Brain Research 1765 (2021) 147489

Contents lists available at ScienceDirect

Brain Research

journal homepage: www.elsevier.com/locate/brainres

Research report

Auditory cues facilitate object movement processing in human extrastriate visual cortex during simulated self-motion: A pilot study

Lucia M. Vaina [a,b,c], Finnegan J. Calabro [a,d], Abhisek Samal [a], Kunjan D. Rana [a], Fahimeh Mamashli [b,e], Sheraz Khan [b,e], Matti Hämäläinen [b,e], Seppo P. Ahlfors [b,e], Jyrki Ahveninen [b,e,*]

[a] Brain and Vision Research Laboratory, Department of Biomedical Engineering, Boston University, Boston, MA, USA
[b] Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital, Charlestown, MA, USA
[c] Harvard Medical School–Department of Neurology, Massachusetts General Hospital and Brigham and Women's Hospital, MA, USA
[d] Department of Psychiatry and Department of Bioengineering, University of Pittsburgh, Pittsburgh, PA, USA
[e] Department of Radiology, Harvard Medical School, Boston, MA, USA

ARTICLE INFO

Keywords:
Auditory
Visual
Crossmodal
Multisensory
Optic flow
Motion processing
Self-motion
Object motion
Magnetoencephalography

ABSTRACT

Visual segregation of moving objects is a considerable computational challenge when the observer moves through space. Recent psychophysical studies suggest that directionally congruent, moving auditory cues can substantially improve parsing object motion in such settings, but the exact brain mechanisms and visual processing stages that mediate these effects are still incompletely known. Here, we utilized multivariate pattern analyses (MVPA) of MRI-informed magnetoencephalography (MEG) source estimates to examine how crossmodal auditory cues facilitate motion detection during the observer's self-motion. During MEG recordings, participants identified a target object that moved either forward or backward within a visual scene that included nine identically textured objects simulating forward observer translation. Auditory motion cues 1) improved the behavioral accuracy of target localization, 2) significantly modulated the MEG source activity in the areas V2 and human middle temporal complex (hMT+), and 3) increased the accuracy at which the target movement direction could be decoded from hMT+ activity using MVPA. The increase of decoding accuracy by auditory cues in hMT+ was significant also when superior temporal activations in or near auditory cortices were regressed out from the hMT+ source activity to control for source estimation biases caused by point spread. Taken together, these results suggest that parsing object motion from self-motion-induced optic flow in the human extrastriate visual cortex can be facilitated by crossmodal influences from auditory system.

1. Introduction

Motion perception is critical for our ability to navigate and act within the environment. In the case of a stationary observer, detecting a moving object within an otherwise still scene is explainable by relatively simple filtering or perceptual grouping mechanisms (Dick et al., 1987; McLeod et al., 1988; Nakayama and Silverman, 1986). However, in most everyday settings, the retinal image constantly changes as a result of the observer's eye, head, and body movements, which produce a complex movement pattern of the visual field, referred to as optic flow (Gibson, 1950; Vaina, 1998). Parsing object motion from retinal translations caused by the observer's self-motion in such settings is a considerable computational feat, whose psychophysical (Burr and Thompson, 2011; Niehorster and Li, 2017; Rushton and Warren, 2005; Rushton et al., 2018; Warren and Rushton, 2009), neurocognitive (Calabro and Vaina, 2012; Pitzalis et al., 2020; Rana and Vaina, 2014), and neuronal (Duffy and Wurtz, 1991; Layton and Fajen, 2016) bases have been extensively studied using unimodal visual designs. In contrast, how motion information from other sensory systems, which is abundantly available in our everyday natural environment (e.g., the sound of an approaching vehicle), modulates cortical bases of visual flow-parsing is not yet fully clear.

The perceptual benefits of multisensory information have been well documented (Stein and Stanford, 2008). In the spatial domain, crossmodal influences are typically much stronger from the visual to auditory system than vice versa, both in the case of stimulus-source localization

* Corresponding author at: MGH/MIT/HMS-Martinos Center, Bldg. 149 13th Street, Charlestown, MA 02129, USA.
E-mail address: jahveninen@mgh.harvard.edu (J. Ahveninen).

https://doi.org/10.1016/j.brainres.2021.147489
Received 4 September 2020; Received in revised form 12 April 2021; Accepted 13 April 2021
Available online 18 April 2021

(Jack and Thurlow, 1973; Kopco et al., 2009) and motion perception (Alais and Burr, 2004; Bertelson and Radeau, 1981; Soto-Faraco et al., 2004). However, crossmodal auditory information can modulate visual motion perception (Cappe et al., 2009; Chien et al., 2013; Schmiedchen et al., 2012; Soto-Faraco et al., 2003), help orienting to relevant visual motion cues (Hanada et al., 2019), and modulate visual cortex activation during motion discrimination tasks (Kayser et al., 2017; Lewis and Noppeney, 2010). Neuroimaging studies, further, suggest that auditory motion cues can modulate visual processing of motion in the human middle temporal complex (hMT+) (Alink et al., 2008; Kayser and Kayser, 2018; Lewis and Noppeney, 2010; Rezk et al., 2020; Strnad et al., 2013; von Saldern and Noppeney, 2013), which is the human homolog of motion sensitive areas including the middle temporal (MT) and middle superior temporal (MST) regions of the monkey brain (Allman and Kaas, 1971; Dubner and Zeki, 1971). Human fMRI studies also suggest that hMT+ is activated by auditory motion stimuli in early blind (Poirier et al., 2006) and blindfolded sighted subjects (Poirier et al., 2005). However, the bulk of the above evidence has concentrated on the stationary observer's perspective; the way auditory motion cues modulate visual parsing of the optic flow that is caused by the observer's self-motion remains unknown.

Our two previous behavioral studies suggest that spatially congruent auditory cues could help identify a moving object in an environment where the observer is in simulated self-forward motion (Calabro et al., 2011; Roudaia et al., 2018). However, the exact level of visual processing at which these crossmodal effects originate is still unknown. For example, although evidence of crossmodal modulation of motion processing exists (see above), an argument has been made that even when the observer is stationary, the behavioral effects of auditory cues on visual motion processing originate at the decision-making rather than at the perceptual level (Meyer and Wuerger, 2001; Wuerger et al., 2012).

Here we used multivariate pattern analyses (MVPA) of MRI-informed magnetoencephalography (MEG) source estimates (for reviews, see Baillet, 2017; Hämäläinen et al., 1993) to examine how auditory cues of object motion in depth facilitate optic flow parsing during the observer's simulated self-motion. A moving target stimulus with a different speed and/or direction was embedded among moving distracter objects (Fig. 1). Our specific hypothesis was that auditory cues enhance the representation of the target stimulus already in the extrastriate visual-cortex area hMT+ (Fig. 2). MEG was considered ideal for estimating the information content in time-varying signals originating in sensory cortices during motion processing. Our previous behavioral studies suggest that in such a simulated flow-parsing task, the duration of the motion stimulus needs to be up to 600–1000 ms to allow optimal detection of targets that are embedded among moving distracter objects (Kozhemiako et al., 2020). In our MVPA analyses, we therefore compared temporally distributed representations of target vs. background motion across in the dynamic MEG activation time courses in visual-auditory and visual task conditions in different areas of visual cortex (Fig. 2).

2. Results

2.1. Behavioral

The behavioral analysis supported the initial hypothesis we put forward in our previous psychophysical study that an auditory cue can improve target identification performance during simulated self-motion (Calabro et al., 2011). The performance was significantly more accurate ($t_9 = 4.87$, FDR-corrected $p < 0.001$) Visual-Auditory Search than in the Visual Search condition. We also confirmed that the participants were able to identify the target object from the four possible alternatives

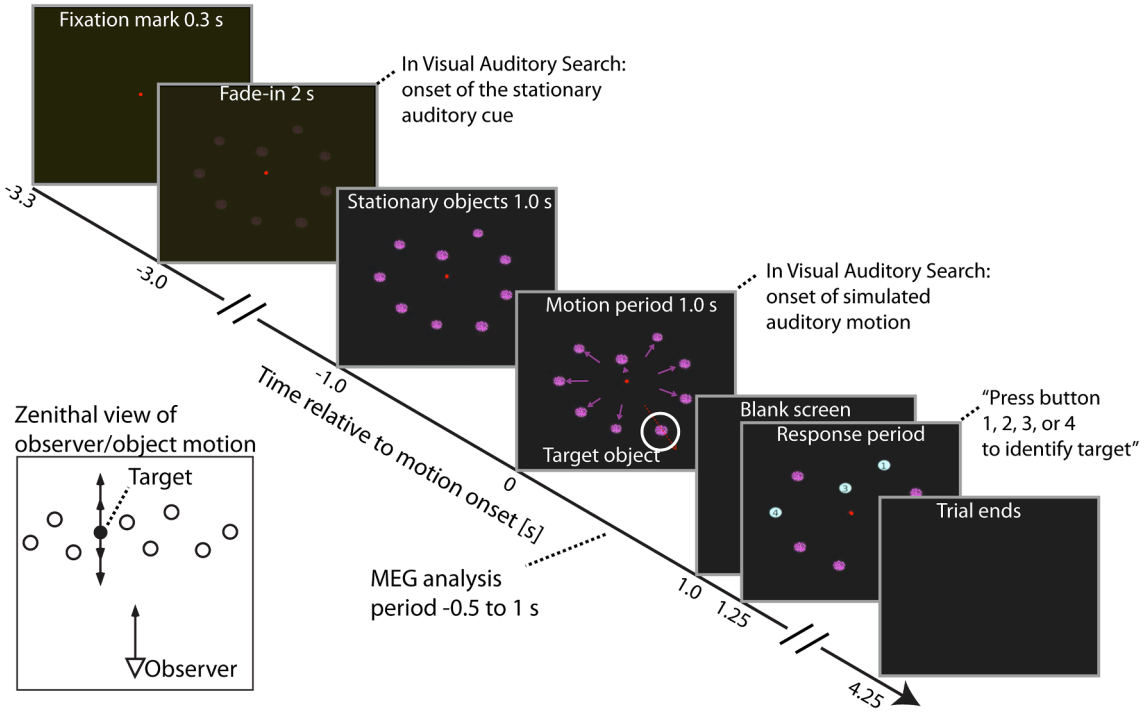

Fig. 1. Schematic view of the stimulus and task design (adapted from Calabro et al., 2011; Calabro and Vaina, 2012). Visual Search: Each trial started with a 300 ms of fixation period, after which 9 textured spheres faded in from the background for 2 s and then remained static for another second. During the subsequent one-second motion period, 8 out of the 9 spheres simulated the forward movement of the observer, and the remaining sphere, the target object, moved forward or backward with speeds different from the observer's motion. The left bottom insert shows a zenithal or "a bird's eye view" of the simulated motion pattern. After the motion ended, a blank screen appeared for 0.25 s, after which all spheres remained static, four of them were labeled with numbers, and subject identified by a button press which them had been the target. Visual-Auditory Search: The visual stimuli and the task were the same as in Visual Search. The amplitude of the continuous auditory cue, which appeared after the 300-ms fixation period, remained fixed until starting to increase or decrease to simulate auditory motion in depth congruent with the motion of the visual target sphere.

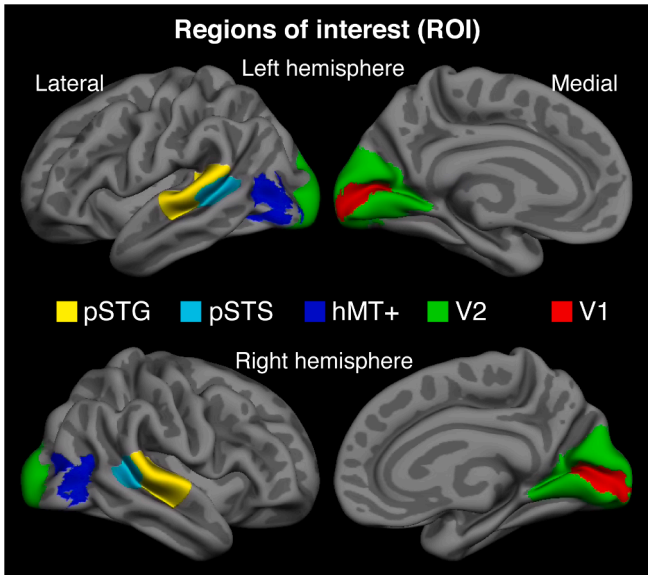

Fig. 2. Regions of interest (ROI) depicted on the "semi-inflated" cortical surface of the FreeSurfer standard brain representation.

significantly above chance level of 25% both in Visual-Auditory Search (mean $P_{Correct}$ ± standard deviation, SD, =0.64 ± 0.11, $t_9 = 11.50$, FDR-corrected $p < 0.001$) and Visual Search (mean $P_{Correct}$ ± SD = 0.54 ± 0.10, $t_9 = 9.24$, FDR-corrected $p < 0.001$) conditions. The individual data points of each subject's performance are shown in Supplementary Fig. 1. Finally, consistently with the behavioral accuracy, behavioral reaction times (RT), which were estimated from the four-object prompt presented after the motion had finished, were significantly faster ($t_9 =$ -5.29, FDR-corrected $p < 0.001$) in the Visual-Auditory Search (mean $RT \pm SD = 672 \pm 148$ ms) than in the Visual Search condition (mean $RT \pm SD = 754 \pm 156$ ms).

2.2. *Univariate analysis of crossmodal modulations of visual cortex motion processing*

Fig. 3 shows the effect of crossmodal auditory cues on the estimated activation time courses in the visual cortex areas V1, V2, and hMT+. Significant crossmodal modulations of visual cortex activity were found in V2 and hMT+. These effects were more prominent for trials in which the target was approaching the observer faster than the effect of simulated self-motion: the cluster-based randomization test indicated significant differences between the Visual Auditory Search vs. Visual Search conditions in V2 (three temporal clusters, each with $p < 0.05$) and two in hMT+ (two clusters; one with $p < 0.01$, another with $p < 0.05$). For trials in which the target was receding, only one significant cluster was found, in hMT+ ($p < 0.05$). No significant effects were found in the area V1 in either case. We also compared motion-related responses between trials with receding vs. approaching targets within the area hMT+; this comparison revealed no significant differences in either Visual or Visual-Auditory Search condition.

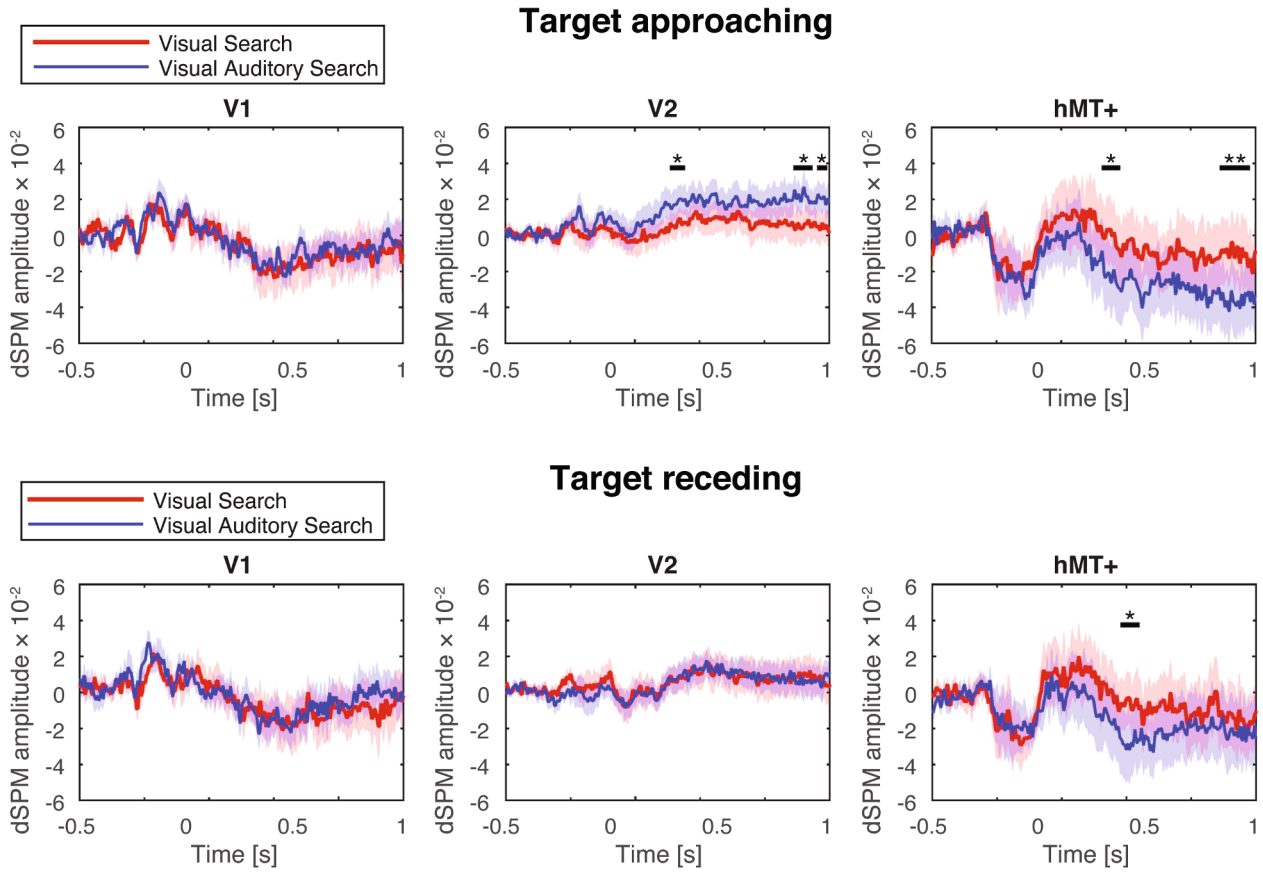

Fig. 3. Comparison of MEG source activity in the visual cortex ROIs across Visual-Auditory Search and Visual Search conditions, separately for trials with approaching vs. receding targets. Significant differences between Visual-Auditory Search vs. Visual Search conditions were found in hMT+ and, in the case of approaching targets, also in V2. * $p < 0.05$, ** $p < 0.01$, cluster-based randomization test. The red and blue shading reflects the standard error of mean across subjects.

2.3. MVPA

Whereas the univariate analyses did not reveal direction-specific effects in hMT+, the MVPA analyses suggested that crossmodal auditory motion cues might, nonetheless, help visual processing of the target stimulus motion that is embedded within the much larger retinal effects due to the observer's self-motion (Figs. 4, 5; Table 1). In the Visual-Auditory Search condition, the direction of the target stimulus motion was decoded significantly above the chance level of 0.5 in the area hMT+ ($t_9 = 3.34$, FDR-corrected $p < 0.05$; numerical details in Table 1). The target motion direction could not be decoded from any of the ROIs in the Visual Search condition. In the comparison across the conditions, the decoding accuracy was significantly higher in the Visual-Auditory Search than in the Visual Search condition in the area hMT+ ($t_9 = 3.59$, FDR-corrected $p < 0.05$; Fig. 4, right panel), but not in V1 or V2. There were no significant correlations between the MVPA decoding accuracy and behavioral measures in the Visual Auditory or Visual Search conditions. An additional control MVPA of posterior parietal cortex during Visual Auditory vs. Visual Search conditions is presented in Supplementary Material.

2.4. Control analyses

Not surprisingly, the direction of the target motion in the Visual-Auditory condition could be classified significantly above the chance level using the estimated MEG source waveforms for the auditory-cortex area pSTG ($t_9 = 9.16$, FDR-corrected $p < 0.001$; numerical details in Table 1). This was also the case in the adjacent area pSTS ($t_9 = 5.95$, FDR-corrected $p < 0.01$; see Table 1), which is a major audiovisual integration site and putative auditory association area. To control for possible crosstalk between ROIs in the MEG source estimates, we performed a control MVPA in which the contribution of pSTG and pSTS to the single-trial activity of hMT+ was regressed out (Fig. 5). The target stimulus motion direction could be decoded significantly above the chance level from the source waveforms of area hMT+ in the Visual-Auditory Search condition even in this case ($t_9 = 3.49$, FDR-corrected $p < 0.05$, see Table 2 for numerical details), but not from those of areas V1 and V2.

We also tested the ROI-specific SVM decoding accuracies in the Visual Control and Auditory Control experiments. Note that in the Visual Control experiment, the motion pattern was much broader and visually clearer than in the two search conditions: the task was to determine whether the entire nine-object pattern was moving towards or away from the observer. The direction of this stimulus was decodable

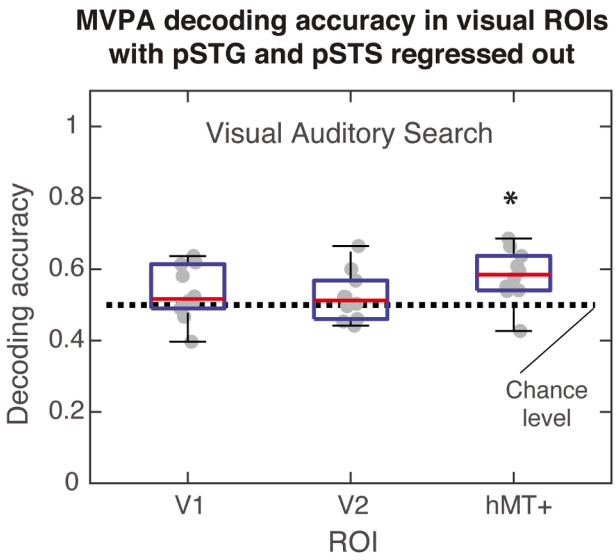

Fig. 5. Control MVPA where the contribution of pSTG and pSTS was regressed out from the visual cortex time courses in Visual-Auditory Search condition. The decoding accuracy of the target direction was significantly above the chance level in the area hMT+, but not in V1 and V2. On each box, the central mark is the median, the bottom box edge is the first quartile, the top edge the third quartile, and the whiskers reflect the 1.5 × inter-quartile range of the group-level decoding accuracy. The gray dots reflect individual observers' data. * FDR-corrected $p < 0.05$ (Benjamini and Hochberg, 1995).

significantly above chance level from all the visual cortex ROIs, including hMT+ ($t_9 = 4.99$, FDR-corrected $p < 0.01$; numerical details in Table 1), V2 ($t_9 = 3.10$, FDR-corrected $p < 0.05$, see Table 1), and V1 ($t_9 = 4.90$, FDR-corrected $p < 0.01$, see Table 1), but not in the auditory area pSTG or pSTS.

In contrast, in the Auditory Control experiment, wherein the observer was asked to attend to the auditory stimulus only and determine its direction, the direction of motion could be decoded significantly above chance level from pSTG ($t_9 = 6.03$, FDR-corrected $p < 0.01$) and also from pSTS ($t_9 = 2.74$, FDR-corrected $p < 0.05$). The MVPA effects in hMT+ ($t_9 = 0.98$, FDR-corrected $p = 0.27$) and the other visual cortex ROIs remained non-significant after FDR correction in the Auditory Control experiment. However, there was a trend-level effect in the area V2. We therefore also analyzed the decoding accuracies from visual cortex ROI time courses from which the contribution of pSTG and pSTS

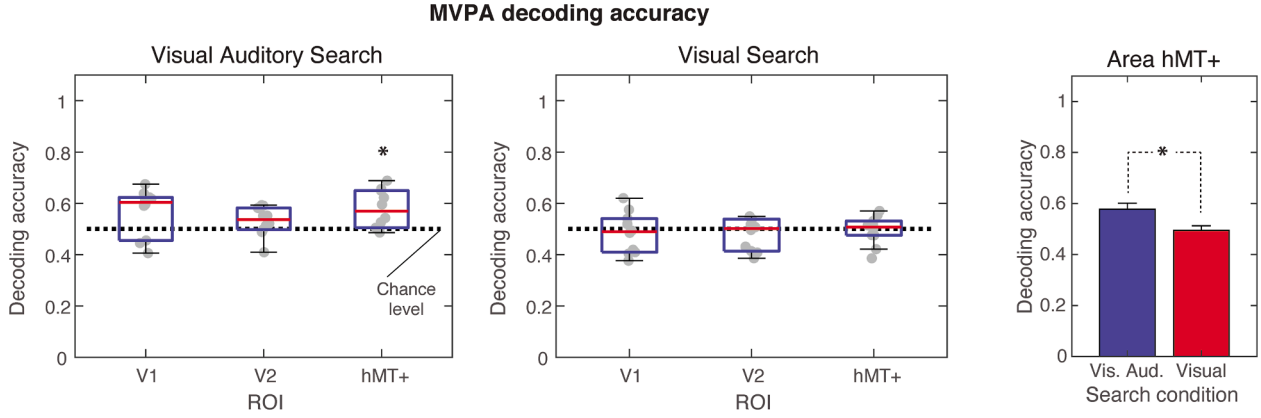

Fig. 4. MVPA results in visual cortices. (**Left, Middle**) The decoding accuracy of the target direction was significantly above chance level during Visual-Auditory Search condition in the area hMT (left), but not in any of the ROIs during the Visual Search experiment (middle panel). On each box, the central mark is the median, the bottom box edge is the first quartile, the top edge the third quartile, and the whiskers reflect the 1.5 × inter-quartile range of the group-level decoding accuracy. The gray dots reflect individual observers' data. * FDR-corrected $p < 0.05$ (**Right**) In the area hMT+, there was a significant difference between the decoding accuracies between the Visual Auditory and Visual Search conditions. Error bars of the show the standard error of mean. * FDR-corrected $p < 0.05$.

Table 1
Accuracy of MVPA decoding of target motion direction using estimate MEG source waveforms for the 5 regions of interest. Group-mean and standard errors of mean (SEM), as well as the *t*-value, effect size (Cohen's *d*), and *p*-value (FDR corrected) are given.

Condition	ROI	Mean	SEM	*t*(9)	Cohen's *d*	FDR-corrected *p*	
Visual Search	hMT+	0.49	0.02	−0.31	−0.10	0.6629	
	V1	0.48	0.03	−0.61	−0.19	0.7482	
	V2	0.48	0.02	−1.02	−0.32	0.8321	
	pSTS	0.51	0.03	0.22	0.07	0.5494	
	pSTG	0.49	0.04	−0.18	−0.06	0.6349	
Visual-Auditory Search	hMT+	0.58	0.02	3.34	1.05	0.0181	*
	V1	0.57	0.03	2.23	0.70	0.0642	
	V2	0.53	0.02	1.67	0.53	0.1449	
	pSTS	0.66	0.03	5.95	1.88	0.0010	**
	pSTG	0.74	0.03	9.16	2.90	0.0001	***
Auditory Control	hMT+	0.52	0.02	0.98	0.31	0.2703	
	V1	0.52	0.03	0.62	0.20	0.3809	
	V2	0.56	0.03	2.38	0.75	0.0540	
	pSTS	0.57	0.03	2.74	0.87	0.0330	*
	pSTG	0.66	0.03	6.03	1.91	0.0010	**
Visual Control	hMT+	0.62	0.02	4.99	1.58	0.0025	**
	V1	0.59	0.02	4.90	1.55	0.0025	**
	V2	0.58	0.03	3.10	0.98	0.0206	*
	pSTS	0.53	0.04	0.92	0.29	0.2781	
	pSTG	0.50	0.03	−0.04	−0.01	0.6238	

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

Table 2
MVPA decoding accuracy in the control analyses where the contribution of pSTG and pSTS was regressed out of the visual cortex ROI time courses. Group-mean and standard errors of mean (SEM), as well as the respective statistical parameters, are shown.

Condition	ROI	Mean	SEM	*t*(9)	Cohen's *d*	FDR-corrected *p*	
Visual-Auditory Search	hMT+	0.58	0.02	3.49	1.10	0.0165	*
	V1	0.53	0.02	1.41	0.44	0.2005	
	V2	0.52	0.02	1.05	0.33	0.2599	
Auditory Control	hMT+	0.50	0.02	−0.04	−0.01	0.6238	
	V1	0.50	0.02	−0.16	−0.05	0.6349	
	V2	0.52	0.02	1.31	0.41	0.2149	

*$p < 0.05$

was regressed out. In contrast to the MVPA results in the Visual-Auditory Search condition (Fig. 5), which were not affected with this regression procedure, the results of this analysis were clearly non-significant in all the visual cortex ROIs (Table 2).

3. Discussion

Little is known about how crossmodal auditory cues influence visual cortex processing of moving objects when the observer passages through the 3D space. We addressed this question using a combination of behavioral methods and MRI-informed MEG source modeling analyses. During MEG recordings, participants were asked to identify a target object that moved either forward or backward within a visual scene. The display consisted of nine identically textured objects, eight of which simulated the observer's forward translation. Our results demonstrate that auditory motion cues not only improve behavioral detection of moving target stimuli during the observer's simulated self-motion, but that they also modulate representations of the target motion direction in hMT+, an extrastriate visual cortex area that has been previously implicated in optic flow parsing (Layton and Fajen, 2016; Pitzalis et al., 2020; Rushton et al., 2018).

The present behavioral findings are consistent with our previous psychophysical studies, which suggested that auditory motion cues facilitate object movement detection within a dynamic scene that simulates the observer's self-motion (Calabro et al., 2011; Roudaia et al., 2018). Our behavioral findings are also in line with previous psychophysical evidence that auditory cues modulate visual motion processing within an otherwise stationary scene (Cappe et al., 2009; Chien et al., 2013; Schmiedchen et al., 2012; Soto-Faraco et al., 2003). Here, the auditory-to-visual performance modulation, further, emerged in a high-load setting where the single target object motion needed to be parsed from the much stronger retinal motion pattern of the eight non-target objects. This is consistent with the expectation that auditory cues benefit visuospatial processing specifically when the task-relevant visual stimulus is embedded among competing stimuli (Van der Burg et al., 2008, 2011) or presented in a peripheral location (Ahveninen et al., 2019; Teramoto et al., 2012).

The area hMT+, which showed the strongest crossmodal MEG effects among the visual cortex ROIs in the present study, refers to the human homolog of the combination of areas MT and MST of non-human primates. Whereas MT is densely populated by direction-selective neurons with relatively focal receptive fields (Allman and Kaas, 1971; Dubner and Zeki, 1971), the area MST includes neurons that are sensitive to larger-field optic flow effects such as those involved in the observers self-motion (Duffy and Wurtz, 1991). In their neural model, Layton and Fajen (2016) proposed that that visual flow parsing during self-motion is based on a feedback-feedforward loop that involves both MT and MST. There is also evidence of multisensory MST neurons that integrate visual and vestibular information to support the observer's heading perception (DeAngelis and Angelaki, 2012). Although functional imaging studies of auditory effects on optic flow parsing are lacking, human neuroimaging activations in or near hMT+ have been reported to be modulated by auditory inputs during more conventional motion processing tasks (Alink et al., 2008; Kayser and Kayser, 2018; Lewis and Noppeney, 2010; Strnad et al., 2013; von Saldern and Noppeney, 2013). One possible explanation for our behavioral result that auditory cues facilitate detection of moving objects from dynamic optic flow, observed here and in our previous studies (Calabro et al., 2011; Roudaia et al., 2018), could thus be crossmodal enhancement of motion sensitivity of hMT+.

In contrast to the long-held view that multisensory integration occurs at higher levels only (Barlow, 1972; Konorski, 1967), more recent evidence suggests that crossmodal auditory modulation of visual processing already in or near the human visual cortex areas V1 and V2 (Giard and Peronnet, 1999; Murray et al., 2016; Raij et al., 2010). One of the proposed explanations for these previously-reported early effects is direct lateral connectivity from auditory cortices to early visual areas, documented in non-human primates (Falchier et al., 2002, 2010; Majka et al., 2019; Rockland and Ojima, 2003). However, direct monosynaptic connections between auditory cortices and the areas MT or MST in primates

appear to be sparse (Palmer and Rosa, 2006). It is therefore possible that the auditory modulations of hMT+ observed here reflect enhanced feedback from other higher-level areas, which support multisensory integration (Smiley et al., 2007) and attentional orienting to the relevant object within the scene (Green et al., 2005; Hillyard et al., 2016). Such a feedback effect could allow a more efficient flow parsing and, consequently, a more precise representation of the target object's motion relative to its dynamic background. However, future studies using measures of effective connectivity (e.g., dynamic causality modeling) are clearly needed to determine the flow of bottom-up vs. top-down types of inputs that underlie the present multisensory effects.

More specifically, in the spatial domain, the auditory system is thought to provide a fast and non-specific alerting mechanism, which supports orienting of visual attention to relevant objects or events that are beyond the observer's immediate focus (Ahveninen et al., 2019; Hillyard et al., 2016; McDonald et al., 2000). One of the key aspects of for this role could be assumed to be the rapid detection of approaching, potentially harmful objects or events (Ghazanfar et al., 2002). Previous studies on multisensory integration of motion information suggest that auditory cues provide a much stronger influence on the perception of approaching/looming than receding objects (Cappe et al., 2009). Consistently with these premises, the present evidence of auditory modulations of MEG source activity in the areas hMT+ and V2 were significant only in trials where the target object was approaching the observer (Fig. 3). The enhanced classification accuracy in our MVPA analyses at the area hMT+ could therefore have been contributed by direction-specific enhancement of feedback from higher areas that are sensitive to auditory looming, such as pSTS (Tyll et al., 2013).

In the present MEG MVPA analyses, we specifically aimed at making sure that the higher decoding accuracy in Visual-Auditory Search than in Visual Search reflects genuine crossmodal information transfer to human visual cortices. To this end, we used a linear regression analysis to exclude the contribution of direct auditory-cortical contribution from single-trial hMT+ activity. In this analysis, we also considered the potential crosstalk from pSTS, which is a well-established multisensory integration site. In line with our main findings, the direction of visual-auditory target motion could be classified from the hMT+ activity even when the contribution of pSTG and pSTS was regressed out (Fig. 5). The validity of our results is also supported by the analysis of Control Experiments, in which the decoding accuracy of the motion direction of the auditory cue alone was not significantly above chance level in the visual cortex ROIs. Thus, the combination of visual and auditory information was necessary for the target motion to be decodable from the area hMT+. Further studies using direct neurophysiological measures of neuronal firing activity, as opposed to local-field potentials or their MEG, EEG, or electrocorticography (EcoG) correlates, are however needed to get a more complete picture of the exact neuronal mechanisms of the present result.

There were no significant correlations between the MVPA decoding accuracy and behavioral measures in the Visual Auditory or Visual Search conditions at the group level. The signal to noise ratio (SNR) of this correlation analysis could have been reduced by the fact that the present task was not related purely on motion perception but required also spatial working memory and decision making (across four possible behavioral responses). It is also noteworthy that MVPA decoding accuracy is a statistic that is affected by a number of factors beyond the strength of the neuronal representation, per se, which may increase the between-subject variability and reduce the power of the correlation analysis between MVPA and behavioral results. Yet another consideration is that the present flow parsing task was based on a relatively simple virtual design where the subject did not actually move. Future studies with either more realistic virtual reality stimulation devices that provide richer depth cues or studies utilizing wireless EEG devices that allow brain recordings during the participant's real self-motion could be helpful.

The present MVPAs were computed to decode temporally distributed representations of the target's motion relative to its dynamic background. A limitation of this analysis is that whereas it answers to the question whether a certain ROI contains information of the properties of the target stimulus, it does not reveal the exact time course of the evolution of such representations. However, it should be noted that it takes a while until a human observer can parse a moving target object from a dynamic pattern that simulates the retinal effects of one's self-motion. While the integration of global motion patterns is achievable based on motion patterns briefer than 200 ms (Lee and Lu, 2010), our previous tests suggest that in the present flow parsing setup, human observers need to sample at least 600 ms of motion to allow above-chance-level behavioral performance across all target speeds in the unimodal condition (Kozhemiako et al., 2020). Given this notion, we utilized the entire 1000-ms motion period from each ROI for our MVPAs. Notably, in this analysis, the primary features of interest were the dynamic variations in the signal. In other words, the data were not collapsed or averaged across the time; all available temporally dynamic information was taken into account to make the predictions of the target stimulus direction. Another limitation of our study is that in the present task version, we kept track only of the position of the correct target stimulus, which prevented us from analyzing how the information content of the MEG signal relates to the exact position of the stimulus that the subject chose. A related potential issue is that we did not use dynamic eye tracking in our experiment: We cannot thus rule out the possibility that upon detecting the differently moving target, some subjects could have inadvertently moved their fixation from the center of the screen to the target, which could have modulated the downstream motion processing in such trials. Further, because the trial-specific position coordinates of targets were not available from the two first subjects of our sample, we did not attempt to decode the spatial position of the target, in addition to the target direction. Finally, the present study sample consisted of 10 participants, which could be considered suboptimal for the generalizability of the results (Friston et al., 1999). Because of this limitation, our analysis was concentrated in relatively focused ROI-based approach instead of more open-ended exploration of different cortical networks involved in motion processing. Future studies with larger number of participants are needed to achieve more complete picture of the broader brain networks involved in motion processing.

In conclusion, our MEG source modeling analyses and MVPA results suggest that the population-level neuronal effects, which accompany crossmodal enhancement of detecting a moving object during the observer's simulated self-motion, involve motion processing areas in or near the hMT+ of human extrastriate visual cortex.

4. Methods and materials

Ten participants (3 females) with normal or corrected-to-normal vision, normal hearing, and mean age of 28.6 years (standard deviation 15.7 years) participated in the study. All procedures were approved by the Institutional Review Boards of Massachusetts General Hospital and Boston University, and all participants gave Informed Consent to participate in this research study. All participants were naïve as to the purpose of the experiment.

4.1. Stimuli and tasks

Fig. 1 shows the task design, which was modified for MEG experiments from our previous psychophysical studies (Calabro et al., 2011; Calabro and Vaina, 2012). The participants were asked to identify a target object that moved either forward or backward within a visual scene, which included nine identically textured high-contrast objects (28.3 cd/m^2 on a 0.3 cd/m^2 background) that simulated forward observer translation (Rana and Vaina, 2014). The stimuli were projected to an MEG-compatible screen, placed 80 cm away from the eyes of the participant, using a LP350 DLP projector (InFocus, Wilsonville, OR) at a resolution of 1024 × 768 pixels with refresh rate of 75 Hz. The room was

darkened. The mean initial diameter of the objects, which were distributed within a 25 × 25 × 60 cm^3 volume centered at a distance of 80 cm, was 1.5° of visual angle. To prevent occlusion between spheres, the display area was nominally divided into nine equally sized wedges, each of which contained one sphere displayed at a random eccentricity of up to 9°. Participants were instructed to maintain fixation on a red circle placed at the center of the MEG-compatible video display. The task timeline was as follows: 1) The fixation mark appeared and was shown alone for 300 ms; 2) During the next 2000 ms, the nine spheres were faded in from 0% to 100% of the final contrast; 3) The spheres remained stationary for another 1000 ms; 4) A 1000-ms stimulus-motion period started thereafter: 8 of the nine spheres were moved and scaled consistent with forward observer translation of 3 cm/s, such that the radial velocity was up to 1.66°/s for the most eccentric spheres, or 0.84°/s for spheres of mean eccentricity, whereas the remaining (target) sphere had an independent forward or backward motion vector of 2, 4, 6 or 8 cm/s within the scene; 5) After the motion period, a blank screen was shown for 250 ms, after which the 9 spheres were displayed static for 3000 ms. The target and three other randomly selected spheres were labeled with numerals 1, 2, 3, and 4. In a four-alternative forced choice (4AFC) task, the observers were asked to identify the target sphere by pressing the corresponding button 1, 2, 3, or 4 of a MEG-compatible response pad with their right-hand fingers. No feedback was provided. After 3000 ms the static spheres disappeared from the screen and a new trial started.

In separate conditions, the above task was presented only visually (Visual Search) or with a co-localized auditory motion cue (Visual-Auditory Search). The auditory cue was a pure tone of frequency 300 Hz whose amplitude changed to simulate movement within the scene in the same forward or backward direction as the visual target sphere. Our working hypothesis and the idea of this manipulation was that this kind of general auditory cue would strengthen the representation of the motion of the visual target sphere among the non-target spheres. The auditory cue was delivered through foam earpieces that were connected to three-meter long plastic tubes, connected to a speaker system outside of the magnetically shielded room (Unides ADU1b, Helsinki, Finland). The change in amplitude was modeled as a sound source at an initial distance of 4.1 m (69 dB SPL, 30 ms ramps), whose amplitude was reduced or increased by approximately 10 dB SPL, to simulate auditory movement towards or away from the observer at 3.5 m/s (Fig. 1). The sound level change as a function of distance was approximated by using a least-squares fit to measurements of sound levels at various distances from a constant sound source.

In additional control conditions, we presented the visual motion pattern that simulated the observer motion only ("Visual Control") or the auditory cue atop the stationary object view alone ("Auditory Control"). The Visual Control task was otherwise similar to the other visual tasks, but during the 1000-ms motion period, the observer motion was simulated in either forward or backward direction (sphere expansion or contraction). In a two-alternative forced choice (2AFC) task, the observers reported the direction of the motion by pressing predefined keys on the response keypad. In the Auditory Control task, the participant was presented with the auditory cue using the same task timeline as in the Visual-Auditory Search task and the participants' task was to press one button if the simulated sound source was receding and another if it was approaching. The auditory cue in this control task was exactly the same than in the main experiment. The visual display was otherwise similar to the main tasks, but the nine objects remained stationary and no target-indices appeared.

Each participant participated in one MEG session where they were presented two consecutive runs of Visual Search task (160 trials in total), two runs of Visual-Auditory Search task (160 trials in total), evenly interleaving each target speed (20 trials per speed) in a random order in each run. These main task runs were preceded by two runs of Visual Control (160 trials, equiprobable forward/backward motion) and two runs of Auditory Control (160 trials, equiprobable forward/backward motion). Prior to the MEG scanning session, all participants practiced the psychophysical task in the laboratory (at Boston University) and they practiced the task until their performance was significantly above chance ($p < 0.01$).

4.2. Data acquisition.

MEG data were acquired inside a magnetically shielded room (Imedco AG, Switzerland) using a whole-head VectorView MEG system (Elekta-Neuromag, Finland), which has 306 sensors arranged in 102 triplets of two orthogonal planar gradiometers and one magnetometer. The signals were bandpass filtered between 0.5 and 200 Hz and sampled at 600 Hz. The position and orientation of the head with respect to the MEG sensor array was recorded with help of four head position indicator coils. To allow co-registration of the MEG and MRI data, the locations of three fiduciary points (nasion and pre-auricular points) that define a head-based coordinate system, a set of points on the head surface, and the sites of the head position indicator coils were digitized using a Fastrak digitizer (Polhemus) integrated with the Vectorview system. Vertical and horizontal electro-oculogram (EOG) was also recorded to monitor eye-movements and blinks. The MEG data were synchronized with the stimulus presentation by sending a transistor transistor-logic (TTL) trigger pulse from the experimental computer to the MEG device. The onset of the TTL pulse was synchronized with the timing of visual stimulus presentation on the screen, by using our in-house Matlab software "BraviShell", an extension of Psychtoolbox.

T1 weighted structural MRI scans were acquired on a separate day using an 8-channel phase array head coil in a 3 T scanner (Siemens-Trio, Erlangen, Germany; distance factor 50%; slices per slab 128; FOV 256; FOV phase 100; slice thickness 1.33 mm, TR 2530 ms, TE 3.39 ms).

4.3. Data analysis

Responses were recorded for each trial in every participant. A trial was treated as having no response, and discarded, if the participant did not respond within the 3000 ms allocated for response. All 10 participants performed above chance level (25%) in all task conditions.

The MEG data were preprocessed using MNE software (Gramfort et al., 2014). The data were downsampled to 300 samples/s at passband of 0.5–100 Hz. The MEG data was divided into epochs of 1.5 s, ranging from 500 ms before to 1000 ms after the onset of the visual or visual-auditory motion. Cardiac and ocular artefacts were first attenuated by a signal space projection (SSP). Epochs with residual artefacts (>4pT/m peak-to-peak gradiometer, >4pT in magnetometers) were rejected. Across all participants, the smallest number of acceptable trials per condition (target approaching vs. receding) was 77 out of the maximum of 80 trials.

Cortical sources of the MEG signals were estimated by calculating ℓ_2 minimum-norm estimates (MNE) (Hämäläinen et al., 1993; Lin et al., 2006). To co-register each individual participant's MEG and structural MRI data the fiduciary landmarks and digitized points were aligned with the reconstructed skin surface from the MRI by applying an iterative closest point algorithm. Based on structural segmentation of the individual MRIs using FreeSurfer (http://surfer.nmr.mgh.harvard.edu/) and the MEG sensor locations, forward solutions were computed using a single-compartment boundary element model (Hämäläinen and Sarvas, 1989). The source space consisted of current dipoles placed on the cortical surfaces at ~ 9,000 locations per hemisphere. The noise covariance matrix was estimated from the raw MEG data during a period of 300 ms before the motion onset.

Single-trial noise-normalized source estimates were averaged within five *a priori* regions-of-interest (ROI) from early visual, multisensory, and auditory cortex areas presumed to be associated with processing of motion (Fig. 2). 1) Three ROIs representing visual cortex areas V1, V2, and hMT+ were defined based on the surface-based *ex vivo* atlas of FreeSurfer, and then coregistered from the standard brain

representation to each individual participant's cortical surface representations. 2) Processing of auditory motion towards and away the observer was assumed to take place in posterior non-primary auditory cortices (Ahveninen et al., 2014; Kopco et al., 2012, 2019), corresponding to the posterior half of superior temporal gyrus label of the FreeSurfer Desikan atlas (Desikan et al., 2006), referred here to as pSTG. 3) A ROI encompassing the putative multisensory area posterior superior temporal sulcus (pSTS) was defined based on the Freesurfer Desikan atlas.

The present stimulus consisted of a set of objects distributed across both visual fields. Previous studies suggest that optic flow parsing in such a setting involves a global analysis of retinal motion (Warren and Rushton, 2008). Therefore, we used the information from both left- and right-hemispheric aspects of each ROI in our main MVPAs.

4.4. Analysis of behavioral data

The proportions of correct behavioral responses (P_{Correct}) during the MEG recordings were analyzed statistically using t-tests, corrected for multiple comparisons using the false discovery rate (FDR) procedure of (Benjamini and Hochberg, 1995). In addition to verifying that the observers performed significantly above chance level (25%) in the 4AFC task, we compared the P_{Correct} values between the Visual-Auditory Search and Visual Search conditions.

4.5. Univariate analysis of MEG source estimates

The main focus of the present study was the analysis of crossmodal auditory influences on visual motion processing during the observer's simulated self-motion. As the auditory cue was very robust in comparison to the motion of the target object, which was embedded among the 8 distracters that simulated the retinal effects of the observers self-motion, our analysis concentrated on the visual cortex and any effects in or near the auditory cortex were analyzed for control purposes only.

Crossmodal modulations of early visual cortex areas were analyzed by comparing the activations during the stimulus motion during the Visual-Auditory Search and Visual Search conditions. Given potential direction-specific influences by auditory cues on visual motion processing (Cappe et al., 2009), the crossmodal effects of auditory cues were analyzed separately for approaching vs. receding target trials. Statistical analyses of ROI time courses of the trial-averaged source estimates, which were pooled across hemispheres, were conducted using the Fieldtrip cluster-based randomization test (cluster-forming threshold $p < 0.05$) (Maris and Oostenveld, 2007).

4.6. Multivariate pattern analysis of optic flow parsing

Machine learning techniques were used to determine whether auditory motion cues enhanced visual cortex representation of the moving target object embedded within the optic flow caused by the observer's simulated self-motion. While constant perceptual sensitivity of motion perception can be achieved with samples shorter than 200 ms (Lee and Lu, 2010), our previous study using the present set up suggests that integration over longer temporal windows (from 600 to 1000 ms) is needed for above-chance-level optic flow parsing (Kozhemiako et al., 2020). Therefore, to dissociate the direction of the target object from motion of the rest of the objects based on MEG data, we used the entire 0–1000 ms motion period from trial-specific ROI time courses in our linear two-class support vector machine (SVM) classifier analyses.

To increase the signal-to-noise ratio, in each participant, we first randomly sampled without replacement and sub-averaged 19 subsamples of 4 trials per class (amounting to 76 trials, i.e., one less than the smallest number of accepted trials across all participants/conditions) to be entered in the linear SVM algorithm with a linear kernel and a fixed regularization parameter $c = 1$. In each of our 19 cross-validation cycles, 18 of the two-class data subsamples were used for training the SVM classifier to predict the condition labels of the remaining two-class data subsample that was the test data. The training and testing procedures were repeated 1000 times with random assignment of individual trials to the sub-averaged feature vectors, to yield the average decoding accuracy of target motion direction from each of the five ROIs, separately for Visual Search vs. Visual-Auditory Search conditions. The MVPA analyses of the control conditions were similar, except that in Visual Control the classification was done on the direction of the observer motion and in Auditory Control based on the direction of the auditory cue.

We also conducted a control analysis to verify that auditory modulations of visual motion processing in the area hMT+ cannot be explained by the MEG source estimate crosstalk from auditory areas or from the known multisensory integration areas in pSTS. To this end, we regressed out the source activations from the bilateral pSTG and pSTS from the left and right hMT+ time courses, in a trial-by-trial fashion, before conducting the MVPA analysis. To deal with potential multicollinearity, we regressed out the contribution of pSTG from the pSTS single-trial time courses before the full regression analysis.

In each MVPA analysis, we used one-sample t-tests to determine whether the SVM classification accuracy was significantly higher than the chance level of 50%. In addition, we compared the decoding accuracies of Visual-Auditory Search vs. Visual Search conditions in the visual cortex ROIs using paired t-tests. To manage the multiple comparisons, the statistical significance was determined using the FDR procedure of (Benjamini and Hochberg, 1995) applied on the vector of p-values of all MVPA-related tests.

CRediT authorship contribution statement

Lucia M. Vaina: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Supervision, Writing - original draft. **Finnegan J. Calabro:** Conceptualization, Formal analysis, Writing - review & editing. **Abhisek Samal:** Formal analysis, Writing - review & editing. **Kunjan D. Rana:** Conceptualization, Data curation, Formal analysis, Investigation, Writing - review & editing. **Fahimeh Mamashli:** Formal analysis, Writing - review & editing. **Sheraz Khan:** Formal analysis, Writing - review & editing. **Matti Hämäläinen:** Conceptualization, Formal analysis, Funding acquisition, Software, Writing - review & editing. **Seppo P. Ahlfors:** Formal analysis, Funding acquisition, Writing - review & editing. **Jyrki Ahveninen:** Formal analysis, Funding acquisition, Writing - original draft.

Acknowledgements

This work was supported by the National Science Foundation grant 1545668 (LMV), and by the National Institutes of Health grants R01DC016765, R01DC016915, R01DC017991, and P41EB015896. The authors declare that there are no conflicts of interest regarding the publication of this article. We thank Nao Matsuda and Dr. Kaisu Lankinen for their support and advice.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.brainres.2021.147489.

References

Ahveninen, J., Kopco, N., Jääskeläinen, I.P., 2014. Psychophysics and neuronal bases of sound localization in humans. Hear Res. 307, 86–97.
Ahveninen, J., Ingalls, G., Yildirim, F., Calabro, F.J., Vaina, L.M., 2019. Peripheral visual localization is degraded by globally incongruent auditory-spatial attention cues. Exp Brain Res. 237, 2137–2143.
Alais, D., Burr, D., 2004. No direction-specific bimodal facilitation for audiovisual motion detection. Brain Res Cogn Brain Res. 19, 185–194.
Alink, A., Singer, W., Muckli, L., 2008. Capture of Auditory Motion by Vision Is Represented by an Activation Shift from Auditory to Visual Motion Cortex. The Journal of Neuroscience. 28, 2690–2697.

Allman, J.M., Kaas, J.H., 1971. A representation of the visual field in the caudal third of the middle tempral gyrus of the owl monkey (Aotus trivirgatus). Brain Res. 31, 85–105.
Baillet, S., 2017. Magnetoencephalography for brain electrophysiology and imaging. Nat Neurosci. 20, 327–339.
Barlow, H.B., 1972. Single units and sensation: a neuron doctrine for perceptual psychology? Perception. 1, 371–394.
Benjamini, Y., Hochberg, Y., 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. 57, 289–300.
Bertelson, P., Radeau, M., 1981. Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. Attention, Perception, & Psychophysics. 29, 578–584.
Burr, D., Thompson, P., 2011. Motion psychophysics: 1985–2010. Vision Res. 51, 1431–1456.
Calabro, F.J., Soto-Faraco, S., Vaina, L.M., 2011. Acoustic facilitation of object movement detection during self-motion. Proc Biol Sci. 278, 2840–2847.
Calabro, F.J., Vaina, L.M., 2012. Interaction of cortical networks mediating object motion detection by moving observers. Exp Brain Res. 221, 177–189.
Cappe, C., Thut, G., Romei, V., Murray, M.M., 2009. Selective integration of auditory-visual looming cues by humans. Neuropsychologia. 47, 1045–1052.
Chien, S.E., Ono, F., Watanabe, K., 2013. A transient auditory signal shifts the perceived offset position of a moving visual object. Front Psychol. 4, 70.
DeAngelis, G.C., Angelaki, D.E., 2012. Visual–Vestibular Integration for Self-Motion Perception. In The Neural Bases of Multsensory Processes. Vol., M.M. Murray, M.T. Wallace, ed.ˆeds. CRC Press/Taylor & Francis, Boca Raton, FL.
Desikan, R., Segonne, F., Fischl, B., Quinn, B., Dickerson, B., Blacker, D., Buckner, R., Dale, A., Maguire, R., Hyman, B., Albert, M., Killiany, R., 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage. 31, 968–980.
Dick, M., Ullman, S., Sagi, D., 1987. Parallel and serial processes in motion detection. Science. 237, 400–402.
Dubner, R., Zeki, S.M., 1971. Response properties and receptive fields of cells in an anatomically defined region of the superior temporal sulcus in the monkey. Brain Res. 35, 528–532.
Duffy, C.J., Wurtz, R.H., 1991. Sensitivity of MST neurons to optic flow stimuli. I. A continuum of response selectivity to large-field stimuli. J Neurophysiol. 65, 1329–1345.
Falchier, A., Clavagnier, S., Barone, P., Kennedy, H., 2002. Anatomical evidence of multimodal integration in primate striate cortex. J Neurosci. 22, 5749–5759.
Falchier, A., Schroeder, C.E., Hackett, T.A., Lakatos, P., Nascimento-Silva, S., Ulbert, I., Karmos, G., Smiley, J.F., 2010. Projection from visual areas V2 and prostriata to caudal auditory cortex in the monkey. Cereb Cortex. 20, 1529–1538.
Friston, K.J., Holmes, A.P., Worsley, K.J., 1999. How many subjects constitute a study? Neuroimage. 10, 1–5.
Ghazanfar, A.A., Neuhoff, J.G., Logothetis, N.K., 2002. Auditory looming perception in rhesus monkeys. Proc Natl Acad Sci U S A. 99, 15755–15757.
Giard, M.H., Peronnet, F., 1999. Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. J Cogn Neurosci. 11, 473–490.
Gibson, J.J., 1950. The perception of the visual world. Vol. Houghton Mifflin, Boston, MA.
Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., Parkkonen, L., Hämäläinen, M.S., 2014. MNE software for processing MEG and EEG data. Neuroimage. 86, 446–460.
Green, J.J., Teder-Salejarvi, W.A., McDonald, J.J., 2005. Control mechanisms mediating shifts of attention in auditory and visual space: a spatio-temporal ERP analysis. Exp Brain Res. 166, 358–369.
Hämäläinen, M., Hari, R., Ilmoniemi, R., Knuutila, J., Lounasmaa, O., 1993. Magnetoencephalography -theory, instrumentation, and applications to noninvasive studies of the working human brain. Rev Mod Phys. 65, 413–497.
Hämäläinen, M.S., Sarvas, J., 1989. Realistic conductivity geometry model of the human head for interpretation of neuromagnetic data. IEEE Trans Biomed Eng. 36, 165–171.
Hanada, G.M., Ahveninen, J., Calabro, F., Yengo-Kahn, A., Vaina, L.M., 2019. Cross-modal cue effects in motion processing. Multisens Res. 32, 45–65.
Hillyard, S.A., Stormer, V.S., Feng, W., Martinez, A., McDonald, J.J., 2016. Cross-modal orienting of visual attention. Neuropsychologia. 83, 170–178.
Jack, C.E., Thurlow, W.R., 1973. Effects of degree of visual association and angle of displacement on the "ventriloquism" effect. Percept Mot Skills. 37, 967–979.
Kayser, S.J., Philiastides, M.G., Kayser, C., 2017. Sounds facilitate visual motion discrimination via the enhancement of late occipital visual representations. Neuroimage. 148, 31–41.
Kayser, S.J., Kayser, C., 2018. Trial by trial dependencies in multisensory perception and their correlates in dynamic brain activity. Sci Rep. 8, 3742.
Konorski, J., 1967. Integrative activity of the brain; an interdisciplinary approach., Vol. University of Chicago Press, Chicago.
Kopco, N., Lin, I.F., Shinn-Cunningham, B.G., Groh, J.M., 2009. Reference frame of the ventriloquism aftereffect. J Neurosci. 29, 13809–13814.
Kopco, N., Huang, S., Belliveau, J.W., Raij, T., Tengshe, C., Ahveninen, J., 2012. Neuronal representations of distance in human auditory cortex. Proc Natl Acad Sci U S A. 109, 11019–11024.
Kopco, N., Doreswamy, K.K., Huang, S., Rossi, S., Ahveninen, J., 2019. Cortical auditory distance representation based on direct-to-reverberant energy ratio. Neuroimage. 208, 116436.
Kozhemiako, N., Nunes, A.S., Samal, A., Rana, K.D., Calabro, F.J., Hamalainen, M.S., Khan, S., Vaina, L.M., 2020. Neural activity underlying the detection of an object movement by an observer during forward self-motion: Dynamic decoding and temporal evolution of directional cortical connectivity. Prog Neurobiol. 101824.
Layton, O.W., Fajen, B.R., 2016. A Neural Model of MST and MT Explains Perceived Object Motion during Self-Motion. J Neurosci. 36, 8093–8102.
Lee, A.L., Lu, H., 2010. A comparison of global motion perception using a multiple-aperture stimulus. J Vis. 10 (9), 1–16.
Lewis, R., Noppeney, U., 2010. Audiovisual synchrony improves motion discrimination via enhanced connectivity between early visual and auditory areas. J Neurosci. 30, 12329–12339.
Lin, F.H., Belliveau, J.W., Dale, A.M., Hämäläinen, M.S., 2006. Distributed current estimates using cortical orientation constraints. Hum Brain Mapp. 27, 1–13.
Majka, P., Rosa, M.G.P., Bai, S., Chan, J.M., Huo, B.X., Jermakow, N., Lin, M.K., Takahashi, Y.S., Wolkowicz, I.H., Worthy, K.H., Rajan, R., Reser, D.H., Wojcik, D.K., Okano, H., Mitra, P.P., 2019. Unidirectional monosynaptic connections from auditory areas to the primary visual cortex in the marmoset monkey. Brain Struct Funct. 224, 111–131.
Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. J Neurosci Methods. 164, 177–190.
McDonald, J.J., Teder-Salejarvi, W.A., Hillyard, S.A., 2000. Involuntary orienting to sound improves visual perception. Nature. 407, 906–908.
McLeod, P., Driver, J., Crisp, J., 1988. Visual search for a conjunction of movement and form is parallel. Nature. 332, 154–155.
Meyer, G.F., Wuerger, S.M., 2001. Cross-modal integration of auditory and visual motion signals. Neuroreport. 12, 2557–2560.
Murray, M.M., Thelen, A., Thut, G., Romei, V., Martuzzi, R., Matusz, P.J., 2016. The multisensory function of the human primary visual cortex. Neuropsychologia. 83, 161–169.
Nakayama, K., Silverman, G.H., 1986. Serial and parallel processing of visual feature conjunctions. Nature. 320, 264–265.
Niehorster, D.C., Li, L., 2017. Accuracy and Tuning of Flow Parsing for Visual Perception of Object Motion During Self-Motion. Iperception. 8, 2041669517708206.
Palmer, S.M., Rosa, M.G., 2006. A distinct anatomical network of cortical areas for analysis of motion in far peripheral vision. Eur J Neurosci. 24, 2389–2405.
Pitzalis, S., Serra, C., Sulpizio, V., Committeri, G., de Pasquale, F., Fattori, P., Galletti, C., Sepe, R., Galati, G., 2020. Neural bases of self- and object-motion in a naturalistic vision. Hum Brain Mapp. 41, 1084–1111.
Poirier, C., Collignon, O., Devolder, A.G., Renier, L., Vanlierde, A., Tranduy, D., Scheiber, C., 2005. Specific activation of the V5 brain area by auditory motion processing: an fMRI study. Brain Res Cogn Brain Res. 25, 650–658.
Poirier, C., Collignon, O., Scheiber, C., Renier, L., Vanlierde, A., Tranduy, D., Veraart, C., De Volder, A.G., 2006. Auditory motion perception activates visual motion areas in early blind subjects. Neuroimage. 31, 279–285.
Raij, T., Ahveninen, J., Lin, F.H., Witzel, T., Jääskeläinen, I.P., Letham, B., Israeli, E., Sahyoun, C., Vasios, C., Stufflebeam, S., Hämäläinen, M., Belliveau, J.W., 2010. Onset timing of cross-sensory activations and multisensory interactions in auditory and visual sensory cortices. Eur J Neurosci. 31, 1772–1782.
Rana, K.D., Vaina, L.M., 2014. Functional roles of 10 Hz alpha-band power modulating engagement and disengagement of cortical networks in a complex visual motion task. PLoS One. 9, e107715.
Rezk, M., Cattoir, S., Battal, C., Occelli, V., Mattioni, S., Collignon, O., 2020. Shared Representation of Visual and Auditory Motion Directions in the Human Middle-Temporal Cortex. Curr Biol. 30 (2289–2299), e8.
Rockland, K.S., Ojima, H., 2003. Multisensory convergence in calcarine visual areas in macaque monkey. Int J Psychophysiol. 50, 19–26.
Roudaia, E., Calabro, F.J., Vaina, L.M., Newell, F.N., 2018. Aging Impairs Audiovisual Facilitation of Object Motion Within Self-Motion. Multisensory Research. 31, 251–272.
Rushton, S.K., Warren, P.A., 2005. Moving observers, relative retinal motion and the detection of object movement. Curr Biol. 15, R542–R543.
Rushton, S.K., Niehorster, D.C., Warren, P.A., Li, L., 2018. The Primary Role of Flow Processing in the Identification of Scene-Relative Object Movement. J Neurosci. 38, 1737–1743.
Schmiedchen, K., Freigang, C., Nitsche, I., Rubsamen, R., 2012. Crossmodal interactions and multisensory integration in the perception of audio-visual motion – a free-field study. Brain Res. 1466, 99–111.
Smiley, J.F., Hackett, T.A., Ulbert, I., Karmas, G., Lakatos, P., Javitt, D.C., Schroeder, C. E., 2007. Multisensory convergence in auditory cortex, I. Cortical connections of the caudal superior temporal plane in macaque monkeys. J Comp Neurol. 502, 894–923.
Soto-Faraco, S., Kingstone, A., Spence, C., 2003. Multisensory contributions to the perception of motion. Neuropsychologia. 41, 1847–1862.
Soto-Faraco, S., Spence, C., Kingstone, A., 2004. Cross-modal dynamic capture: congruency effects in the perception of motion across sensory modalities. J Exp Psychol Hum Percept Perform. 30, 330–345.
Stein, B.E., Stanford, T.R., 2008. Multisensory integration: current issues from the perspective of the single neuron. Nat Rev Neurosci. 9, 255–266.
Strnad, L., Peelen, M.V., Bedny, M., Caramazza, A., 2013. Multivoxel pattern analysis reveals auditory motion information in MT+ of both congenitally blind and sighted individuals. PLoS One. 8, e63198.
Teramoto, W., Hidaka, S., Sugita, Y., Sakamoto, S., Gyoba, J., Iwaya, Y., Suzuki, Y., 2012. Sounds can alter the perceived direction of a moving visual object. J Vis. 12.
Tyll, S., Bonath, B., Schoenfeld, M.A., Heinze, H.J., Ohl, F.W., Noesselt, T., 2013. Neural basis of multisensory looming signals. Neuroimage. 65, 13–22.
Vaina, L.M., 1998. Complex motion perception and its deficits. Curr Opin Neurobiol. 8, 494–502.

Van der Burg, E., Olivers, C.N., Bronkhorst, A.W., Theeuwes, J., 2008. Pip and pop: nonspatial auditory signals improve spatial visual search. J Exp Psychol Hum Percept Perform. 34, 1053–1065.
Van der Burg, E., Talsma, D., Olivers, C.N., Hickey, C., Theeuwes, J., 2011. Early multisensory interactions affect the competition among multiple visual objects. Neuroimage. 55, 1208–1218.
von Saldern, S., Noppeney, U., 2013. Sensory and striatal areas integrate auditory and visual signals into behavioral benefits during motion discrimination. J Neurosci. 33, 8841–8849.
Warren, P.A., Rushton, S.K., 2008. Evidence for flow-parsing in radial flow displays. Vision Res. 48, 655–663.
Warren, P.A., Rushton, S.K., 2009. Optic flow processing for the assessment of object movement during ego movement. Curr. Biol. 19, 1555–1560.
Wuerger, S.M., Parkes, L., Lewis, P.A., Crocker-Buque, A., Rutschmann, R., Meyer, G.F., 2012. Premotor cortex is sensitive to auditory-visual congruence for biological motion. J. Cogn. Neurosci. 24, 575–587.