regression coefficients for horizontal change-in-position ($B_{\Delta H}(t)$) and the constant term ($C(t)$) (grand

average across the population, black vs. green traces). The vertical change-in-position term ($B_{\Delta V}(t)$) was slightly more variable but also quite consistent across tasks.

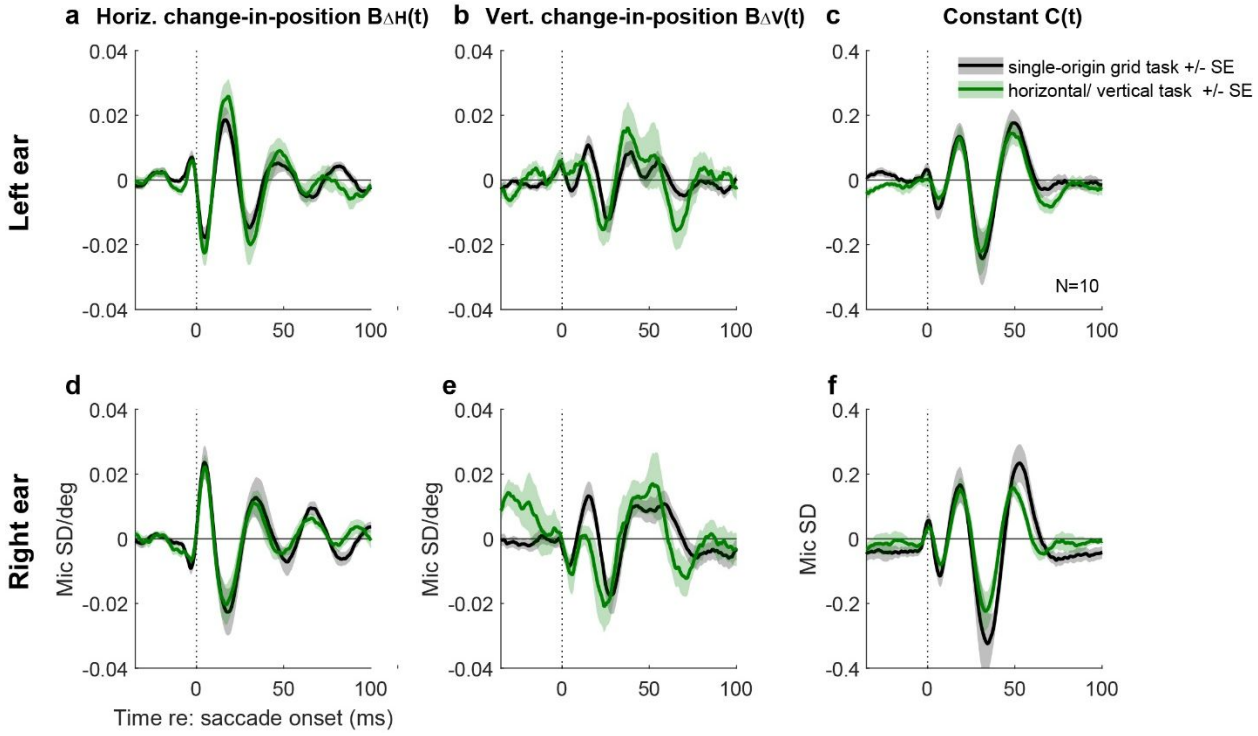


Figure 5. Different tasks generate similar regression coefficient curves. Grand average of the regression results for the single-origin grid (black lines) and horizontal/vertical (green lines) tasks. The lines and shading represent the average and standard error of the coefficient values across the same 10 subjects for the two tasks.

Given the consistency of the regression coefficient values between the single-origin grid and horizontal/vertical tasks, we surmised that it should be possible to use the *coefficient* values from one task to predict the EMREO *waveforms* in the other. Specifically, we used the time-varying regression values from the horizontal/vertical task to predict the observed waveforms from the single origin grid task.

The black traces in Figure 6 show the grand average microphone signals associated with each target in the single-origin grid task. The location of each trace corresponds to the physical location of the associated target in the grid task (similar to Figure 2). The superimposed predicted wave forms (red traces) were generated from

the $B_{\Delta H}(t)$, $B_{\Delta V}(t)$, and C(t) regression coefficients fit to only the horizontal/vertical data, then evaluated at each target location and moment in time to produce predicted curves for each of the locations tested in the grid task.

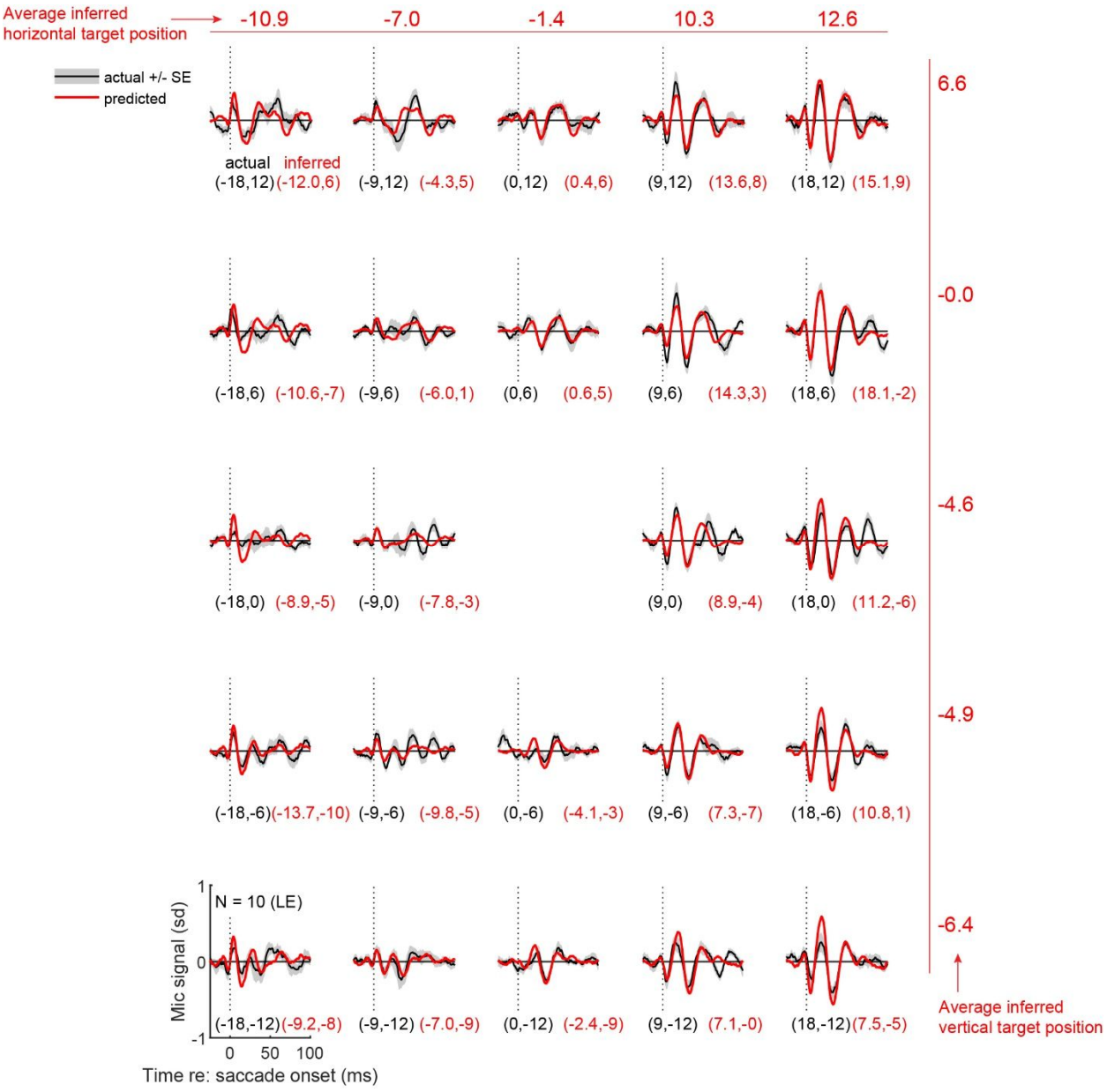


Figure 6. Regression coefficients fit to microphone recordings from the horizontal/vertical-saccade task can be used to predict the waveforms observed in the grid task and their corresponding target locations. Combined results for all N=10 participants' left ears. The black traces indicate the grand average of all the individual participants' mean microphone signals during the single-origin grid task, with the shading indicating +/- the standard error across participants. The red traces show an estimate of the EMREO at that target location based only on regression coefficients

measured from the horizontal/vertical task. Black values in parentheses are the actual horizontal and vertical coordinates for each target in the grid task. Corresponding red values indicate the inferred target location based on solving a multivariate regression which fits the observed grid task microphone signals in a time window (-5 to 70 ms with respect to saccade onset) to the observed regression weights from the horizontal/vertical task for each target location. The averages of these values in the horizontal and vertical dimensions are shown across the top and right sides. See Figure 7 for additional plots of the inferred vs actual target values, and Supplementary Figure 3 for corresponding right-ear data.

Overall, there is good correspondence between the predicted EMREO oscillations and the observed EMREO from actual microphone recordings, including the oblique target locations that were not tested in the horizontal/vertical task. This illustrates two things: 1) the EMREO is reproducible across task contexts, and 2) the horizontal and vertical change-in-position contributions interact in a reasonably linear way, so that the EMREO signal observed for a combined horizontal-vertical saccade can be predicted as the sum of the signals observed for purely horizontal and purely vertical saccades with the corresponding component amplitudes.

Given that it is possible to predict the microphone signal from one task context to another, it should also be possible to decode the target location and its associated eye movement from just the simultaneously-recorded microphone signal. To do this, we again used the weights from the horizontal/vertical task data for the regression equation:

$$Mic(t) = B_{\Delta H}(t)\,\Delta H + B_{\Delta V}(t)\,\Delta V + C(t) \qquad \text{Eq 2}$$

We then used the Mic(t) values observed in the single-origin grid task to solve this system of multivariate linear equations across the time window -5 to 70 ms with respect to the saccade (a time period in which the EMREO appears particularly consistent and substantial in magnitude) to generate the "read out" values of ΔH and ΔV associated with each target's actual ΔH and ΔV. We conducted this analysis on the left ear and right ear data separately. The left ear results of this analysis are seen in each of the individual panels of Figure 6; the black values (e.g. -18, 12) indicate the actual horizontal and vertical locations of the target, and the associated red values indicate the inferred location of the target. Across the top of the figure, the numbers indicate the average inferred horizontal location, and down the right side, the numbers indicate the

306 average inferred vertical location. These results indicate that, on average, the targets
307 can be read out in the proper order, but the spatial scale is compressed: the average
308 read-out values for the +/-18 degree horizontal targets are +/- ~11-12 degrees, and the
309 averages for the vertical +/- 12 degree targets are +/- ~6-7 degrees. Similar findings
310 applied to the right ear data (Supplementary Figure 3).

311 Plots of these target readouts in both horizontal and vertical dimensions for both
312 ears are shown in Figure 7A-F. Figure 7A shows the inferred location of the target (red
313 dots) connected to the actual location of the target (black dots) using the data from
314 Figure 6, i.e the left ear readout, and Figure 7B-C show regressions of these target
315 readouts as a function of the horizontal and vertical locations. Figure 7D-F show the
316 corresponding results for the right ears. Altogether, these figures illustrate that the
317 readout accuracy is better in the horizontal than in the vertical dimensions.
318 Quantitatively, the r^2 values for the horizontal dimension were 0.89 (LE) and 0.91 (RE),
319 and the corresponding values for the vertical dimension were 0.61 (LE) and 0.67 (RE).
320 Slopes were also closer to a value of 1 for the horizontal dimension (0.71, LE; 0.77, RE)
321 than for the vertical dimension (0.51, LE, 0.51, RE). for left and right ears alone are
322 shown in Figure 7A-F.

323 Given that it is known that the brain uses binaural computations for
324 reconstructing auditory space, we wondered whether the accuracy of this read-out could
325 be improved by combining signals recorded in each ear simultaneously. We first
326 considered a binaural difference computation, subtracting the right ear microphone
327 recordings from the left, thus eliminating the part of the signal that is common between
328 the two ears. Figure 7G shows the results. Generally, the horizontal dimension is well
329 ordered whereas the vertical dimension continues to show considerable shuffling. This
330 can also be seen in Figure 7H and 7I, which show the relationship between the inferred
331 target location and the true target location, plotted on the horizontal and vertical
332 dimension, respectively. The correlation between inferred and actual target is higher in
333 the horizontal dimension (r^2 0.95) than the vertical dimension (r^2 0.41), which is actually
334 worse than the monaural readouts. This makes sense because the binaural difference
335 computation serves to diminish the contribution from aspects of the signal that are in
336 phase across the two ears, such as the dependence on vertical change in eye position.
337 We then reasoned that improvement in the vertical readout could be achieved by

instead averaging (rather than subtracting) the signals across the two ears, and indeed this is so. Averaging across the two ears produces an improved vertical readout (r^2 0.73, Figure 7K). A hybrid readout operation in which the horizontal location is computed from the binaural difference, and the vertical location is computed from the binaural average, produces a modest improvement in the overall reconstruction of target location (Figure 7J). Overall, these results parallel human sound localization which relies on a binaural difference computation in the horizontal dimension (and is more accurate in that dimension), vs. potentially monaural or averaged spectral cues for the vertical dimension (which is less accurate) (Blauert, 1997; Groh, 2014).

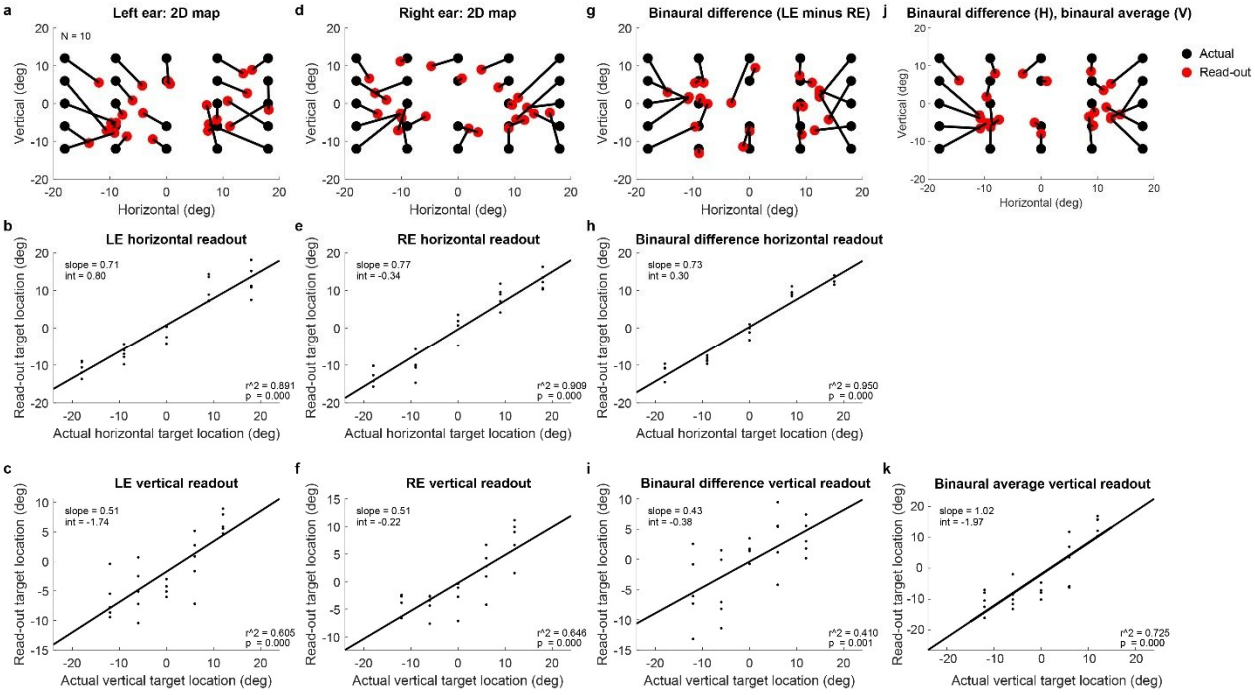


Figure 7. Multiple ways of reading out target location from the ear canal recordings. As in Figure 6 and Supplementary Figure 3, the relationship between EMREOs and eye movements was quantitatively modelled using Eq 2 and the ear canal data recorded in the horizontal/vertical task. Inferred grid task target location was "read out" by solving equation (2) for ΔH and ΔV *using the coefficients as fit from the horizontal/vertical task and the microphone values as observed in the single-origin grid task; see main text for details. a. Inferred target location (red) compared to actual target location (black), based on the left ear (same data as in Figure 6). b. Horizontal component of the read-out target vs the actual horizontal component (left ear microphone signals). c. Same as (b) but for the vertical component. d-f. Same as A-C but for the right ear. g-i, Same as a-c and d-f but computed using the binaural difference between the microphone signals*

(left ear – right ear). j., k.. A hybrid read-out model (j) using binaural difference in the horizontal dimension (h) and binaural average in the vertical dimension (k).

Discussion

Sound locations are inferred from head-centered differences in sound arrival time, loudness, and spectral content, but visual stimulus locations are inferred from eye-centered retinal locations (Blauert, 1997; Groh, 2014). Information about eye movements with respect to the head/ears is critical for connecting the visual and auditory scenes to one another (Groh and Sparks, 1992). This insight has motivated a number of previous neurophysiological studies in various brain areas in monkeys and cats, all of which showed that changes in eye position affected the auditory response properties of at least some neurons in the brain area studied (Inferior colliculus: (Groh et al., 2001; Zwiers et al., 2004; Porter et al., 2006; Bulkin and Groh, 2012a, b) ; auditory cortex: (Werner-Reiss et al., 2003; Fu et al., 2004; Maier and Groh, 2010) ; superior colliculus: (Jay and Sparks, 1984, 1987b, a; Hartline et al., 1995; Zella et al., 2001; Populin et al., 2004; Lee and Groh, 2012) ; frontal eye fields: (Russo and Bruce, 1994; Caruso et al., 2019) ; intraparietal cortex: (Stricanne et al., 1996; Cohen and Andersen, 2000; Mullette-Gillman et al., 2005, 2009)).

These findings raised the question of where signals related to eye movements first appear in the auditory processing stream. The discovery of EMREOs (Gruters et al., 2018) introduced the intriguing possibility that the computational process leading to visual-auditory integration might be manifest in the most peripheral part of the auditory system. Here we show that the signals present in the ear exhibit the properties necessary for playing a role in this process: these signals carry information about the horizontal and vertical components of eye movements, and display signatures related to both change-in-eye-position and the absolute position of the eyes in the orbits. Because of the parametric information present in the EMREO signal, we are able to predict EMREOs in one task from the eye movements recorded in another, and even predict the target of eye movements from the simultaneous EMREO recording.

Our present observations raise two key questions: what causes EMREOs and how do those mechanisms impact hearing/auditory processing? The proximate cause of

EMREOs is likely to be one or more of the known types of motor elements in the ear[1]: the middle ear muscles (stapedius and tensor tympani), which modulate the motion of the ossicles (Mendelson, 1957; Hung and Dallos, 1972; Gelfand, 1984), and the outer hair cells, which modulate the motion of the basilar membrane (Brownell et al., 1985). One or more of these elements may be driven by descending brain signals originating from within the oculomotor pathway and entering the auditory pathway somewhere along the descending stream that ultimately reaches the ear via the 5th (tensor tympani), 7th (stapedius muscle), and/or 8th nerves (outer hair cells) (see refs: Galambos, 1956; Liberman and Guinan, 1998; Cooper and Guinan, 2006; Guinan, 2006; Mukerji et al., 2010; Guinan, 2014) for reviews). Efforts are currently underway in our laboratory to identify the specific EMREO generators/modulators (Schlebusch et al., 2019, 2020).

Uncovering the underlying mechanism should shed light on another question. Does the temporal pattern of the observed EMREO signal reflect the time course and nature of that underlying mechanism's impact on auditory processing? It is not clear how an oscillatory signal like the one observed here might contribute to hearing. However, it is also not clear that the underlying mechanism is, in fact, oscillatory. Microphones can only detect signals with oscillatory energy in the range of sensitivity of the microphone. It is possible that the observed oscillations reflect ringing associated with a change in some mechanical property of the transduction system, and that change could have a non-oscillatory temporal profile (Figure 8A). Of particular interest would be a ramp-to-step profile in which aspects of the middle or inner ear shift from one state to another during the course of a saccade and hold steady at the new state during the subsequent fixation period. This kind of temporal profile would match the time course of the saccade itself.

[1] We note that EMREOs are unlikely to be due to the actual sound of the eyes moving in the orbits. Our original study, Gruters et al (2018) showed that when microphone recordings are aligned on saccade offset (as opposed to onset, as we did here), EMREOs continue for at least several 10's of ms after the eyes have stopped moving. We also have unpublished observations in patients with various hearing abnormalities; EMREOs are altered in such patients despite normal eye movements.

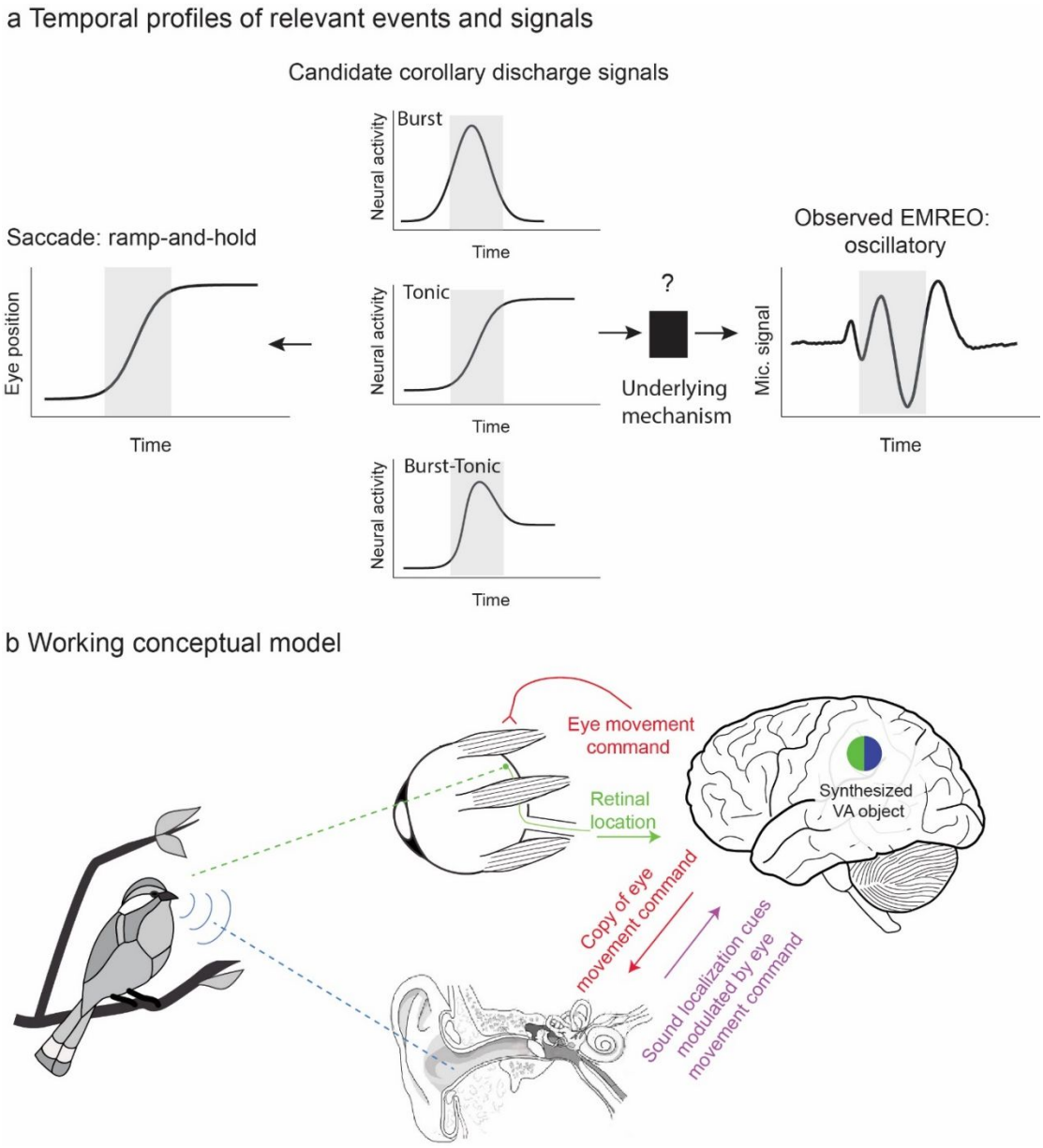


Figure 8. Temporal profiles of relevant signals and working conceptual model for how EMREOs might relate to our ability to link visual and auditory stimuli in space. A. Temporal profiles of signals. The EMREO is oscillatory whereas the eye movement to which it is synchronized involves a ramp-and-hold temporal profile. Candidate source neural signals in the brain might exhibit a ramp-and-hold (tonic) pattern, suggesting a ramp-and-hold-like underlying effect on an as-yet-unknown peripheral mechanism, or could derive from other known temporal profiles including bursts of activity time-locked to saccades. B. Working conceptual model. The brain causes the eyes to move by sending a command to the eye muscles. Each eye movement shifts the location of visual stimuli on the retinal surface. A copy, possibly a highly transformed one, of this eye movement command is sent to the ear, altering ear mechanics in some unknown way. When a sound occurs, the ascending signal to the brain will depend on the combination of its location in head-centered space (based on the physical values of

binaural timing and loudness differences and spectral cues) and aspects of recent eye movements and fixation position. This hybrid signal could then be read-out by the brain.

Available eye movement control signals in the oculomotor system include those that follow this ramp-and-hold temporal profile, or tonic activity that is proportional to eye position throughout periods of both movement and fixation. In addition to such tonic signals, oculomotor areas also contain neurons that exhibit burst patterns, or elevated discharge in association with the saccade itself, as well as combinations of burst and tonic patterns (for reviews, see Fuchs et al., 1985; Takahashi and Shinoda, 2018). It remains to be seen which of these signals or signal combinations might be sent to the auditory periphery and where they might come from. The paramedian pontine reticular formation (PPRF) is a strong candidate for a source, having been implicated in providing corollary discharge signals of eye movements in visual experiments (Sparks et al., 1987) (see also Guthrie et al., 1983), and containing each of these basic temporal signal profiles (Fuchs et al., 1985; Takahashi and Shinoda, 2018). Regardless of the source and nature of the descending corollary discharge signal, the oscillations observed here should be thought of as possibly constituting a biomarker for an underlying, currently unknown, mechanism, rather than necessarily the effect itself.

Despite these critical unknowns, it is useful to articulate a working conceptual model of how EMREOs might facilitate visual and auditory integration (Figure 8B). The general notion is that, by sending a copy of each eye movement command to the motor elements of the auditory periphery, the brain keeps the ear informed about the current orientation of the eyes. If, as noted above, these descending oculomotor signals cause a ramp-to-step change in the state of tension of components within the EMREO pathway, time-locked to the eye movement and lasting for the duration of each fixation period, they would effectively change the transduction mechanism in an eye position/eye movement dependent fashion. In turn, these changes could affect the latency, gain, or frequency-filtering properties of the response to sound. Indeed, intriguing findings from Puria and colleagues (Cho et al., in revision) have recently indicated that the tension applied by the middle ear muscles likely affects all three of these aspects of sound transmission throughout the middle ear. In short, the signal sent to the brain in response to an incoming sound could ultimately reflect a mixture of

459 the physical cues related to the location of the sound itself - the ITD/ILD/spectral cues -
460 and eye position/movement information.

461 Most neurophysiological studies report signals consistent with a hybrid code in
462 which information about sound location is blended in a complex fashion with information
463 about eye position and movement, both within and across neurons (Jay and Sparks,
464 1984, 1987b; Groh et al., 2001; Porter et al., 2006; Maier and Groh, 2010; Lee and
465 Groh, 2012; Caruso et al., 2019). Computational modeling confirms that, in principle,
466 these complex signals can be "read out" to produce a signal of sound location with
467 respect to the eyes (Groh et al., 2001). However, substantive differences do remain
468 between the observations here and such neural studies, chiefly in that the neural
469 investigations have focused primarily on periods of steady fixation. A more complete
470 characterization of neural signals time-locked to saccades is therefore needed (Porter et
471 al., 2007; Bulkin and Groh, 2012a).

472 Note that this working model differs from a spatial attention mechanism in which
473 the brain might direct the ears to "listen" selectively to a particular location in space.
474 Rather, under our working model, the response to sounds from any location would be
475 impacted by peripheral eye movement/position dependence in a consistent fashion
476 across all sound locations. However, such a system could well work in concert with
477 top-down attention, which has previously been shown to impact outer hair cells even
478 when participants are required to fixate and not make eye movements (Delano et al.,
479 2007; Harkrider and Bowers, 2009; Srinivasan et al., 2012; Srinivasan et al., 2014;
480 Walsh et al., 2014; Wittekindt et al., 2014; Walsh et al., 2015).

481 Another question concerns whether EMREOs might actually impair sound
482 localization, specifically for brief sounds presented during an eye movement. We think
483 the answer to this is no. Boucher et al (Boucher et al., 2001) reported that
484 perisaccadic sound localization is quite accurate, which suggests that EMREOs (or their
485 underlying mechanism) do not impair perception. This is an important insight because
486 given the rate at which eye movements occur - about 3/sec – and with each associated
487 EMREO signal lasting 100 ms or longer (due to extending past the end of saccades, as
488 explored by Gruters, Murphy et al. 2018), it would be highly problematic if sounds could
489 not be accurately detected or localized when they occur in conjunction with saccades.

If there is indeed a step-ramp system underlying the observed oscillations, then transduction of all sounds will be affected, regardless of when they occur with respect to saccades.

Overall, how brain-controlled mechanisms adjust the signaling properties of peripheral sensory structures is critical for understanding sensory processing as a whole. Auditory signals are known to adjust the sensitivity of the visual system via sound-triggered pupil dilation (Bala and Takahashi, 2000), indicating that communication between these two senses is likely to be a two-way street. The functional impact of such communication at low-level stages is yet to be fully explored and may have implications for how individuals compensate when the information from one sensory system is inadequate, either due to natural situations such as noisy sound environments or occluded visual ones, or due to physiological impairments in one or more sensory systems.

Methods

General

Healthy human subjects that were 18 years of age or older with no known hearing deficits or visual impairments beyond corrected vision were recruited from the surrounding campus community (N=16; 8 female, 8 male; female-male ratio was also equal in subgroups tested on different tasks). If subjects were unable to perform the saccade task without vision correction, they were excluded from the study. All study procedures involving subjects were approved by the Duke Institutional Review Board, and all subjects received monetary compensation for their participation.

Acoustic signals in both ear canals were measured simultaneously with Etymotic ER10B+ microphone systems coupled with ER2 earphones to allow calibrations of the microphones (However, note that no auditory stimuli were used during any of the saccade tasks in the current study) (Etymotic Research, Elk Grove Village, IL). A low-latency audio interface (Focusrite Scarlett 2i2, Focusrite Audio Engineering Ltd., High Wycombe, UK) was used for audio capture and playback through the Etymotic hardware at a sampling rate of 48kHz. Eye tracking was performed with an Eyelink 1000 system sampling at 1000Hz. Stimulus presentation and data acquisition were

controlled through custom scripts and elements of The Psychophysics Toolbox in MATLAB, with visual stimuli presented on a large LED monitor.

In all experiments, eye position and microphone data were recorded while subjects performed silent, visually-guided saccade tasks. Experimental sessions were carried out in a darkened, acoustically isolated chamber made anechoic with the addition of acoustic damping wall tiles. Subjects were seated 70 cm from the screen, and a chin rest was used to maintain head position and minimize movement. Experimental sessions were subdivided into multiple runs, approximately 5 minutes each. This provided subjects with the opportunity to take a brief break from the experiment if needed to maintain alertness or to address any possible discomfort from maintaining their posture. Each run typically consisted of approximately 125 trials and fixation positions and saccade targets were presented in pseudorandom order.

Before each experimental session, the eye-tracking system was calibrated using the calibration routine provided with the Eyelink system to register raw eye-tracking data to gaze locations on the stimulus presentation screen. If the subject requested an adjustment to the chin rest or left the recording chamber for a break, the eye-tracking calibration was repeated. Before each run, the microphone system was calibrated to ensure that each microphone had a frequency response that was similar to the pre-recorded frequency response of the microphones when placed in a volume that approximated the size and geometry of the human ear canal - a 3ml syringe cut to accept the Etymotic earpieces. The syringe stopper was pulled to 1.25 cm^3 to approximate the volume of the average adult human ear canal. A small amount of gauze (.25cm^3) was added to the volume to emulate the attenuation caused by the soft tissue of the ear canal. The calibration routine played tones from 10 to 1600Hz, at a constant system output amplitude. As the purpose of this calibration was to compare microphone function in a control volume with that in an earpiece just placed in a subject, the weighting of the tones was not specifically calibrated. If the input-output results of the same tone sequences were consistent between ears and matched the overall shape of the syringe calibration curves, microphone placement was considered successful. No sounds were delivered during periods of experimental data collection.

Task descriptions

All tasks followed the same stimulus timing sequence: initial fixation points were displayed on screen for 750ms and then removed as the saccade targets were presented for 750ms (Figure 1A). Fixation and target locations were indicated by green dots. Subjects were instructed to fixate on the initial fixation locations until targets were presented on the screen, then to saccade to the targets and fixate on the targets until they changed from green to red for the last 100ms of the target presentation (the color cue was intended to help subjects maintain fixation through the end of the target presentation). Inter-trial-intervals were jittered 350±150ms. This was done to minimize the potential impact of non-saccade related noise signals that may be periodic (i.e. heartbeat, external acoustic and electromagnetic sources).

In the five-origin grid task (Figure 1B), participants performed saccades to multiple targets from five different initial eye positions in a plus-shaped configuration at -9º, 0º, and +9º horizontally and at -6 º, 0º, and 6 º of elevation as shown. Twenty five saccade targets ranged from -18º to +18º in 9º degree increments horizontally and from -12º to +12 º in 6 º increments vertically.

In the horizontal/vertical task (Figure 1D), participants performed saccades to targets along the vertical and horizontal axes from a central fixation. Vertical targets ranged from -12º to +12º in 3º increments and horizontal targets ranged from -18º to +18º in 3º increments.

In the single-origin grid task (Figure 1C), participants made saccades to 24 distinct targets of varying vertical and horizontal placement combinations from a central fixation. Horizontal location components ranged from -18º to +18º in 9º increments and vertical location components ranged from -12º to +12º in 6º increments.

Preprocessing analysis

Saccade-microphone synchronization:

Microphone data was synchronized to the onset the saccade from the fixation point to the target. This was defined based on the third derivative of eye position, or jerk. The first peak in the jerk represents the moment when the change in the eye acceleration is greatest. Prior to each differentiation, a lowpass discrete filter with a 7ms window was

used to smooth the data and reduce the effects of noise and quantization error. This filter was normalized, such that its output to a constant series of values equaled those values.

Trial exclusion criteria:

Trials were excluded based on saccade performance and the quality of microphone recordings. Exclusion criteria used for eye tracking: 1) if subjects made a sequence of two or more saccades to achieve the target; 2) if the eye tracking signal dropped out during the trial (e.g. due to blinks); 3) if the eye movement was slow and drifting, rather than a saccade; 4) if the saccade curved by more than 4.5º (subtended angle); or 5) subjects failed to maintain 200ms of fixation before and after the saccade. 6) If eye tracking dropped samples that prevented the calculation of the saccade onset time. On average these saccade-related exclusion criteria resulted in the exclusion of about 12% of the trials.

Prior to any further analysis, microphone data was downsampled from 48 kHz to 2 kHz sampling rate to reduce processing time given that the previously observed eye-movement related signals of interest are well below 100 Hz (Gruters et al., 2018). Exclusion based on noise in the microphone recordings was minimal. Within each block of trials, the mean and standard deviation of the RMS values for each trial was calculated. Individual trials were excluded if the microphone signal on that trial contained any individual values that were more than 10 standard deviations away from that mean. This typically resulted in the exclusion of < ~2% of the trials, after application of the eye position screen described above.

Z scoring

To facilitate comparison across subjects, sessions, and experiments, all microphone data reported in this study was z-scored within blocks and prior to the application of the exclusion criteria described above. The mean and standard deviation of the microphone values in a window -150 to -120 ms prior to saccade onset were used as the normalization baseline period.

614 Regression Analyses

615 Regression was used to assess how EMREOs vary with both eye position and
616 eye movement. The microphone signal at each moment in time *Mic(t)* was fit as
617 follows:

618 $$Mic(t) = B_H(t)\ H + B_{\Delta H}(t)\ \Delta H + B_V(t)\ V + B_{\Delta V}(t)\ \Delta V + C(t) \quad (1)$$

619 where H and V correspond to the initial horizontal and vertical eye position and ΔH and
620 ΔV correspond to the respective changes in position associated with that trial. The
621 slope coefficients B_H, $B_{\Delta H}$, B_V, $B_{\Delta V}$ are time-varying and reflect the dependence of the
622 microphone signal on the respective eye movement parameters. The term $C(t)$
623 contributes a time-varying "constant" independent of eye movement metrics, and can be
624 thought of as the best fitting average oscillation across all eye positions and
625 displacements.
626 The term $C(t)$ was included for all regressions, but other parameters were
627 omitted when not relevant. Specifically, for the single-origin grid tasks and horizontal-
628 vertical tasks, the model used vertical and horizontal saccade displacement ($B_{\Delta H}(t)\ \Delta H$,
629 $B_{\Delta V}(t)\ \Delta V$) as regression variables but not $B_H(t)\ H$ or $B_V(t)\ V$ as initial position did not
630 vary for those tasks. The analysis produced values for the intercept and variable
631 weights (or slopes), their 95% confidence intervals, R^2, and p-value for each time point.
632 For most analyses, the measured eye positions/changes-in-eye positions were
633 used as the independent variables, so as to incorporate any variability due to scatter in
634 fixation or saccade endpoint. For the target readout analysis described in Figures 6 and
635 7, the horizontal and vertical positions of the targets, rather than the associated eye
636 movements, were used.

637 # References

Bala AD, Takahashi TT (2000) Pupillary dilation response as an indicator of auditory discrimination in the barn owl. J Comp Physiol [A] 186:425-434.

Barczak A, Haegens S, Ross DA, McGinnis T, Lakatos P, Schroeder CEJCr (2019) Dynamic modulation of cortical excitability during visual active sensing. 27:3447-3459. e3443.

Blauert J (1997) Spatial hearing. Cambridge, MA: MIT Press.

Boucher L, Groh JM, Hughes HC (2001) Afferent delays and the mislocalization of perisaccadic stimuli. Vision Research 41:2631-2644.

Brownell WE, Bader CR, Bertrand D, de Ribaupierre Y (1985) Evoked mechanical responses of isolated cochlear outer hair cells. Science 227:194-196.

Bulkin DA, Groh JM (2012a) Distribution of eye position information in the monkey inferior colliculus. J Neurophysiol 107:785-795.

Bulkin DA, Groh JM (2012b) Distribution of visual and saccade related information in the monkey inferior colliculus. Front Neural Circuits 6:61.

Caruso VC, Pages DS, Sommer MA, Groh JM (2019) Compensating for a shifting world: A quantitative comparison of the reference frame of visual and auditory signals across three multimodal brain areas. bioRxiv:669333.

Caruso VC, Pages DS, Sommer MA, Groh JM (2021) Compensating for a shifting world: evolving reference frames of visual and auditory signals across three multimodal brain areas. J Neurophysiol 126:82-94.

Cho NH, Ravicz ME, Puria S (in revision) Human Middle-Ear Muscle Pulls Change Tympanic-Membrane Shape and Low-Frequency Middle-Ear Transmission Delays and Magnitudes. Hear Res in revision.

Cohen YE, Andersen RA (2000) Reaches to sounds encoded in an eye-centered reference frame. Neuron 27:647-652.

Cooper NP, Guinan JJ, Jr. (2006) Efferent-mediated control of basilar membrane motion. J Physiol 576:49-54.

Delano PH, Elgueda D, Hamame CM, Robles L (2007) Selective attention to visual stimuli reduces cochlear sensitivity in chinchillas. J Neurosci 27:4146-4153.

Fu KM, Shah AS, O'Connell MN, McGinnis T, Eckholdt H, Lakatos P, Smiley J, Schroeder CE (2004) Timing and laminar profile of eye-position effects on auditory responses in primate auditory cortex. J Neurophysiol 92:3522-3531.

Fuchs AF, Kaneko CR, Scudder CA (1985) Brainstem control of saccadic eye movements. Annu Rev Neurosci 8:307-337.

Galambos R (1956) Suppression of auditory nerve activity by stimulation of efferent fibers to cochlea. J Neurophysiol 19:424-437.

Gallagher L, Diop M, Olson ES (2021) Time-domain and frequency-domain effects of tensor tympani contraction on middle ear sound transmission in gerbil. Hear Res 405:108231.

Gelfand SA (1984) The contralateral acoustic reflex. . In: The acoustic reflex: Basic principles and clinical applications pp 137-186. New York, NY: Academic Press.

Groh JM (2014) Making space: how the brain knows where things are. Cambridge, MA: Harvard University Press.

Groh JM, Sparks DL (1992) Two models for transforming auditory signals from head-centered to eye- centered coordinates. Biological Cybernetics 67:291-302.

Groh JM, Trause AS, Underhill AM, Clark KR, Inati S (2001) Eye position influences auditory responses in primate inferior colliculus. Neuron 29:509-518.

Gruters KG, Murphy DLK, Jenson CD, Smith DW, Shera CA, Groh JM (2018) The eardrums move when the eyes move: A multisensory effect on the mechanics of hearing. Proc Natl Acad Sci U S A 115 E1309-E1318.

Guinan JJ, Jr. (2006) Olivocochlear efferents: anatomy, physiology, function, and the measurement of efferent effects in humans. Ear Hear 27:589-607.

Guinan JJ, Jr. (2014) Cochlear mechanics, otacoustic emisssions, and medial olivocochlear efferents: Twenty years of advances and controversies along with areas ripe for new work. In: Perspectives on Auditory Research, pp 229-246. New York: Springer.

Guthrie BL, Porter JD, Sparks DL (1983) Corollary discharge provides accurate eye position information to the oculomotor system. Science 221:1193-1195.

Harkrider AW, Bowers CD (2009) Evidence for a cortically mediated release from inhibition in the human cochlea. J Am Acad Audiol 20:208-215.

Hartline PH, Vimal RL, King AJ, Kurylo DD, Northmore DP (1995) Effects of eye position on auditory localization and neural representation of space in superior colliculus of cats. Exp Brain Res 104:402-408.

Hebrank J, Wright D (1974a) Are two ears necessary for localization of sound sources on the median plane? J Acoust Soc Am 56:935-938.
Hebrank J, Wright D (1974b) Spectral cues used in the localization of sound sources on the median plane. J Acoust Soc Am 56:1829-1834.
Hung IJ, Dallos P (1972) Study of the acoustic reflex in human beings. I. Dynamic characteristics. J Acoust Soc Am 52:1168-1180.
Jay MF, Sparks DL (1984) Auditory receptive fields in primate superior colliculus shift with changes in eye position. Nature 309:345-347.
Jay MF, Sparks DL (1987a) Sensorimotor integration in the primate superior colliculus. I. Motor convergence. J Neurophysiol 57:22-34.
Jay MF, Sparks DL (1987b) Sensorimotor integration in the primate superior colliculus. II. Coordinates of auditory signals. J Neurophysiol 57:35-55.
Lee J, Groh JM (2012) Auditory signals evolve from hybrid- to eye-centered coordinates in the primate superior colliculus. J Neurophysiol 108:227-242.
Liberman MC, Guinan JJ, Jr. (1998) Feedback control of the auditory periphery: anti-masking effects of middle ear muscles vs. olivocochlear efferents. J Commun Disord 31:471-482; quiz 483; 553.
Macpherson EA, Sabin AT (2013) Vertical-plane sound localization with distorted spectral cues. Hear Res 306:76-92.
Maier JX, Groh JM (2010) Comparison of gain-like properties of eye position signals in inferior colliculus versus auditory cortex of primates. Frontiers in Integrative Neuroscience 4:121-132.
Mendelson ES (1957) A sensitive method for registration of human intratympanic muscle reflexes. J Appl Physiol 11:499-502.
Metzger RR, Mullette-Gillman OA, Underhill AM, Cohen YE, Groh JM (2004) Auditory saccades from different eye positions in the monkey: implications for coordinate transformations. J Neurophysiol 92:2622-2627.
Middlebrooks JC, Green DM (1991) Sound localization by human listeners. Annu Rev Psychol 42:135-159.
Mukerji S, Windsor AM, Lee DJ (2010) Auditory brainstem circuits that mediate the middle ear muscle reflex. Trends Amplif 14:170-191.
Mullette-Gillman OA, Cohen YE, Groh JM (2005) Eye-centered, head-centered, and complex coding of visual and auditory targets in the intraparietal sulcus. J Neurophysiol 94:2331-2352.
Mullette-Gillman OA, Cohen YE, Groh JM (2009) Motor-related signals in the intraparietal cortex encode locations in a hybrid, rather than eye-centered, reference frame. Cerebral Cortex 19:1761-1775.
O'Connell MN, Barczak A, McGinnis T, Mackin K, Mowery T, Schroeder CE, Lakatos PJI (2020) The role of motor and environmental visual rhythms in structuring auditory cortical excitability. 23:101374.
Populin LC, Tollin DJ, Yin TC (2004) Effect of eye position on saccades and neuronal responses to acoustic stimuli in the superior colliculus of the behaving cat. J Neurophysiol 92:2151-2167.
Porter KK, Metzger RR, Groh JM (2006) Representation of eye position in primate inferior colliculus. J Neurophysiol 95:1826-1842.
Porter KK, Metzger RR, Groh JM (2007) Visual- and saccade-related signals in the primate inferior colliculus. Proceedings of the National Academy of Sciences of the United States of America 104:17855-17860.
Russo GS, Bruce CJ (1994) Frontal eye field activity preceding aurally guided saccades. J Neurophysiol 71(3):1250-1253.
Schlebusch SN, Cooper MW, Kaylie DM, King CD, Murphy DLK, Shera CA, Groh JM (2019) Changes in saccade-related eardrum oscillations after surgical denervation of the stapedius muscle. Soc Neurosci Abstr.
Schlebusch SN, Cooper MW, Kaylie DM, King CD, Murphy DLK, Shera CA, Groh JM (2020) Changes in saccade-related eardrum oscillations after surgical denervation of the stapedius muscle. . Association for Research in Otolaryngology Abstracts.
Sparks DL, Mays LE, Porter JD (1987) Eye movements induced by pontine stimulation: interaction with visually triggered saccades. J Neurophysiol 58:300-318.

Srinivasan S, Keil A, Stratis K, Woodruff Carr KL, Smith DW (2012) Effects of cross-modal selective attention on the sensory periphery: cochlear sensitivity is altered by selective attention. Neuroscience 223:325-332.

Srinivasan S, Keil A, Stratis K, Osborne AF, Cerwonka C, Wong J, Rieger BL, Polcz V, Smith DW (2014) Interaural attention modulates outer hair cell function. Eur J Neurosci 40:3785-3792.

Stricanne B, Andersen RA, Mazzoni P (1996) Eye-centered, head-centered, and intermediate coding of remembered sound locations in area LIP. J Neurophysiol 76:2071-2076.

Takahashi M, Shinoda Y (2018) Brain Stem Neural Circuits of Horizontal and Vertical Saccade Systems and their Frame of Reference. Neuroscience 392:281-328.

Walsh KP, Pasanen EG, McFadden D (2014) Selective attention reduces physiological noise in the external ear canals of humans. I: auditory attention. Hear Res 312:143-159.

Walsh KP, Pasanen EG, McFadden D (2015) Changes in otoacoustic emissions during selective auditory and visual attention. J Acoust Soc Am 137:2737-2757.

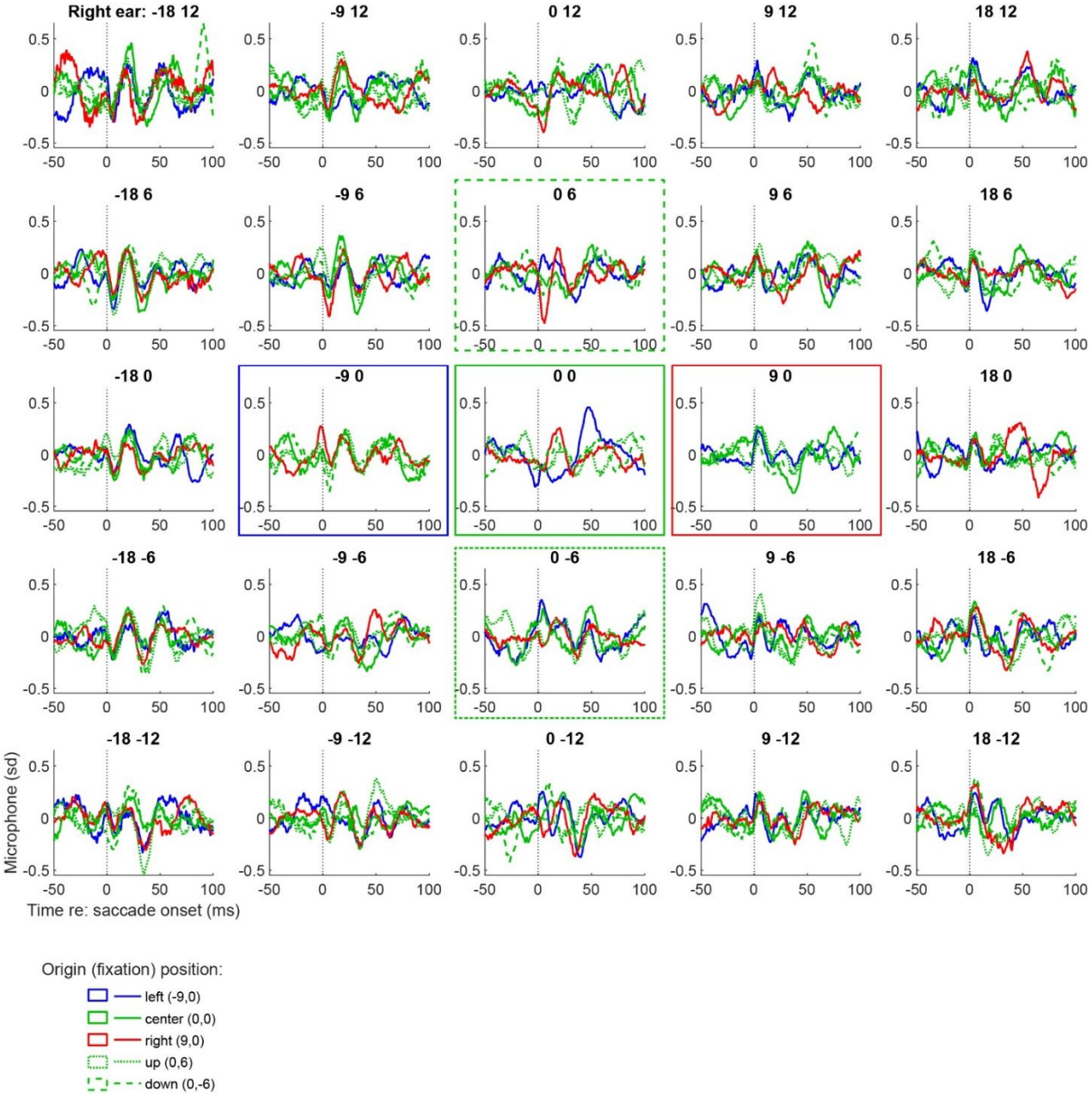
Werner-Reiss U, Kelly KA, Trause AS, Underhill AM, Groh JM (2003) Eye position affects activity in primary auditory cortex of primates. Current Biology 13:554-562.

Wittekindt A, Kaiser J, Abel C (2014) Attentional modulation of the inner ear: a combined otoacoustic emission and EEG study. J Neurosci 34:9995-10002.

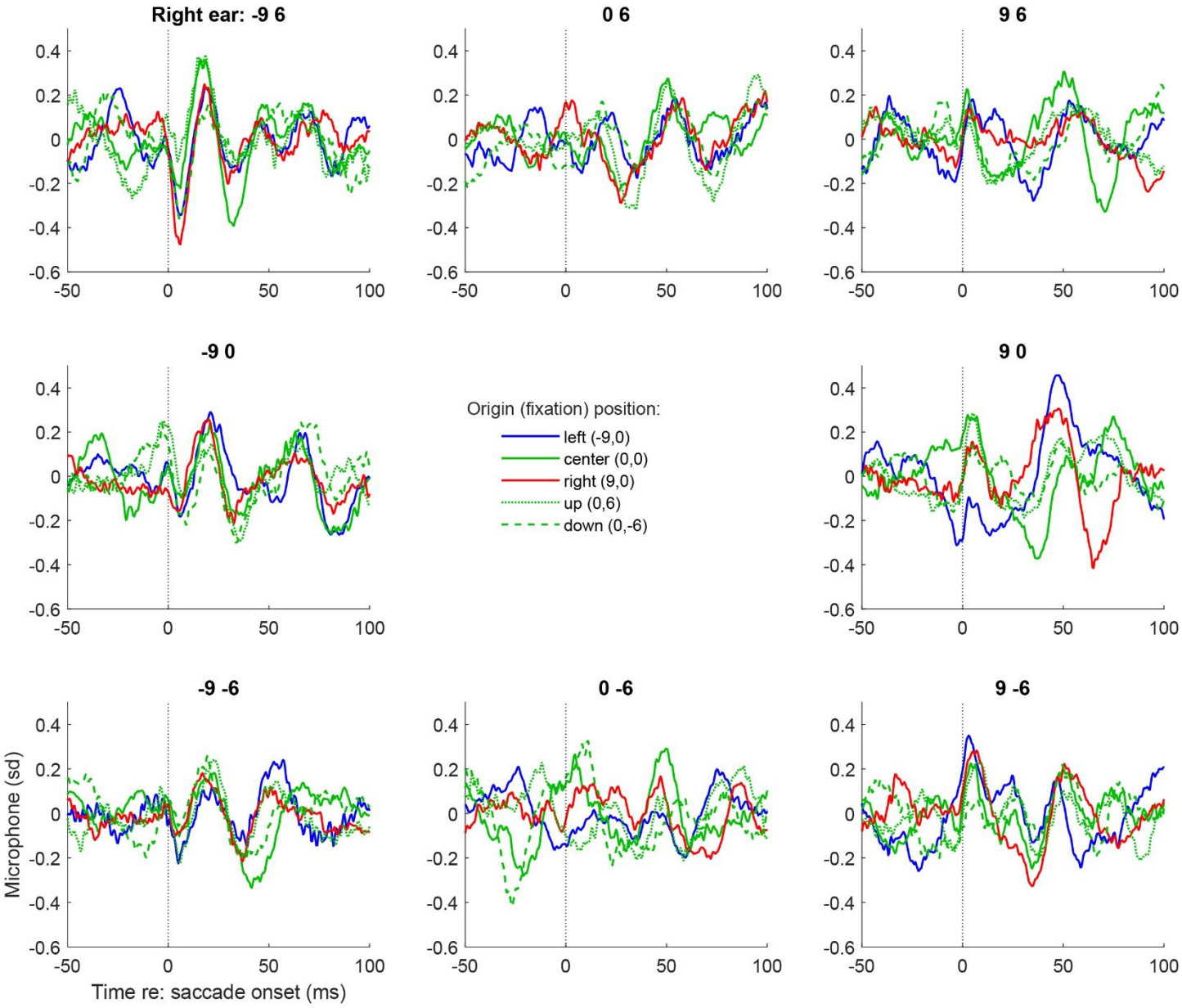
Zella JC, Brugge JF, Schnupp JW (2001) Passive eye displacement alters auditory spatial receptive fields of cat superior colliculus neurons. Nat Neurosci 4:1167-1169.

Zwiers MP, Versnel H, Van Opstal AJ (2004) Involvement of monkey inferior colliculus in spatial hearing. J Neurosci 24:4145-4156.

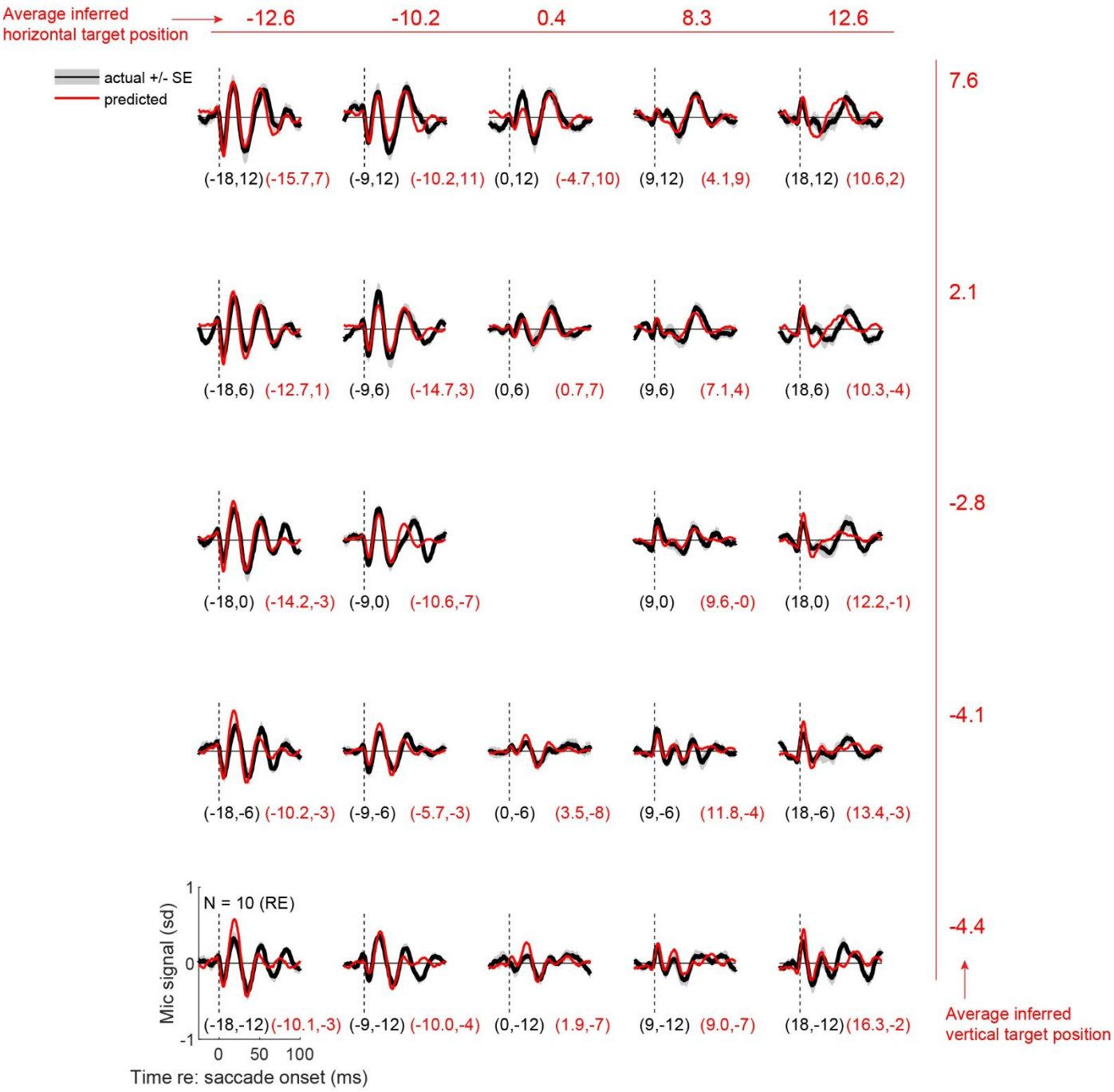
Supplementary Figures



Supplementary Figure 1. *Grand average EMREOs recorded during the 5-origin grid task in right ears of ten subjects. Same format as Figure 2: Each panel is the grand average EMREO signal (average of the individual subject averages) that is generated when a saccade is made to that location on the screen e.g. the top right panel involves saccades to the top right target location. The color and line styles of each trace correspond to the initial fixation point as indicated by boxes of the same color and line style; e.g. all red oscillations are generated during a simultaneous saccade that originated from the right fixation point.*



Supplementary Figure 2. *Grand average EMREOs as a function of target location with respect to the fixation position, for N=10 right ears. Same format as Figure 3. The data shown are the same as (a subset of) those shown in Supplementary Figure 1, but here each panel location corresponds to a particular target location defined with respect to the fixation point. The color/linestyle indicate the associated fixation position, as in Supplementary Figure 1.*



Supplementary Figure 3. Same as Figure 6, but for right ear data. See main text for details.