Multisensory guidance of orienting behavior

Joost X. Maier *, Jennifer M. Groh

Center for Cognitive Neuroscience, Department of Neurobiology, Department of Psychology and Neuroscience, Duke University, LSRC B203, Durham NC 27708, USA

**A R T I C L E   I N F O**

Article history:
Received 11 December 2008
Received in revised form 15 May 2009
Accepted 20 May 2009
Available online xxxx

**K e y w o r d s:**
Reference frame
Representational format
Superior colliculus
Auditory cortex
Multisensory
Orienting behavior

**A B S T R A C T**

We use both vision and audition when localizing objects and events in our environment. However, these sensory systems receive spatial information in different coordinate systems: sounds are localized using inter-aural and spectral cues, yielding a head-centered representation of space, whereas the visual system uses an eye-centered representation of space, based on the site of activation on the retina. In addition, the visual system employs a place-coded, retinotopic map of space, whereas the auditory system’s representational format is characterized by broad spatial tuning and a lack of topographical organization. Auditory spatial information must be transformed from a head-centered rate code to an eye-centered map to match the frame of reference used by the visual system, or vice versa. Here, we review a number of studies that have focused on the neural basis underlying such transformations in the primate auditory system. Although, these studies have found some evidence for such transformations, many differences in the way the auditory and visual system encode space exist throughout the auditory pathway. We will review these differences at the neural level, and will discuss them in relation to differences in the way auditory and visual information is used in guiding orienting movements.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

In our dynamic environment, we constantly shift our gaze to potentially relevant objects and events. These objects and events may be detected by one or more of the spatial sensory systems: vision, hearing and touch. This is advantageous when one modality fails, for example, vision in darkness or audition in a noisy environment. The ability to use multiple senses in combination can offer advantages, from faster behavioral responses to more accurate detection (for review see Rowe, 1999). In this paper, we will review data relating to the representation of spatial information in the primate auditory system, and the transformations auditory and visual signals undergo in order to access a common oculomotor pathway.

2. Reference frame

An important consideration in coordinating orienting behavior is the frame of reference of spatial information provided by the different senses. For example, receptive fields on the retina move with the eye, providing eye-centered spatial information about light sources. In contrast, sound localization relies on inter-aural time and level differences, as well as on spectral filtering properties of the pinnae (for review see Middlebrooks and Green, 1991), providing ear- or head-centered spatial information about sound sources. Because the eyes can move independently with respect to the head, reference frames for visual and auditory spatial information do not maintain a fixed relationship to each other. It seems that the misalignment between visual and auditory signals must be resolved in order to perform common behaviors such as making eye movements to sound sources, or localizing auditory–visual objects and events. It is generally assumed that the brain accomplishes this by converting spatial sensory information into a frame of reference that enables the execution of appropriate behavior. In this section, we will review evidence for such coordinate transformations in different areas along the auditory pathway.

In order to orient the eyes to the location of a sound source, head-centered auditory information would seem to have to be converted into a frame of reference suitable for accessing the oculomotor pathway. Conventionally, this is assumed to be an eye-centered frame of reference, based on evidence regarding the oculomotor representation in the superior colliculus, which is reviewed in detail below. To accomplish this coordinate transformation, information about the spatial location of the sound source with respect to the head must be combined with information about the position of the eyes. In theory, the location of a sound source in eye-centered coordinates (S_E) is given simply by the difference between the

---

Abbreviations: PPC, posterior parietal cortex; SC, superior colliculus
* Corresponding author.
E-mail address: joost.maier@gmail.com (J.X. Maier).
location in head-centered coordinates ($S_E$) and eye position ($E_H$) with respect to the head:

$$S_E = S_H - E_H$$

Technically, this is an overly simplified view because the relevant variables involve rotations of the eyes and thus involve horizontal, vertical and torsional components (for review see Kliger et al., 2003), but it will suffice for the purposes of discussion.

Electrophysiological studies in monkeys have investigated the encoding of auditory space in different areas along the auditory pathway. The rationale behind these studies is that if auditory space is encoded in an eye-centered frame of reference, auditory spatial tuning curves should shift proportional to changes in eye position. On the other hand, if auditory space is encoded in a head-centered frame of reference, auditory spatial tuning curves should be invariant to changes in eye position (Fig. 1A and B).

The main finding of all these studies is that eye position has a profound effect on the response of a portion of spatially sensitive neurons in areas along the auditory pathway of primates (Fig. 1C and D), namely, the inferior colliculus (Groh et al., 2001; Porter et al., 2004b; Zwiers et al., 2004) and the auditory cortex (Fu et al., 2004; Werner-Reiss et al., 2003). However, the frame of reference in which these areas represent sound source location does not seem to conform to descriptions in terms of head- or eye-centered-ness. Instead, eye position influences auditory responses in a complex way, resulting in spatial tuning curves that shift with the eyes, but in a partial and nonspecific way. We will term such representations hybrid because they fit poorly with both eye- and head-centered descriptions.

Because hybrid reference frames appear to be the rule rather than the exception, it is reasonable to assume that this type of coding has functional relevance for computing coordinate transformations. Indeed, computational models of multisensory coordinate transformations show that hybrid spatial representations can be used to provide auditory spatial location in effector-specific reference frames (Pouget and Sejnowski, 1997; Pouget et al., 2002; Xing and Andersen, 2000). For example, the oculomotor system could in principle use a hybrid representation of auditory space to read out eye-centered location of sound sources (Groh et al., 2001; Zwiers et al., 2004).

The studies discussed above indicate that the representation of auditory spatial information along the auditory pathway is neither eye- nor head-centered. This way of coding space is different from the eye-centered, retinotopic, frame of reference used by the visual system (for review see Wandell et al., 2007). However, the results leave open the possibility that the observed hybrid frame of reference for auditory space represents an intermediate step towards a complete coordinate transformation, and that eye-centered representations of sound source location are present at different stages of processing, especially within the oculomotor pathway. We will return to this question in Sections 1.4 and 1.5.

3. How does eye position information reach the auditory pathway?

The source of the eye position signal that is present along the auditory pathway is still unknown. It is also unknown whether the eye position effects observed in the different areas along the auditory pathway have one common source, or that they have different origins. Information about the position of the eyes could in principle come from two sources: proprioceptive signals from the eye muscles (‘inflow’), or efference copy signals of motor commands (‘outflow’). Proprioceptive eye position signals are present in monkey somatosensory cortex (Wang et al., 2007). Such eye position signals could reach primary and secondary auditory cortex through feedforward projections from somatosensory areas in the insula (Hackett et al., 2007; Schroeder et al., 2001), which receive direct projections from somatosensory cortex, and could from there be conveyed to the inferior colliculus. Efference copies of movement commands could arise from the superior colliculus, conveyed to different areas in the auditory pathways, including the inferior colliculus, auditory cortex and parietal cortex (for review see Sparks and Hartwich-Young, 1989).

4. Representational format

Auditory and visual systems differ not only in the coordinate system they use to represent space, but also in the representational format used for encoding spatial information (Groh, 2001). Visual areas are thought to represent visual space as a place code: each neuron in the population responds best to a narrow range of preferred stimulus locations, depending on the location of that neuron in the topographic map. Tuning curves of such neurons are non-monotonically shaped (Fig. 2A), and information is conveyed by the location of active neurons in a topographic map.

In contrast, studies that looked at the representational format used by the auditory system have yielded less clear results. Although, in some species (guinea pig, Binns et al., 1992; barn owl, Knudsen and Konishi, 1978) the inferior colliculus seems to contain a place-coded map of auditory space, numerous studies have failed to find a map of auditory space in the auditory system of primates. Instead, a variety of representational formats, without any obvious topographic organization, has been proposed to encode auditory space. Single auditory neurons have been shown to carry information about sound source location via temporal patterns in their spiking activity (Middlebrooks et al., 1994), spike latency (Brugge et al., 1998) and spike rate (Groh et al., 2003). For example, latency of the first spike occurring after presentation of an auditory stimulus varies systematically with sound source azimuth in cat auditory cortex neurons (Brugge et al., 1998). In the auditory cortex of rhesus monkeys, firing rate of single neurons is usually broadly tuned to sound source location – neurons respond to sounds presented in large regions of space (Benson et al., 1981; Recanzone et al., 2000; Werner-Reiss and Groh, 2008; Woods et al., 2006). Depending on the range of sound locations tested, these tuning curves are often monotonically shaped, with more contralateral sounds evoking stronger responses (Fig. 2B and C). The resulting distribution of preferred locations at the population level is heavily skewed towards contralateral locations (Werner-Reiss and Groh, 2008; Woods et al., 2006). A similar pattern of spatial sensitivity was found in monkey inferior colliculus neurons (Fig. 2C, (Groh et al., 2003; Zwiers et al., 2004)). This suggests that firing rates of neurons in the primate auditory system might encode information about sound source location using a rate code: stimulus location is conveyed by the rate of activity, rather than the location of active neurons in a topographic map, or the temporal patterns of the activity (for review see Porter and Groh, 2006a).

It is important to note that different representational formats (temporal and rate codes) are not mutually exclusive, and it is likely that a combination of temporal- and rate-based information is used for encoding sound source location. Indeed, using activity recorded from cat and ferret auditory cortex neurons, Nelken et al. (2005) showed that sound source location could be much better estimated when considering spike rate in conjunction with first spike latency measures, as opposed to either measure in isolation. However, the amount of information carried by single neurons is limited, and recently, models have been proposed that suggest that information about sound source location is encoded in the activity of populations of neurons (Miller and Recanzone, 2009; Stecker et al., 2009).
et al., 2005). For example, Miller and Recanzone (2009) showed that the firing rate of large populations of broadly tuned neurons recorded from monkey auditory cortex contains enough information to account for human sound localization data.

5. The oculomotor pathway: intraparietal cortex

Both visual and auditory signals converge on the lateral, medial, and ventral banks of the posterior parietal cortex (PPC) in primates,
a region that is commonly considered to be an association area and postulated to play a role in either the attentional selection of stimuli that may serve as saccade targets (for review see Colby et al., 1995) or more directly in the planning of saccades or other intended movements (for review see Andersen and Buneo, 2002). Recent work has suggested that both visual and auditory signals undergo a transformation to hybrid coordinates by the time they reach this structure (Mullette-Gillman et al., 2005, 2008; Schlack et al., 2005; Stricanne et al., 1996). For example, Mullette-Gillman et al. (2005, 2008) report that not only are auditory signals encoded in a continuum that ranges from eye- to head-centered with hybrid response patterns predominating, but so are visual signals (Fig. 3). This is a marked departure from the prior view of PPC as containing a predominantly eye-centered representation of visual space (Andersen et al., 1985, 1990; Batista et al., 1999).

Do these intraparietal signals play a role in guiding saccades? Evidence suggests that they might: microstimulation in the PPC evokes saccades with short latency (Shibutani et al., 1984). However, it is unknown what is exactly encoded in PPC activity. If PPC simply controls saccades to visual and auditory targets, the difference in reference frame between PPC and the superior colliculus (see below), downstream from PPC, is rather puzzling. On the other hand, PPC is not just involved in controlling simple eye movements, but also in goal-directed movements of other effectors, such as head movements, and visually- and auditory-guided reaches and grasps (Buneo et al., 2002; Cohen and Andersen, 2000). Accordingly, neuronal activity in PPC often shows a mix of eye-, head- and hand-related signals (Batista et al., 1999; Brotchie et al., 1995; Buneo et al., 2002; Cohen and Andersen, 2000), resulting in hybrid response patterns.
6. The oculomotor pathway: superior colliculus

Within the oculomotor pathway, the superior colliculus (SC) has received most scrutiny for its role in integrating visual and auditory signals (for review see Sparks, 1986; Stein and Meredith, 1993). In this section, we will review frame of reference and representational format of auditory-, visual- and saccade-related spatial information. The primate SC receives converging input from both cortical and subcortical sensory and motor systems, and projects to oculomotor nuclei in the brain stem (for review see Sparks, 2002). Microstimulation of the SC in monkeys evokes reliable gaze shifts with short latencies that involve saccades of both the eyes and head (Freedman et al., 1996; Robinson, 1972), and inactivation impairs these orienting movements (Hikosaka and Wurtz, 1985; Lee et al., 1988). Moreover, electrophysiological recordings have shown that the SC of monkeys contains neurons that are sensitive to the spatial location of visual (Cynader and Berman, 1972), auditory (Jay and Sparks, 1984, 1987a) and somatosensory (Groh and Sparks, 1996a,b) stimuli (Wallace et al., 1996). Given these anatomical and functional characteristics, the SC is thought to play an important role in the multisensory guidance of orienting movements.

6.1. Reference frame

Deep layers of the SC contain motor neurons that discharge just before and during the execution of saccades. Activity of these neurons is not tuned to shifting gaze to a particular location in space, but to saccades of a particular direction and amplitude (Schiller and Stryker, 1972; Wurtz and Goldberg, 1971). This was supported by the finding that microstimulation of a particular site in the superior colliculus evokes saccades of a fixed amplitude and direction, independent of initial eye position (Robinson, 1972; Schiller and Stryker, 1972). These findings suggest that the motor layers of the SC contain an eye-centered map of gaze displacement. The function of such a representation might be to provide a motor error signal: the shift that is required to direct gaze to a particular target, independent of the position of the eyes (for review see Sparks, 1986). Visual, auditory and somatosensory pathways all converge on this motor map (Groh and Sparks, 1996a; Jay and Sparks, 1987b). Accordingly, it has been suggested that visual, auditory and somatosensory spatial information needs to be encoded in a common, eye-centered, frame of reference at the level of the superior colliculus (Groh and Sparks, 1992, 1996b; Jay and Sparks, 1987a). Superficial layers of the SC in the monkey receive projections from retinotopic visual areas (for review see Sparks and Hartwich-Young, 1989), forming an eye-centered map of visual space (Cynader and Berman, 1972). Auditory information presumably reaches the SC via auditory and multisensory areas such as the inferior colliculus, auditory cortex and parietal cortex (for review see Sparks and Hartwich-Young, 1989). Jay and Sparks (1987a) investigated the frame of reference of auditory space in the monkey SC. This study used the same strategy as the studies investigating the frame of reference in various auditory areas described above: monkeys were presented with sounds originating at different locations in space, while varying their gaze direction. They found that auditory spatial receptive fields shift consistently in the direction of gaze. However, the receptive field shift was not proportional to the gaze shift: for a gaze shift of 24°, the average receptive field shift was only about 12° (see Fig. 4). These findings indicate that auditory space is not represented in an eye-centered frame of reference in the SC, and that on the basis of this representation of auditory space, unlike the retinotopic map of visual space, the SC would seem unable to provide an accurate motor error signal.

6.2. Representational format

The primate SC contains a place code for both visual target location and motor error signals: individual neurons have circumscribed visual receptive fields (Cynader and Berman, 1972) and/or movement fields (Wurtz and Goldberg, 1971), and microstimulation evokes saccades whose metrics depend on the site of stimulation (Robinson, 1972; Schiller and Stryker, 1972). However, little is known about the encoding of auditory space. Studies in the cat have shown that at least some neurons have circumscribed receptive fields (Middlebrooks and Knudsen, 1984). In the monkey, no quantitative or population analyses have been reported so far. However, the examples that have been published have included both monotonic and non-monotonic, broadly tuned neurons (Jay and Sparks, 1987a,b; Wallace et al., 1996), indicating uncertainty about whether the primate SC employs a place-coded map of auditory space.

7. Gaze orienting behavior

Primates are readily able to localize both light and sound sources by making eye, head, or combined eye-head movements with reasonable accuracy (humans, Carlile et al., 1997; Frens and Van Opstal, 1995; Makous and Middlebrooks, 1990; Populin, 2008; monkeys, Metzger et al., 2004; Populin, 2006). However, there are important differences between orienting to visual and auditory targets. For example, gaze shifts to visual and auditory...
targets have different temporal dynamics: auditory saccades are on average slower than visual ones (Goldring et al., 1996; Lueck et al., 1990), and some studies reported that when localizing sound sources, subjects often make an initial saccade of a fixed amplitude, followed by a more accurate corrective saccade (Jay and Sparks, 1990; Lueck et al., 1990; Zambarbari et al., 1982). Frens and Van Opstal (1995) also observed that saccades to sound sources are corrected towards the location of the target during execution, whereas gaze shifts to visual targets usually consist of a single, straight saccade. Onset latencies of auditory and visual saccades also differ: the latency of auditory saccades, unlike visual saccades, decreases with increasing amplitudes (Frens and Van Opstal, 1995; Goldring et al., 1996; Yao and Peck, 1997; Zambarbari et al., 1982).

An important aspect of auditory saccade behavior that relates to the reference frame of neural representation of sound source location is the influence of initial eye position on saccade accuracy. Several studies have investigated the effects of initial eye position on sound localization behavior in humans (Hartnagel et al., 2007; Lewald, 1997, 1998; Lewald and Ehrenstein, 1996; Razavi et al., 2007) and found modest effects. However, these studies used manual- or laser-pointing as a behavioral measure of localization ability. One study has quantitatively evaluated this issue in monkeys making saccades to visual and auditory stimuli from different initial eye positions, and found that the accuracy of auditory saccades was no more affected by eye position than were saccades to visual stimuli, suggesting that eye position is largely compensated for prior to the issuance of a motor command to move the eyes (Metzger et al., 2004). A related study in humans reported that saccades to sounds are accurate even when eye movements to an irrelevant visual stimulus intervene after sound presentation and before the saccade to the sound (Boucher et al., 2001).

Another issue that has been considered in the context of gaze shifts to auditory targets is the role of head movements. In principle, the freedom to move one’s head during sound presentation could enhance the accuracy of sound localization because interaural and spectral cues could be sampled multiple times with the head at different angles with respect to the sound (Perrett and Noble, 1997a,b; Wightman and Kistler, 1999). Indeed, there is evidence that the accuracy of gaze shifts to auditory targets is increased when subjects are free to move their heads (Brown et al., 1982; Populin, 2006).

Taken together, behavioral findings indicate that auditory and visual spatial information can both be used to guide reasonably accurate saccades despite variation in initial eye position. However, auditory saccades are usually less accurate than visual saccades and have a slightly lower peak velocity. This is not surprising given that these different sensory systems have evolved to detect different types of signals that can be used differently in guiding orienting behavior, necessitating synergy of processing. For example, the auditory system can detect stimuli throughout the full 360 degree range of our auditory system is much larger compared to our visual system. Also, auditory signals, unlike visual signals, are usually not blocked by obstacles. Therefore, the auditory system might provide a coarse initial estimate regarding the location of an object out of view. An initial eye (or head and body) movement guided by the auditory system might bring gaze to the general vicinity of the target, followed by more accurate localization guided by the visual system. To our knowledge, there is no experimental evidence for such sequential involvement of the auditory and visual systems, possibly due to the fact that targets are usually presented within the subjects’ visual field. However, there is evidence for interactions between auditory and visual signals from studies that show increased accuracy and decreased reaction times for saccades to bimodal (auditory–visual) targets (Colonnus and Arndt, 2001; Corneil et al., 2002; Perrott et al., 1990), indicating synergy of processing.

8. Conclusion

In this paper, we have reviewed electrophysiological and behavioral data relating to the way the auditory and visual system use spatial information to guide orienting behavior. As a result of the way the auditory and visual sensory organs are organized, auditory and visual information enters the central nervous system in different frames of reference. In addition, the representational format used to encode space also differs between the two sensory systems. Electrophysiological studies in monkeys as well as theoretical studies have often focused on how the brain reconciles these differences. The available neural data show that in the oculomotor system, at the level of the SC, the representations of auditory and visual space are still discrepant. Importantly, differences between auditory and visual saccades exist at the behavioral level as well. Therefore, this apparent discrepancy may be partly resolved by considering neural responses in relation to behavior. Our understanding of the neural representation of space has advanced largely independent of our understanding of its relation to behavior: most of the electrophysiological studies discussed above were performed in the absence of relevant behavior and/or under head-fixed conditions, thus preventing natural orienting behavior consisting of combined eye-head movements. Although, the available behavioral data show that the oculomotor system resolves some of the discrepancies between the representation of auditory and visual space (e.g., saccades to auditory targets largely compensate for initial eye position), many differences between orienting behavior to auditory and visual targets exist. Recording neural activity while subjects are performing orienting movements will shed light on how apparent discrepancies between neural activity evoked by auditory and visual signals might relate to differences in orienting behavior.

Acknowledgments

The authors thank Dave Bulkin for valuable discussions. This work was supported by NS50942-01, NSFO15634, EY016478 to JMG.

References


