‒ ‒

Multisensory spatial representations in eye-centered coordinates for reaching

‒ ‒

Pouget, Alexandre; Ducom, Jean Christophe; Torri, Jeffrey; Bavelier, Daphné

Brief article

Multisensory spatial representations in eye-centered coordinates for reaching

Alexandre Pouget[*], Jean-Christophe Ducom,
Jeffrey Torri, Daphne Bavelier

Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY 14627, USA

Abstract

Humans can reach for objects with their hands whether the objects are seen, heard or touched. Thus, the position of objects is recoded in a joint-centered frame of reference regardless of the sensory modality involved. Our study indicates that this frame of reference is not the only one shared across sensory modalities. The location of reaching targets is also encoded in eye-centered coordinates, whether the targets are visual, auditory, proprioceptive or imaginary. Furthermore, the remembered eye-centered location is updated after each eye and head movement. This is quite surprising since, in principle, a reaching motor command can be computed from any non-visual modality without ever recovering the eye-centered location of the stimulus. This finding may reflect the predominant role of vision in human spatial perception. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Multisensory spatial representations; Eye-centered coordinates; Reaching

1. Introduction

In order to reach for an object currently in view, our brain must compute the set of joint angles of our shoulder, arm and hand (a.k.a. the joint coordinates) that bring the fingers to the location of the target. This involves combining the retinal coordinates of the object – provided by the visual system – with posture signals such as the position of the eyes in the orbit and the position of the head with respect to the trunk. As illustrated in Fig. 1, this process can be broken down into several intermediate transformations in which the position of the object is successively recoded into a series of intermediate frames of reference (Soechting & Flanders, 1992).

* Corresponding author. Department of Brain and Cognitive Sciences, 402 Meliora Hall, University of Rochester, Rochester, NY 14627, USA. Tel.: +1-716-275-0760; fax: +1-716-442-9216.
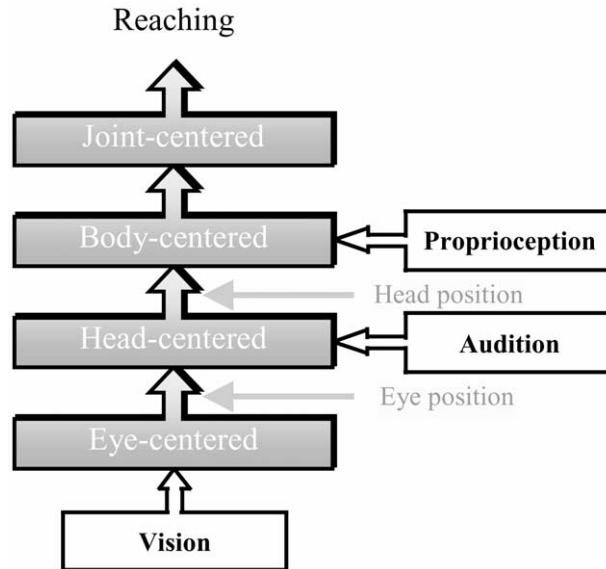 *E-mail address:* alex@bcs.rochester.edu (A. Pouget).

Reaching

Fig. 1. Coordinate transforms for multisensory motor transformations. In order to reach for a visual stimulus, the joint coordinates of the stimulus must be computed from its retinal coordinates, a process which can be decomposed into a series of intermediate transformations. Other modalities could enter this overall transformation at the level corresponding to the coordinates used in early stages of processing (e.g. head-centered coordinates for audition). This view predicts that the first frame of reference shared across all modalities is close to joint coordinates.

Humans can also reach for objects which are only heard or touched. Thus, similar coordinate transformations can be performed for other sensory modalities. Here we ask how the transformations for the various sensory modalities are integrated. One possibility is for each modality to enter the transformation at the level of its natural frame of reference, i.e. the one used in the early stages of processing (Fig. 1). For instance, audition could enter at the level of the head-centered frame of reference since the spatial position of a sound is computed by comparing the sound's arrival time and pressure differential across the two ears. Other modalities, such as touch and proprioception, might enter even later in the transformation, perhaps at the level of the body-centered frame of reference.

According to this scheme, the first representation shared by all modalities would have a frame of reference close to the one used for reaching, such as a joint-centered frame of reference. We present a series of experiments suggesting a different scenario. It appears that all modalities go through a stage in which the position of the object is encoded in eye-centered coordinates. This is surprising for the non-visual modalities since none of these modalities use an eye-centered frame of reference in the early stages of cortical processing and since the eye-centered frame of reference is not required – from a mathematical point of view – for tasks such as pointing.

Our study is based on an experiment by Bock (1986) showing that the eye-centered coordinates of visual targets influence the accuracy of hand pointing. His experiment
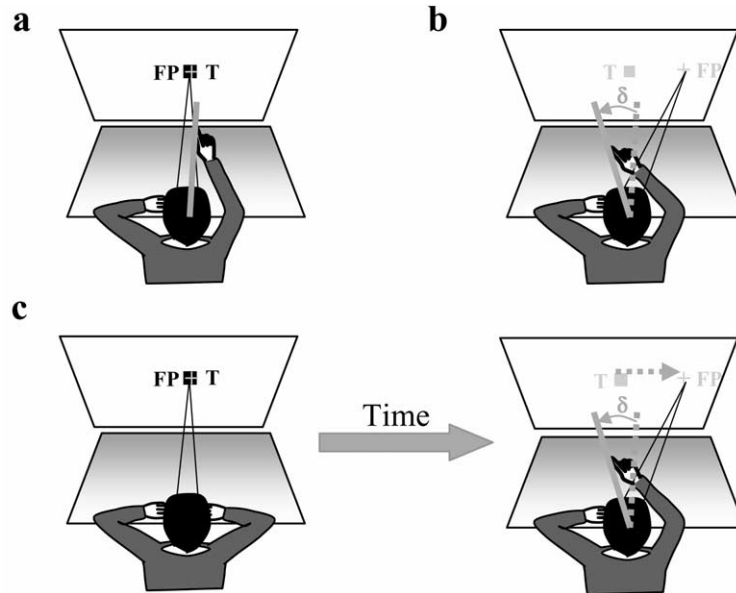
Fig. 2. Overshoot of a visual target as a function of retinal eccentricity. In all conditions, subjects are pointing without visual feedback from their hand. (A) Mean position of the hand when the subject points at a target (black square, T) located at $0°$ on the retina. (B) Mean position of the hand when the same subject points at a target located at $-10°$ on the retina (the retinal location of the target is set to $-10°$ by moving the fixation point, FP, $10°$ to the right). The mean position of the hand tends to be further to the left (the pointing direction from (A) is indicated as a dotted line). We refer to this shift (δ) as an overshoot because it indicates that the subject overestimates the retinal eccentricity of the target. The overshoot illustrated in this figure has been amplified for visual clarity – actual overshoots are of the order of a few degrees. The same overshoot has been found when subjects point to the remembered location of the target suggesting that the position of targets for pointing is memorized in eye-centered (retinal) coordinates (Enright, 1995; Henriques et al., 1998). (C) An eye movement (dotted arrow) is intervened between the offset of the target and the onset of pointing, such that the target is presented at $0°$ on the retina (left plot) but its updated position is $-10°$ by the time pointing is initiated (right plot, the gray square indicates the extinguished target). Under these conditions, the position of the hand reflects the updated retinal location, even though the target is no longer visible.

involved asking subjects to point to a visual target without visual feedback from their hand and with the eyes maintaining fixation at a point distinct from the pointing target. Bock found that, for retinal eccentricity within the $\pm10°$ range, subjects overshot the target by an amount related to its retinal eccentricity, that is to say, the pointing bias – the difference between the position of the target and the position of the hand – *increased* as the retinal eccentricity of the target *increased* (Fig. 2A,B). Beyond that range, the amplitude of the bias was found to saturate. Subsequently, Enright (1995) and Henriques, Klier, Smith, Lowy, and Crawford (1998) have shown that the same overshoot is observed when a delay is introduced between the offset of the target and the onset of the pointing movement. This suggests that the spatial location of visual targets for reaching is stored in an eye-centered representation since the amplitude of the overshoot depends on the retinal eccentricity of the target.

To remain spatially accurate, a representation using eye-centered coordinates must update the location of objects after each eye movement (Goldberg & Bruce, 1990). For instance, if one fixates an object and then moves one's eyes 10° to the left, the eye-centered position of the object changes from 0° to 10° to the right. More generally, if $\mathbf{R}_1$ is the eye-centered position of an object and the eyes move by $\mathbf{E}_1$, the new eye-centered position of the object is approximately $\mathbf{R}_2 \simeq \mathbf{R}_1 - \mathbf{E}_1$ (see Westheimer, 1957 for details). To determine whether spatial memory performs a similar remapping, Enright (1995) and Henriques et al. (1998) asked subjects to perform a saccade between the offset of the target and the onset of the pointing. The overshoot was found to reflect the retinal location of the stimulus after the saccade, demonstrating that spatial memory updates the eye-centered coordinates of the stimulus (Fig. 2C). Therefore, it appears that the coordinates used to remember the location of visual targets are centered on the eyes and are updated after each eye movement. These conclusions are further supported by recent neurophysiological recordings showing that the motor field of the majority of neurons in the reach area of the parietal cortex is defined in eye-centered coordinates and that neural activity in this area is updated after each eye movement (Batista, Buneo, Snyder, & Andersen, 1999).

In this study, we explored whether audition uses eye-centered coordinates for reaching. Previous work by Lewald and Ehrenstein (1996) suggests that this might be the case. Using a paradigm similar to the one used by Bock, they reported a pointing bias for auditory targets. However, their experiment did not involve memorized targets, making unclear whether eye-centered coordinates are used and updated to remember auditory targets. To address this question, we used an experimental design similar to the one developed by Henriques et al. (1998) (Fig. 3) with auditory targets. Like Henriques et al., we tested whether the update takes place with eye movements. We also used head movements to determine whether the update generalizes to the movement of other body parts besides the eyes. Finally, we investigated whether the use of eye-centered coordinates for reaching generalizes to proprioceptive and mental imagery targets.

2. Methods

In all experiments, subjects were asked to point at the perceived location of a target in the absence of visual feedback from their hand. Depending on the experiment, the target could be a visual stimulus, a sound, their right foot or the perceived straight-ahead direction. Targets were always located straight-ahead with respect to the subject's trunk. Two conditions were tested. In the static condition (Fig. 3a), subjects had to maintain fixation throughout the trial. The fixation point was located at either −10°, 0° or 10° (the minus sign refers to left positions), meaning that they had to point to a target with a retinal eccentricity of, respectively, 10°, 0° and −10°. In the dynamic condition (Fig. 3b), subjects always started with fixation at 0°. On one-third of the trials, they maintained fixation at 0° and initiated the pointing responses after the stimulus offset as in the static condition. On the other two-thirds of the trials, the fixation point moved to −10° or 10° shortly after the offset of the target, requiring subjects to perform a saccade. The pointing response was initiated after the completion of the saccade.
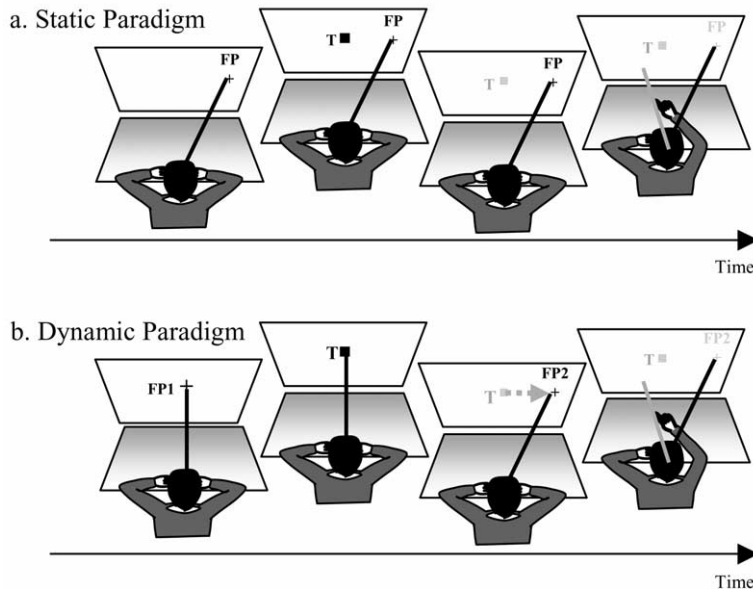
Fig. 3. Experimental procedure for the static (a) and dynamic (b) conditions. (A) Trials started with the appearance of the fixation point at 0°, −10° or 10° (shown here at 10°). Five hundred milliseconds after subjects started fixating, the pointing target was presented straight-ahead for 700 ms. Subjects were asked to wait until the disappearance of the fixation point before initiating their pointing movement to the remembered location of the target (extinguished target indicated in light gray). This occurred 700 ms after the offset of the target. They were also asked to maintain fixation during pointing even though the fixation point was no longer visible (the extinguished target and fixation point are indicated in gray). (B) In the dynamic condition, subjects viewed first the fixation point at 0°, and then the target at 0°. During the memory phase (after the disappearance of the target), the fixation point moved to −10° or 10° on two-thirds of the trials (the other one-third of the trials were identical to the static condition with fixation at 0°). Subjects were instructed to make a saccade to the new location of the fixation point and to maintain fixation there. The remaining part of the trial followed as in the static condition.

2.1. Subjects

All subjects (aged 18–35 years; all right-handed except two left-handers in the auditory paradigm) were healthy and naive as to the purpose of the studies. We ran new sets of subjects every time we tested a new modality to prevent a training effect across modalities. When static and dynamic paradigms were used, as was the case for the visual and auditory experiments, the same subjects were run in both conditions beginning with the static condition and then the dynamic condition 48–72 h later. The number of subjects who participated in each experiment is given in Fig. 4.

2.2. Equipment

A head-mounted display (HMD; Virtual Research V8; diagonal FOV = 60°; 640 × 480 pixels resolution at 60 Hz) was used for displaying visual stimuli binocularly. The position
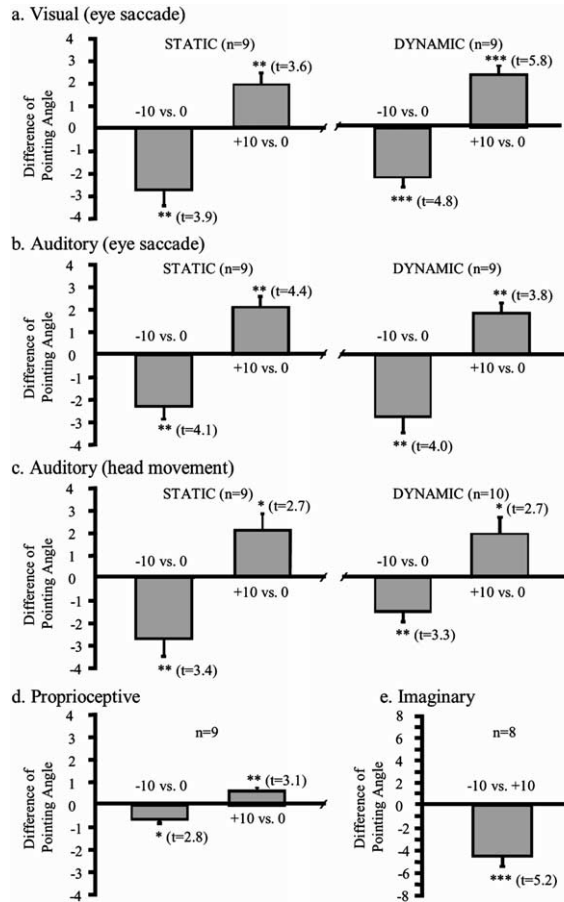
Fig. 4. Results for visual, auditory, proprioceptive and imaginary targets. In each case, we plot the difference in hand position when the eye-centered position of the target is moved (i) from −10° to 0° and (ii) from 10° to 0°. If the subjects overshoot the target, the difference in hand position is negative when the eye-centered position of the target is moved from −10° to 0° and positive when the eye-centered position of the target is moved from 10° to 0°. (A) Significant overshoots were found for pointing to visual targets in the static and dynamic conditions. (B) A similar pattern was found for auditory targets in the static and dynamic conditions using eye movements. (C) Overshoots for auditory targets were also observed when using head movements in the dynamic conditions. (D) Smaller but significant overshoots were found when subjects were asked to point to their right foot. (E) Finally, subjects also showed an overshoot in the imaginary condition in which they were asked to point to the perceived straight-ahead direction with respect to their trunk. Note that subjects were not tested with fixation straight-ahead in this case since the fixation point would have provided them with the trunk-centered straight-ahead direction they were asked to imagine during this task. Together, these results indicate that reaching relies on the eye-centered coordinates of the stimulus, independently of whether its location is specified visually, auditorily, proprioceptively or through imagery. Furthermore, these results indicate that the eye-centered coordinates of the objects are updated after each eye movement for visual and auditory targets, and also after head movements for auditory targets. The number of asterisks in each histogram indicates the one-tailed P value (*$0.01 < P < 0.05$, **$0.001 < P < 0.01$ and ***$P < 0.001$), n refers to the number of subjects, and t to Student's t value.

of the right eye was monitored by an infrared tracking system (ISCAN) mounted inside the HMD with a sampling rate of 60 Hz and an accuracy of 0.5° for horizontal movements and 1° for vertical movements. The positions and orientations of the head and of the pointing finger (index of the dominant hand) were recorded by a Polhemus Fastrak system using electromagnetic fields with a sampling rate of 20 Hz. The translational resolution is 0.0002 inches/inch of range with an accuracy of 0.03 inches root mean square (RMS). The angular resolution is ±0.025° with an accuracy of 0.15°.

Auditory targets were generated using one Audix PH3-S speaker ($4.7 \times 7.5 \times 4.7$ inches, 4 Ω, 20 W) positioned 52 inches from the subject's chin. The speaker was hidden behind a black curtain to prevent subjects from seeing it upon entering the experimental room when the light was still on.

2.3. Experimental paradigm

Subjects were seated in complete darkness with their pointing hand (dominant hand) resting on the table in front of them. This position was determined on an individual basis at the beginning of the experiment and subjects were constrained to start from the same resting position (i.e. within a 5×5 inch square at table level) for the extent of the experiment. The sequence of events and their timing are shown in Fig. 3. The fixation cross-hair was always red and subtended 1°. In the visual condition, green squares subtending a visual angle of 1° were used as visual targets. Subjects were asked to point to the place they thought the green square was presented. In the auditory condition, the target sounds consisted of 20 ms bursts of white noise at 89 dB separated by 10 ms silent gaps, for a total duration of 700 ms. Subjects were asked to point to the place they thought the sound was coming from. In the proprioceptive condition, the subject's right foot was placed straight-ahead on a foot holder at the beginning of the experiment. Movements were restrained by blocking the foot in a comfortable position for the subject during the whole length of the experiment. The subject's task was to point to the location of their right foot. In the imaginary condition, subjects were asked to point straight-ahead with respect to their trunk in the absence of any visual feedback.

When eye movements were manipulated, the eye position was enforced by a fixation cross-hair. If the eyes deviated more than ±0.7° for 0° eccentricity and ±1.5° for +10° or −10° eccentricity, the trial was discarded and replaced. Similarly, in these trials, the head was positioned to point straight-ahead with respect to the trunk, and movements were restricted by an adjustable chin-rest. If the head nevertheless moved (±2° in azimuth and ±3° in elevation), the trial was canceled and replaced.

When head movements were required, subjects were instructed to keep their eyes at 0° with respect to their head. In the static condition, this was attained by requiring subjects to first fixate a cross-hair in the center of the HMD (0° with respect to the head) and second, to rotate their head to align the cross-hair with the word "READY" which initially appeared at −10°, 0°, or 10°. This was achieved in the HMD by sliding the word "READY" in the opposite direction as the head movement. The sequence of events associated with a trial was then initiated if the subject kept head and eye positions within the values mentioned above. In the dynamic condition, the sequence of events was similar to the eye saccade dynamic condition except that subjects were allowed a longer time to perform the head

movement (1.8 s instead of 700 ms). As subjects moved their head, the final fixation point was seen sliding in the opposite direction toward the center of the HMD. Since this condition was harder than the others, the azimuth of the head was allowed to deviate $\pm 3°$ (instead of $\pm 2°$) for a trial to be included.

Subjects had no visual feedback from their hand. The subjects were asked to point to the target with a straight arm, with the elbow fully extended and locked. The pointing period lasted for 2.5 s and was ended by displaying the message "Done" at the final fixation location. This signaled the subject to lower the arm to the resting position. Each session consisted of a block of 60 trials following an initial training phase of approximately ten trials.

2.4. Data analysis

The final pointing angle was computed using the coordinates measured by the head receiver $\mathbf{H} = (x_h, y_h, z_h)$ and the pointing finger receiver $\mathbf{I} = (x_i, y_i, z_i)$. The last five values of the vectors $\mathbf{H}$ and $\mathbf{I}$ (about 0.25 s of pointing time) were averaged in order to reduce fluctuations over time. The pointing angle is then given by the following equation:

$$\alpha = a\tan((\langle x_i \rangle - \langle x_h \rangle)/(\langle y_i \rangle - \langle y_h \rangle))$$

where the brackets are used to indicate the mean over the last five values. Subjects were quite consistent in their pointing angle. The standard deviation of the pointing angle when pointing at $0°$ was $0.56°$ over the four experiments testing this position. Subjects that exhibited a large variance in their pointing (i.e. deviated by more than 3.5 times the average $0.56°$ standard deviation) were discarded. This only happened in the experiment requiring head movements (two subjects) which all subjects reported to be more difficult.

Pointing angles were computed for each retinal location of the target ($0°$, $-10°$ and $10°$). Pairwise one-tailed t-tests were then performed to compare pointing angles when the retinal location was non-zero to the $0°$ baseline condition.

3. Results

As in previous studies using this paradigm, we plotted our results in terms of the difference in hand position when the retinal location of the target was $0°$ vs. $-10°$ or $+10°$ (or equivalently, when the subject's eyes fixated at $0°$ vs. when their eyes fixated at either $+10°$ or $-10°$, Fig. 2).

First, we confirmed the findings of earlier studies for remembered visual targets (Enright, 1995; Henriques et al., 1998). In the static condition, subjects overshot the targets proportionally to the retinal eccentricity (Fig. 4A). For example, subjects tended to point further to the left when the retinal position of the target was moved from $0°$ to $10°$ left (Fig. 2). Furthermore, in the dynamic condition, the amplitude of the overshoot reflected the position of the visual target after the saccade (Fig. 4A), indicating a remapping of the remembered target to its new retinotopic position. In all cases, the amplitude of the overshoot is within the $1.5–3°$ range, which is comparable to the value reported in previous studies (Enright, 1995; Henriques et al., 1998). Next, we tested a new set of

subjects with auditory targets and observed a similar overshoot (Fig. 4B) in both conditions, static and dynamic, suggesting that the position of auditory targets is also remembered and updated in eye-centered coordinates.

An eye-centered representation must not only be updated after eye movements but also after head movements if they result in a change of gaze direction. To determine whether the updating takes place after head movements, we ran an additional experiment using auditory targets in which the retinal position of the target was manipulated by moving the subject's head while keeping their eyes at $0°$ with respect to their head. Once again, we found that the overshoot was proportional to the retinal location of the stimulus in both the static and the dynamic conditions, demonstrating that the eye-centered location of auditory targets is also remapped after head movements (Fig. 4C).

Our next experiments investigated whether the use of eye-centered coordinates to point to the location of objects generalizes to other modalities. To address this issue, we tested subjects with proprioceptive and imaginary targets. For the proprioceptive targets, subjects had to point to the tip of their right foot, placed straight-ahead with respect to their trunk. Unlike the visual and auditory trials, this condition did not involve spatial memory since this proprioceptive input was continuously available throughout the trial. Consequently, only the static condition was run. For the imaginary target, we tested whether subjects would show the overshoot for a target which has no physical existence and must be internally generated. Subjects were asked to point to the perceived straight-ahead direction with respect to their trunk. Again, only the static condition was run since this target was also continuously available to the subject, although not at the sensory level as in the proprioceptive condition. Results are shown in Fig. 4D,E. Once again, a significant overshoot was observed for both types of targets.

These findings indicate that the position of reaching targets is represented in eye-centered coordinates regardless of the sensory modality. Before we discuss the implications of these results, we need to address two potential problems with the methodology we have used. First, in our experiments, subjects were always required to point straight-ahead. One might therefore worry that the bias is specific to the straight-ahead direction and does not generalize to peripheral locations. We think this is very unlikely because the pointing bias has been reported for peripheral visual targets in previous studies (Bock, 1986; Enright, 1995). Moreover, we have used peripheral auditory targets in the static condition in pilot studies and found a significant bias (data not shown).

A second problem is that the eye-centered position of the target is manipulated by changing the position of the eyes. Accordingly, it is possible to argue that the overshoot is due to the position of the eyes and not the eye-centered position of the target. This can be tested by asking subjects to point to a target located, say, $10°$ to the left of the subject's body, and systematically varying eye position. The interesting comparison is when the subject fixates straight-ahead vs. $10°$ to the right, corresponding to eye-centered locations for the target of $10°$ vs. $20°$ to the right. If the eye-centered position matters, the overshoot should not increase because the overshoot saturates beyond the $\pm 10°$ range, as mentioned in Section 1. If the position of the eyes matters, the amplitude of the overshoot should increase. Bock performed this experiment and found the former to be true, hence confirming that the eye-centered position of the target is the critical variable.

4. Discussion

These data indicate that the position of reaching targets is represented in eye-centered coordinates regardless of the sensory modality. Furthermore, it appears that, for visual and auditory targets, these representations are updated after each eye or head movement. Note that our data do not rule out the existence of other representations for spatial memory using other frames of reference. It only indicates that the eye-centered reference frame is among the ones shared across modalities.

If proprioception uses an eye-centered frame of reference to encode limb position, as we are suggesting, one would expect that the perceived location of the pointing hand would also be biased. In fact, it is possible to account for our results by assuming that subjects *underestimate* the eye-centered position of their pointing hand as opposed to *overestimate* the eye-centered position of the target. This would explain why the overshoot was found to be smaller in the foot pointing experiments since a bias in the proprioceptive system would also apply to the estimation of the foot position. This hypothesis still suggests that proprioception uses an eye-centered frame of reference to encode the location of a target, a surprising result per se, but it would not tell us whether this conclusion generalizes to audition and visual imagery.

However, the notion that only proprioception uses eye-centered coordinates is inconsistent with the results of recent neurophysiological recordings in the parietal lobe. Two studies (Batista et al., 1999; Cohen & Andersen, 2000) have reported neurons in the reach area of the parietal lobe which respond before reaching movements toward visual and auditory targets and whose receptive, memory and motor fields are defined in eye-centered coordinates. This indicates that, in monkeys, vision and audition are combined into a common neural representation using eye-centered coordinates for reaching. The transformation of the position of auditory targets from head-centered to retinotopic coordinates appears to start early since auditory cells with head-centered receptive fields are modulated by the position of the gaze in the inferior colliculus (Groh, Trause, Underhill, Clark, & Inati, 2001) and in the primary auditory cortex (Trause, Werner-Reiss, Underhill, & Groh, 2000). These observations are consistent with our findings although more work is required before a causal relationship can be established.

Our experiments are not the first ones to suggest that auditory space can be remapped in eye-centered coordinates. It is well established that such a remapping takes place in the oculomotor system, such as when subjects are asked to foveate a sound source (Jay & Sparks, 1987; Stricanne, Andersen, & Mazzoni, 1996). Unlike reaching, however, eye movements are specified in eye-centered coordinates; it is therefore natural to find that auditory targets are remapped in eye-centered coordinates in the oculomotor system. By contrast, it is mathematically possible to compute a reaching motor command from the head-centered location of a sound – or any non-visual stimuli – without ever recovering its eye-centered coordinates (Fig. 1). Thus, it is surprising that auditory targets are remapped in eye-centered coordinates in the context of reaching.

To conclude, it appears that all sensory modalities use eye-centered coordinates for reaching. The choice of this particular frame of reference may reflect the pivotal role played by the visual system in human spatial perception.

Acknowledgements

References

Batista, A., Buneo, C., Snyder, L., & Andersen, R. (1999). Reach plans in eye-centered coordinates. *Science*, *285*, 257–260.

Bock, O. (1986). Contribution of retinal versus extraretinal signals towards visual localization. *Experimental Brain Research*, *64*, 476–482.

Cohen, Y., & Andersen, R. (2000). Reaches to sounds encoded in an eye-centered reference frame. *Neuron*, *27*, 647–652.

Enright, J. (1995). The nonvisual impact of eye orientation on eye-hand co-ordination. *Vision Research*, *35*, 1611–1618.

Goldberg, M., & Bruce, C. (1990). Primate frontal eye fields. III. Maintenance of spatially accurate saccade signal. *Journal of Neurophysiology*, *64*, 489–508.

Groh, J. M., Trause, A. S., Underhill, A. M., Clark, K. R., & Inati, S. (2001). Eye position influences auditory responses in primate inferior colliculus. *Neuron*, *29*, 509–518.

Henriques, D., Klier, E., Smith, M., Lowy, D., & Crawford, J. (1998). Gaze centered remapping of remembered visual space in an open-loop pointing task. *Journal of Neuroscience*, *18*, 1583–1594.

Jay, M. F., & Sparks, D. L. (1987). Sensorimotor integration in the primate superior colliculus. I. Motor convergence. *Journal of Neurophysiology*, *57*, 22–34.

Lewald, J., & Ehrenstein, W. (1996). The effect of eye position on auditory localization. *Experimental Brain Research*, *104*, 1586–1597.

Soechting, J. F., & Flanders, M. (1992). Moving in three-dimensional space: frames of reference, vectors, and coordinate systems. *Annual Review in Neuroscience*, *15*, 167–191.

Stricanne, B., Andersen, R., & Mazzoni, P. (1996). Eye-centered, head-centered, and intermediate coding of remembered sound locations in area lip. *Journal of Neurophysiology*, *76*, 2071–2076.

Trause, A.S., Werner-Reiss, U., Underhill, A.M., Groh, J.M. (2000). Effect of eye position on auditory signals in primate auditory cortex. Society for Neuroscience Abstracts, New Orleans.

Westheimer, G. (1957). Kinematics of the eye. *Journal of the Optical Society of America*, *47*, 967–974.