Perception of Auditory, Visual, and Egocentric Spatial Alignment Adapts Differently to Changes in Eye Position

Qi N. Cui,1 Babak Razavi,1,2,4 William E. O’Neill,1,3 and Gary D. Paige1,4

1 Departments of Neurobiology and Anatomy and 2 Biomedical Engineering, 3 Center for Navigation and Communication Sciences, and 4 Center for Visual Science, University of Rochester Medical Center, Rochester, New York

Submitted 9 June 2009; accepted in final form 13 October 2009

Cui QN, Razavi B, O’Neill WE, Paige GD. Perception of auditory, visual, and egocentric spatial alignment adapts differently to changes in eye position. J Neurophysiol 103: 1020–1035, 2010. First published October 21, 2009; doi:10.1152/jn.00500.2009. Vision and audition represent the outside world in spatial synergy that is crucial for guiding natural activities. Input conveying eye-in-head position is needed to maintain spatial congruence because the eyes move in the head while the ears remain head-fixed. Recently, we reported that the human perception of auditory space shifts with changes in eye position. In this study, we examined whether this phenomenon is 1) dependent on a visual fixation reference, 2) selective for frequency bands (high-pass and low-pass noise) related to specific auditory spatial channels, 3) matched by a shift in the perceived straight-ahead (PSA), and 4) accompanied by a spatial shift for visual and/or bimodal (visual and auditory) targets. Subjects were tested in a dark echo attenuated chamber with their heads fixed facing a cylindrical screen, behind which a mobile speaker/LED presented targets across the frontal field. Subjects fixated alternating reference spots (0, 20°) horizontally or vertically while either localizing targets or indicating PSA using a laser pointer. Results showed that the spatial shift induced by ocular eccentricity is 1) preserved for auditory targets without a visual fixation reference, 2) generalized for all frequency bands, and thus all auditory spatial channels, 3) paralleled by a shift in PSA, and 4) restricted to auditory space. Findings are consistent with a set-point control strategy by which eye position governs multimodal spatial alignment. The phenomenon is robust for auditory space and egocentric perception, and highlights the importance of controlling for eye position in the examination of spatial perception and behavior.

INTRODUCTION

We routinely use both vision and audition to navigate through our environment. It is important for these two senses to provide a spatially and temporally congruent map of the external world if we are to accurately integrate spatial information across the senses and properly identify objects in the environment that evoke both an image and an associated sound. Vision and audition, however, differ considerably in how they process spatial information (Knudsen and Brainard 1995). Whereas visual space is represented in eye-centered coordinates by a direct topographic projection of the outside world onto the retina, auditory space is represented in head-centered coordinates centrally constructed from spatially dependent cues contained within sounds registered by the two ears. The auditory cues include interaural differences in time and intensity (ITD and IID) for azimuth (horizontal) and pinna-dependent spectral cues for elevation (vertical) (cf. Middlebrooks and Green 1991 for review).

From these diverse encoding methods, visual and auditory space must be integrated into a common reference frame to be useful and synergistic. Moreover, because the retinae move with the eyes while the ears remain head-fixed, every eye movement realigns visual and auditory space, leading to a frequently shifting spatial mismatch between the sensory modalities. The brain must accurately account for changes in eye position to maintain space-constancy and a coherent multisensory map of the external world. This presumably occurs by using an eye-in-head signal to transform audition and vision onto a common coordinate scheme with reference to the head and ultimately in space relative to the body (Cohen and Andersen 2000; Groh and Sparks 1992; Jay and Sparks 1987).

How accurately is eye movement integrated with visual and auditory input to maintain space constancy? Might eye movement also influence the perception of auditory and visual space in the process of matching coordinate schemes to a unified spatial representation? A number of behavioral studies in human and nonhuman primates have shown that sound localization shifts in response to changes in eye position, possibly reflecting an imperfect cross-sensory calibration of eye and head position signals. However, the methodology used in these studies is inconsistent, and the reported magnitude, direction, and spatial uniformity of eye position effects on sound localization vary widely. Sound localization has been assessed over a wide range of brief ocular fixation durations (1–20 s), and the magnitude of reported localization errors ranges from 3–37% of ocular angle (Bohlander 1984; Getzmann 2002; Lewald 1997, 1998; Lewald and Ehrenstein 1996, 1998; Lewald and Getzmann 2006; Metzger et al. 2004; Weerts and Thurlow 1971; Yao and Peck 1997). Neurophysiological studies in cats and nonhuman primates have also implicated eye position in the modulation of auditory spatial cues in the superior colliculus (SC) (Hartline et al. 1995; Jay and Sparks 1984; Peck et al. 1995; Populin et al. 2004; Van Opstal et al. 1995; Zella et al. 2001), inferior colliculus (IC) (Groh et al. 2001; Zwiers et al. 2004), frontal eye fields (FEF) (Russo and Bruce 1994), and auditory cortex (Fu et al. 2004; Werner-Reiss et al. 2003). Nevertheless, inconsistent findings and diverse experimental conditions preclude a unifying explanation of how eye position influences spatial hearing, or how space constancy is maintained between the senses despite eye movements.

Recently, we reported that the human perception of auditory space shifts dramatically in the direction of a new and prolonged eccentric eye position (Razavi et al. 2007). This auditory spatial adaptation develops exponentially in time (with a
time constant of ~1 min), approaches ~40% of sustained eye eccentricity, and extends over a wide range of frontal auditory space.

In this study, we sought to further characterize the effect of eye position on sound localization to define more completely the properties of this novel spatial adaptation phenomenon and clarify its potential implications for other aspects of spatial perception and function. Several questions were addressed regarding spatial localization and eye position.

**Is eye eccentricity alone, without a visual or other sensory fixation reference, sufficient for producing auditory spatial adaptation?**

In the absence of a visual reference, extraretinal eye position signals could be derived centrally from feedforward motor commands (efference copy of eye movement) to the extraocular muscles (EOMs) (Bridgeman 1995), with subsequent maintenance of eye position by the oculomotor integrator (Arnold and Robinson 1997; Cannon and Robinson 1987), and/or from proprioceptive input from EOM afferents (Leigh and Zee 1999; Richmond et al. 1984; Steinbach 1987). Although experimental evidence has traditionally favored efference copy as the source of eye position signals necessary for visually guided behavior on a moment-to-moment basis (Ruskell 1999), there is no shortage of studies postulating a role for EOM proprioceptors in positional awareness as well (Bridgeman and Stark 1991; Gauthier et al. 1990; Skavenski 1972; Zella et al. 2001).

Past studies that examined the effect of eye position on sound localization have largely provided visual targets for ocular fixation, thereby supplying retinal information along with extraretinal eye position signals. To determine whether the eye position–related auditory spatial shift also makes use of a retinal signal, we examined whether eccentric eye position causes a physiological adaptation of auditory space in the absence of continuous retinal feedback from a fixation target or other visual reference in the frontal field of space.

**Is the influence of eye position selective for, or differentially applied to, ITD, IID, or spectral cues, and if so, can we infer the level in the neuraxis and spatial processing stream at which eye position exerts its effect?**

Auditory localization in azimuth (Az) is a construct of two separate frequency-dependent channels: interaural time differences (ITD), attributable to spatial separation of the two ears, and interaural intensity differences (IID), resulting from the wavelength-dependent directional filtering and “shadowing” effect of the pinna, head, and body (Grantham 1984; Yost and Dye 1988). In humans, IIDs provide reliable localization cues at frequencies over ~3 kHz, whereas ITD cues are unambiguous at frequencies under ~1 kHz and typically dominate when available (Klumpp and Eady 1956; Macpherson and Middlebrooks 2002; Wightman and Kistler 1992; Zwischlocki and Feldman 1956). Elevation (El), in contrast, is only reliable for complex broadband sounds with frequency content above ~4 kHz, where the pinna generates vertical position-dependent spectral notches and peaks (Butler and Humanski 1992; Hebrank and Wright 1974; Roffler and Butler 1968). Monaural/pinna cues also play a part in azimuthal sound localization and help to resolve front-back confusion in the horizontal plane (Belendiuk and Butler 1975; Musicant and Butler 1984). These distinct spatial cues (channels), initially processed along distinct parallel central auditory pathways (Heffner and Masterton 1990), are ultimately integrated into unified representations of auditory space within the CNS (Brainard et al. 1992).

Although previous examinations of eye position influences on sound localization used auditory stimuli of various frequencies, intensities, and/or durations (Bohlander 1984; Getzmann 2002; Lewald 1997, 1998; Lewald and Ehrenstein 1996, 1998; Lewald and Getzmann 2006; Metzger et al. 2004; Weerts and Thurlow 1971; Yao and Peck 1997), a systematic examination of the influence of frequency-specific localization cues on this phenomenon has not been conducted. It is conceivable that eye position selectively affects particular spatial channels for sound localization, especially in light of studies showing that ITD dominates ILD cues for localization when conflicts arise (Wightman and Kistler 1992). To directly address whether eye position effects are global or local in the spectral domain, we used band-limited and broadband auditory targets.

**Does the subjective perception of straight-ahead, and therefore the egocentric spatial orientation reference, adapt to eye position in similar fashion to auditory space?**

We have suggested previously that prolonged maintenance of eccentric eye position might serve to realign the perception of egocentric “straight-ahead” along with co-registration of not only auditory but also multimodal space (Razavi et al. 2007). Although eccentric eye position has indeed been shown to shift perceived straight-ahead, the magnitude and direction of the shift are disputed (Bohlander 1984; Lewald and Ehrenstein 2000; Weerts and Thurlow 1971). To clarify this effect, we devised a paradigm to quantify the dynamics and magnitude of the shift in perceived straight-ahead in conjunction with the adaptation of sound localization that occurs with prolonged eccentric eye position.

**Does visual spatial localization shift in tandem with sound localization? If differences in shift direction and/or magnitude exist between the two sensory modalities, how is spatial perception of multi-sensory targets affected?**

Although past studies have shown a small change in visual spatial localization in response to eye position (Ebenholtz 1976; Hill 1972; Lewald and Ehrenstein 2000; Metzger et al. 2004; Morgan 1978; Yao and Peck 1997), in addition to ambiguities in direction and magnitude, it is unclear whether these changes represent a true adaptation of visual spatial perception to prolonged eccentric eye position or instead reflect a simple reference frame misalignment (Bock 1986; Henriques et al. 1998). In this experiment, visual targets were localized along with auditory targets to establish the effect of eye position on visual and auditory space within the same experimental session. By also examining the localization of concurrently presented auditory and visual (bimodal) targets occupying the same spatial location, we asked the question: If confronted with two sensory spatial maps that likely respond differently to sustained eccentric fixation, how will the brain accommodate this discrepancy?

The answers to these four questions are crucial for determining the nature of eye position effects on auditory spatial
processing, spatial reference frames (visual and auditory), and global spatial perception, and have important implications for how we register and interact with a complex and changing spatial environment.

METHODS

Subjects

Eleven human subjects (females = 5; males = 6; mean age = 30 yr) participated in these experiments, five of which participated in all four parts of the study. All subjects were recruited from the University of Rochester community and were experienced with previous sound localization studies in our laboratory.

All subjects were in good general health and self-sufficient. None expressed a history of neurological, ophthalmological, otological, or other sensory-motor dysfunction. Best-corrected visual acuity was 20/20 or better binocularly, with normal stereo-acuity (Titmus test) and visual fields (tangent screen, paying particular attention to the region of visual space tested in this study). All subjects were free of defects in cranial nerve, cerebellar/coordinational, and somatosensory functions. Standard videonystagmography (VNG) showed normal oculomotor and vestibular function (including caloric tests).

In addition, subjects met the following criteria: 1) absence of a conductive hearing loss indicated by tympanometry and bone conduction thresholds; 2) an audiometric configuration within the range defined operationally as “normal hearing” (hearing thresholds between 1 and 20 dB HL at octave intervals between 0.25 and 8 kHz); and 3) speech perception in quiet scores of 88–100% on a word recognition test (NU-6).

The study was performed with approval from the University of Rochester Research Subjects Review Board in accordance with the 1964 Declaration of Helsinki. All subjects gave informed consent and were compensated for their participation.

Target apparatus and positioning

The Spatial Localization Laboratory (Fig. 1A) has been described previously (Razavi et al. 2007; Zwiers et al. 2003). The experimental chamber consists of a fully enclosed, 3.0 × 3.7 × 2.7-m room lined with 7.6-cm-thick acoustic foam (Sonex, Illbruck, Minneapolis, MN) and vinyl backing to attenuate echoes and extrinsic sounds, respectively. The ambient background noise within the chamber is ~35 dBA SPL. Subjects sat on an adjustable ergonomic chair, facing the center of a 1.8-m-high cylindrical screen of acoustically transparent black speaker cloth at 2-m distance. The screen also served as a projection medium to register all visual targets projected onto it from the front or back, as detailed below. The head was fixed using a rigid bite-bar of dental impression compound molded over a steel plate custom made for each subject (Fig. 1B). The bite-bar was secured to an adjustable tripod ball head mounted on top of a rigid platform in front of the chair at chest level. The metal platform and its supporting structures were covered with a synthetic fleece to minimize acoustic reflections.

The head was positioned to match the subject’s cyclopean eye (the midpoint between the 2 eyes) with the geometric origin of the cylindrical screen and target space. During the first session for each subject, Reid’s baseline, a line extending from the inferior margin of the orbit to the superior border of the external auditory meatus, was aligned with the horizontal plane at the level of the target space’s vertical center. By recording the position of the aligned bite-bar using a Polhemus FASTRAK (Polhemus, Colchester, VT), a three-dimensional (3-D) individualized baseline position was obtained to ensure consistent positioning of subjects over multiple experimental sessions.

All auditory targets were presented by a single 8-cm-diam, two-way coaxial loudspeaker (model PCx 352, Blaupunkt, Hildesheim, Germany). Visual targets were projected onto the back of the cylindrical screen by a miniature red LED mounted at the center of the speaker, yielding a 3-mm projected diameter (~0.1° subtended angle). The speaker/LED assembly was mounted on a two-axis servo-controlled robotic arm hidden behind the cylindrical screen. This system enabled rapid, accurate, and precise positioning of the speaker in cylindrical space.
coordinates, providing an essentially unlimited and continuous array of targets within the range of \( \pm 65^\circ \) Az by \( \pm 25^\circ \) El.

**Stimulus characteristics**

All auditory stimuli were synthesized digitally using SigGenRP software and presented using TDT System II hardware (Tucker-Davis Technologies, Alachua, FL). Auditory targets consisted of 150-ms bursts of Gaussian white noise (equalized, 75 dB SPL) with a rise and fall time of 10 ms, while visual stimuli similarly consisted of 150 ms LED flashes, both repeating at 5 Hz. To mask potential spatially dependent loudness cues and any amplitude cues related to room location, stimulus level was randomly varied between 70 and 75 dB SPL in 1-dB steps from trial to trial. Stimuli used in these experiments included broadband (BB; 0.1–20 kHz), low-pass (LP; 0.1–1 kHz), and high-pass (HP; 3–20 kHz) band-limited noise to selectively assess the importance of ITD, IID, and spectral spatial cues relevant to sound localization.

ITDs can be perceived in high-frequency carriers if they are amplitude modulated, have abrupt onsets, or narrow bandwidths that generate low-frequency envelope cues. The HP stimulus used here provides little or no usable ITD cues. Regarding onset cues, the 10 ms, \( \cos^2 \) rise-fall time of the noise bursts is equivalent to an AM rate of 50 Hz. Studies using amplitude modulated high-frequency tones show that ITD thresholds are outside the ecological range generated by the head at modulation rates of <100 Hz (Henning 1974; Nuetzel and Hafer 1981). Regarding ongoing envelope cues, the bandwidth of the HP noise (17 kHz, over 2 octaves) is too wide to generate useful low-frequency envelope ITD cues (McFadden and Pasanen 1976). Thus the HP and LP noise bursts used in these experiments selectively emphasized IID/spectral and ITD auditory spatial cues, respectively.

To avoid the potential for subjects to predict target location based on cues derived from the robotic arm motion, we routinely used two tactics. First, the robotic arm was programmed to position the speaker/LED in two steps: initially to a random location and then to the intended target position. This approach nullifies the use of speaker travel time as a potential predictor. Second, nondirectional Gaussian white noise (65 dB SPL) was presented between trials during speaker movements to mask mechanical noise emanating from the robotics. This provides little or no useable ITD cues. Regarding onset cues, the 10 ms, \( \cos^2 \) rise-fall time of the noise bursts is equivalent to an AM rate of 50 Hz. Studies using amplitude modulated high-frequency tones show that ITD thresholds are outside the ecological range generated by the head at modulation rates of <100 Hz (Henning 1974; Nuetzel and Hafer 1981). Regarding ongoing envelope cues, the bandwidth of the HP noise (17 kHz, over 2 octaves) is too wide to generate useful low-frequency envelope ITD cues (McFadden and Pasanen 1976).

**Fixation targets**

To control eye position during spatial localization trials, fixation targets were utilized. Five red laser-LEDs were mounted to the room’s ceiling above the subject’s head and projected fixation spots (3-mm projected diameter, \( \sim 0.1' \) subtended angle) onto the cylindrical screen at center (Ctr; 0° Az, 0° El), ±20° left and right of Ctr (L20° and R20°), and ±20° up and down from Ctr (U20° and D20°).

**Response measures**

**VISUALLY GUIDED LASER POINTING.** A two-axis cylindrical joystick was mounted beneath the platform holding the bite-bar. It housed a laser-LED pointer that projected its beam onto the screen (1-mm projected diameter, <0.03° subtended angle). The joystick provided no position-related tactile cues and was coupled to two orthogonal optical encoders (resolution <0.1°) that measured its angles in both Az and El. The utility and advantages of visually guided (laser) pointing for sound localization have been addressed previously (Razavi et al. 2007).

For each target presentation, the response endpoint was registered by the subject (Fig. 1B) with a key press, at which time the target and pointer positions were recorded and transformed into Az and El after a cylindrical coordinate scheme centered at the subjects’ cyclopean eye (Razavi et al. 2007). Subjects used their preferred hand to manipulate the laser pointer and the other to register the response. Subjects were instructed to localize targets quickly but accurately (response time was typically \( \sim 4 \) s). In general, no feedback was available with regard to localization performance during all experiments.

**VISUAL CALIBRATION.** At the end of each experiment, subjects localized an array of visual targets at 10° intervals within a rectangular range comparable to that of the experimental (auditory) targets (\( \pm 43^\circ \) Az, \( \pm 20^\circ \) El). Visual localization of speaker LED targets has the benefits of minimal noise, high accuracy, and a flat spatial gain (\( \sim 1.0 \)) and served to ensure optimal system performance while verifying the subject’s ability to perform the task.

**Data acquisition**

Experiments were controlled by customized software written in the Visual C++ programming language (Microsoft). This software communicated directly with the robotic arm (target positioning) control system, TDT System II (stimulus and masker presentation), and the fixation lasers. Target and pointer position was recorded each time the response button was pressed which terminated each trial. An Excel (Microsoft) spreadsheet served as an operator interface for the software and provided a user-friendly environment for implementing experimental paradigms and storing all data.

**Experimental protocol**

To characterize the effect of eye position on spatial localization over time, an Alternating Fixation paradigm (Razavi et al. 2007) was used in each of the four experiments. In all cases, head-fixed subjects maintained ocular fixation on one of five red laser-LED spots projected onto the screen (at Ctr, L20°, R20°, U20°, or D20°) and used peripheral vision to guide the pointer to localize spatial targets. Trials began with the presentation of the auditory and/or visual stimuli (Fig. 2A). During a “continuous target” task, localization began immediately on stimulus presentation. For paradigms involving visual target localization, a “transient target” (memory) task was used, in which all stimuli were briefly presented (5 bursts and/or flashes of 150 ms for 1 s) and...
extinguished before the initiation of localization. The subject signaled the response endpoint with the key press. The onset of the masking noise and the repositioning of the speaker followed immediately in the intertrial interval. The masker was switched off once the speaker reached its new target location and the next trial began without delay.

For every experimental paradigm, two sessions (on different days) of \( \pm 240 \) trials were parsed into four to five separate but contiguous fixation epochs (based on the number of trials per epoch and epoch duration). Sessions began and ended with an epoch of central fixation (0° Az, 0° El), interjected by alternating epochs of eccentric ocular fixation at ±20° Az or El (Fig. 3A). The order of eccentricity was reversed between sessions. The paradigm produced controlled gaze shifts of 20° (e.g., 0 to ±20°) and 40° (e.g., +20 to −20°) and served to maintain fixation during epochs. Each epoch lasted 4–14 min depending on the number of trials it contained and subject response times. Each session lasted 30–50 min.

Auditory and visual targets consisted of 21 locations within, and 22 beyond, ±10° Az by ±10° El, totaling 43 locations over ±25° Az by ±25° El (Fig. 1C). All trial conditions, stimuli, and speaker locations were randomly interleaved from trial to trial within each epoch to prevent prediction. Four separate experimental paradigms were used to answer questions regarding the effect of eye position on spatial localization.

GAZE IN DARKNESS WITH INTERMITTENT FIXATION REFERENCE (9 SUBJECTS: 5 FEMALES AND 4 MALES). In previous studies using the Alternating Fixation paradigm (Razavi et al. 2007), subjects localized BB (0.1–20 kHz) auditory targets while holding gaze on continuously fixation location in the dark while continuing to localize auditory targets. Two sessions of 161 trials were parsed into five epochs and tested for each reference condition (continuous and intermittent, for a total of 4 sessions per subject).

SOUND LOCALIZATION ACROSS SPATIAL CHANNELS (8 SUBJECTS: 3 FEMALES AND 5 MALES). To directly assess the influence of sustained eccentric eye position on sound localization for targets that specifically emphasized different spatial channels, we used BB, HP (3–20 kHz, to emphasize IID cues), and LP (0.1–1 kHz, to emphasize ITD cues) noise stimuli during epochs of sound localization. In different Alternating Fixation experiments, epochs alternated in either Az (at Ctr, L, or R20°) or El (at Ctr, U, or D20°). Two sessions of 240 trials were parsed into five epochs and tested for both Az and El (total of 4 sessions).

PERCEIVED STRAIGHT-AHEAD (7 SUBJECTS: 4 FEMALES AND 3 MALES). The subjective PSA was assessed in two ways. First, interleaved with trials of BB auditory localization, subjects were asked to maintain fixation (at Ctr, L, and R20°) while guiding the laser pointer to indicate gaze-fixed PSA (head-referenced) in otherwise dark silence. Second, at the end of each epoch, and at the beginning

![Graph](Image link)

**FIG. 3.** Sound localization during horizontal gaze shifts with continuous or intermittent visual fixation reference. A: sound localization accuracy (error between response and target location) in Az is shown during either continuous (left) or intermittent (right) presentation of a fixation reference for a representative subject. Each session consists of 5 continuous epochs in which fixation was cued by laser-projected spots that alternated between center, right 20°, and left 20° (thin gray trace). Responses are offset by average pointer error during the initial epoch of central fixation. Errors for individual trials (filled circles) show a shift in localization over time in the direction of fixation. The exponential model (black trace; see equation at right) predicted time constants (\( \tau \)) and shift magnitudes (\( a \)) for each epoch.

B: average shift magnitude in response to 40 and 20° gaze shifts (each pooled for L and R) during continuous and intermittent fixation references, calculated as end-epoch averages (Δ accuracies between epochs) and model estimates (a). Values are comparable between calculation methods and reference conditions (\( P < 0.05 \)). Error bars indicate SD across 9 subjects. The format and conventions here apply to similar figures except as noted.
and end of each session, the fixation spot was extinguished and subjects were instructed to indicate gaze-guided PSA with both their eyes and the laser pointer (6 trials per session). These two types of PSA trials assessed, respectively, the change in PSA over time in response to sustained eye position and its immediate residual effects after breaking fixation. In both cases, subjects were instructed to draw an imaginary line straight-ahead from the tip of their nose to the screen and align the pointer with this location. Two sessions of 222 trials were parsed into four epochs.

MULTISENSORY TARGET LOCALIZATION (9 SUBJECTS: 5 FEMALES AND 4 MALES). Preliminary experiments from our laboratory failed to show a spatial shift in the localization of memorized visual targets in response to eccentric eye position (Razavi et al. 2005). It is conceivable that visual targets do in fact respond to eye eccentricity but require a longer fixation time than was allotted in our preliminary studies or that a concurrent shift in auditory space is necessary to induce a similar response in visual targets. To examine these possibilities and the effects of sustained eye position on multisensory spatial localization, BB auditory, visual (speaker LED), and bimodal (spatially and temporally congruent auditory and visual targets) stimuli were localized during our typical Alternating Fixation paradigm in Az (Ctr, L, or R20°). With bimodal targets, subjects were instructed to attend to both auditory and visual stimuli but to “only localize the sound target.” To ensure subjects did not simply match the laser pointer to an ongoing target-LED during localization, all stimuli were transiently presented (5 x 150-ms bursts and/or flashes over 1 s) and extinguished before the initiation of localization (“transient target” task). Two sessions of 216 trials were parsed into four epochs.

Task and eye position monitoring

To ensure that subjects maintained the intended ocular eccentricity while performing spatial localization tasks, eye position was monitored by electrooculography (EOG) at all times during experimental sessions while subjects maintained gaze on visual fixation targets. In previous experiments (Razavi et al. 2007) using EOG to monitor eye position during the same task used here, subjects had little trouble maintaining fixation and only rarely made eye movements. Interruptions in fixation typically involved saccades away from, and then immediately back to, the fixation spot, and occurred in 0–8% (median, 2%) of trials in a given experimental session. Importantly, the adaptation of auditory space by eye position remained unaffected by such brief breaks in fixation.

During one of our paradigms, the Gaze in Darkness with Intermittent Fixation Reference condition, subjects maintained eccentric gaze largely in darkness in the absence of a visual fixation spot. For this case, a CCD camera-based eye tracker system (EL-MAR, Toronto, Ontario, Canada) was used to monitor and record eye position binocularly. In contrast to the EOG, the advantage of the EL-MAR eye-tracker lies in its lack of drift over time and thus its ability to record absolute horizontal and vertical eye positions with high accuracy and sensitivity (~0.2°) over prolonged periods (DiScenna et al. 1995). The disadvantages of the EL-MAR eye-tracker are that it must be worn as a headgear and is substantially bulkier and heavier than the simple skin electrodes of the EOG. Further, additional time is necessary for calibration and set-up, and the visual field is restricted to ±45° Az by ±30° El. For this particular task, the EL-MAR tracker was important for continuous eye position monitoring in darkness. For all other tasks, the EOG method was sufficient and far less obtrusive. Also, because the EL-MAR headgear has the potential to alter auditory spatial cues, sound localization controls were performed with and without the apparatus, and results proved unaffected by the headgear in the range of Az relevant to the experiment.

For either the EOG or the EL-MAR method, inadvertent eye movements were always noted by the experimenter, and subjects were reminded to maintain fixation on the projected reference spot. Subjects were also encouraged to report (by room intercom) instances when fixation was interrupted; voluntary reporting promoted subject vigilance and emphasized the importance of maintaining steady fixation. Such self-reports were very rare across multiple experimental sessions. In addition to eye-position monitoring, real-time tracking of the laser pointer and speaker positions also ensured that subjects were performing the tasks correctly.

Data analysis

Data were analyzed using Matlab (The MathWorks), Excel, SigmaPlot, and SYSTAT statistical software. Response accuracy values (pointer error relative to target position) across the spatial target array were sorted by stimulus type (e.g., auditory, visual, and bimodal) and fixation condition (e.g., at Ctr, R20, L20, U20, and D20°). Values for each session were offset by the average pointer error during the initial central fixation epoch, which served as a zero reference for the session. This offset was small, typically between 1 and 4° across subjects, and within 2° for a given subject across sessions.

The effect of eye eccentricity was quantified as the difference between end-epoch averages (Δ accuracy; differences in mean accuracy between the final trials of the current epoch and that of the previous epoch) for each stimulus. Depending on epoch duration, the last 10 or 12 trials of each epoch were included in the analysis, such that all localization trials have approached the asymptote of spatial shift (cf. Fig. 3A).

To further quantify both the dynamics and magnitude of auditory spatial shift for each epoch (after each refixation), accuracy values were modeled parametrically using the first-order exponential equation

\[ y(t) = y_0 + a(1 - e^{-\tau}) \]  

yielding a time constant (τ) for the adaptation (with 1/τ as its rate) to eccentric fixation and a model estimate of shift magnitude (a) from starting point (y₀, based on the final shift magnitude of the previous epoch) to asymptote (y∞ = y₀ + a) (Razavi et al. 2007).

The slope of the linear regression of response versus target location was used to quantify spatial gain (SG) for the response accuracies across target positions; SG = 1 represents perfect performance, whereas values >1.0 and <1.0 indicate over- and undershoot of target position, respectively.

Because localization performance was imperfect (i.e., SG ≠ 1) and could differ between subjects, target types (e.g., auditory vs. visual versus PSA), or conditions (e.g., eye-fixed vs. eye-free), SG acted as a confounding variable that added noise to our data (particularly in the auditory periphery) and degraded the model fit. A 10% overshoot in SG, for example, results in a 4° error in localization accuracy for a target at 40°. To isolate the effect of eye position on localization accuracy across space, and particularly shift magnitude over time, it was useful to normalize all data points within each session by subtracting epoch- and stimulus-specific SG (regression slope) from all raw data sets before modeling. In other words, normalized accuracy = response location – SG (target location). We showed earlier that SG is not influenced by prolonged eccentric gaze (Razavi et al. 2007). Normalization thereby eliminated the effect of SG on individual localization responses and focused the analysis on the shift.

In sum, the change in localization performance as a result of eye position (shift magnitude) was quantified in two ways: nonparametrically as end-epoch average (Δ accuracy) and parametrically as model estimate (a). Because the end-epoch average includes localization responses across a roughly balanced range of target eccentricities, the influence of SG across these trials is averaged out, leaving any shift intact. This is based on our determination that end-epoch average calculations with and without SG subtraction proved comparable within subjects. That the end-epoch average closely reflects the model fit at roughly the same points in the data set re-emphasizes the selectivity of eccentric gaze on overall spatial shift.
For all four experimental paradigms, the two methods generated calculations of shift magnitudes that proved statistically indistinguishable within experiments ($P \geq 0.28$, paired Student’s $t$-test), which indicates that the subtraction of SG most likely did not affect model estimates of final shift magnitude. Moreover, the values of shift magnitude generated by the two methods are highly correlated. A linear regression of end-epoch averages versus model estimates of shift magnitude for the Gaze in Darkness with Intermittent Fixation Reference experiment resulted in a slope of 0.97 ($R^2 = 0.61$). Results will emphasize either or both approaches as appropriate, and figures provide both for reference.

Algebraic average (mean) and SD were used as the measures of central tendency and variability, respectively, across subject populations for all summary statistics. Comparison between epochs, stimuli, and experimental paradigms were quantified using standard parametric statistics (Student’s $t$-test and ANOVA with Tukey’s HSD, as appropriate), and whenever possible, used paired differences (individual or pooled target locations) within subjects (i.e., paired-sample Student’s $t$-test). $P$ values were also calculated to assign statistical significance to regression and model parameters (e.g., SG, $a$, and $\tau$) whenever regression analysis and modeling were used to parameterize relationships between variables.

RESULTS

Gaze in darkness with intermittent fixation reference

To address the effects of eye eccentricity on auditory localization with and without a persistent visual fixation reference, subjects localized BB auditory targets while holding gaze on either continuously or intermittently presented fixation spots in two separate Alternating Fixation sessions. The fixation spot in the intermittent condition was presented every fifth trial during speaker movement, always between trials, and extinguished before target presentation. Thus targets were always presented and localized in darkness in the intermittent condition. The time course of the shift in sound localization after changes in eye position under these conditions was analyzed as described above.

After a change in eye position, sound localization immediately shifted in the direction of gaze, rising rapidly at first, but slowing to approach a steady state (Fig. 3A). For both fixation conditions, the dynamic (rate) and static (magnitude) components of the auditory spatial shift proved variable from subject to subject but remained similar between conditions. Initial gaze shifts away from central fixation exhibited the longest time constant (slow), whereas final gaze shifts returning to center exhibited the shortest time constant (fast). As a measure of overall precision, the average SD of the localization error was statistically comparable between the two reference conditions [2.89 ± 0.76 and 3.04 ± 0.96° (SD), respectively; $P = 0.49$].

Model-derived estimates of shift magnitude (see $a$ in Eq. 1) were also statistically indistinguishable between fixation reference conditions. In both conditions, model estimate of shift magnitude was greater for a 40° change in fixation ($\sim 27\%$ of gaze shift; $a = 10.81 \pm 3.59$ and 10.49 ± 3.81° for continuous and intermittent, respectively; $P = 0.59$) than for a 20° change in fixation ($\sim 30$ and 28%; $a = 5.98 \pm 2.43$ and 5.63 ± 2.12°; $P = 0.56$). The end-epoch averages ($\Delta$ accuracy; calculated as differences between the final 10 trials of each epoch) also proved statistically indistinguishable between reference conditions ($P \geq 0.56$; Fig. 3B).

Certain features of the temporal dynamics were also common for both continuous and intermittent reference conditions. The time constant was directionally asymmetric and individualized, being longer when the eyes moved away from center ($\tau = 57 \pm 32$ and 95 ± 51, respectively) than when they returned to center (32 ± 24 and 52 ± 36), and slower for a 40° (49 ± 39 and 97 ± 78 s, respectively) than for a 20° change in eye position (37 ± 25 and 77 ± 44 s). Although these general observations were not significant statistically ($P > 0.05$), in contrast to shift magnitude, time constants proved consistently greater for the intermittent reference condition compared with the continuous one ($\tau = 84 \pm 53$ and 43 ± 23 s, respectively; $P < 0.01$).

These results definitively show that the auditory spatial shift was preserved in the near absence of a visual fixation reference in the intermittent condition. That is, despite the lack of visual feedback, sound localization still shifted by the same magnitude and direction as when subjects fixated a persistent visual reference.

In the continuous condition, fixation was monitored using EOG and proved steady over the course of each epoch. Brief saccadic interruptions to fixation proved rare. In a previous study using the same experimental paradigm (Razavi et al. 2007), records of EOG were inspected off-line in great detail, and it was shown that the average shift magnitude for the three trials before and after breaks in fixation were within 1 SD of each other. Thus the overall auditory spatial shift remained effectively unaltered by brief eye movements intruding on steady fixation.

Other than the intermittent presentation of the fixation spot, the Alternating Fixation paradigm used in our study was identical to the one in the Razavi et al. (2007) study, and the statements above can be reasonably applied to our findings as well. Off-line analysis of eye-tracker (EL-MAR) data in the intermittent condition indicates that subjects had little trouble maintaining eye position in the dark as instructed, although there were occasional variances within a range of 2–15° (median ~3°) around the intended point of fixation. Saccades were typically centripetal, small, and infrequent, occurring in only 0–21% (median ~6%) of trials in a given experimental session. Fixation proved somewhat steadier for L than for R fixation epochs, possibly accounting for the slight difference in shift magnitude between L and R fixations. Importantly, however, average eye position across epochs remained near the desired 20° L or R in darkness (−19.4 and 18.7° for L and R20° gaze, respectively) in between presentation of the fixation spot. In other words, the overall quality of fixation during experimental tasks had little to no effect on the overall eye position during alternating fixation, and auditory space shifted over time under all fixation conditions.

Sound localization across spatial channels

To examine the effect of eye position on specific auditory spatial channels, high-pass (HP; 3–20 kHz) and low-pass (LP; 0.1–1 kHz) targets were interleaved and localized along with broadband (BB; 0.1–20 kHz) targets while fixation alternated ±20° in either Az or El (Fig. 4A). Whereas BB sounds activated all auditory spatial channels during localization, the HP stimulus emphasized IID and pinna-dependent spatial cues, whereas the LP stimulus contained primarily ITD cues, thus...
allowing for the examination of frequency and channel-specific effects of sustained eye position.

As before, the shift in auditory spatial perception for HP and LP targets was estimated using a first-order exponential model (Eq. 1) to quantify the magnitude ($\alpha$) and dynamics ($\tau$) of the shift. For this paradigm, 12 trials of BB localization were included and interleaved only at the end of each epoch. This experimental design minimized session time and eye fatigue while still enabling the nonparametric quantification of eye position effects as end-epoch average ($\Delta$ accuracy; the difference in localization error between the final 12 trials of the current epoch and previous epoch) for all three types of auditory targets.

After each fixation change in either Az or El, sound localization shifted exponentially for all spatial channels in the direction of ocular eccentricity (Fig. 4A, gray trace). The magnitude of the auditory spatial shift was highly variable from subject to subject and often directionally asymmetric.

LOCALIZATION IN AZ. In Az, model estimates of shift magnitude proved indistinguishable between HP and LP targets. Shift magnitudes were greater for a 40° change in fixation ($\alpha = 12.02 \pm 4.34$ and $11.33 \pm 3.89°$ for HP and LP targets, respectively) than for a 20° change ($\alpha = 7.07 \pm 3.02$ and $7.00 \pm 2.36°$), but corresponding to roughly the same proportion of gaze shift (30–35%). End-epoch averages for a 40° change in fixation were also similar among all three target types in Az ($\Delta$ accuracy $= 12.43 \pm 4.61$, $12.82 \pm 4.45$, and $12.69 \pm 4.25°$ for BB, HP, and LP stimuli, respectively; $P = 0.72$; Fig. 4B). Not surprisingly, the end-epoch average for the BB targets was comparable to that for the continuous fixation reference condition ($P = 0.44$). As might be expected because of the paucity of ITD cues in the HP stimulus, Az precision (i.e., variance, the average intrasubject SD of end-epoch average estimates of shift magnitude) was higher for HP targets ($5.26 \pm 2.28°$) than for the BB and LP targets ($3.38 \pm 0.99$ and $3.64 \pm 1.06°$, respectively; $P < 0.01$).

Time constants were also comparable between LP and HP targets ($\tau = 78 \pm 37$ vs. $65 \pm 27$ s, respectively; $P = 0.17$). Epochs when the eyes returned to center were significantly faster ($\tau = 31 \pm 22$ and $50 \pm 44$ s for HP and LP targets, respectively) than when they shifted eccentrically ($\tau = 136 \pm 92$ and $98 \pm 35$ s for HP and LP targets, respectively; $P < 0.05$), suggesting a preference for recouping the original spatial coordinates after an adaptive shift. Although the experimental design did not allow for the assessment of BB temporal dynamics within this paradigm, when compared with BB targets in the continuous fixation reference condition ($\tau = 43 \pm 23$ s; see above), a weak difference was noted only for LP targets ($P < 0.05$).

LOCALIZATION IN EL. Whereas SG was comparable for all three auditory targets in Az ($-1.20$), SG in El differed significantly between LP and both HP and BB target types. This result is to be expected, because the lack of high-frequency spectral cues in LP noise functioned to impede its localization in the vertical dimension. Interestingly, despite the fact that EL SG for LP targets was only ~49% of HP targets (0.53 vs. 1.19, respectively), model estimates ($\alpha$) of the shift for HP and LP targets remained statistically indistinguishable, reaching...
16.56 ± 3.60 and 16.54 ± 4.19° (41%; \( P = 1.00 \)) for a 40° El change in fixation and 8.53 ± 3.57 and 9.43 ± 3.66° for a 20° change in fixation (43 and 47%; \( P = 0.34 \)), respectively. End-epoch averages in El were similarly comparable, approaching ~43% of ocular eccentricity for a 40° change in fixation across all three auditory targets (Δ accuracy = 17.30 ± 3.98, 17.41 ± 4.97, and 16.90 ± 4.09° for BB, HP, and LP stimuli, respectively; \( P = 0.81 \); Fig. 4B). In other words, although subjects were unable to localize the vertical position of LP targets accurately, all targets were still perceived to be shifted in the direction of sustained eye position, suggesting that shift magnitude can change independently of SG. As might be expected because of the lack of spectral cues in the LP stimulus, precision/variance of end-epoch averages in El was higher for LP targets (6.68 ± 1.42°) than for the BB and HP targets (4.90 ± 1.39 and 4.93 ± 0.87°, respectively; \( P < 0.01 \)). Finally, shift magnitudes for all three targets in El differed significantly from their counterparts in Az (\( P < 0.05 \)), pointing to a difference in the effect of eye position on horizontal and vertical localization not shown previously.

In contrast to Az, the time constants of the shift in El proved variable. \( T \) for LP targets was greater than for HP targets (106 ± 44 vs. 68 ± 32 s, respectively); however, the difference was just short of significance (\( P < 0.06 \)). A faster dynamic for LP targets was noted for epochs when the eyes returned to center (\( \tau = 26 ± 18 \) s) than when they moved eccentrically (\( \tau = 131 ± 91 \) s; \( P < 0.01 \)). Although a similar trend was noted for HP targets (\( \tau = 37 ± 61 \) vs. 97 ± 80 s, respectively), these differences were also not significant (\( P = 0.17 \)). The shift in LP target localization occurred at a slower rate for epochs with a ~40° change in gaze (\( \tau = 208 ± 105 \) s) compared with all other epochs (\( \tau = 73 ± 31 \) s; \( P < 0.01 \)). This finding might be related to an initial upward localization bias during central fixation for LP (6.13 ± 9.27°) but not for BB (−0.01 ± 8.55°) and HP (2.77 ± 6.07°) targets.

Perceived straight-ahead

Given the profound effect that eye position had on auditory localization, we wondered whether adaptation of auditory space was also accompanied by, or even a result of, a more general shift in the perception of egocentric space. This paradigm directly addressed PSA during auditory spatial adaptation to changes in eye position. Subjects performed an Alternating Fixation task that interleaved trials of PSA assessment with localization of BB sound. Perceptual changes in straight-ahead were quantified in two ways. First, subjects were asked to maintain fixation on the reference spot and guide the laser pointer with peripheral vision to indicate the location of PSA (gaze-fixed condition). Second, while the reference spot remained off at the beginning and ends of each session and during epoch transitions, subjects foveally guided the laser pointer to indicate the location of PSA (gaze-guided condition).

GAZE-FIXED PSA. The auditory spatial adaptation was paralleled by an even greater shift in the perception of straight-ahead while the eyes remained fixed and eccentric. The end-epoch average (Δ accuracy; including the last 11 trials of each epoch) of gaze-fixed PSA reached 58% of ocular eccentricity for 40° gaze shifts in Az (23.28 ± 8.47°), far exceeding that for auditory targets under the same conditions (39%; 15.63 ± 4.04°; \( P < 0.05 \)). The dynamics and magnitude of auditory and PSA shifts for 40° changes were quantified using the model of Eq. 4. Model estimates of the shift \((a)\) averaged 23.82 ± 7.49° for gaze-fixed PSA and 15.74 ± 3.54° for auditory targets (Fig. 5, A and B). The temporal dynamics of the shift in gaze-fixed PSA proved comparable to sound localization (\( \tau = 74 ± 59 \) and 69 ± 53 s, respectively; \( P = 0.87 \); Fig. 5A).

Interestingly, the end-epoch average calculated for BB auditory targets in this paradigm was larger than that for the continuously presented fixation reference paradigm (Δ accuracy = 39 and 26%, respectively; \( P < 0.005 \)).

GAZE-GUIDED PSA. Even when subjects were free to make eye movements, the position of perceived straight-ahead persisted in the direction of previously maintained gaze. The shift magnitude of gaze-guided PSA trials was ~25% of gaze (or 9.84 ± 6.02°), and proved to be less than the end-epoch averages of the gaze-fixed condition (\( P < 0.05 \)).
Multisensory target localization

A remaining question is whether auditory spatial adaptation with changes in eye position persists when visual targets are also included. This query was motivated by the well-known effect of visual capture, a correlate of the familiar ventriloquism effect, whereby the localization of auditory targets is biased in the direction of a modestly displaced concurrent visual object (Alais and Burr 2004; Battaglia et al. 2003; Bertelson and Radeau 1981; Recanzone 1998). To evaluate the interaction between eye position effects and audio-visual spatial interactions, auditory, visual, and bimodal (concurrent auditory and visual) targets were randomly interleaved and localized during an Alternating Fixation paradigm (Fig. 6A). Because all sessions included visual targets, stimuli (five 150-ms sound bursts and/or flashes over a 1-s period) were presented and extinguished before the initiation of localization in a “transient target” memory task (Fig. 2A).

The effect of eye eccentricity was quantified as end-epoch averages (Δ accuracy between the final 10 trials of the current and previous epoch) for auditory, visual, and bimodal targets. It reached 26% of eccentric gaze (10.38 ± 3.50°) for auditory targets in response to 40° horizontal changes in fixation (Fig. 6B), a value comparable to the continuous fixation reference condition noted above (P = 0.47).

Localizing of visual targets. Transforming eye-to-head-centered coordinates (i.e., room center = 0, 0 regardless of eye position), this shift in sound localization appeared to be accompanied by a small but significant shift in the memorized location of visual targets toward the direction of gaze (~7%; Δ accuracy = 3.02 ± 2.62°; P < 0.01). On further analysis, however, this shift in visual localization proved to be caused by the different reference frames of visual versus auditory space.

When the eyes shift from a central to an eccentric target, the images that previously fell on the central retina now occupy the peripheral retina. In contrast, because our subjects were head-fixed, the relationship between auditory targets and the ears remains unaffected by ocular shift. Thus by virtue of the fact that auditory space is head-referenced while visual space is eye-referenced (i.e., 0 indicates foveation), space as perceived by the two sensory modalities becomes misaligned during eccentric fixation.

We reanalyzed the visual localization data to address any effect of changing visual reference frames that might erroneously result in a shift of visual localization caused by eye movements. When localization errors of memorized visual targets were remapped with respect to their position on the retina during central and eccentric fixation, response profiles (accuracy and gain) proved comparable for targets that occupied the same retinal eccentricities regardless of actual target positions with respect to the room (and the head). A remapping of memorized auditory targets, in contrast, retained the localization shift and failed to show a similar alignment of response profiles (Fig. 7). This result supports the notion that Δ accuracy calculations for visual targets, unlike those for auditory, do not reflect an eye position–dependent gradual shift in localization but are...
instead manifestations of diminished localization accuracy and spatial distortions as visual images are projected onto peripheral instead of central retina.

LOCALIZATION OF BIMODAL TARGETS. When localizing bimodal targets, subjects were instructed to “point to the sound.” The end-epoch average for bimodal targets was 50% of auditory targets alone averaged across the subject population (14%; 5.59° ± 4.02° for a 40° gaze shift; Fig. 6B). However, when auditory and visual Δ accuracies were normalized across a gradient from 0 (fully vision-dominant) to 1 (fully auditory-dominant), bimodal end-epoch averages proved highly individualized, falling on a continuum between the two extremes (average 0.35, and significantly different from 0 or 1; P ≤ 0.01; Fig. 8). Individual values were also unrelated to the magnitude of the auditory or visual shift or to the differences in magnitude between the two. Thus, the large intersubject differences in shift magnitude for bimodal targets reflect both individual abilities to identify the location of auditory targets and the tendency for sound localization to be captured by vision.

As with other Alternating Fixation paradigms, we modeled the dynamic shift in localization over time using Eq. 1. Resulting estimates of temporal dynamics proved variable but comparable between auditory and bimodal targets (τ = 63 ± 43 and 80 ± 77, respectively; P = 0.15). For both auditory and bimodal targets, model estimates of shift were greater for a 40° change in fixation (~28 and 16%; a = 11.16 ± 3.63 and 6.56 ± 3.63°, respectively; P < 0.005) than for a 20° change (~27 and 19%; a = 5.35 ± 1.99 and 3.79 ± 2.04°; P < 0.005; Fig. 6A).

DISCUSSION

This study addressed the congruence of auditory and visual spatial perception in the presence of eye movements that misalign spatial coordinates between the two sensory modalities. Previously, we reported that the perception of sound location adapts to sustained eccentric eye position through a generalized shift in its direction (Razavi et al. 2007). This effect of eye position is not restricted to visually guided pointing, because the adaptation also occurs when pointing with a stick in the dark in the absence of visual feedback (Razavi et al. 2005). The experiments reported here address the extent to which a true physiological adaptation of auditory

![Figure 7](image-url) localization accuracy in Az for visual (left) and auditory (right) targets during fixation at either center, left 20°, or right 20°. Target positions in Az are plotted with respect to both room-referenced (top) and retinal-referenced (bottom) coordinates. For head-fixed subjects, room-referenced coordinates are also head-referenced. In contrast, retinal-centered coordinates (0° indicates the fovea) change with eye position. Differences between the 3 fixation positions disappear when visual targets are remapped with respect to their location on the retina; auditory targets retain the same shift in target localization regardless of the frame-of-reference.

![Figure 8](image-url) Auditory and visual influences on the eye position-dependent shifts of bimodal targets. For each subject in the Multi-Sensory Target Localization paradigm, the mean shift magnitude for bimodal targets is normalized as a fraction between 0 (V; matched to visual localization shift alone) and 1 (A; matched to auditory localization shift alone). The normalized bimodal shifts for each subject (light gray bars) show a continuum of magnitudes from those that resemble visual targets alone to those that mimic auditory targets alone. The average bimodal shift across all 9 subjects (dark gray bar; 0.35 ± 0.27) shows a slight bias toward visual targets.
space by eye position persists across several important challenges and whether it encompasses other sensory modalities.

Specifically, we re-evaluated the effect of eye position on spatial perception in four experiments designed to determine whether 1) auditory adaptation occurs without the presence of visual inputs; 2) adaptation to eye position is selective for individual auditory spatial channels (i.e., ITD, IID, and Spectral); 3) egocentric straight-ahead (PSA) shifts alongside auditory spatial perception; and 4) a change in eye position also affects visual spatial perception.

Experimental results showed that 1) the adaptation in auditory space is a response to eye position alone, because the absence of visual input does not impact the magnitude of the shift; 2) the adaptation to eye position extends equally to specific frequency bands that isolate auditory spatial channels; 3) PSA adapts to a new eye position in a similar manner to sound localization; and 4) memorized visual targets do not adapt to sustained eccentric eye position. Both predicted (model estimate) and actual (end-epoch average) values of auditory shift magnitude proved statistically comparable for each of the four experiments, thereby validating our parametric modeling approach to data analysis. Furthermore, because linear regression estimates of SG were subtracted from all data sets during the calculation for the predicted but not the actual shift magnitude, the results strongly suggest that the two factors of SG and shift are independent variables. We conclude that our previous assertion of an eye position–dependent adaptation of a physiological “spatial zero” indeed holds and includes nonauditory perception of straight-ahead in addition to sound localization.

Gaze in darkness with intermittent fixation reference

Past studies examining the effect of eye position on auditory localization, including our own, have generally provided a visual fixation reference to maintain the intended ocular eccentricity (Getzmann 2002; Lewald 1997, 1998; Lewald and Ehrenstein 1998, 2000; Lewald and Getzmann 2006; Metzger et al. 2004; Razavi et al. 2007; Weerts and Thurlow 1971; Yao and Peck 1997). The possibility remains that the visual fixation reference (i.e., the retinal signal) functions to facilitate the adaptation, and the auditory spatial shift is in fact not a response to eye position alone. We examined the role of visual inputs in the adaptation of auditory space to eye position by restricting fixation spot presentation to occasional bursts in between localization trials and found that the lack of a continuous fixation reference had no observable effect on final shift magnitude. That the adaptation proceeds unabated in darkness is particularly notable during the initial period immediately after an eye movement, when spatial shift is most rapid. It is worth emphasizing that the fixation spot in the intermittent reference condition was only present during the intertrial interval and was always extinguished before stimulus presentation, so that localization always occurred during darkness.

While the final magnitude of the adaptation is comparable between reference conditions, the shift during the intermittent condition developed more slowly (larger time constants). In other words, although the same shift magnitude was ultimately reached regardless of the state of fixation, more time was required to do so when gaze was maintained in the dark. The lack of a consistent retinal input in the intermittent condition resulted in a less consistent and weakly attenuated average eye position (−1° for 20° eccentricity) and might have functioned to elongate and modify the temporal dynamics of the shift.

In a study by Lewald and Ehrenstein (1996), human subjects under headphones were asked to adjust the IID of noise stimuli to “point” the lateralized sound image toward the median plane after a very brief period of eccentric gaze (±2 s). The auditory median plane was shown to shift in the direction of gaze, albeit with a slight decrease in magnitude when gaze was maintained in the dark. Based on our findings, it is clear that the slower temporal dynamics of the shift in the intermittent condition would indeed result in a difference in shift magnitude between reference conditions after such a brief duration of eccentric gaze, but this difference would disappear if the measurements were carried out to asymptote. Thus this previously reported difference in shift magnitude between reference conditions is only a transient feature of spatial adaptation.

Interestingly, monitoring of eye position in the intermittent reference condition showed an occasional loss of binocular fixation during eccentric gaze, where one eye remained on the remembered target position while the more eccentric eye drifted away, often toward center. Although infrequent, this convergence might have contributed to a less robust and consistent eye position signal, thus prolonging the time course of auditory adaptation in the absence of a fixation reference. This observation is corroborated by subject reports of occasionally seeing two images on reintroduction of the fixation spot, possibly as a result of eye fatigue after prolonged eccentric gaze in darkness.

Despite differences in temporal dynamics, we can reasonably conclude that the influence of eye position on human sound localization does not require a retinal image and the spatial adaptation is in response to oculomotor signals related to eye position alone.

Sound localization across spatial channels

ITD, IID, and spectral cues, following initial processing along parallel but separate neural channels (Heffner and Masterton 1990), ultimately converge to resolve spatial ambiguities and provide a unified map of auditory space (Brainard et al. 1992). Eye position modulation of neural responses to auditory spatial stimuli has been found in the inferior colliculus (Groh et al. 2001; Zwiers et al. 2004), superior colliculus (Hartline et al. 1995; Jay and Sparks 1984; Peck et al. 1995; Populin et al. 2004; Van Opstal et al. 1995; Zella et al. 2001), the auditory cortex (Fu et al. 2004; Werner-Reiss et al. 2003), frontal eye fields (Russo and Bruce 1994), and the intraparietal cortex (Mullette-Gillman et al. 2005; Stricanne et al. 1996). By examining the effect of eye eccentricity on individual spatial channels, we sought to demarcate the scope of the eye position effect in auditory spatial perception and ascertain potential loci of adaptation within the auditory pathway.

Our results indicate an auditory spatial adaptation to ocular eccentricity that is generalized for all band-limited sounds carrying interaural differences and pinna cues, and by extension, all auditory spatial channels. Both magnitude and dynamics of the adaptation are statistically comparable between the relevant frequency bands within Az and El. Shift magnitudes in El, however, surpassed that of Az. Because the oculomotor range is greater in Az (±60°) than in El (+44 and −50°) (see
Von Noorden 1985), and sound localization in the two dimensions is achieved very differently, it is possible that a 20° gaze shift constitutes a greater adjustment in El than in Az, thereby effecting a greater adaptive response.

Furthermore, even though El SG and localization accuracy for LP targets are poor because of lack of high-frequency spectral cues in the stimulus, shift magnitudes remain comparable for all three auditory targets. This finding directly supports the notion that SG for Az is independent of that for El and that, because localization accuracy (or lack thereof) does not affect the spatial response to eye position in either Az or El, the adaptive shift (or “set-point”) is independent of SG. Importantly, the results show that the adaptation of auditory perception to eye position crosses all spatial channels and functions independently of channel-dependent factors impacting spatial differentiation in Az or El. Therefore we propose that the locus of adaptation most likely resides within the neuraxis at, or downstream from, the level where all auditory channels are integrated into a unified map of auditory space, i.e., rostral to the IC.

Perceived straight-ahead

To determine whether sustained eccentric eye position affects not just sound localization but also a more global determination of egocentric spatial “zero” as we had previously hypothesized, PSA was assessed in two ways. Gaze-fixed trials examined changes in PSA while the eyes remained fixed and eccentric. Gaze-guided trials quantified the “after-effect” of sustained eccentric gaze immediately after the eyes were released from fixation. While the eyes remained eccentric during the gaze-fixed condition, the auditory spatial shift induced by a change in ocular eccentricity was paralleled by an even greater shift in the subjective perception of straight-ahead. This perceptual shift developed quickly after a change in eye position with temporal dynamics and time constants comparable, although not identical, to those for sound localization.

Immediately after breaking fixation and resuming eye movements in the gaze-guided condition, the shift in PSA became attenuated, but nevertheless remained deviated in the direction of previously maintained gaze and deviated sound localization. These findings add nuance to the perception of PSA and how it is assessed. However, given the environmental cues and postural references to objective straight-ahead provided by the bite-bar, laser pointer, and the subject chair, the fact that PSA changed at all, much less robustly so, is a testament to the potency of the adaptation. Although our finding of a PSA shift in the direction of gaze corresponds to other estimates of subjective center reported previously (Bohlander 1984; Lewald and Ehrenstein 2000; Weerts and Thurlow 1971), this experiment, to the best of our knowledge, marks the first time PSA has been assessed to quantify both the magnitude and time course of the change in response to a sustained change in eye position.

An interesting observation during PSA experiments is that, whereas the shift magnitudes of auditory targets are comparable among the other three experimental paradigms, the shift of auditory localization interleaved with PSA assessments proved greater than that for auditory trials in the continuous fixation reference experiment. Because subject instructions and experimental set-up were identical in both experiments, the main difference lies in the frequent assessments of straight-ahead in the PSA experiment. It is conceivable that the continuous attention paid to PSA, which was greatly biased toward eye position, functioned to facilitate/enhance the auditory spatial adaptation.

In a prior preliminary assessment of gaze-fixed PSA in concert with sound localization, PSA was assessed with less frequency (8 of 121 trials) over the course of a single session. As was the case with auditory targets, the PSA shift in this study proved to be greater than that shown previously (3.63 ± 1.36°), despite similar experimental conditions. It is possible that here too, the greatly augmented assessment frequency in this study served to highlight and facilitate the perceptual shift in PSA over time.

Multisensory target localization

Past studies that have assessed the interaction between eye position and visual spatial perception in both humans and primates have generated largely inconsistent and equivocal findings (Bock 1986, 1993; Ebenholtz 1976; Henriques et al. 1998; Hill 1972; Lewald 1998; Lewald and Ehrenstein 2000; Metzger et al. 2004; Morgan 1978; Yao and Peck 1997). In this experiment, we interleaved localization of transiently presented auditory, visual, and bimodal (visual and auditory) targets to quantify the effect of a change in eye position on memorized targets of both modalities individually and in combination. The inclusion of bimodal targets, where visual and auditory stimuli were presented together from the same spatial location, served to examine perceptual interactions between vision and audition. We focused on sound localization, thereby allowing for the potential influence of visual capture during bimodal target presentation.

Our findings showed definitively that sustained eccentric eye position (≤10 min) does not evoke a spatial shift in visual target localization. An apparent small head-centered “shift” in the perception of visual targets during eccentric fixation disappears when localization errors are remapped in eye-centered coordinates (i.e., where 0.0 is in line with the fovea). The shift in auditory space, being a true adaptation to a changing eye position, remains unaffected by such remapping (see Fig. 7). Our results echo those of Henriques et al. (1998), in that the differences in visual target localization with changes in fixation represent a reference frame misalignment instead of a true adaptation of visual space to eye position. The possibility remains, however, that visual space adapts to a changing eye position very slowly and expresses a temporal dynamic governed by a large time constant. Its reduced rate of adaptation would render it seemingly immutable during the relatively short duration of fixation used in this and previous experiments. We cannot therefore state definitively that visual spatial localization will not be affected should eccentric eye position be maintained for hours or days instead of minutes.

Visual localization of continuously presented LED targets show near perfect accuracy during central fixation (Dobreva et al. 2005). Here, localization of memorized visual targets demonstrated sinusoidal spatial distortion with a central SG overshoot and peripheral undershoot (see Fig. 7). Stick (Lewald 1998; Lewald and Ehrenstein 2000) and arm pointing (Bock 1986; Henriques et al. 1998) toward both continuous and memorized visual targets also showed similar distortions. In-
Interestingly, this distortion occurs during both eye-fixed and eye-guided (foveal) pointing (Dobreva et al. 2005; Henriques et al. 1998; Razavi et al. 2005) and seems to be present in humans but not rhesus macaques (Metzger et al. 2004). The distortion is time dependent and builds over 5–10 s of delay before target localization (Dobreva et al. 2005). In combination, these findings suggest that the source of the visual spatial distortion lies not with the experimental condition or method of localization but might correspond to an inherent misrepresentation in human visual spatial memory that is normally corrected by immediate visual and/or motor feedback. Finally, the fact that this spatial distortion tracks along with eye position with respect to the head serves to explain the difference in visual spatial topography shown previously; differences in accuracy are comparable between target locations occupying the same retinal position (Lewald 1998; Lewald and Ehrenstein 2000; Yao and Peck 1997).

In response to concurrent auditory and visual (bimodal) target presentation, the eye position–dependent shift in auditory spatial perception became significantly attenuated but did not disappear entirely. The attenuation in shift magnitude of auditory targets with bimodal presentation likely shows a form of visual capture, a correlate of the ventrokesoq effect, whereby the visual component of the bimodal stimulus serves as a spatial attractor for the auditory target, thereby reducing the adaptation of auditory space to changing fixation. That this occurs for targets falling on the peripheral retina shows an influence of visual over auditory space away from spatial regions proximal to fixation and therefore the fovea.

Although subjects were apparently unable to ignore vision in favor of sound despite instructions to only localize the auditory target, vision did not dominate completely in all subjects because the shift magnitudes of bimodal targets are highly individualized. A spectrum of responses was noted, ranging from those that closely emulate the visual to those that heavily favor the auditory. The effect of visual capture has been shown to be dependent on the improved fidelity of the visual stimulus over auditory in spatial processing (Alais and Burr 2004; Battaglia et al. 2003; Ernst and Banks 2002; Heron et al. 2004). When localizing bimodal targets under fixation, the spatial representation of the visual stimulus is compromised by lower acuity in the retinal periphery and thereby loses some of its inherent advantage over auditory localization. This may explain why auditory targets were able to escape visual capture in some subjects despite the spatial and temporal congruence in target presentation.

Summary and implications

In a previous effort to conceptualize this eye position–dependent adaptation of auditory space (Razavi et al. 2007), we advanced the model of a physiological “set-point” controller that dictates and adjusts our perceived auditory “zero” (i.e., straight-ahead) with respect to the head. This controller receives eye position inputs and induces a corresponding time-dependent shift in auditory spatial perception in the direction of ocular eccentricity. Given a long time constant of ~1 min, brief shifts in eye position typical of natural behavior would incur little perceived change in auditory spatial alignment, especially given the lower precision of spatial hearing in comparison with vision.

This spatial set-point adaptation may have a clinical correlate in the phenomenon of rebound nystagmus that likely exists in mild form in most or all healthy subjects (Leigh and Zee 1999; cf. p. 210). Rebound nystagmus is evoked by an eccentric change in eye position that stabilizes over time. After returning the eyes back to the primary position, nystagmus arises in the opposite direction and gradually declines again, as if a rapidly adapting “set-point” controller continuously realigns the “zero” point in space with the new prolonged gaze direction. We also hypothesized that this set-point adaptor governs other spatial behaviors and influences other sensory and motor modalities, possibly with different temporal dynamics and magnitudes of change. One example is the difference noted above between the shift in sound localization and that of perceived straight-ahead during the same experimental sessions.

From the current set of experiments, we validated and enhanced the concept of a set-point model driven by eye position by showing a large shift in auditory space in response to changes of eye position that is independent of both retinal signals and the specific auditory spatial channels crucial to sound localization. Visual space, however, did not behave similarly, but we cannot rule out the possibility of a large adaptation time constant in excess of our experimental time frame.

Under natural conditions, gaze-guided behaviors typically involve a combination of eye, head, and trunk realignments ending with the eyes recentered within the orbit. Consequently, prolonged eccentric ocular deflections are relatively rare (Freedman and Sparks 1997; Goossens and Van Opstal 1997; Populin et al. 2002; Razavi et al. 2006; Tomlinson and Bahra 1986). However, because spatial adaptations to changes in eye position do arise over time, this phenomenon gains relevance whenever experimental conditions do not control for, or directly involve, eye movements in the assessment of sound localization and cross-modal interaction.

In the classic paradigm of prism adaptation, for example, because the visual world is shifted under optical prisms, the eyes must also deflect in the same direction to perceive the same field of targets (Lackner 1976; Redding et al. 2005; Wooster 1923). Even though the eyes are free to move, average eye position is deviated over time (Kalil and Freedman 1966; McLaughlin and Webster 1967). Given our results, we might expect this change in eye position alone to cause a shift in sound localization independent of audio–visual interaction. Indeed, preliminary data from our laboratory support this hypothesis and suggest that the adaptation of auditory space to a new average eye position accounts for a large portion of prism effects within the first hour of exposure (Cui et al. 2008).

Last, we consider the impact of eye position–dependent auditory spatial adaptation on the sound localization literature. Most of what is known about sound localization comes from experiments in which eye position was unknown and uncontrolled. Regardless of the pointing technique (e.g., head, hand, stick, or laser pointing), subjects were free to move their eyes naturally and direct them toward the perceived location of a sound source. As we have shown, such “ocular pointing” conditions create dramatic increases in spatial gain (overshoot) beyond that seen when the eyes are fixed and pointing is done with extrafoveal vision. Our prior experiments have shown that the overshoot during ocular pointing is the result of auditory
spatial adaptation to changes in eye position (Razavi et al. 2007). Moreover, this oculomotor adaptation applies to the egocentric perception of straight-ahead as well. These observations in turn raise the possibility, if not likelihood, that the effect of uncontrolled eye movements extends to phenomena involving other aspects of perception and orienting behaviors that entail eye-head-body coordination, such as visually guided reaching and other activities that interact with the external environment.

ACKNOWLEDGMENTS

We thank M. Gira for technical assistance; M. Rozanski, E. Knoth, L. Bachus, and J. Housel for help in gathering and analyzing the data; A. Glade for implementing pilot versions of the Alternating Fixation task; and S. H. Seidman for valuable insights on data analysis.

GRANTS

This work was supported by National Institutes of Health (NIH) Grants R01-AG-16319, P30-DC-05409, and P30-EY-01319. Q. N. Cui was supported by NIH Grant F30-DC-009372; B. Razavi received support from NIH Grant T32-EY-07125; and both Q. N. Cui and B. Razavi were supported by Training Grant T32-GM-07356.

REFERENCES


The influence of pinnae-based spectral cues on music


