Behavioral/Systems/Cognitive

Dynamic Sound Localization during Rapid Eye–Head Gaze Shifts

Joyce Vliegen, Tom J. Van Grootel, and A. John Van Opstal

Department of Medical Physics and Biophysics, Institute for Neuroscience, Radboud University Nijmegen, 6525 EZ Nijmegen, The Netherlands

Human sound localization relies on implicit head-centered acoustic cues. However, to create a stable and accurate representation of sounds despite intervening head movements, the acoustic input should be continuously combined with feedback signals about changes in head orientation. Alternatively, the auditory target coordinates could be updated in advance by using either the preprogrammed gaze–motor command or the sensory target coordinates to which the intervening gaze shift is made ("predictive remapping"). So far, previous experiments cannot dissociate these alternatives. Here, we study whether the auditory system compensates for ongoing saccadic eye and head movements in two dimensions that occur during target presentation. In this case, the system has to deal with dynamic changes of the acoustic cues as well as with rapid changes in relative eye and head orientation that cannot be preprogrammed by the audiomotor system. We performed visual–auditory double-step experiments in two dimensions in which a brief sound burst was presented while subjects made a saccadic eye–head gaze shift toward a previously flashed visual target. Our results show that localization responses under these dynamic conditions remain accurate. Multiple linear regression analysis revealed that the intervening eye and head movements are fully accounted for. Moreover, elevation response components were more accurate for longer-duration sounds (50 msec) than for extremely brief sounds (3 msec), for all localization conditions. Taken together, these results cannot be explained by a predictive remapping scheme. Rather, we conclude that the human auditory system adequately processes dynamically varying acoustic cues that result from self-initiated rapid head movements to construct a stable representation of the target in world coordinates. This signal is subsequently used to program accurate eye and head localization responses.

Key words: auditory system; human; reference frames; gaze control; models; remapping

Introduction

Unlike the eye, the ear does not possess a topographical representation of the external world. Instead, points on the basilar membrane respond to specific sound frequencies, thus providing a tonotopic code of sounds. To localize sounds, the auditory system relies on implicit cues in the sound-pressure wave. Binaural differences in sound arrival time and sound level vary systematically in the horizontal plane (azimuth), whereas direction-dependent spectral filtering by the head and pinnae [head-related transfer functions (HRTFs)] encodes positions in the vertical plane (elevation) (Oldfield and Parker, 1984; Wightman and Kistler, 1989; Middlebrooks, 1992; Blauert, 1997; Hofman and Van Opstal, 1998).

However, adequate sound localization behavior cannot rely exclusively on acoustic input (Pöppel, 1973). In humans, the acoustic cues define a head-centered reference frame. Therefore, accurate eye movements toward sounds require a coordinate transformation of the target into eye-centered motor commands, which necessitates information about eye position in the head (Jay and Sparks, 1984, 1987). Furthermore, in everyday life, eye and head positions change continuously, both relative to the target sound and to each other. To ensure accurate acoustic orienting of eyes and head, the audiomotor system should account for these changes (Goossens and Van Opstal, 1999).

In typical free-field localization experiments, the eyes and head start pointing straight ahead. Under such conditions, eye- and head-centered and world-coordinate reference frames coincide, and a craniocentric target representation suffices to localize sounds and guide eye–head movements. To dissociate the different reference frames, Goossens and Van Opstal (1999) used an open-loop double-step paradigm (see Fig. 1A), in which the auditory gaze shift was made after an intervening eye–head saccade toward a visual target (ΔG_1). Saccades toward the sound reached the actual spatial target location (II), suggesting that the initial craniocentric target coordinates (T_H) were combined with the first eye–head movement. Although this supports the hypothesis of a reference frame in world coordinates for sounds, an important alternative explanation, advanced in the visuomotor literature (Duhamel et al., 1992; Colby et al., 1995; Walker et al., 1995; Umeno and Goldberg, 1997), cannot be ruled out. In this so-called "predictive remapping scheme" the craniocentric target location is updated either by previous efference information of the primary gaze shift (ΔG_1) or by the visual target vector (FV)

(see Fig. 1*A*). Note that these three different hypotheses yield nearly equivalent performance in the classical double-step task (compare II, III).

The present study extends these experiments in two important ways. First, by presenting the sound during eye–head gaze shifts, the binaural and spectral acoustic cues are no longer static but vary in an extremely complex way with head velocity.

Second, the audiomotor system is denied any previous information about either the upcoming target location or the subsequent changes in eye and head orientation, which renders the acoustic cue dynamics entirely unpredictable. This poses a serious problem for the predictive remapping model, according to which the craniocentric target location is updated on the basis of the (preprogrammed) full first gaze-displacement vector, rather than on the partial gaze shift after target presentation. This allows for a clear dissociation of the different schemes (Fig. 1*B*, compare II, III).

Our data show that, in contrast to the prediction of the predictive remapping models, the audiomotor system remains accurate, also under these dynamic conditions. These results demonstrate that the system is capable to create, and adequately use, a stable representation of sounds in world coordinates.

Materials and Methods

Subjects
Five subjects (one female and four males; age, 25–46) participated in the experiments. All had normal hearing and were experienced in the type of sound-localization experiments conducted in our laboratory. All subjects had normal vision, except for JO (an author), who is amblyopic in his right, recorded eye. Oculomotor and head–motor responses of subjects were all within the normal range. Subjects MW and RK were kept naive about the purpose of this study. Subjects JO, JV, and TG participated in all experiments; subject RK only participated in the first target configuration (see below); subject MW only participated in the second target configuration, so that each experiment contains data from four subjects.

Apparatus
Experiments were conducted in a completely dark, sound-attenuated room (length times width times height: $3 \times 3 \times 3$ m^3) in which the four walls, floor, ceiling and all other large objects were covered with black sound-absorbing foam that eliminated acoustic reflections down to 500 Hz (Schulpen Schuim, Nijmegen, The Netherlands). The ambient background noise level in the room was 35 dBA sound pressure level (SPL) (measured with a BK-414 microphone and BK-2610 amplifier; Brüel and Kjær, Norcross, GA). Subjects were seated comfortably on a chair in the center of the room with support in their back and lower neck. They faced an acoustically transparent thin-wire hemisphere with a radius of 0.85 m, the center of which coincided with the center of the subject's head. On this hemisphere, 85 red/green light-emitting diodes (LEDs) were mounted at seven visual eccentricities, $R = [0, 2, 5, 9, 14, 20, 27, 35]$ degrees, relative to the straight-ahead viewing direction (defined in polar coordinates as $[R, \Phi] = [0, 0]$ degrees), and at 12 different directions, given by $\Phi = [0, 30,.., 330°]$, where $\Phi = 0°$ is rightward from the center location, and $\Phi = 90°$ is upward. The hemisphere was covered with thin black silk to hide the speaker completely from view (Hofman and Van Opstal, 1998).

Auditory stimuli emanated from a mid-range speaker that was attached to the end of a two-link robot that consisted of a base with two nested L-shaped arms, each driven by a stepping motor (VRDM5; Berger-Lahr, Lahr, Germany). The speaker could move quickly (within 3 sec) and accurately (within 0.5°) to practically any point on a virtual hemisphere at a radius of 0.90 m from the subject's head. To prevent sounds generated by the stepping motors from providing potential clues to the subject about either the location or displacement of the speaker, the robot always made a random dummy movement of at least 20° away from the previous location, before moving to its next target position.
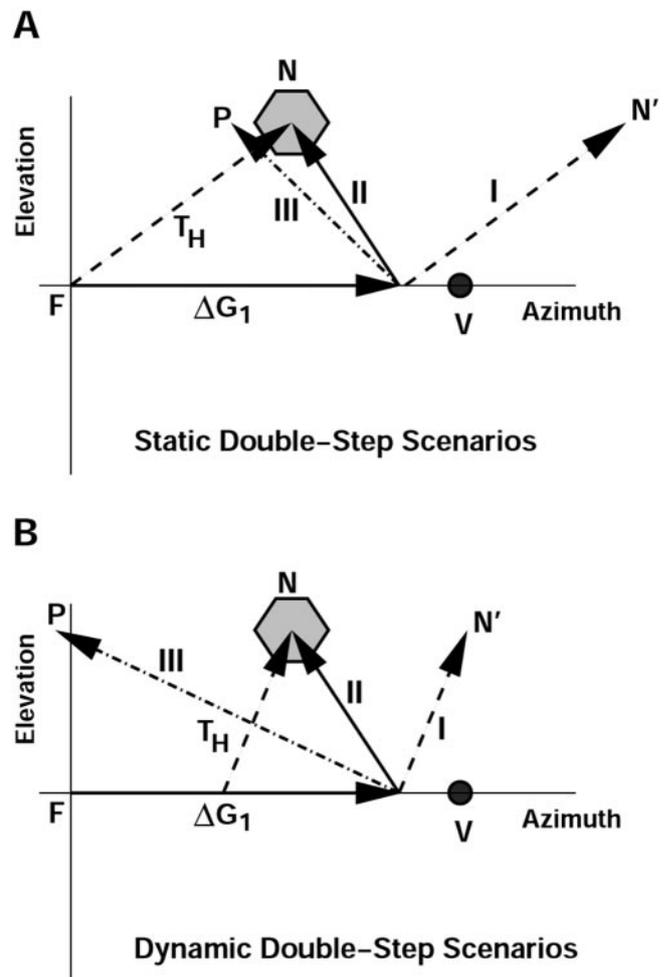

Figure 1. Three models for how the audiomotor system could behave in the double-step paradigm. *A*, Static double-step trial in which the sound (N) is presented before the first gaze shift (ΔG_1). The noncompensation model (I) predicts that the auditory target (N) is kept in a fixed craniocentric reference frame (T_H). Thus, after making the first gaze shift to V (the visual target), the second movement is directed to the location at N′. In the dynamic feedback model (II), the eye–head motor response to the sound fully accounts for the actual intervening gaze shift, ΔG_1. The response is given by $\Delta G_2 = T_H - \Delta G_1$ and is directed to N. In the visual-predictive remapping model (III), the system uses the predicted first gaze shift, specified by the required movement, FV. The second saccade is preprogrammed as $\Delta G_2 = VN = T_H - FV$. Any localization error of the first movement will not be accounted for. Thus, the response is directed to P rather than to N. *B*, Predictions of the same three models for the dynamic double step, in which the sound is presented during the first gaze shift. This yields a different head-centered target location, T_H. Model III uses the preprogrammed full first gaze shift to update the head-centered target location, instead of the partial gaze displacement after sound presentation (as in model II), thereby directing the response to P rather than to N. Note that, in contrast to model II, models I and III now predict different responses than in the static paradigm and that the predictions of models II and III are now better dissociated.

Previous studies in our group have verified that this procedure guaranteed that sounds from the stepping motors did not provide any consistent localization cues (Frens and Van Opstal, 1995; Goossens and Van Opstal, 1997).

Stimuli
Auditory stimuli were digitally generated with Matlab software (Math-Works, Natick, MA). Signals consisted of 50 msec duration broadband (0.2–25 kHz) Gaussian white noise, with 0.5 msec sine-squared onset and offset ramps, and were stored on disk at a 50 kHz sampling rate. After receiving a trigger, the stimulus was passed through a 12 bit digital–analog converter (Data Translation DT2821; output sampling rate, 50 kHz), bandpass filtered (Krohn-Hite model 3343; 0.2–20 kHz), and

passed to an audio amplifier (Luxman A-331) that fed the signal to the robot's speaker (AD-44725; Philips, Eindhoven, The Netherlands). The intensity of the auditory stimuli was fixed at 55 dBA SPL (measured at the position of the subject's head). Visual stimuli consisted of red LEDs with a diameter of 2.5 mm (which subtended a visual angle of 0.2° at a 0.85 m viewing distance) and an intensity of 0.15 cd/m^2.

Measurements

Head and eye movements were measured with the magnetic search-coil induction technique (Collewijn et al., 1975). Subjects wore a lightweight helmet (~150 gm), consisting of a narrow strap above the ears, which could be adjusted to fit around the head, and a second strap that ran over the head. A small coil was mounted on the latter. Subjects also wore a scleral search coil on one of their eyes. In the room, two orthogonal pairs of 3 × 3 m^2 square coils were attached to the side walls, floor, and ceiling to create the horizontal (30 kHz) and vertical (40 kHz) oscillating magnetic fields that are required for this recording technique. Horizontal and vertical components of head and eye movements were detected by phase-lock amplifiers (models 128A and 120; Princeton Applied Research), low-pass filtered (150 Hz), and sampled at 500 Hz per channel before being stored on disk.

Two personal computers controlled the experiment. One, PC-486, was equipped with the hardware for data acquisition (Metrabyte DAS16), stimulus timing (Data Translation DT2817), and digital control of the LEDs (Philips I2C). The other, PC-486, controlled the robot movements and generated the acoustic stimuli after receiving a trigger from the DT2817.

Experimental paradigms

Calibration of eye and head. Each experimental session started with three runs to calibrate the eye and head coils (Goossens and Van Opstal, 1997). Before the calibration, subjects were asked to keep their heads in a neutral, comfortable straight-ahead position and adjust a dim red LED mounted at the end of a thin pliable aluminum rod that was attached to their helmet (at a distance of ~0.40 m in front of the subject's eyes) such that it was approximately aligned with the center LED of the hemisphere. This rod LED was illuminated only in the second and third calibration sessions and was off during the actual localization experiments.

First, eye position in space ("gaze") was determined. During this calibration, subjects kept their heads still in the straight-ahead position and fixated the LEDs on the hemisphere with their eyes. Targets ($n = 37$) were presented once, in a fixed counterclockwise order, at the center location ($R = 0$), followed by three different eccentricities, $R = [9, 20, 35]$ degrees, and all 12 directions. When subjects fixated the target, they pushed a button to start data acquisition, while keeping their eyes at that location for at least 1000 msec.

In the second calibration run, the eye-in-head offset position was determined. To that end, subjects fixated the dim red LED on the helmet rod (rather than the LED on the hemisphere) while keeping their heads in the straight-ahead position. This procedure kept their eyes at a fixed orientation in the head. When the subject assumed the neutral head posture, he or she pushed a button to start 1000 msec of data acquisition. This procedure was repeated 10 times. In between trials, subjects were asked to freely move their head before re-assuming the neutral position.

The third calibration run served to calibrate the coil on the head. Now subjects had to fixate the dim red LED at the end of the head-fixed rod with their eyes and align this rod LED with the same 37 LED targets on the hemisphere as in the eye calibration run. In this way, the eyes remained at the same fixed offset position in the head as in the second calibration. When the subject pointed to the target, he or she started 1000 msec of data acquisition by pushing a button.

After the calibration runs were completed, the experimental localization sessions started. One experimental session consisted of at least four different blocks of trials: (1) visual single-step localization; (2) visual–visual double-step localization; (3) auditory single-step localization; and (4) visual–auditory double-step localization. Blocks of one modality were always presented together, and the single-step block was always presented first. After these four blocks, additional visual–auditory double-step blocks could be performed until the subject wanted to stop.

Here, we will focus on the auditory single- and double-step experiments only. Results of the visual eye–coordination experiments will be presented elsewhere. All calibration and experimental sessions were performed in complete darkness.

Auditory single-step paradigm. To determine a subject's baseline localization behavior, a single-step localization experiment was performed. Each trial started with the presentation of a fixation LED. During fixation, subjects had their eyes and head approximately aligned. After 800 msec, this LED was switched off, and 50 msec later an auditory stimulus was presented at a peripheral location. Subjects were asked to point to the apparent location of the stimulus as quickly and as accurately as possible by redirecting their gaze line to the perceived peripheral target location. Because stimuli were always extinguished well before the initiation of the eye and head movement, the subject performed under completely open-loop conditions.

To enable a direct comparison of the single-step responses with the second gaze shifts from the double-step paradigms (see below), we designed the single-step experiment such that the initial visual fixation targets of this experiment were the same as the first peripheral visual targets in the double-step experiments. Also the sound locations of the single step experiment were the same as those in the double-step experiments.

There were two different stimulus configurations. The first consisted of a central visual fixation target at $[R, \Phi] = [0, 0]$ degrees and 10 different auditory target positions (relative to the straight-ahead direction) with $[R, \Phi] = [14, 0], [14, 180], [20, 0], [20, 90], [20, 180], [20, 270], [27, 60], [27, 120], [27, 240]$, or $[27, 300]$ degrees. Target locations were selected in random order. One block consisted of 20 trials. In the second configuration, the initial fixation target was at either $[R, \Phi] = [20, 90]$ or $[20, 270]$ degrees (pseudorandomly chosen with both fixation targets occurring equally often). Auditory targets were presented at a randomly selected position within a circle of $R = 35°$ around the straight-ahead direction, but always at least 10° away from the initial fixation target. A total of 24 trials were presented in one block.

Visual–auditory double-step paradigms. We used both a static double-step target condition in which the second target was presented before initiation of the first eye–head movement and a dynamic condition in which the second target was presented during the first eye–head movement. The latter paradigm is adopted from the classical saccade-triggered visuomotor paradigm of Hallett and Lightstone (1976). The visual–auditory double-step paradigm is illustrated in Figure 2. First, a fixation target (F) is presented for 800 msec. Then, after 50 msec of complete darkness, a visual target (V) is flashed for 50 msec (Fig. 2A).

The timing of the second auditory target (N) was varied, resulting in three different conditions. (1) In the nontriggered (static) condition, the auditory target was presented after a fixed delay of 50 msec after extinction of the peripheral visual target. In this condition, both targets were presented before the first gaze-shift onset, which typically started at a latency of ~200 msec after the visual stimulus flash. (2) In the early-triggered (dynamic) condition, the auditory target was triggered as soon as the head velocity in the direction of the visual target exceeded 40°/sec. In this way, the timing of the auditory stimulus fell early in the first head movement and often was presented while the gaze line (the eye in space) was still moving. (3) In the late-triggered (dynamic) condition, the auditory target was triggered 50 msec after head velocity in the direction of the visual target exceeded 60°/sec. In this way, stimulus presentation fell approximately halfway through the first head movement and typically close to the moment of the peak velocity of the head (Goossens and Van Opstal, 1997).

Two different stimulus configurations were used (Fig. 2B). The first configuration (subjects JO, JV, RK, and TG) consisted of an eccentric fixation target at $[R, \Phi] = [35, 0]$ degrees or $[35, 180]$ degrees (pseudo-randomly chosen as above), a visual target at $[R, \Phi] = [0, 0]$ degrees, and an auditory target at 10 possible target positions at polar coordinates $[R, \Phi] = [14, 0], [14, 180], [20, 0], [20, 90], [20, 180], [20, 270], [27, 60], [27, 120], [27, 240]$, or $[27, 300]$ degrees. Target locations were selected in random order. One block consisted of 20 nontriggered and 20 early-triggered trials (randomly interleaved).

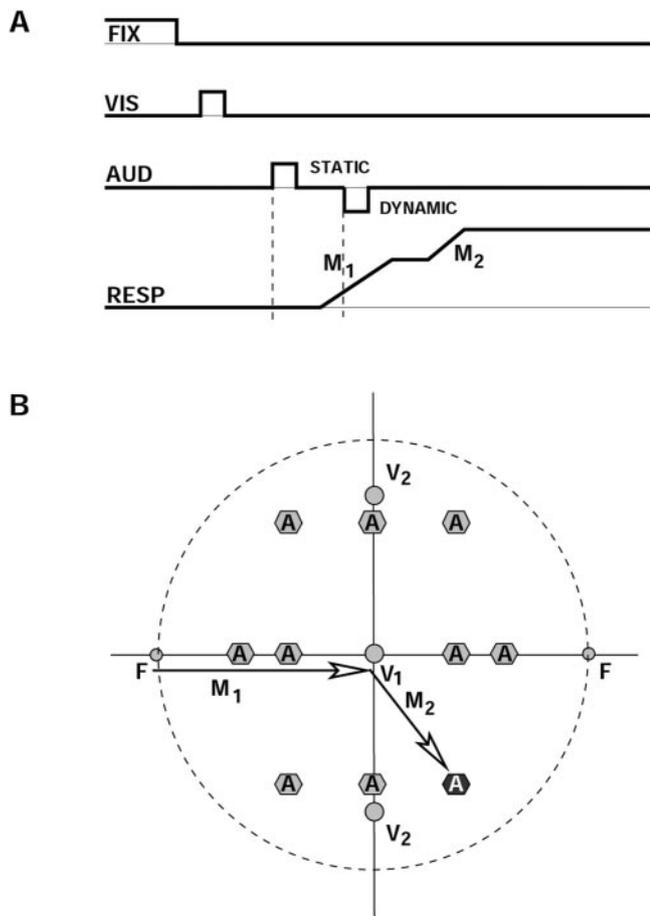Because in this configuration the peripheral visual target was always at

Figure 2. Double-step paradigms. *A*, Temporal order of the different targets in the static and dynamic double-step trials. M_1 and M_2, First and second eye–head movement; FIX, fixation; VIS, visual; AUD, auditory; RESP, response. *B*, Spatial layout of the target configurations. F, Initial fixation positions; V_1, visual target in the first double-step series in which M_1 is a purely horizontal movement; V_2, visual target in the second double-step series in which M_1 is an oblique gaze shift; A, potential auditory target locations in the first target configuration; dashed circle, area within which the auditory targets were selected for the second target configuration.

Data analysis

After calibration, the coordinates of auditory and visual target locations, as well as the eye and head positions and movement displacement vectors, were all expressed in a double-pole azimuth–elevation coordinate system in which the origin coincides with the center of the head (Knudsen and Konishi, 1979). In this system, the azimuth angle, α, is defined as the angle within the horizontal plane with the vertical midsaggital plane, whereas the elevation angle, ϵ, is defined as the direction within a vertical plane with the horizontal plane through the subject's ears. The straight-ahead direction is defined by $[\alpha, \epsilon] = [0, 0]$ degrees. The relationship between the $[\alpha, \epsilon]$ coordinates and the polar $[R, \Phi]$ coordinates defined by the LED hemisphere (see above) was described by Hofman and Van Opstal (1998).

Calibration of the data. The raw eye position data and the corresponding known LED positions from the first calibration run were used to train two three-layer back propagation neural networks that mapped the raw eye position signals to calibrated azimuth/elevation angles of eye position in space (gaze). The networks compensated for minor cross talk between the horizontal and vertical channels and for small nonhomogeneities and nonlinearities in the magnetic fields.

Calibration of the head-coil fixations was obtained in the following way. First, the calibrated eye position data from the second calibration session, with the head in the neutral position, were determined and averaged to yield an average eye-in-head offset gaze position, G_0. Then, the raw eye position data obtained from the head-coil calibration run were calibrated with the eye-coil calibration networks from the first calibration run. Subsequently, the static head position data were corrected for the mean offset in eye-in-head position according to: $H = G - G_0$, where H represents the position of the head in space, as measured with the eye coil. Finally, the head-coil data were calibrated by mapping the raw head position signals on the calibrated eye-coil data with an additional set of two neural networks (Goossens and Van Opstal, 1997). In the calibrated response data, we identified head and gaze saccades with a custom-written computer algorithm that applied separate velocity and mean acceleration criteria to vectorial saccade onset and offset, respectively. Markings were visually checked and corrected, if deemed necessary. To ensure unbiased detection criteria, the experimenter was denied any information about the stimulus. Responses with a first saccade latency shorter than 80 msec (considered to be predictive) or longer than 800 msec (potentially caused by inattentiveness of the subject) were discarded from additional analysis. To ensure that the static trials were indeed static, we checked whether first head-saccade latency in those trials exceeded 150 msec (offset time of auditory target relative to onset visual target). This requirement was met for all trials (for an example, see Fig. 5).

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

The bootstrap method was applied to obtain confidence limits for the optimal fit parameters in the regression analyses. To that end, 100 data sets were generated by random selections of data points from the original data. Bootstrapping thus yielded a set of 100 different fit parameters. The SDs in these parameters were taken as an estimate for the confidence levels of the parameter values obtained in the original data set (Press et al., 1992).

To determine whether two (non-Gaussian) data distributions were statistically different, we applied the Kolmogorov–Smirnov (KS) test. This test provides a measure (d-statistic) for the maximum distance between the two distributions, for which the significance level, p, that the distributions are the same can be readily computed. If $p < 0.05$, the two data sets were considered to correspond to different distributions. For data expressed as two-dimensional distributions (e.g., the azimuth–elevation end points in Fig. 7), we computed the two-dimensional KS statistic to measure their mutual distance and its significance level (Press et al., 1992).

The bin-width (BW) of histograms (see Figs. 5 and 7) was determined by $BW = \text{Range}/\sqrt{N}$, where Range is the difference between the largest

the same position, eight additional catch trials were included in the experiment to prevent the subject from making a predictive movement to the visual target position. In catch trials, the visual target was at either $[R, \Phi] = [35, 30], [35, 150], [35, 210],$ or $[35, 330]$ degrees, and the second (auditory) target was presented at either $[R, \Phi] = [20, 90]$ or $[20, 270]$ degrees (pseudorandomly chosen with all positions occurring equally often).

In the second double-step configuration (subjects JO, JV, MW, and TG), the initial fixation target was again at $[R, \Phi] = [35, 0]$ or $[35, 180]$ degrees, but now the peripheral visual target was at either $[R, \Phi] = [20, 90]$ or $[20, 270]$ degrees (both pseudorandomly chosen as above). This resulted in a first gaze shift with a horizontal as well as a considerable vertical component, in contrast to the first target configuration in which the first gaze shift was always purely horizontal. The auditory target was presented at a randomly selected position within a homogeneous area of $R = 35°$ around straight-ahead but always at least 10° away from the visual target. This block consisted of 48 late-triggered trials, but if, after four experimental blocks, the subject was capable of doing additional experiments, we repeated this visual–auditory block with a reduced number of 24 trials.

In all experimental localization sessions, subjects were free to move their head and eyes to localize both targets. They were asked to localize the stimulus as quickly and as accurately as possible, by fixating the perceived stimulus location with their eyes, but they were not given specific instructions about the movements of their head.
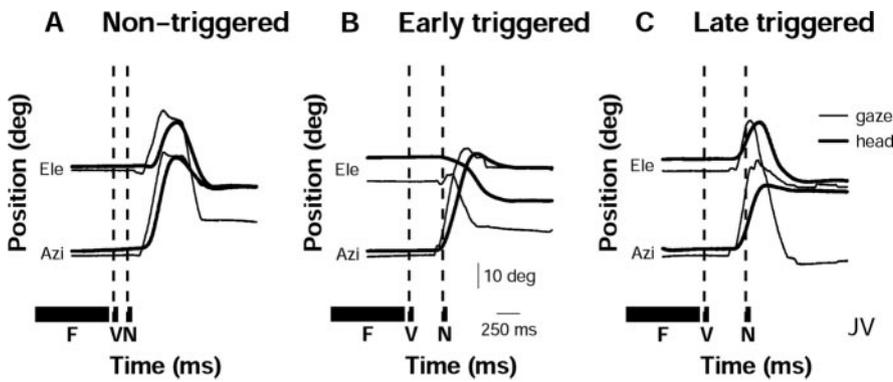
Figure 3. Head (thick lines) and gaze (thin lines) double-step responses as a function of time for azimuth and elevation components. F, V, and N indicate the time of presentation of the fixation target, the visual target, and the auditory target, respectively. *A,* Trial from the nontriggered static condition in which the second auditory target is presented before initiation of the primary head and gaze movement. *B,* Trial from the early-triggered dynamic double-step condition. Here, the auditory target is presented early in the saccade. *C,* Trial in the late-triggered dynamic condition. Here, the auditory target falls halfway through the first head saccade. Data are from subject JV.
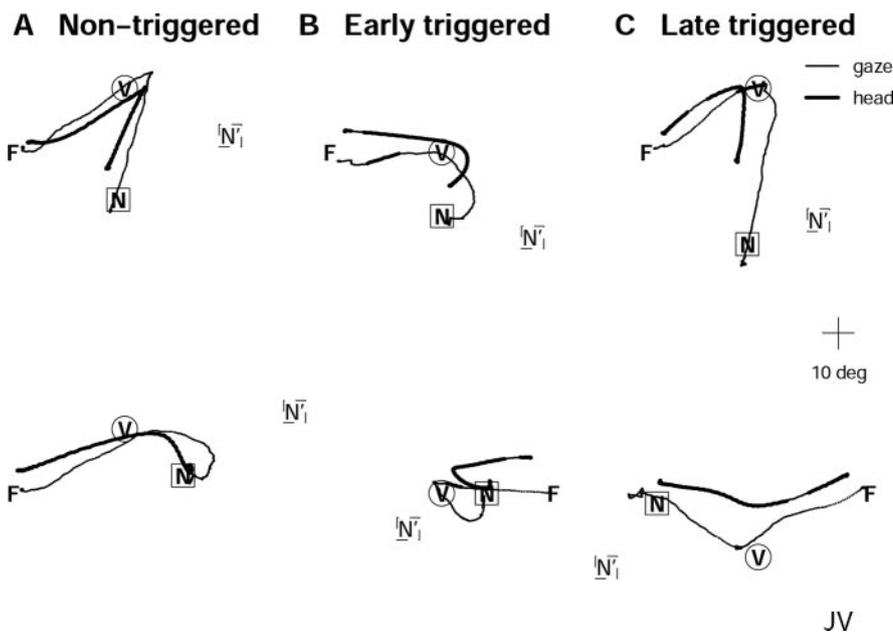

Figure 4. Head (thick lines) and gaze (thin lines) double-step response traces in space. F, V, and N indicate the positions of the fixation target, the visual target, and the auditory target, respectively. *A,* Two representative trials from the nontriggered condition. *B,* Two trials from the early-triggered condition. The target presentation epoch is indicated by a change in line thickness. *C,* Two trials from the late-triggered condition. If the second saccade would be based purely on the initial head-motor error, the responses would be directed toward the dashed square (N′). For the dynamic conditions, the initial target re-head position is defined as the target position relative to the head at sound onset. Note that the responses are directed toward the veridical location of the sound. Data are from subject JV. Responses in top row are the same as for Figure 3.

and smallest values (excluding the two most extreme points), and *N* is the number of included points.

Results

Double-step response behavior

Figure 3 shows three typical examples of head and gaze traces as a function of time of subject JV elicited in the double-step experiments, one for the static condition (Fig. 3*A*) and two for the dynamic conditions [early triggered (Fig. 3*B*) and late triggered (Fig. 3*C*)]. In the static double-step condition, the visual and the auditory target are both presented and extinguished before the initiation of the visually evoked head and gaze movement. For the two dynamic conditions, the auditory target, which is triggered by the head move-

ment, falls either early in (Fig. 3*B*), or halfway through (Fig. 3*C*), the first head saccade. For all three conditions, gaze saccades are faster and larger than head saccades, which is a typical pattern for eye–head coordination (Goossens and Van Opstal, 1997). At the end of the second gaze shift, the VOR ensures that gaze-in-space remains stable, despite the ongoing movement of the head.

Figure 4 shows six typical examples of two-dimensional spatial head and gaze trajectories of subject JV for the static condition (Fig. 4*A*), for the early-triggered condition (Fig. 4*B*), and for the late-triggered condition (Fig. 4*C*). The dashed squares (N′) indicate the spatial locations to which the second gaze shift would be directed if it were only based on the initial head-centered acoustic input. However, these examples show that head and gaze responses are both directed toward the actual stimulus location. Gaze approaches the auditory target more closely than the head, which tends to undershoot the vertical target component (Fig. 4, top).

Head and eye movements during sound stimuli

The aim of the triggered double-step experiments was to ensure considerable and variable head movements during the presentation of brief acoustic stimuli. To verify that the head and eye were indeed moving substantially during sound presentation, Figure 5 shows all two-dimensional head (left) and eye (right) movement traces of subject JO during the 50 msec acoustic noise burst pooled for the two dynamic triggering conditions (Fig. 5*A*). The onsets of all movements are aligned in (0, 0) degrees for ease of comparison. Note that the majority of head displacements during the brief stimulus were on the order of ≥10° (Fig. 5*A*, left). Typically, the eye moved much faster in an eye–head gaze shift (Fig. 3). Therefore, in the late-triggered double steps the eye often reached the visual target location, whereas the head was still moving. In those cases, the vestibular-ocular reflex (VOR) kept gaze at its new position. Yet, for the majority of dynamic trials, the eye-in-space moved substantially also during sound presentation (Fig. 5*A*, right), especially for the early-triggered condition (horizontal traces). The head and eye movement amplitude in the dynamic condition, averaged across subjects, was 8.0 ± 3.0° and 5.0 ± 3.0°, respectively.

Figure 5*B* shows histograms of the mean (black) and peak (light gray) head (left) and eye (right) velocities during sound presentation in both the dynamic and static (dark gray; only mean velocity shown) double-step conditions for this subject. As required, the eyes and head were not moving in the static double-step trials. In the dynamic conditions, however, there is a large range of both the mean and peak head velocities. The mean head velocity is ~150°/sec; the peak head velocity is, on average, 200°/
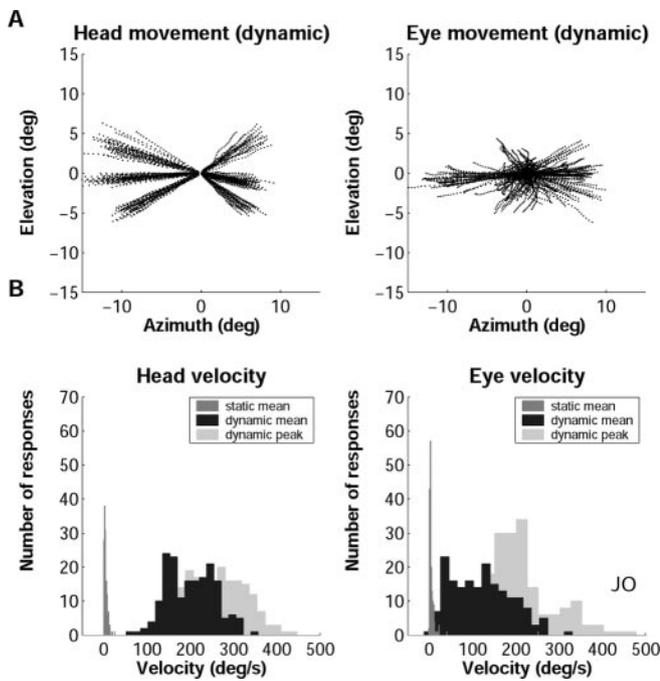
Figure 5. Properties of ongoing head and eye movements during presentation of the auditory target. *A*, Two-dimensional head (left) and eye (right) movement traces during stimulus presentation in the dynamic condition (early- and late-triggered trials pooled). *B*, Head and eye mean and peak velocity during the 50 msec stimulus presentation for both the static (dark gray histogram; only mean velocity shown) and the dynamic (black and light gray histograms for mean and peak velocities, respectively) condition. Note large trial-to-trial variability in eye and head movement kinematics for the dynamic double-step trials. Data are from subject JO. deg, Degrees.

Figure 6. Eye-in-head positions at the onset of the second auditory-evoked gaze shift (E_0 in Eq. 1). The eye is typically eccentric in the head, so that gaze-in-space and head-in-space are not aligned at the start of the second gaze shift. Points within the square correspond to eye positions with azimuth and elevation components <10°. deg, Degrees.

sec. As a result, the acoustic cues vary considerably from trial to trial and in an unpredictable way. Moreover, in many trials, the eyes also moved substantially with respect to the sound.

Although at the start of a double-step trial the eyes and head were approximately aligned, this is no longer the case after the first gaze shift. To illustrate the trial-to-trial variability in eye–head misalignment at the onset of the auditory-evoked gaze shift, Figure 6 shows the distribution of eye-in-head positions pooled for all subjects across trials. The shaded central square indicates trials for which both the horizontal and vertical eye position eccentricity was <10° (see also Fig. 9). Note that the misalignment between the eye and head can be as large as 30°, although for the majority of trials the eye stays within 10° of the center of the oculomotor range.

Sound-localization errors
To compare response accuracy for the different stimulus conditions, Figure 7 shows the two-dimensional distributions of the end points of second gaze saccades for static (filled circles) and dynamic (gray triangles) double-step trials (early- and late-triggered data pooled, as they were statistically indistinguishable), as well as for the single-step localization responses (open dots). In this figure, all auditory target locations have been aligned with the origin of the azimuth–elevation coordinate system. Gaze end positions are plotted as undershoots (azimuth and elevation <0) or overshoots (azimuth and elevation >0) with respect to the target coordinates. The static double-step data are summarized by the black histograms, and the corresponding dashed lines indicate their medians. The dynamic double-step data are represented by the gray histograms, and the continuous
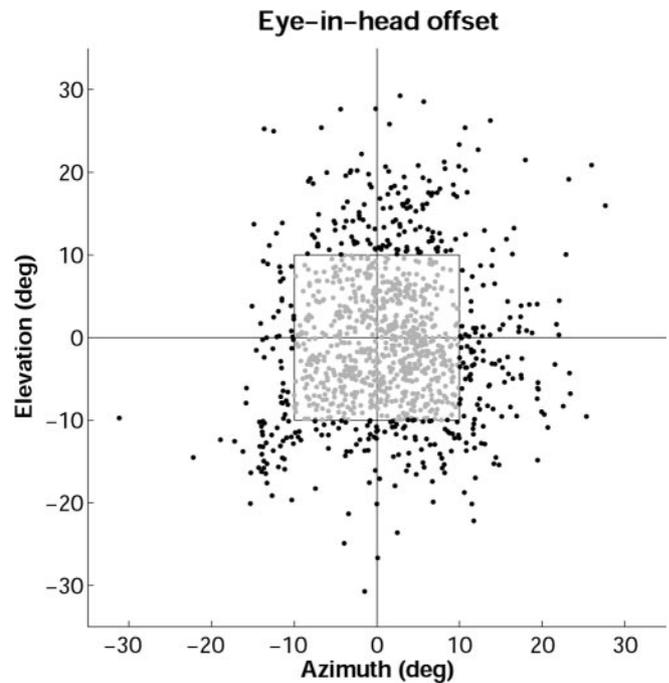
lines show their median values. The medians of the single-step condition are indicated by black dotted lines.

Quite remarkably, the response distributions for the single-step localization trials and the static and dynamic double-step trials are very similar. The mean unsigned errors and SDs for the three conditions are virtually the same. The three pairwise two-dimensional KS tests (Press et al., 1992) indicated that the end point distributions were statistically indistinguishable, except for the single-step versus the nontriggered double-step comparison (single step vs nontriggered double steps: $p < 0.05$, d = 0.25; single step vs triggered double steps: $p = 0.09$, d = 0.17; nontriggered vs triggered double steps: $p = 0.10$, d = 0.14). Table 1 summarizes the mean unsigned errors for the different conditions, pooled for all subjects. Note also that for all conditions the response distributions are broader for elevation than for azimuth response components (all three KS tests on azimuth vs elevation, $p < 0.001$). Such a difference in response accuracy is typical for human sound localization performance to single steps and underscores the different neural mechanisms for the extraction of the spatial acoustic cues. This feature appears to be preserved also in the static and dynamic double-step localization trials.

Regression analysis: sound reference frame
To test in a quantitative way to what extent the intervening eye and head movements of the first gaze shift are accounted for in planning the eye–head saccade to the auditory targets, we performed multiple linear regression on the second auditory-guided gaze displacement. In this analysis, ΔG_2, which is the displacement of the eye in space from its starting position at the end of the first gaze shift, was described by a linear combination of the initial sound location in head-centered coordinates, $T_{H,ini}$, the subsequent displacement of the head during the first gaze shift, ΔH_1, and the position of the eye in the head after the first gaze shift, E_0, according to the following equation:
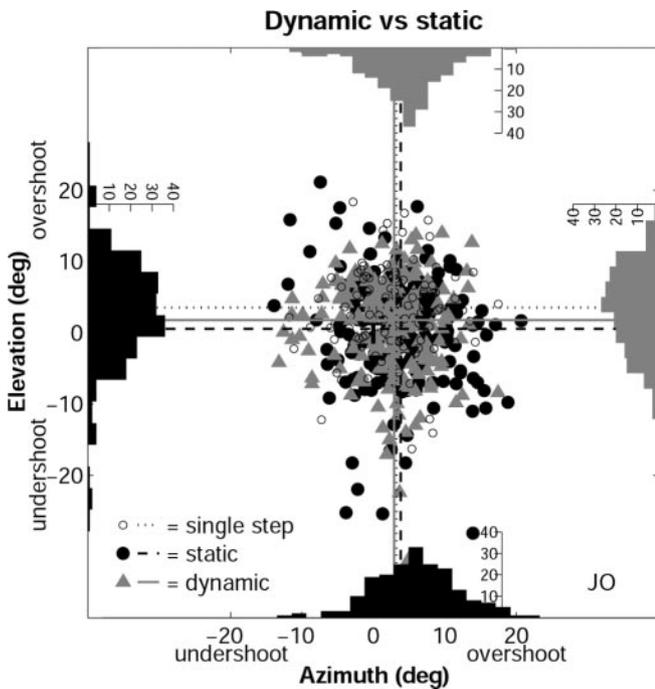
Figure 7. End points of second gaze saccades in azimuth and elevation plotted relative to the acoustic target position. The latter (T) is aligned with (0, 0) degrees; gaze responses are expressed as undershoots or overshoots with respect to the target location. Histograms show the respective response distributions for the static (black; filled black circles) and triggered dynamic (gray; gray triangles) double-step responses. The dashed lines indicate means of the static double steps; solid lines indicate means of the dynamic double steps. Note similarities in the distributions. Open dots correspond to gaze end points toward single-step targets, and dotted lines indicate their means. Data are from subject JO. deg, Degrees.

Table 1. Mean and SD of saccade end point errors for the single-step and double-step paradigms

Condition	Azimuth (degrees)	Elevation (degrees)	KS test	n
Single steps	1.8 ± 5.8	3.1 ± 9.8	$p < 0.001$	432
Nontriggered double steps	3.4 ± 6.4	3.0 ± 12.1	$p < 0.001$	657
Triggered double steps	1.2 ± 6.3	3.3 ± 11.6	$p < 0.001$	651

The 1D KS test was performed on ranked azimuth versus elevation distributions within each stimulus condition. Data are pooled for all five subjects and recording sessions.

$$\Delta G_2 = a \cdot T_{H,ini} + b \cdot \Delta H_1 + c \cdot E_0 + d, \quad (1)$$

in which (a, b, c) are dimensionless response gains, and d (in degrees) is the response bias. Equation 1 was applied separately to the azimuth and elevation response components.

Note that if the audiomotor system would not compensate for the intervening eye–head gaze shift but instead would keep the sound in the initial head-centered coordinates determined by the acoustic cues, the regression should yield $a = 1$ and $b = c = d = 0$ (indicated by model I in Fig. 1A). Full compensation for the first gaze shift requires that $a = 1$, $b = c = -1$, and $d = 0$ (model II in Fig. 1A), in which case Equation 1 simply reduces to $\Delta G_2 = T_{H,ini} - \Delta G_1$. For the static, nontriggered double-step responses, the first head displacement (ΔH_1) is defined as the entire head displacement, whereas for the triggered double-step trials it is the portion of the head displacement that followed the sound onset (Fig. 1B). The head-centered location of the sound is determined by the head position in space at sound onset. Data from the early-triggered and late-triggered experiments were pooled. The resulting gains (a, b, c) of the regression, averaged across subjects,
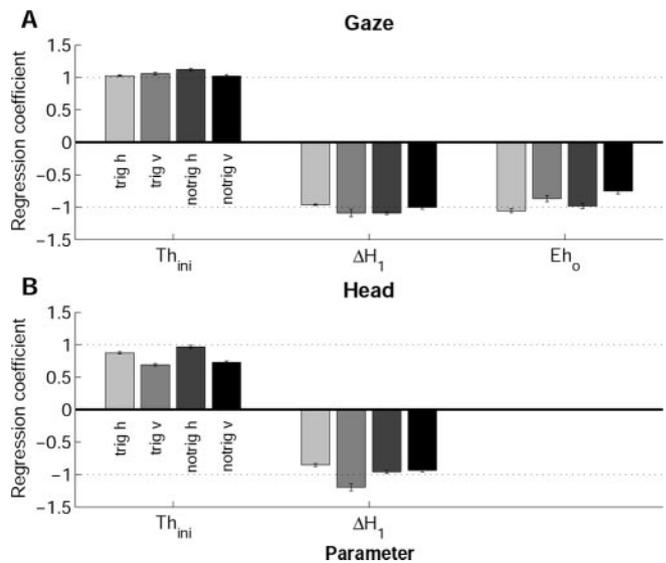

Figure 8. *A*, Regression coefficients of Equation 1 for second gaze saccades (ΔG_2), averaged across subjects and recording sessions. *B*, Regression coefficients of Equation 2 for second head saccades (ΔH_2). Different double-step conditions (dynamic/static) and response directions (horizontal/vertical) are represented by the different gray-coded bars. The dotted lines at the values of +1.0 and −1.0 correspond to ideal compensation for the intervening movements.

are summarized in Figure 8A for the different conditions and response components (results for individual subjects are provided in Supplementary Table IIA). The gain-coefficient (a) for the craniocentric target location is close to +1.0 for all conditions and response components. Moreover, the response gains for head displacement, as well as for eye-in-head position, are close to the optimal values of −1.0. The coefficient for eye-in-head tends to be slightly lower than −1.0. Because we did not systematically control eye position offset, it varied between subjects; some subjects made relatively large head movements, causing their eyes to remain closer to the center of the oculomotor range. Because there were no subjects who over-compensated eye-in-head position, the average across subjects tended to be lower than −1.0. The offsets (d) were always close to 0° and are not shown.

This result implies that subjects fully compensate for the intervening eye–head movement, even under dynamic localization conditions.

A similar multiple regression analysis was performed on the second head movement vector, ΔH_2, in response to the auditory target. In that case, the head response was described by the following equation:

$$\Delta H_2 = a \cdot T_{H,ini} + b \cdot \Delta H_1 + d. \quad (2)$$

The results (averaged across subjects) are shown in Figure 8B. Note that also for the head, the fitted gains (a, b) are close to the ideal values of +1.0 and −1.0, respectively. The target elevation gain (a in Eq. 2) for the head responses was found to be lower than for the eye (Eq. 1). This probably reflects a robust motor strategy to withhold the head from making large movements against gravity (André-Deshays et al., 1988). Results of this analysis for the individual subjects are provided in Supplementary Table IIB.

Regression analysis: motor error frames
In generating a gaze shift toward an auditory target, it is not trivial that eye and head both move toward the target, especially if eye and head are not aligned. For that to happen, the world target coordinates need to be transformed into oculocentric and head-

centered coordinates, respectively. Alternatively, both could be driven by the same motor error signal, either by an oculomotor gaze error signal (as in the so-called common-gaze control model for eye and head) (Vidal et al., 1982; Guitton, 1992; Galiana and Guitton, 1992) or by a (acoustically defined) head motor error signal. The difference between these two reference frames is determined by the position of the eye in the head, which varies considerably and unpredictably from trial to trial and can be as large as 30° (Fig. 6). To investigate this point, we subjected the data to a normalized multiple linear regression in which the auditory-evoked head movement, ΔH_2, and the gaze shift, ΔG_2, are each described as a function of gaze motor error, GM, and head motor error, HM:

$$\Delta H'_2 = p \cdot GM' + q \cdot HM', \tag{3a}$$

$$\Delta G'_2 = p \cdot GM' + q \cdot HM'. \tag{3b}$$

In Equations 3a and 3b, head motor error (HM) was determined as the difference between the auditory target in space and the head position in space at the start of the second gaze shift. Gaze motor error (GM) was taken as the difference between the auditory target location and the eye position in space at the start of the gaze shift (i.e., the retinal error of the sound). These response variables were transformed into their (dimensionless) z-scores: $x' = (x - \mu_x)/\sigma_x$, where μ_x is the mean of variable x and σ_x is its variance. In this way, the variables can be directly compared, and p and q are the (dimensionless) partial correlation coefficients for gaze motor error and head motor error, respectively. If $p > q$, the head (or eye) is driven predominantly by an oculocentric gaze error signal. If $q > p$, the head (or eye) rather follows the head-centered motor error signal. In case $p > q$ (or $p < q$), for both equations, eye and head are considered to be driven by the same error signal. To allow for a meaningful dissociation of the oculo-centric and head-centered reference frames, we only incorporated trials for which the absolute azimuth or elevation component of eye-in-head position exceeded 10° (those positions falling outside the square in Fig. 6), and the directional angle between the head and gaze motor error vectors was at least 15°. In this way, we incorporated a sufficient number of data points for three subjects.

Figure 9 shows the regression coefficients on the pooled data from all subjects for all conditions. It can be seen (Fig. 9A) that for head movement, the coefficients for head motor error are larger than those for gaze motor error (for all conditions, $p < 0.01$, apart from the triggered vertical condition, in which the difference failed to reach significance). This suggests that the head is indeed driven by a craniocentric motor command. Conversely, the eye-in-space is clearly driven by gaze motor error, because for all conditions, $p > q$ (Fig. 9B) (for all conditions, $p < 0.01$). These data therefore show that the audiomotor system is capable to dynamically transform the auditory target coordinates into the appropriate motor reference frames. Data for individual subjects are summarized in Supplementary Tables IIIA and IIIB. The values for p and q vary somewhat between subjects and conditions, especially for the head movements, in which for 2 of 16 conditions $p > q$. We have no obvious explanation for this variability.

Quantitative model tests
In Introduction, we described four different models to predict the coordinates of the second gaze shift in a visual–auditory double-step paradigm (Fig. 1). In particular, it was argued that the results from the nontriggered double-step trials could be explained
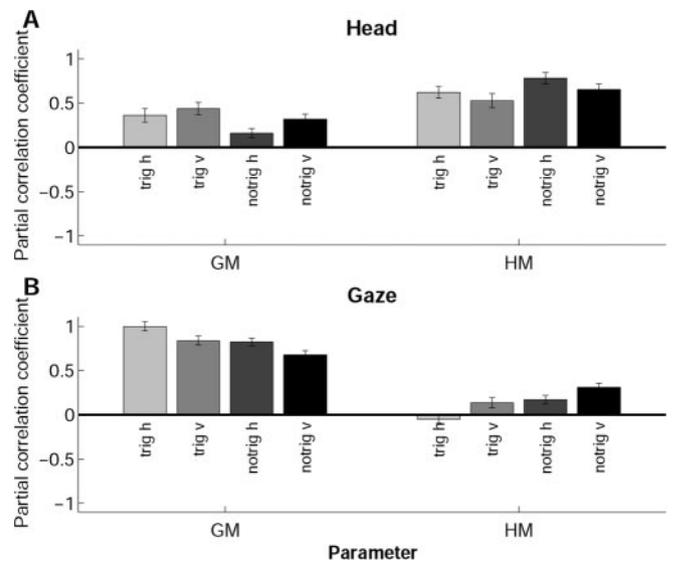

Figure 9. Partial correlation coefficients for the regression on the second head saccade (ΔH_2) (A) and the second gaze saccade (ΔG_2) (B), which are described as a function of the gaze (GM) and head (HM) motor errors (Eqs. 3a and 3b). The different gray-coded bars represent the two different conditions (dynamic/static) and response components (azimuth/elevation). Data are pooled across all subjects and recording sessions. Note that eye and head are mainly driven by motor commands expressed in their own reference frames.

equally well by two conceptual models. In the dynamic feedback scheme (model II), the instantaneous head and eye movements are incorporated in the computation of the auditory spatial coordinates. In contrast, the predictive remapping scheme (model III) uses previous (static) information of the upcoming gaze shift to update the auditory target location. To test whether the results from the triggered double-step experiments could indeed dissociate these models, we computed the predicted second gaze displacement for the different schemes from the recordings.

The predictive remapping model was tested in two different ways: in the first version (visual predictive), we used the initial retinal error vector for the first gaze shift, FV, as the predictive signal for remapping (indicated by model III in Fig. 1A). In the second version (motor predictive), we instead took the actual first gaze displacement (ΔG_1) to update the auditory target. This leads to the following two predictive remapping models:

$$\Delta G_2 = a \cdot T_{H,ini} + b \cdot FV + c, \tag{4a}$$

$$\Delta G_2 = a \cdot T_{H,ini} + b \cdot \Delta G_1 + c. \tag{4b}$$

Note that Equations 1 and 4b predict the same gaze shift for the nontriggered double-step experiment if the first gaze shift is fully accounted for (i.e., when $b = c = -1$ in Eq. 1). Also, when the first gaze shift brings the eye close to the extinguished visual target location, vectors FV and ΔG_1 will be very similar, as will Equations 4a and 4b (Fig. 1A). However, if the first gaze shift misses the visual target location, Equations 4a and 4b yield different predictions.

For the triggered double-step experiments, the head-centered auditory target coordinates were taken relative to the position of the head in space at sound onset (Fig. 1B). The head-displacement signal for the model of Equation 1 is then given by the subsequent displacement after sound onset. Note, however, that for the predictive remapping schemes the preprogrammed signals in Equations 4a and 4b are the same for the nontriggered and triggered double-step conditions because they relate to infor-
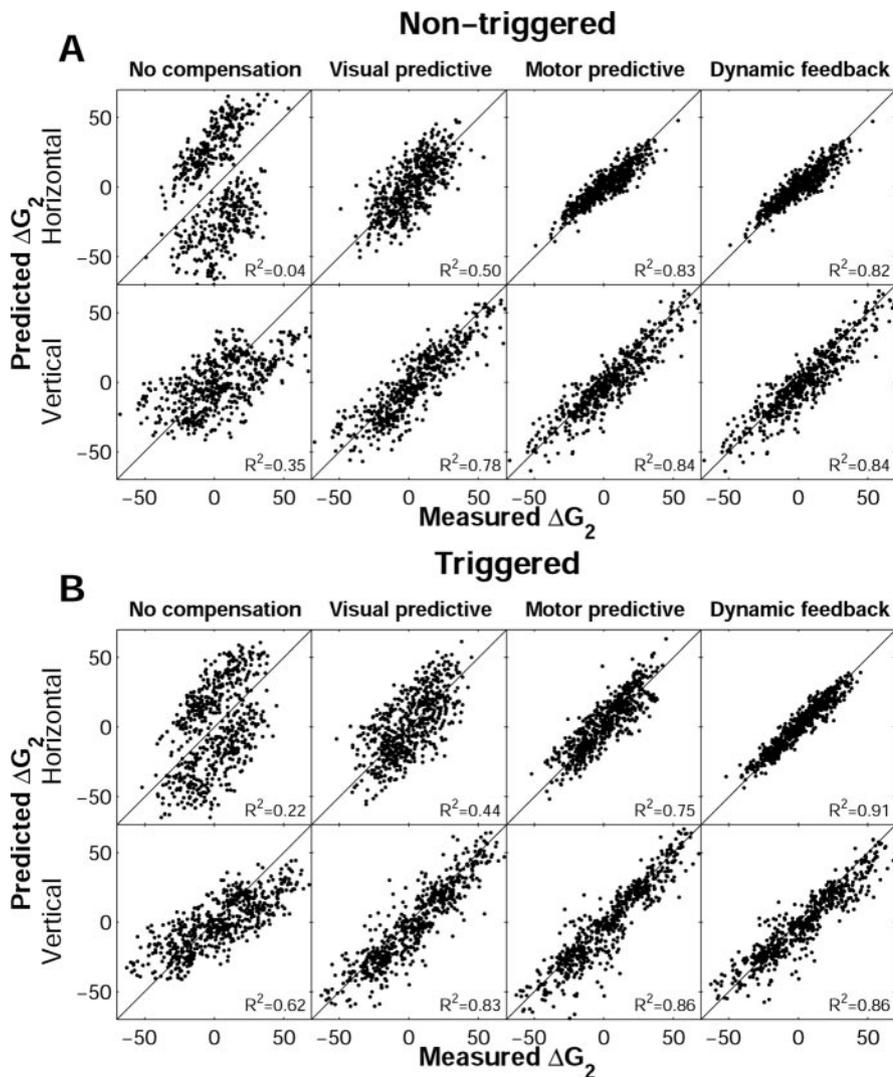
Figure 10. Predicted auditory-evoked gaze shifts (ΔG_2; ordinate) for the four models described in Results, plotted against measured responses (abcissa). Data are pooled across subjects and recording sessions. *A*, Static double-step condition, for horizontal (top row) and vertical (bottom row) response components. *B*, Dynamic double-step condition for both response components. If the model would predict ΔG_2 perfectly, data points would fall on the unity line, and R^2 would be 1. R^2 values are given in the bottom right corner of all panels. The predictions of the dynamic feedback model are superior to the other models.

mation about the first gaze displacement before it was actually generated.

Figure 10 shows the predicted gaze displacement, ΔG_2, for each of the four models, plotted against the measured gaze shift for the azimuth and elevation response components (pooled for all subjects and sessions), together with the R^2 values. Figure 10 *A* shows the results for the nontriggered double-step conditions, whereas Figure 10 *B* gives the predictions for the triggered double steps. As expected, the noncompensation model (left column) does not yield a good description of the measured data for either double-step condition. The predictive remapping model based on retinal error (visual predictive; Eq. 4a) (Fig. 10, second column) performs slightly better but is clearly inferior to the predictive remapping model that is based on the actually programmed first gaze shift (motor predictive; Eq. 4b) (Fig. 10, third column). In the nontriggered condition, performance of the motor-predictive model is equal to the dynamic feedback model (Fig. 10 *A*, right column). In the triggered double-step condition, however, the motor predictive model bases the updated craniocentric target location on the preprogrammed, full, first gaze shift,

whereas the dynamic feedback hypothesis updates the craniocentric target location with the partial gaze shift after the auditory target presentation (Fig. 1). In this condition, the dynamic feedback model provides the best prediction of the measurements (Fig. 10 *B*).

Short- versus long-duration sounds
Recent experiments have indicated that the auditory system needs a minimum duration ($\sim$20–40 msec) of broadband input to build a stable perception of sound-source elevation. For shorter sound durations, the elevation gain decreases systematically with either decreasing stimulus duration (Hofman and Van Opstal, 1998; Vliegen and Van Opstal, 2004) or increasing stimulus level (MacPherson and Middlebrooks, 2000; Vliegen and Van Opstal, 2004). The former phenomenon was proposed to be attributable to a neural integration process that improves its elevation estimate by accumulating spectral evidence about the current HRTF through consecutive short-term (few milliseconds) "looks" at the acoustic input.

So far, experiments that have studied the influence of sound duration have been performed with a stationary head during stimulus presentation. Because high-velocity ($>$200°/sec), two-dimensional head movements sweep the acoustic input across a multitude of different HRTFs on a short time scale, it is conceivable that the resulting dynamic changes in spectral input could interfere with the integrity of the neural integration process.

Suppose, however, that self-generated head movements would somehow enhance the performance of the short-term cue-extracting mechanisms. Accurate localization of elevation during rapid eye–head movements could then also be explained by a strategy that incorporates only a brief portion of the sound, say the first few milliseconds, while bypassing the neural integration stage. If true, short stimuli ($<$10 msec) should be localized better when presented during rapid head movements than without head movements. Moreover, there should be no benefit of longer stimulus durations during head movements.

To test these predictions, we repeated the single-step and static and dynamic double-step experiments with four subjects by presenting very short (3 msec) and longer (50 msec) acoustic stimuli (randomly interleaved across trials; late-triggered conditions only). Figure 11 summarizes the results as cumulative error distributions for the elevation response components for the two different stimulus durations (short, solid lines; long, lines through symbols) and three spatial-temporal target configurations (different gray codes: single-step, black; static double step, dark gray; dynamic double step, light gray). The figure shows that localization performance is quite comparable for the three conditions (single-step, static, and dynamic double steps), although the single-step trials yielded slightly more accurate responses
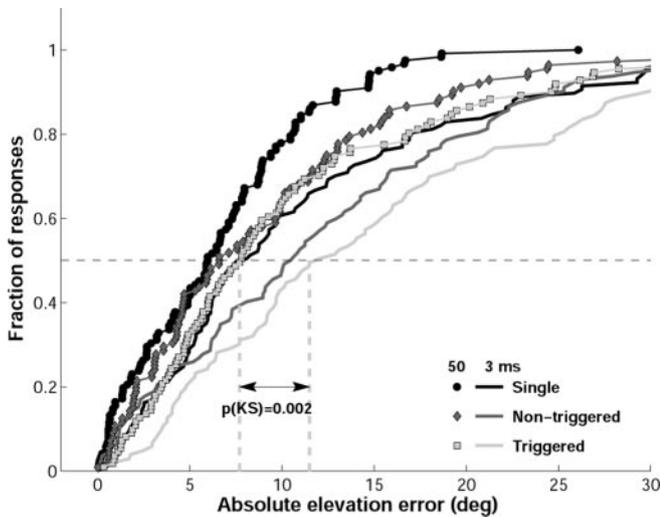
Figure 11. Comparison of gaze-elevation responses for single-step, static, and dynamic double-step conditions for very brief (3 msec) and longer (50 msec) sound durations. Data are pooled for all four subjects (JO, JV, MW, and RK), and ranked to create cumulative distributions of absolute elevation response errors for each of the three conditions and two stimulus durations (see legend for explanation of symbols and line styles). Note the larger errors in the 3 msec data when compared with the 50 msec data for all three conditions ($p < 0.025$; KS test). Thus, responses to the 50 msec stimuli are more accurate than to the 3 msec stimuli. The result for the triggered double-step responses is highlighted ($p = 0.002$). The horizontal dashed line at 50% is the median of the response distributions. For both stimulus durations, responses to the triggered and nontriggered double steps are slightly more inaccurate than to the single-step targets ($p < 0.05$).

than the two double-step conditions ($p < 0.05$; KS test). Thus, the self-generated head movements for short- and long-duration stimuli did clearly not enhance localization performance in the double-step experiments.

More importantly, however, for all three conditions the short stimuli yielded larger localization errors than the longer stimuli (KS statistic for 3 vs 50 msec data: single step: $p = 0.008$, d = 0.21; nontriggered double steps: $p = 0.02$, d = 0.20; triggered double steps: $p = 0.002$, d = 0.25). The median differences in absolute response errors were 2.1° for the single-step data, 4.0° for the nontriggered condition, and 4.3° for the triggered condition (indicated by arrowheads). The differences were negligible for the azimuth response components for all six conditions ($p > 0.05$; data not shown).

Discussion
Summary
Our results show that eye–head localization responses to brief acoustic noise bursts are equally accurate under fundamentally different stimulation conditions (Fig. 7, Table 1). In the (open-loop) dynamic double-step experiment, eye and head position, as well as the acoustic localization cues, varied at a high and variable speed during sound presentation (Fig. 5). Nevertheless, the intervening eye and head movements were, on average, fully compensated under all conditions tested (Fig. 8). Considering the complexities of the underlying coordinate transformations, this is quite a remarkable result. We also found that for all localization conditions, eyes and head both made goal-directed movements toward the sound (Fig. 9), which further strengthens the idea that they are each driven by motor commands in their own appropriate reference frame, rather than by a common signal. Finally, improved response accuracy for longer sound stimuli is preserved during rapid head movements (Fig. 11).

Comparison to other studies
In a recent study, Kopinska and Harris (2003) measured the lateralization perception of dichotic auditory targets. First, a sound was presented with head and body upright and eyes looking straight ahead. Then, subjects reproduced the memorized sound location in the head by adjusting the binaural level difference after adopting a new horizontal eye, head, or body posture. Neither eye position in the head nor whole-body orientation in space affected localization judgments. In contrast, head orientation on the body, and body orientation with the head fixed, had a small but systematic effect on response accuracy. The authors concluded that acoustic stimuli are expressed in a body-centered frame of reference.

Lewald and Ehrenstein (1998) and Lehwald et al. (2000) came to a similar conclusion when reporting small shifts in the localization of sounds after changes in horizontal head orientation.

Lewald and colleagues (1998, 2000) also reported systematic shifts in the perceived midline after changes in eye position. However, in dichotic (Kopinska and Harris, 1998) and free-field localization tasks (Goossens and Van Opstal, 1999; present study), effects of eye position were absent.

The studies by Kopinska and Harris (1998) and Lewald and colleagues (1998, 2000) all suggested that the sound-localization errors were caused by an inaccurate representation of the head-on-body signal. Because these experiments did not vary body and head orientation with respect to gravity, the static posture changes did not result in a tonic stimulation of the otoliths. In contrast, Goossens and Van Opstal (1999) found that saccadic eye movements made to free-field noise bursts were not systematically affected by static changes in vertical head tilt. The present data further extend these findings to dynamic localization conditions.

Note that the localization responses to pure tones under static head tilts did vary with head orientation. However, this effect was shown to depend strongly on the frequency of the sound, which suggested that a signal about head orientation is incorporated at a level in which acoustic input is still tonotopically represented (Goossens and Van Opstal, 1999).

Lewald et al. (2000) observed systematic undershoots for horizontal head pointing to free-field sounds. However, because they did not measure eye position, it is possible that their subjects actually pointed with their eyes, even when asked to point with their nose. Indeed, undershoots disappeared when subjects used a visual reference.

An important difference between our study and the previous studies resides in immediate, reflexive open-loop gaze orienting to brief sounds in our experiments, versus voluntary, perceptual and closed-loop localization tasks to long-duration stimuli in the other studies. It is conceivable that the mechanisms underlying action (rapid orienting) and perception (voluntary judgments) use different computational strategies, weightings, and neural pathways to update the frames of reference (see also below).

Note that because the subject's body was stationary in our experiments, we cannot distinguish body-centered from world-coordinate representations. Yet, we expect that rapid sound localization behavior will compensate for intervening changes in body orientation, too.

Implications for models
Single-step localization performance is explained by at least three different mechanisms (Fig. 1). The first model does not account for the intervening gaze shift under double-step conditions and can be readily dismissed because of the static double-step results.
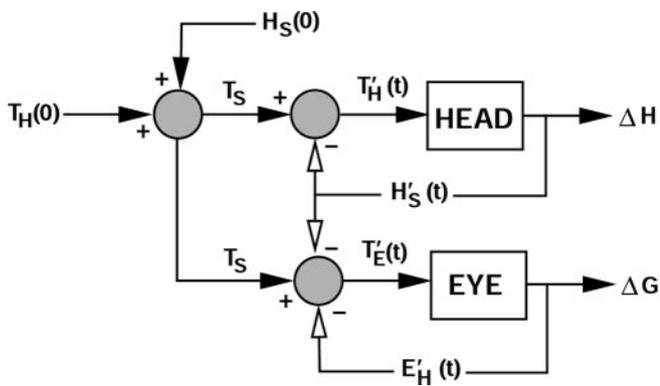
Figure 12. Conceptual scheme underlying accurate dynamic gaze control toward acoustic targets. Two stages are discerned. First, head-centered target location at target onset, $T_H(0)$, is added to head orientation in space at sound onset, $H_S(0)$, to create a stable representation of the sound in world coordinates, T_S. This target representation is kept in memory until a new target is selected. In the second stage, adequate motor commands for the eyes and head are generated by remapping the sound into head-centered ($T'_H(t)$) and eye-centered ($T'_E(t)$) target locations, respectively. This latter process requires signals about instantaneous eye position in the head, $E'_H(t)$, and about head orientation in space, $H'_S(t)$.

The latter condition still allows alternative possibilities to explain accurate localization behavior.

The predictive remapping interpretation is inspired by the neurophysiology underlying visuomotor behavior and was proposed by Goldberg and colleagues. Their studies convincingly demonstrated predictive visual responses in neurons within the posterior parietal cortex (Duhamel et al.,1992; Colby et al., 1995), frontal eye fields (Umeno and Goldberg, 1997), and superior colliculus (Walker et al., 1995). This activity preceded the occurrence of a saccade that would, after its completion, bring the stimulus into the visual receptive field of the cell. These predictive visual responses can be regarded to update the retinal coordinates of visual stimuli through an impending eye movement (efference copy) or, alternatively, by the retinal stimulus location evoking that eye movement. Predictive transformations could underlie the perception of a stable visual environment despite intervening saccades (transsaccadic integration) but could also explain the fast and accurate programming of subsequent eye movements in, for example, open-loop double steps or remembered target sequences.

Here, we propose that a similar mechanism might be used for the sound-localization system. During head movements, the system should be able to dissociate changes in acoustic cues attributable to self-motion from those that arise as a result of target motion. Furthermore, the localization perception of sound-sources needs to incorporate changes in head orientation to maintain spatial constancy and accuracy. A predictive mechanism that remaps perceived sound locations by impending head movements (or, alternatively, previous sound-source locations) could thus underlie the perception of a stable acoustic environment.

An alternative explanation for accurate double steps to visual or auditory targets, however, is that the target location is continuously updated, rather than beforehand. In this proposal, the target is mapped into a reference frame in world coordinates as soon as stimulus information becomes available and is kept in memory for as long as this information is required (Fig. 12). This transformation requires, in its simplest form, dynamic feedback about absolute eye position in the head, about head orientation on the body, and body orientation in space, rather than about impending displacement signals.

The predictive models and the dynamic feedback mechanism yield near-identical predictions for the static double-step trials (Figs. 1A, 10A). However, in the dynamic double-step paradigm, these schemes predict quite different updated target locations (Fig. 1B). Hallett and Lightstone (1976) showed that saccadic eye movements toward visual targets, flashed in mid-flight during an intervening saccade, remain accurate. While the predictive schemes yield an updated target location on the basis of wrong gaze-displacement information (Fig. 1B), only the dynamic feedback scheme predicts such accurate localization behavior. This is indeed the result of our sound-localization experiments (Fig. 10B).

Interestingly, responses for early-triggered and late-triggered double steps were equally accurate. This finding contrasts with experiments from the visuomotor literature showing that visual stimuli, presented around the onset of a saccadic eye movement, are systematically miss-localized (Dassonville et al., 1995; Schlag and Schlag-Rey, 2002).

Our data suggest that although predictive remapping could underlie the process of transsaccadic integration to form a stable spatial perception of the (visual and acoustic) environment, it is not adequate to update the target coordinates under dynamic spatial orienting tasks. Thus, we propose that the mechanisms underlying transsaccadic integration (presumably subserving spatial perception of the sensory scene) and target updating (dedicated to goal-directed actions toward specific stimuli) are quite different. Whereas the former may rely on an upcoming motor command, the latter needs to account for instantaneous changes in eye and head orientation. This is in agreement with recent visuomotor studies that show an effect of saccadic adaptation on the perceived target location, but not on the actual eye saccades, toward the target (Bahcall and Kowler, 1999; Tanaka, 2003).

Dynamic localization cues

Perrett and Noble (1997a,b) observed that in the absence of pinna cues, elevation localization improved with horizontal head movements, provided low frequencies were present in the signal. They suggested that the system uses dynamic changes in binaural timing differences to localize sound-source elevation. Other studies showed that slow head movements (Hofman et al., 2002) and self-induced stimulus motion (Wightman and Kistler, 1999) for long-duration broadband sounds can resolve front–back confusions. In contrast, Goossens and Van Opstal (1999) reported that rapid two-dimensional head movements did neither improve nor deteriorate the localization of pure tones with a duration of 500 msec, despite the systematic head movement-related changes in sound level that result when the tone sweeps across the different HRTFs. Thus, head movements are beneficial for some localization conditions but not for others.

Fast head movements during broadband sounds may also pose potential problems to the auditory system, because of the resulting rapid variations of spectral localization cues. These changes could interfere with the need to improve the elevation estimate through the integration of multiple short-term looks of the otherwise stable stimulus spectrum.

The results of the stimulus–duration experiments (Fig. 11) show, however, that there was neither an advantage nor a disadvantage of head movements because there was no difference between static and dynamic conditions. Moreover, the results indicate that the neural integration process also functions during fast head movements because elevation performance was significantly better for long versus short stimulus durations.

Taken together, the data from the current experiments sup-

port the possibility that the high-velocity and variable changes in head orientation are already incorporated at the stage of dynamic neural integration and that the outcome of this process could be the target in world-centered, rather than in head-centered, coordinates.

References

André-Deshays C, Berthoz A, Revel M (1988) Eye-head coupling in humans. I. Simultaneous recording of isolated motor units in dorsal neck muscles and horizontal eye movements. Exp Brain Res 69:399–406.

Bahcall DO, Kowler E (1999) Illusory shifts in visual direction accompany adaptation of saccadic eye movements. Nature 400:864–866.

Blauert J (1997) Spatial hearing: the psychophysics of human sound localization. Cambridge, MA: MIT.

Colby CL, Duhamel JR, Goldberg ME (1995) Oculocentric spatial representation in parietal cortex. Cereb Cortex 5:470–481.

Collewijn H, van der Mark F, Jansen TC (1975) Precise recording of human eye movements. Vision Res 15:447–450.

Dassonville P, Schlag J, Schalag-Rey M (1995) The use of egocentric and exocentric location cues in saccadic programming. Vision Res 35:2191–2199.

Duhamel JR, Colby CL, Goldberg ME (1992) The updating of the representation of visual space in parietal cortex by intended eye movements. Science 255:90–92.

Frens MA, Van Opstal AJ (1995) A quantitative study of auditory-evoked saccadic eye movements in two dimensions. Exp Brain Res 107:103–117.

Galiana HL, Guitton D (1992) Central organization and modeling of eye-head coordination during orienting gaze shifts. Ann NY Acad Sci 656:452–471.

Goossens HHLM, Van Opstal AJ (1997) Human eye-head coordination in two dimensions under different sensorimotor conditions. Exp Brain Res 114:542–560.

Goossens HHLM, Van Opstal AJ (1999) Influence of head position on the spatial representation of acoustic targets. J Neurophysiol 81:2720–2736.

Guitton D (1992) Control of eye-head coordination during orienting gaze shifts. Trends Neurosci 15:174–179.

Hallett PE, Lightstone AD (1976) Saccadic eye movements towards stimuli triggered by prior saccades. Vision Res 16:99–106.

Hofman PM, Van Opstal AJ (1998) Spectro-temporal factors in two-dimensional human sound localization. J Acoust Soc Am 103:2634–2648.

Hofman PM, Vlaming MS, Termeer PJ, Van Opstal AJ (2002) A method to induce swapped binaural hearing. J Neurosci Methods 113:167–179.

Jay MF, Sparks DL (1984) Auditory receptive fields in primate superior colliculus shift with changes in eye position. Nature 309:345–347.

Jay MF, Sparks DL (1987) Sensorimotor integration in the primate superior colliculus. II. Coordinates of auditory signals. J Neurophysiol 57:35–55.

Knudsen EI, Konishi M (1979) Mechanisms of sound localization in the barn owl (Tyto alba). J Comp Physiol 133:13–21.

Kopinska A, Harris LR (2003) Spatial representation in body coordinates: evidence from errors in remembering positions of visual and auditory targets after active eye, head, and body movements. Can J Exp Psychol 57:23–37.

Lewald J, Ehrenstein WH (1998) Influence of head-to-trunk position on sound lateralization. Exp Brain Res 121:230–238.

Lewald J, Dörrscheidt GJ, Ehrenstein WH (2000) Sound localization with eccentric head position. Behav Brain Res 108:105–125.

MacPherson EA, Middlebrooks JC (2000) Localization of brief sounds: effects of level and background noise. J Acoust Soc Am 108:1834–1849.

Middlebrooks JC (1992) Narrow-band sound localization related to external ear acoustics. J Acoust Soc Am 92:2607–2624.

Oldfield SR, Parker SP (1984) Acuity of sound localisation: a topography of auditory space. I. Normal hearing conditions. Perception 13:581–600.

Perrett S, Noble W (1997a) The effect of head rotations on vertical plane sound localization. J Acoust Soc Am 102:2325–2332.

Perrett S, Noble W (1997b) The contribution of head motion cues to localization of low-pass noise. Percept Psychophys 59:1018–1026.

Pöppel E (1973) Comment on "visual system's view of acoustic space." Nature 243:231.

Press WH, Flannery BP, Teukolsky SA, Vetterling WT (1992) Numerical recipes in C, Ed 2. Cambridge, MA: Cambridge UP.

Schlag J, Schlag-Rey M (2002) Through the eye, slowly: delays and localization errors in the visual system. Nat Rev Neurosci 3:191–215.

Tanaka M (2003) Contribution of signals downstream from adaptation to saccade programming. J Neurophysiol 90:2080–2086.

Umeno MM, Goldberg ME (1997) Spatial processing in the monkey frontal eye field. I. Predictive visual responses. J Neurophysiol 78:1373–1383.

Vidal PP, Roucoux A, Berthoz A (1982) Horizontal eye position-related activity in neck muscles of the alert cat. Exp Brain Res 46:448–453.

Vliegen J, Van Opstal AJ (2004) The influence of duration and level on human sound localization. J Acoust Soc Am 115:1705–1713.

Walker MF, Fitzgibbon EJ, Goldberg ME (1995) Neurons in the monkey superior colliculus predict the visual result of impending saccadic eye movements. J Neurophysiol 73:1988–2003.

Wightman FL, Kistler DJ (1989) Headphone simulation of free-field listening I: stimulus synthesis. J Acoust Soc Am 85:858–867.

Wightman FL, Kistler DJ (1999) Resolution of front-back ambiguity in spatial hearing by listener and source movement. J Acoust Soc Am 105:2841–2853.