Dynamic sound localization in cats

Janet L. Ruhland, Amy E. Jones, and Tom C. T. Yin

Department of Neuroscience and Neuroscience Training Program, University of Wisconsin, Madison, Wisconsin

Submitted 3 February 2015; accepted in final form 5 June 2015

Ruhland JL, Jones AE, Yin TC. Dynamic sound localization in cats. J Neurophysiol 114: 958–968, 2015. First published June 10, 2015; doi:10.1152/jn.00105.2015.—Sound localization in cats and humans relies on head-centered acoustic cues. Studies have shown that humans are able to localize sounds during rapid head movements that are directed toward the target or other objects of interest. We studied whether cats are able to utilize similar dynamic acoustic cues to localize acoustic targets delivered during rapid eye-head gaze shifts. We trained cats with visual-auditory two-step tasks in which we presented a brief sound burst during saccadic eye-head gaze shifts toward a prior visual target. No consistent or significant differences in accuracy or precision were found between this dynamic task (2-step saccade) and the comparable static task (single saccade when the head is stable) in either horizontal or vertical direction. Cats appear to be able to process dynamic auditory cues and execute complex motor adjustments to accurately localize auditory targets during rapid eye-head gaze shifts.

SOUND LOCALIZATION along the horizontal, or azimuthal, plane requires binaural processing of interaural time and level differences (ITDs and ILDs, respectively) of the incoming acoustic signals. ITDs can be computed from pure tone frequencies up to ~1,200–1,500 Hz via phase locking. Temporal features can also be extracted from complex sounds of higher carrier frequency such as bursts of narrowband noise and sinusoidally amplitude-modulated (SAM) tones if the modulating frequency is not too high. Localizing sounds in elevation relies on the broadband spectral shapes of head-related transfer functions (HRTFs) that result from the direction-dependent filtering properties of the head and pinnae (Tollin and Yin 2009). Both sets of cues are based on the position of the head relative to the sound source.

Human psychophysical studies show that accurate sound localization can occur in situations where the body, head, and ears are moving during sound presentation. Normal hearing listeners typically orient both their head and their gaze toward an auditory target (Fuller 1992; Thurlow et al. 1967). Studies have shown how head movement could be used as a strategy to facilitate sound localization ability, especially in compromised situations where the sound source contains limited frequencies, the perception of the sound location is ambiguous, or the hearing of the listener is impaired (Middlebrooks and Green 1991; Perrett and Noble 1997a, 1997b; Pollack and Rose 1967; Thurlow and Runge 1967; Wallach 1939, 1940; Wightman and Kistler 1998). Wallach (1939) observed that a given ITD or ILD could arise from many points along a “cone of confusion” and that such ambiguities could be reduced by head turns.

Lambert (1974) extended this theory to include using multiple interaural auditory samples while turning the head to accurately determine the distance in depth of a sound source, which can also be ambiguous.

In the conditions described above, the head movement is typically in the direction of the sound of interest. There may also be situations when a listener needs to localize a sound while his/her head is moving toward a different target or simply scanning the environment. The velocity of movement may be high and the direction of attention different or even opposite to the direction of the auditory target to be localized. For gaze movements to sound sources to remain accurate despite the intervention of head and eye-in-head movements, the acoustic cues must be combined with ongoing information related to changes in head and eye-in-head position. The listener needs to keep track of movements in progress when the sound is presented, so that even if the head moves further after the end of the sound a correct localization of the sound is possible. Vliegen et al. (2004) have shown that humans are able to use dynamic acoustic cues and accurately localize sounds delivered during rapid eye-head gaze movements. Apparently, the human auditory system is capable of utilizing such cues, and the oculomotor system is capable of accurately issuing appropriate motor commands despite ensuing head and eye movements.

To our knowledge, the above issues have not previously been addressed in laboratory animals. We studied whether cats are able to utilize similar varying acoustic cues during rapid eye-head gaze shifts. The aim of our study was to compare the ability of cats to localize a sound presented when their head and eyes are moving at high velocity with localization under head-stable conditions. A notable difference between cats and humans is that cats have mobile pinnae, which also rapidly orient to sound sources (Populin and Yin 1998b; Tollin et al. 2009). We hypothesize how this pinna movement might assist accurate sound localization during head movement.

METHODS

Subjects and surgery. All surgical and experimental procedures were reviewed and approved by the University of Wisconsin Animal Care and Use Committee and complied with the guidelines of the National Institutes of Health. Many of our methods and materials have been described previously (Populin and Yin 1998a; Tollin et al. 2005). In four deeply anesthetized adult cats, we implanted a stainless steel head post and fine wire coils (AS632, Cooner Wire, Chatsworth, CA, or S170012A7-FEP, Alan Baird Industries, Ho-Ho-Kus, NJ) on the head and in each eye and ear under aseptic surgical conditions. The head coil was imbedded in the coronal plane in the dental acrylic of the head cap. Coils were placed subcutaneously on the caudal-dorsal aspect of each ear to monitor ear position. Anesthesia was induced with an intramuscular injection of ketamine (20 mg/kg) and maintained throughout the surgery by inhalation of isoflurane (1–2% in 1 l/min O2) via a tracheal cannula. Postoperative analgesia was provided...
were saved to disk by sampling at 500 Hz. Analog outputs of the coil systems (CNC Engineering, Seattle, WA) were covered with sound-absorbing acoustic foam (10.2 cm, Sonex, Ilbruck, Minneapolis, MN) to minimize acoustic reflections. The magnetic search coil technique (Fuchs and Robinson 1966) was used to measure the positions of the eyes, head, and ears, and the analog outputs of the coil systems (CNC Engineering, Seattle, WA) were saved to disk by sampling at 500 Hz.

Targets in these experiments consisted of acoustic or visual stimuli presented from 1 of 19 different locations in the frontal hemisphere distributed along two arcs of 80-cm radius on the horizontal and vertical meridians or at four diagonal locations ±20° in azimuth and elevation from the origin (0°, 0°). Visual stimuli consisted of a 2.0-mm-diameter red (λmax = 635 nm) LED located at the center of each speaker. Acoustic stimuli were delivered from Morel Acoustics speakers (model MDT20) with matched frequency-response characteristics. The speakers themselves were hidden from view behind a black translucent cloth through which illuminated LEDs could be easily seen and sounds heard. The acoustic stimuli were generated digitally with a Tucker-Davis (Alachua, FL) stimulus presentation system and custom-written MATLAB software. Acoustic stimuli consisted of 25-ms broadband noise (~1.5–25 kHz) with 7-ms rise/fall ramps.

During initial training, the heads of the cats were restrained in the center of the coils comprising the magnetic search coil system (Populin and Yin 1998a). After the cats learned the task, the heads were freed but a body restraint helped to maintain the position of the head within the center of the coil system. All aspects of the experiments including selection of the visual or acoustic stimuli, the location of the target speaker and/or LED, the acquisition of the eye position, determination and delivery of reward, etc. were under computer control.

Calibrations. Eye (Populin and Yin 1998a), pinna (Populin and Yin 1998b), and head (Tollin et al. 2005) coils were calibrated as previously described, relying on the cats’ instinct to look at the LEDs when they were suddenly illuminated in the darkened chamber. The head coil was calibrated by mounting a laser pointer on the head and positioning the laser so that it pointed to the speaker at the origin when the cat was fixating at that point. The laser was then adjusted so that the laser (and head) pointed to each of the speakers on the vertical and horizontal meridians while monitoring the output of the head coil.

Calibration of the ear coils was more problematic, since there were no behavioral constraints and the external ear has more degrees of freedom than the eyeball. We exploited the consistent behavior of the cat to bring its pinnae to a “ready” position anticipating the LED located straight ahead as the trial was about to begin (May and Huang 1996; Populin and Yin 1998b). While the cat was working in the chamber, we carefully placed a coil made of malleable copper parallel to the orientation of the ear coil during the time that the cat was fixating straight ahead. The cat was then removed from the chamber, and a coil identical to the one in the ear was placed at the position of the middle of the head in the same orientation. The coil was then rotated in yaw and pitch through 10° increments while the horizontal and vertical components of movement were measured.

Psychophysical procedure and training. All cats were trained with operant conditioning for food reward. They were automatically rewarded under computer control if they maintained their eye position within the square acceptance window centered on the target location for a period of time, typically 650 ms. Acceptance windows were set as described previously (Populin and Yin 1998a).

To determine a subject’s baseline localization behavior, a single-step saccade behavioral task was utilized. Here, the cat was initially required to fixate an LED presented from straight ahead (0°, 0°) and maintain gaze fixation within the acceptance window for a variable period of time. If the cat satisfied this initial fixation condition, then simultaneously the fixation LED was extinguished and an acoustic or visual target was presented from 1 of the 19 locations within ±40°. The cat was then required to make a gaze saccade to the perceived location of the target and maintain fixation within the specified acceptance window for another 600–1,000 ms to receive a food reward.

The experimental task consisted of a dynamic visual-auditory double-step saccade that began like the single-step task except that the initial fixation LED could be at any location (Fig. 1A). Cats were required to localize the fixation LED and then make a gaze shift toward a visual target. During the time the head and eyes were moving toward the visual target, a 25-ms auditory target was presented from a different location. The cat was then required to redirect its gaze to the location of the auditory target and maintain fixation for another 600–1,000 ms. Since the duration and peak velocities of gaze saccades vary with saccade amplitude, we varied the timing of auditory target presentation between 5 and 150 ms after initiation of the saccade to the visual target to attempt to present the sound during maximal head velocity. The timing was set empirically by examining each cat’s head velocity profile for gaze saccades in different directions and amplitudes and adjusted as necessary.

Initial fixation targets for the double-step saccade task were within ±20° of the origin. For three of the four cats the visual targets were located such that the head was primarily moving horizontally (HH) or vertically (HV) during sound presentation, while for the other cat only HH targets were used. The placement of the auditory target could require that the auditory saccade continue in the same direction as the first saccade or change or reverse directions compared with the initial saccade (Fig. 1B).

Various other visual and auditory trial types and durations with initial fixations within ±20° of the origin (0°, 0°) were randomly interleaved with the single-step and double-step saccade trials in order to avoid anticipation of a certain trial type.

Data analysis. One key dependent variable in this experiment was the final horizontal and vertical gaze position at the completion of the saccadic shift to the apparent location of the target. For all trials, we

![Fig. 1A](http://jn.physiology.org/doi/fig/10.1152/jn.00105.2015)

Fig. 1A: example of the 2-step HH (head moving horizontally) task. A fixation target (green) at −20° is presented at time 0. The cat makes a gaze movement (black) toward the fixation target and holds for ~600 ms. Then, a visual target (blue) is presented. During the saccade to the visual target, an auditory target (orange) is briefly presented for 25 ms. The aim of the task is to present the sound during a time of high gaze movement, indicated by double dashed vertical lines (the head shift is not represented in A). HV trials are similar except that the head movement is primarily vertical. B: example of the target locations and trajectory of head movements during a dynamic double-saccade task. Locations of the fixation (green), visual (blue), and auditory (orange) targets are represented by the colored squares. The brief auditory target is presented (yellow shading) during the head movement from the fixation to the visual target. Dashed black arrow represents TH, ini, the sound location in head-centered coordinates (at the time of sound onset); dashed red arrow represents ΔG1, the gaze shift from time of sound onset to the beginning of the final gaze shift (black arrow) ΔG2 from near the visual target to the auditory target.
used a velocity criterion to determine the “end of fixation,” or when gaze movements began, by determining the time at which the magnitude of the velocity trace exceeded 2 standard deviations (SDs) from the mean velocity computed during the initial fixation (Populin and Yin 1998a). During these times, the gaze was expected to be nearly constant and the velocity close to zero. The final gaze position was determined by the position at the time of “return to fixation,” which was computed as the time at which the magnitude of the velocity trace returned to within 2 SD of the baseline mean velocity.

Peak velocity and total movement amplitude of the head in space and the eye-in-head during the 25-ms sound presentation were also determined with custom software.

For the single-step trials, if corrective movements were made within 200 ms of the end of the initial saccade the final position was determined from the return to fixation of the corrective saccade. For two-step trials there were always two saccades, the first toward the visual target and the second to the auditory target. There were many cases in which the velocity did not return to within 2 SD of the baseline between these two saccades. Thus it appears that the cat did not complete one saccade before initiating the second. In these cases we marked the start of the second saccade when there was a sharp change in the amplitude of the velocity trace. As with single-step trials, the final gaze position was determined by the position at the time of return to fixation.

To quantify saccade accuracy and precision, the initial distance of the gaze from the target at time of target onset and the response amplitude were computed for each trial (Tollin et al. 2005). For the single-step trials, the initial distance was defined as the difference between the target-in-space position and the initial gaze position. This is the magnitude of the gaze shift needed to acquire the target position at the time of target onset given the initial gaze position. The response amplitude was defined as the angular magnitude and direction of the final gaze position from the initial gaze position.

For the double-step trials, the initial distance is the same as for single-step trials except that the measurement is taken during a time of rapid gaze shift. The response amplitude is the resultant vector of the two gaze shifts following acoustic target onset, first to the visual target and then to the acoustic target.

To obtain a quantitative measure of the localization performance across all target locations, a linear function was fit to the plots of response amplitude vs. initial distance. Horizontal and vertical components of the target locations were analyzed separately. The coefficients of the fits are indicators of localization performance. The slope of the response-target localization function, referred to here as “gain,” indicates the accuracy with which the cats localized the targets. The gain indicates the fractional overshoot or undershoot, relative to perfect accuracy of 1.0. The SD of the residuals of the fitted function, $\delta$, represents the distribution of behavioral responses about the mean gain and gives a numerical estimate of the inverse of precision (or consistency) of the localization responses, i.e., the more precise the response.

Standard statistical bootstrapping techniques (Efron and Tibshirani 1986) were used to obtain an estimate of the 95% confidence intervals of the gain. For a given stimulus configuration, 1,000 synthetic data sets, containing the same number of trials as the empirical data set, were created by randomly sampling with replacement localization data from individual trials from the empirical data sets for each cat. A linear function was fit to each synthetic data set, resulting in 1,000 measurements of the gain, from which the 95% confidence interval was obtained. As described above, the horizontal and vertical components of the behavioral responses were analyzed separately.

To determine whether or not the localization accuracy and precision of the static condition were statistically comparable to the dynamic condition, a slightly different algorithm of bootstrapping (Moore et al. 2002) was used. Specifically, a null hypothesis was constructed by pooling all the single-step trials together with the two-step trials. For each bootstrapping iteration, two new sets of trials were randomly selected (with replacement) from the pool, and the difference between the two new gains or $\delta$s was computed. A distribution of the gain/$\delta$ difference was formed after 1,000 iterations. If the actual gain/$\delta$ change $(g_g$/$\delta_\Delta)$ fell on the tail of this distribution ($P < 0.05$; a rare event), the difference between static and dynamic tasks was considered significant. If $g_g$/$\delta_\Delta$ fell on the main body of this distribution ($P > 0.05$), the difference was not significant (i.e., null hypothesis approved). Each cat was analyzed separately. The HH trials were analyzed separately from the HV trials.

We also computed localization errors for each trial by measuring the horizontal and vertical angles separating the final gaze position and the absolute position of the target in space. We preserved the direction of the errors so that the average of these signed errors indicates whether and by how much each target was underestimated (errors $< 0^\circ$) or overestimated. Absolute azimuth or elevation error was also calculated for each trial by taking the absolute value of the signed error. The Kolmogorov-Smirnov (KS) test was used to determine whether the two (non-Gaussian) data distributions were statistically different. If $P < 0.05$, the two data sets were considered to correspond to different distributions. We computed the two-dimensional KS statistic for data expressed as two-dimensional distributions (e.g., the azimuth-elevation end points in Fig. 7), to measure their mutual distance and its significance level (Press et al. 1992). Significant differences for normally distributed signed errors were determined with t-tests.

Onset times of gaze, head, and pinna movement following sound target presentation (latencies) were measured at the end of fixation relative to target onset.

To evaluate the extent of compensation of the audiomotor system for the occurrence of intervening eye and head movements, we analyzed the second gaze shift by applying a multiple linear regression analysis to the horizontal and vertical response components, respectively. Parameters were determined on the basis of the least-squares error criterion.

**RESULTS**

These experiments were designed to examine the ability of cats to localize acoustic targets during rapid eye-head gaze movements. The results and statistical analyses are based on the localization performance of four adult female cats. Of these, three cats also had pinna data for analysis. A total of 7,532 trials were analyzed.

**Behavior during double-step response.** Three examples of typical two-step saccade trials showing two-dimensional gaze and head trajectories are shown in Fig. 2. In Fig. 2A the visual target was localized to the right along the horizontal axis, while the auditory target was located at $0^\circ$ azimuth and $30^\circ$ elevation, requiring the cat to quickly reverse direction to obtain the auditory target. In Fig. 2B the head was moving horizontally to the left, and the gaze trajectory required the cat to reverse direction to reach the auditory target. In Fig. 2C the head was moving vertically during sound presentation. The head and the gaze tended to move together toward both the visual and auditory targets. The cat was not required to reach the visual target prior to localizing the auditory target.

Because the cats did not know whether a trial would be a one- or two-step task, they initiated the first saccade to the visual target of a two-step saccade in the same way they initiated the saccade to a single-step visual target. Typically, the position of the head and eye-in-head moved an appreciable distance and obtained a high peak velocity, during which time we presented the 25-ms sound target (yellow highlights in Fig. 2). When the auditory target was presented, the gaze changed...
direction toward the auditory target, followed by the head a short time later.

Eye and head movements during sound presentation. Our goal was to present the second (auditory) target during the high head velocity of the gaze saccade to the first (visual) target. For each cat we measured a distribution of peak velocities that was approximately Gaussian. The mean peak horizontal head velocity during the 25-ms noise burst ranged from 89°/s to 230°/s for the four cats (Fig. 3), while mean peak vertical head velocity varied from 61°/s to 96°/s. The mean head movement amplitude (both horizontal and vertical) during the 25-ms duration of the auditory target duration ranged from 1.0° to 5.9° for the four cats (Table 1). Each of the four cats had different mean amplitudes and peak velocities, yet this did not seem to influence the magnitude of error for each cat, i.e., the cat with the fastest peak velocity during sound presentation did not have the largest errors.

Sound localization errors. A common finding in localization experiments is higher accuracy in azimuth than in elevation even for broadband noise (Goossens and Van Opstal 1997; Makous and Middlebrooks 1990; Tollin et al. 2005, 2013). This was also true in this study, where gains for localization in elevation were statistically lower than gains in azimuth \( P \leq 0.05 \) in all conditions for three of the four cats. The only exception was cat 28 in the HH condition, where the vertical gain was equal to the horizontal. Therefore we kept the azimuth and elevation analyses separate. Previous studies have also shown that although the behavior of different cats tends to be qualitatively similar, they often differed quantitatively.

Figure 4A shows the final horizontal and vertical gaze position for cat 21 for the 19 most extensively tested target locations for the static, HH, and HV conditions (Fig. 4A, left, center, and right, respectively). The responses to the brief sounds were located near each target in azimuth and elevation (good accuracy), with some scatter of response at each location (fair precision).

To quantify these qualitative observations, Fig. 4B shows scatterplots of response amplitude as a function of the distance of the target from the gaze at the time of target onset for the vertical (Fig. 4B, top) and horizontal (Fig. 4B, bottom) response components. For the static, one-step trials (Fig. 4B, left), the response amplitude is the gaze shift toward the target. For the dynamic, two-step trials (Fig. 4B, center and right) the response amplitude is the resultant vector of the two gaze shifts, first to the visual and then the acoustic target. The assumption that gaze shift changed linearly with target eccentricity can be evaluated by the first-order correlation coefficient \( r \), which was between 0.84 and 0.98 for azimuth and between
0.68 and 0.93 for elevation (mean \( r = 0.88 \pm 0.07 \)). The correlation coefficients of the fitted functions for all conditions and all cats were highly significant \( (P < 0.005) \). Importantly, the cats’ responses were similar between the static and dynamic conditions.

### Table 1. Amplitude and peak velocity during sound presentation

<table>
<thead>
<tr>
<th>Cat</th>
<th>Head Amp, °</th>
<th>Eye-in-Head Amp, °</th>
<th>Head Peak Velocity, °/s</th>
<th>Eye-in-Head Peak Velocity, °/s</th>
<th>( n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>21 head horizontal</td>
<td>2.1 ± 4.0</td>
<td>1.1 ± 2.6</td>
<td>89 ± 51</td>
<td>57 ± 41</td>
<td>557</td>
</tr>
<tr>
<td>21 head vertical</td>
<td>1.5 ± 1.1</td>
<td>1.6 ± 1.8</td>
<td>63 ± 30</td>
<td>100 ± 72</td>
<td>1,483</td>
</tr>
<tr>
<td>28 head horizontal</td>
<td>2.7 ± 1.8</td>
<td>2.1 ± 1.7</td>
<td>135 ± 90</td>
<td>126 ± 92</td>
<td>305</td>
</tr>
<tr>
<td>33 head horizontal</td>
<td>4.8 ± 1.7</td>
<td>1.3 ± 1.5</td>
<td>230 ± 130</td>
<td>106 ± 102</td>
<td>580</td>
</tr>
<tr>
<td>33 head vertical</td>
<td>2.1 ± 1.4</td>
<td>1.1 ± 1.1</td>
<td>96 ± 55</td>
<td>81 ± 63</td>
<td>689</td>
</tr>
<tr>
<td>36 head horizontal</td>
<td>5.9 ± 6.1</td>
<td>2.8 ± 2.8</td>
<td>107 ± 53</td>
<td>53 ± 40</td>
<td>955</td>
</tr>
<tr>
<td>36 head vertical</td>
<td>1.0 ± 2.1</td>
<td>0.8 ± 1.8</td>
<td>63 ± 38</td>
<td>28 ± 23</td>
<td>697</td>
</tr>
</tbody>
</table>

Values are means ± SD.

Accuracy of gaze responses in the static and dynamic conditions for each of the four cats is displayed in Fig. 5 and \( \delta \), or 1/precision, in Fig. 6. Average responses in azimuth of all four cats showed slightly higher accuracy (gain = 0.87) in the static condition compared with HH (gain = 0.81) and HV (gain = 0.75).

Fig. 4. A: localization in single-step, HH, and HV auditory tasks. Top: scatterplot of final 2-dimensional gaze position (small symbols) for stimuli presented from 19 target locations (large open symbols). Bottom: mean and SD of final gaze positions. Data are from cat 21. B: accuracy of the vertical (response elevation, top) and horizontal (response azimuth, bottom) components of the responses to the 19 targets. Each point corresponds to a single trial. x-Axis shows the horizontal or vertical component of the distance between the gaze position on each trial and the actual position of the target at the time of target onset. The response amplitude (y-axis) is the corresponding horizontal or vertical component of the gaze shift response to that target position from the initial gaze position following the 1 (static task) or 2 (dynamic task) gaze shifts. Solid black line indicates the linear regression of response amplitude component and the initial distance of the gaze from the target. Gain is the slope of the regression line and represents localization accuracy. Dashed red line indicates a perfect gain of 1.0. \( \delta \) is the residual error in degrees after regression and is an indication of response precision or consistency. \( n \), Number of trials.
conditions. There was also better mean precision ($\delta = 6.1^\circ$) in the static condition compared with HH ($\delta = 6.9^\circ$) and HV ($\delta = 6.2^\circ$). In elevation, mean static accuracy (gain $= 0.67$) was higher than in the HH (gain $= 0.59$) or the HV (gain $= 0.61$) trials. Mean static localization was less precise ($\delta = 5.9^\circ$) than in the HH trials ($\delta = 5.2^\circ$) and more precise than in the HV trials ($\delta = 6.7^\circ$).

Standard statistical bootstrapping techniques (Efron and Tibshirani 1986; Moore et al. 2002) were used to determine statistical differences between the static control (black symbols) and either HH (red) or HV (blue) condition.

Fig. 5. Plots of response accuracy or gain (filled symbols) with associated 95% confidence intervals for the 4 cats to sources in elevation (El) and azimuth (Az). Subjects are identified along the x-axis; for example, “C21El” refers to data from cat 21 for errors in elevation. Asterisks indicate statistically significant differences between the static control (black symbols) and either HH (red) or HV (blue) condition.

We also analyzed accuracy of gaze responses as measured with signed error (Fig. 7). Overall, the cumulative distributions of signed gaze errors in the static condition are similar to those in the HH or HV condition for both horizontal (Fig. 7A) and vertical (Fig. 7B) components. Average responses in azimuth of all four cats to static, HH, and HV conditions showed higher accuracy (mean signed error $= -1.6^\circ$) in the static condition compared with HH (signed error $= -3.2^\circ$) and HV (signed error $= -3.9^\circ$). In elevation, static accuracy (mean signed error $= -2.6^\circ$) was similar to that for HH localization (signed error $= -2.7^\circ$) and HV accuracy (signed error $= -2.6^\circ$). In 8 of 14 cases, the differences were significantly different. In 2 of 8 cases, accuracy for the static condition was worse.

Fig. 6. Plots of response precision for the 4 cats. Same format as Fig. 5 except that $\delta$, or $1/\text{precision}$, of gaze responses in the static and dynamic conditions for each of the 4 cats is displayed.

Fig. 7. Cumulative distribution of signed gaze errors, the distance between final gaze position and target position, for static, HH, and HV conditions for all 4 cats. A: horizontal signed errors. Vertical dashed lines at 0° error connect the x-axis segment to the appropriate distribution. B: same as A for vertical signed errors. Cat 28 did not perform the HV task.
Precision was analyzed using unsigned (absolute) error. Average responses showed better precision ($\delta = 5.2^\circ$) in the static condition compared with HH ($\delta = 9.2^\circ$) and HV ($\delta = 6.2^\circ$). In elevation, static precision ($6.7^\circ$) was similar to HH precision ($6.6^\circ$) and the same as HV precision ($6.7^\circ$). In 9 of 14 cases, differences in precision were significant ($P < 0.05$). Of these, the static had better precision in 7 cases than the dynamic and 2 had worse precision.

In summary, we compared the localization performance under static and dynamic conditions for both horizontal and vertical response components using accuracy (gain), precision ($1/\delta$), and signed and unsigned error. In none of these measures were there consistent differences between the static and dynamic conditions.

Compensation by the audiomotor system of intervening eye and head movements. Following the work of Vliegen et al. (2004) in humans, we performed multiple linear regression on the two gaze shifts to determine how the first gaze shift affects the ability to localize the sound target. From Fig. 1B, the following vector equation defines the two gaze shifts and acoustic target at time of onset:

$$\Delta G_2 + \Delta G_1 = T_{H,ini}$$  

where $\Delta G_1$ is the first gaze shift from the time of sound onset toward the visual target, $\Delta G_2$ is the second gaze shift to the auditory target, and $T_{H,ini}$ is the initial sound location in head-centered coordinates at the time of sound onset.

Furthermore, $\Delta G_1$ can be expressed as the sum of the displacement of the head during the first gaze shift, $\Delta H_1$ and the position of the eye-in-head at the end of the first gaze shift, $E_0$. Thus $\Delta G_2$ can be described by a linear combination according to the following equation, where the gain variables ($a$, $b$, $c$) carry the signs ($+$ or $-$):

$$\Delta G_2 = a \cdot T_{H,ini} + b \cdot \Delta H_1 + c \cdot E_0 + d$$  

HH and HV conditions were analyzed separately. The resulting gains ($a$, $b$, $c$) of the regression, averaged across subjects, are summarized in Fig. 8. If there is full compensation for the first gaze shift in the execution of the second gaze shift, $a = 1$, $b = c = -1$, and $d = 0$. Our results showed that the gains were similar for the HH and HV conditions but compensation was not as complete for the vertical component of the gaze shift as it was for the horizontal component. For example, the average of variable $a$ was 0.78 and 0.81 for the horizontal components, respectively, and 0.57 and 0.58 for the vertical components. Similarly, $b$ was $-0.75$ and $-0.81$ for the horizontal components and $-0.53$ and $-0.58$ for the vertical components; $c$ averaged $-0.88$ and $-0.81$ for the horizontal components and $-0.81$ and $-0.57$ for the vertical components. These results suggest that compensation was more complete for the horizontal than the vertical component of the gaze.

Gaze, head, and pinna latency. In three cats we measured pinna, gaze, and head latencies with respect to sound onset. We analyzed both static and dynamic trials to horizontal targets ipsilateral to the measured ear, because pinna movements of the ipsilateral ear are more consistent (Populin and Yin 1998b) and have shorter latencies than the head (Tollin et al. 2009). In the case of the two-step trials, the gaze, head, and pinna usually had to change or even reverse direction to make a saccade to the auditory target.

In two of the three cats the gaze, head, and pinna latencies to dynamic targets were longer than the latencies to static targets (Table 2). In cat 33 the latencies were shorter in the dynamic conditions. In all cases the pinna latencies were shorter than either the gaze or head latency. In general, the head minus pinna latencies were greater in the dynamic cases, indicating that the pinna was moving relatively faster than the head toward the auditory target than in the static trials. This may reflect the greater mass and inertia of the head than the pinna, requiring more time for the head to change or reverse direction.

As an example of the head, eye, and pinna movements during a typical dynamic localization, Fig. 9 shows horizontal position traces as a function of time for gaze, head, right pinna, and pinna-on-head. The start of the sound presentation is time $0$. The first fixation LED is presented in this trial at $-2,333$ ms. The cat must fixate its gaze within the blue open fixation window for $850$ ms to trigger the first target LED. In this example the cat required $1,188$ ms to acquire the fixation LED at time $-1,145$ ms. There is no fixation requirement for the position of the head or pinna. After the start of the target LED at $-338$ ms, latency of response to the visual cue was $243$ ms for gaze, $231$ ms for head, and $254$ ms for pinna. The auditory target was turned on $-100$ ms after the cat started to move toward the visual target ($-10^\circ$ azimuth as indicated by the arrow labeled LED in Fig. 9A). Within a few tens of milliseconds, the pinna moved toward the auditory target ($+30^\circ$ azimuth, located at the solid black arrow, Fig. 9A, top right).

The response latency of the gaze and head is delayed compared with the pinna, resulting in a pinna-on-head move-

---

**Table 2. Movement latency to sound target**

<table>
<thead>
<tr>
<th></th>
<th>Gaze</th>
<th>Head</th>
<th>Pinna</th>
<th>Head – Pinna</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cat 28 static</td>
<td>54 ± 32</td>
<td>57 ± 31</td>
<td>41 ± 36</td>
<td>15 ± 51</td>
</tr>
<tr>
<td>Cat 28 dynamic</td>
<td>114 ± 83</td>
<td>93 ± 114</td>
<td>62 ± 42</td>
<td>30 ± 38</td>
</tr>
<tr>
<td>Cat 33 static</td>
<td>66 ± 30</td>
<td>51 ± 19</td>
<td>50 ± 15</td>
<td>1 ± 20</td>
</tr>
<tr>
<td>Cat 33 dynamic</td>
<td>44 ± 20</td>
<td>46 ± 20</td>
<td>40 ± 16</td>
<td>6 ± 11</td>
</tr>
<tr>
<td>Cat 36 static</td>
<td>47 ± 20</td>
<td>59 ± 37</td>
<td>33 ± 8</td>
<td>26 ± 36</td>
</tr>
<tr>
<td>Cat 36 dynamic</td>
<td>75 ± 30</td>
<td>79 ± 29</td>
<td>33 ± 11</td>
<td>46 ± 31</td>
</tr>
</tbody>
</table>

Values (in ms) are means ± SD.
ment toward the target of \(\sim 9^{\circ}\). Starting at the time of peak pinna-on-head movement, as the head starts to move toward the target, the pinna moves in an equal and opposite direction to the head because of the vestibuloauricular reflex (VAR; Tollin et al. 2009). During that brief (~100–150 ms) time period (shading in Fig. 9B) the pinna-in-space is relatively stable in space. After this brief stable period, the gaze, head, and pinna proceed to move to the target and all three end up near the acoustic target position of \(+30^{\circ}\). Since the gaze remained in the fixation window (designated by bracket, Fig. 9A) for 600 ms, this was judged to be a successful trial and the cat was rewarded with a small food reward. These data provide evidence that the pinna also responds appropriately with short-latency movements to both the visual and acoustic targets and apparently compensates for the moving head and eyes when orienting to the brief acoustic target.

Figure 10 shows accuracy of pinna movements represented by the cumulative horizontal signed errors, similar to Fig. 7 for gaze errors. Even though the pinnae move independently in time and position from the head and gaze (Fig. 9), the final pinna position reflects accurate localization in two of the three cats. Cat 28’s pinna movements undershot the target by \(\sim 30^{\circ}\) in both static and HH trials. The fact that the distributions of signed errors in the static and dynamic tasks are similar in all three cats indicates that the pinna accuracy in these two conditions is similar.

**DISCUSSION**

The major finding in this study is that cats can localize a brief noise that is presented while the cat’s eyes, head, and pinnae are moving rapidly toward another target in space (Figs. 2 and 3). No consistent differences (improvement or decline) in accuracy or precision were found between the static single-step and dynamic double-step tasks (Figs. 5 and 6). Regression analysis indicates that intervening eye and head movements are largely compensated for during the ongoing movement (Fig. 8). Pinna movements are also comparable during the static and dynamic localization trials, suggesting that pinna movements can also compensate for intervening head movements (Fig. 10). Head and pinna movements may allow multiple samples of acoustic cues to be obtained and integrated with proprioceptive input during localization.

*Comparison to other studies.* Previous studies have examined the ability of humans to localize sound targets presented during head movement. Humans accurately localized a 50-ms broadband noise target presented just before or during a rapid eye and head movement toward a light target (Vliegen et al. 2004). Using regression analysis, Vliegen et al. showed that humans were able to apply the appropriate coordinate transformations to fully compensate for all intervening eye and head movements. We replicated their analysis and found that cats also keep track of and compensate for rapid changes in head, eye, and ear position that may occur during sound presentation (Fig. 8).

Cooper et al. (2008) found that localization of a noise target in elevation using head pointing, presented either in the early or late phase of a head turn, remained accurate to targets in
both the front and rear hemispheres. They attributed the accuracy in elevation to the idea that, for a given elevation, spectral cues do not change very much with changes in horizontal position although experimental measurements of HRTFs showed substantial changes in the position of the midfrequency notch as the positions of the pinnae were varied (Young et al. 1996). Localization of targets in azimuth was much less accurate if the target was located in the rear hemisphere and turned on during the latter part of the head turn. They attributed more accurate localization to targets in the frontal hemisphere to general allocation of attention there. Our measurements are all in the frontal hemisphere.

Researchers have long observed and studied the apparent strong weighting of initial interaural cues (onset ITD and ILD) on the perceived location of a sound target in azimuth (Brown and Stecker 2013; Freymann et al. 2010; Hafter et al. 1983; Hafter and Buell 1990; Hafter and Dye 1983; Stecker and Hafter 2002). The decline in usefulness of interaural information after the signal’s onset was called “binaural adaptation” (Hafter and Buell 1990). However, there are conditions that allow recovery from an adapted binaural system, thus allowing a resampling of interaural information (Hafter and Buell 1990; Stecker and Hafter 2002). Transient changes in the amplitude spectrum introduced by gaps, among other triggers, have been found to allow this resampling. As noted above, Young et al. (1996) have shown how movement of the pinna can produce a steep rising slope of the amplitude of a specific frequency. Improvement in localization accuracy by obtaining a second sample of binaural information during head movement would also require that the listener be able to monitor ongoing changes in head, pinna, and eye position.

Head movements and sound localization. Any given ITD or ILD can originate from a locus of points that Wallach (1939) termed a “cone of confusion.” The ambiguities of localization in this cone of confusion can be reduced by head movements during sound presentation, which would provide a number of different lateral angle determinations for the same sound source. Wallach (1939) also proposed that front/back confusions and elevation ambiguities could be resolved “on the basis of the pinna factor alone, i.e., without head movements. . . .”

Experimental support for improvement in localization by head movements has been shown by a number of studies. Thurlow and Runge (1967) observed a reduction in horizontal errors and front/back confusions for both low- and high-frequency noise and click stimuli following induced (voluntary) head motions. However, Pollak and Rose (1967) observed improvement in accuracy only when the sound stimulus was long (>1 s) and the head was turned toward the sound source, supposedly providing the position for optimal extraction of localization cues. Subsequent studies (Perrett and Noble 1997a, 1997b) provided further support for improved localization with head rotation, especially for low-frequency targets varying in elevation. Wightman and Kistler (1998) observed a reduction in front/back errors for both free-field and virtual-sound targets, either when the listeners moved their own heads or when the listeners controlled movement of the sound source. So movement of the head and pinna may result in a “transient change in the amplitude/phase spectrum” (Hafter and Buell 1990), allowing resampling of an adapted binaural system.

However, we found that accuracy of localization was similar in our dynamic task compared with the head-stable task. Perhaps the rapid pinna movement to sound targets in our cats in the static task results in beneficial resampling of binaural cues, a situation that may also be enhanced during the head and pinna movement of our dynamic task.

Dynamic sound localization cues in elevation. Humans (Cooper et al. 2008; Vliegen et al. 2004) and cats (present study) do not show any significant improvement or decline in localization of noise targets in elevation presented during rapid head movement, compared with when sound is presented to a stable head. Vliegen et al. (2004) did find that increasing noise duration from 3 ms to 100 ms resulted in improved localization accuracy to targets in elevation presented during head movement, as it does when the head is stable. They attribute this to additional neural integration during fast head movements (Vliegen et al. 2004). That is, the auditory system will make a final location estimate based on multiple short-term location estimates (Hofman and Van Opstal 1998).

Spatial updating of gaze and ears. Our behavioral results confirm the findings of Vliegen et al. (2004) in human subjects by showing that cats are also able to compensate for movements of the head while localizing sound sources. A novel aspect of our work is to show that the pinna movements in the dynamic tasks were similar to those in the static task. This suggests that there is a common neural machinery for keeping track of the head position even when it is changing rapidly and to provide the appropriate compensation to circuits responsible for pinna movement as well as gaze movements.

Because the head, pinnae, and eyes are changing position during sound presentation, it is necessary that the brain have information about those changes in location. Behavioral and physiological evidence with two-step tasks similar to that employed here have provided firm evidence that the oculomotor system can compensate for perturbations in eye position using feedback from a corollary discharge or efference copy signal carrying current eye position. Hallett and Lightstone (1976) showed that human subjects can accurately saccade to a briefly flashed remembered second visual target following an intervening saccade to a first visual target. Since the second saccade originates from the location of the first target and not from the original fixation point, the retinotopic location of the second target does not correspond to the trajectory of the second gaze shift and can even be in the opposite direction. Monkeys can also execute such two-step visual gaze shifts accurately (Mays and Sparks 1980). Furthermore, if the two-step task is executed by delivering a brief electrical pulse to the deep layers of the superior colliculus (SC) to produce the intervening saccade, monkeys are also able to compensate for the electrically evoked movement. In this case the pulse is delivered shortly after the presentation of a visual target but before the eye has a chance to move, which moves the eyes to a new position corresponding to the location of the SC that is stimulated. The monkey will then make an appropriate saccade to the visual target even though the original retinotopic signal of the visual target does not match the saccade (Mays and Sparks 1980; Sparks and Mays 1983).

Sommer and Wurtz (2002, 2004a, 2004b) have described a neural substrate for a corollary discharge signal that projects from the SC to the medial dorsal (MD) nucleus of the thalamus and then to the cortical frontal eye fields. Cells in the MD nucleus have the requisite physiological responses expected of the corollary discharge. Importantly, when inactivated by in-
jections of the GABA agonist muscimol, monkeys have deficits on the Hallett and Lightstone (1976) two-step task. In addition, studies have shown that the oculomotor circuitry in the paramedian pontine reticular formation (PPRF) and abducens nucleus contain cells that code current eye position in the so-called “tonic” cells (see Fuchs et al 1985 for review). These studies show that a corollary discharge for eye position is used for spatial updating during voluntary movement of the eyes.

Similarly, there is indirect evidence that head position is also monitored. Experiments in human subjects using two-step gaze shifts with combined head and eye movements show compensation for intervening gaze shifts (Herter and Guittion 1998; Vliegen et al. 2005). Furthermore, human subjects can make saccades to remembered positions following torsional movements of the head, suggesting use of a head movement signal (Klier et al. 2005; Medendorp et al. 2002).

On the other hand, little is known about the motor control of pinna movements and possible efference copy signals. In the cat the pinna muscles are controlled by motoneurons in the facial nucleus (Populin and Yin 1995). Anatomical studies show that the pinna motoneuron pools in the facial nucleus receive input indirectly from the SC through the paralemiscal region of the midbrain (Henkel and Edwards 1978). Our findings that the pinna movements in the dynamic task are similar to those in the static task suggest that pinna movements are also compensated by efference copy signals of unknown origin. A possible related finding is the somatosensory input from the dorsal column (serving cervical and pinna muscles) and spinal trigeminal nuclei to the dorsal cochlear nucleus (Kanold and Young 2001).

ACKNOWLEDGMENTS

We thank J. Sekulski for helping with computer programming and Dr. Yan Gai for comments on the manuscript.

GRANTS

This work was supported by National Institute on Deafness and Other Communication Disorders Grant DC-07177.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

AUTHOR CONTRIBUTIONS


REFERENCES


