

Current Biology

Shared Representation of Visual and Auditory Motion Directions in the Human Middle-Temporal Cortex

Highlights

- Cross-modal decoding of visual and auditory motion directions in hMT⁺/V5
- Motion-direction representation is, however, not abstracted from the sensory input
- We reveal a multifaced representation of multisensory motion signals in hMT⁺/V5

Authors

Mohamed Rezk, Stephanie Cattoir, Ceren Battal, Valeria Occelli, Stefania Mattioni, Olivier Collignon

Correspondence

mohamed.rezk@uclouvain.be (M.R.),
olivier.collignon@uclouvain.be (O.C.)

In Brief

Rezk et al. find that hMT⁺/V5, classically considered uniquely visual, also contains information about auditory motion direction in a format that is partially aligned across the senses. The representation of motion directions is, however, not abstracted from the sensory input, thus revealing the multifaced multisensory nature of hMT⁺/V5.



Article

Shared Representation of Visual and Auditory Motion Directions in the Human Middle-Temporal Cortex

Mohamed Rezk,^{1,3,*} Stephanie Cattoir,² Ceren Battal,¹ Valeria Occelli,² Stefania Mattioni,¹ and Olivier Collignon^{1,2,*}¹Institut de Recherche en Sciences Psychologiques (IPSY) and Institute of Neuroscience (IoNS), Louvain Bionics, Crossmodal Perception and Plasticity Lab, Université catholique de Louvain (UCL), 1348 Louvain-la-Neuve, Belgium²Center for Mind/Brain Sciences (CIMEC), University of Trento, 38068 Rovereto, Italy³Lead Contact*Correspondence: mohamed.rezk@uclouvain.be (M.R.), olivier.collignon@uclouvain.be (O.C.)<https://doi.org/10.1016/j.cub.2020.04.039>

SUMMARY

The human occipito-temporal region hMT⁺/V5 is well known for processing visual motion direction. Here, we demonstrate that hMT⁺/V5 also represents the direction of auditory motion in a format partially aligned with the one used to code visual motion. We show that auditory and visual motion directions can be reliably decoded in individually localized hMT⁺/V5 and that motion directions in one modality can be predicted from the activity patterns elicited by the other modality. Despite shared motion-direction information across the senses, vision and audition, however, overall produce opposite voxel-wise responses in hMT⁺/V5. Our results reveal a multifaceted representation of multisensory motion signals in hMT⁺/V5 and have broader implications for our understanding of how we consider the division of sensory labor between brain regions dedicated to a specific perceptual function.

INTRODUCTION

In primates, the visual cortex evolved a region in the middle-temporal (MT) cortex that is highly specialized to process visual motion [1–3]. One hallmark organizational feature of this region, first disclosed using electrophysiological recording in animals, is the presence of cortical columns that are preferentially tuned to a specific direction/axis of visual motion [4]. Motion preference [5] and directional selectivity [6, 7] were also probed using functional magnetic resonance imaging (fMRI) in the human homolog of the MT cortex (hereafter hMT⁺/V5). These studies promoted hMT⁺/V5 as one cornerstone region in the visual system showing reliable functional preference for processing visual motion direction.

Is hMT⁺/V5 uniquely visual? It was recently suggested that hMT⁺/V5 might also be involved in processing motion information from non-visual modalities [8–13]. If true, this has major implication on how we classically consider the division of sensory labor between brain regions dedicated to a specific perceptual or cognitive function [14, 15]. This would suggest that regions typically considered as purely unisensory may in fact show some level of abstraction by implementing canonical perceptual computations across the senses [16–18]. However, the idea that hMT⁺/V5 involves in motion processing other than for vision remains controversial and has notably been challenged by studies suggesting that the overlap between brain activity elicited by visual and non-visual motion is an artifactual byproduct of smoothing fMRI activity maps or by using group-level rather than individually defined hMT⁺/V5 [8, 10–12, 19–23].

Unlike univariate techniques, multivariate pattern analyses (MVPA) were able to dissociate information about visual motion direction in hMT⁺/V5 [24, 25]. Interestingly, planes of auditory motion were successfully decoded in hMT⁺/V5 [8, 26]. Although these results suggest that auditory motion information may indeed be represented in hMT⁺/V5, whether the representation of motion direction is similar across sensory modalities remains unsolved. In other words, the idea that hMT⁺/V5 can share an abstracted representation of directional motion across the senses has never been tested. In this context, MVPA and, in particular, cross-modal classification methods represent an appealing application of machine learning to operationalize the presence of abstracted brain representations [27]. The goal of this study was to use fMRI to characterize the representation of visual and auditory motion in individually defined hMT⁺/V5.

RESULTS

Experiment 1: Motion Localizers

We used separate visual and auditory motion localizers to functionally and individually define regions that are preferentially recruited during visual and auditory motion processing, and their connectivity with other brain regions. Last, we used MVPA to decode the motion and static conditions in each localizer.

Visual and Auditory Motion Localizers

Description of the procedures used to define auditory (planum temporale [PT] and MTa) and visual (hMT⁺/V5) motion-selective regions individually is detailed in the STAR Methods.



Beta Parameter Estimates in Auditory and Visual Motion Selective Regions

Beta parameter estimates were extracted from 6 regions of interest (ROIs) (bilateral hMT⁺/V5, MTa, and PT) for the motion and static conditions in the visual and the auditory motion localizers. Due to the variability in the anatomical locations of the peak coordinates of hMT⁺/V5 [28, 29] and MTa across participants, and the possible illusory overlap between the visual and auditory responsive regions arising from smoothing the functional data or using group-level averaging [20], we extracted the beta parameter estimates from the individually defined ROIs using unsmoothed data. The beta parameter extraction was performed in both group-level and individually defined ROIs (Figure 1E). The beta parameter estimates from the visually defined ROIs in the visual localizer and the auditory-defined ROIs in the auditory localizer are for illustration purposes only. No statistical analyses were performed to avoid circularity as these ROIs were defined from contrasting the motion condition versus the static condition in their respective localizers.

In hMT⁺/V5, we observed a significant difference between the auditory motion condition and the auditory static condition in the group-level-defined right hMT⁺/V5 ($p = 0.0071$) but not in the group-level-defined left hMT⁺/V5 ($p = 0.1536$) nor in the individually defined left ($p = 0.1819$) or right hMT⁺/V5 ($p = 1$). For the auditory-defined regions (MTa and PT) during visual motion processing in the visual motion localizer, we found enhanced recruitment during the visual motion condition in the individually defined left ($p = 0.0148$) and right MTa ($p = 0.0052$) and in the group-level right MTa ($p = 0.0074$). In both individually defined and group-level PT, there was no significant difference between the visual motion condition and the visual static condition ($p > 0.05$).

Functional Connectivity Analysis

We ran psychophysiological interactions (PPI) analyses [31] to investigate the connectivity profile of hMT⁺/V5 and MTa during processing auditory and visual motion. In the visual motion localizer, visually defined hMT⁺/V5 showed enhanced task-based connectivity with the bilateral middle occipito-temporal areas, the lingual gyrus, and the primary visual areas. Auditory-defined left MTa showed increased connectivity with the cuneus and the lingual gyrus during visual motion processing (Figure 2A). In contrast, for the auditory motion localizer, hMT⁺/V5 did not show any connectivity with the primary visual cortex during motion processing (when compared to static). Similarly, MTa did not show enhanced connectivity with earlier visual areas; however, it showed enhanced task-based connectivity with the superior temporal gyrus and PT (Figure 2B; Table S3).

Motion versus Static Decoding

The univariate analysis revealed a non-significant trend for preferential auditory motion processing, when compared to static, in individually defined hMT⁺/V5. The lack of significant voxel-wise univariate activation does not constitute evidence for the absence of auditory motion information in hMT⁺/V5. We relied on the higher sensitivity of MVPA to try to classify the motion and the static blocks in both the group-level and the individually defined hMT⁺/V5, MTa, and PT in each motion localizer independently (see STAR Methods for analysis details). No statistical

analysis was performed for the (visual motion versus visual static) decoding in hMT⁺/V5 to avoid circularity as this region was defined from the univariate analysis by contrasting visual motion versus visual static. Similarly, the (auditory motion versus auditory static) decoding was not tested in the auditorily defined ROIs (MTa and PT), and the statistical analysis was limited to hMT⁺/V5. The (auditory motion versus auditory static) decoding was significantly above chance (50%) in the group-level and in the individually defined bilateral hMT⁺/V5 ($p < 0.0001$). The (visual motion versus visual static) decoding was significant in the group-level and in the individually defined bilateral MTa ($p < 0.0001$) and not significant in PT (Figure 3).

EXPERIMENT 2: DIRECTIONAL MOTION DECODING

The participants discriminated the direction of auditory and visual motion with high accuracy inside the scanner (see the STAR Methods and Supplemental Information for the behavioral analysis).

Within-Modality Decoding

We tested whether the classification of the four motion directions in the individually defined ROIs (hMT⁺/V5, MTa, and PT) was above-chance level (25%; Figure 4D; Table S4). Non-parametric significance testing (false discovery rate [FDR]-corrected) revealed that, in vision, motion-direction decoding was above chance in the left ($p = 0.0041$) and the right hMT⁺/V5 ($p = 0.0048$) and in the left MTa ($p = 0.0401$). In the auditory modality, above-chance classification was observed in right hMT⁺/V5 ($p = 0.0086$), left MTa ($p = 0.0002$), right MTa ($p = 0.0004$), left PT ($p = 0.00006$), and right PT ($p = 0.00006$).

Interestingly, within-axis auditory decoding was less reliable than across axes auditory decoding in hMT⁺/V5 (Figures S1 and 4D), which could be in line with the previously suggested large-scale axis-of-motion organization in hMT⁺/V5 [6]. Similarly, multi-dimensional scaling (MDS) [32] showed that opposite directions within the same axis were clustered closer to each other when compared to motion directions in the different axis (Figure 4F).

Cross-Modal Decoding

To test whether there was shared motion-direction information across modalities in regions that contained both visual and auditory motion-direction information, we performed a cross-modal decoding analysis where the classifier was trained on one modality and tested on the other modality (training in vision and testing in audition, and vice versa; FDR-corrected). We found significant visual and auditory motion-direction information in right hMT⁺/V5 and left MTa; therefore, these two regions were considered candidates to potentially contain shared representation of motion direction across modalities. Successful cross-modal decoding was observed in the right hMT⁺/V5 ($p = 0.0052$) but not in left MTa ($p = 0.5432$) (Figure 4E). In the binary within-modality axes decoding (vertical versus horizontal), visual and auditory motion information was found in hMT⁺/V5 bilaterally. When performing cross-modal decoding (Figure 4E), shared axis-of-motion-direction information was again found in the right hMT⁺/V5 ($p = 0.0029$) and not in the left hMT⁺/V5 ($p = 0.0966$). In addition to our ROI approach, we implemented a whole-brain

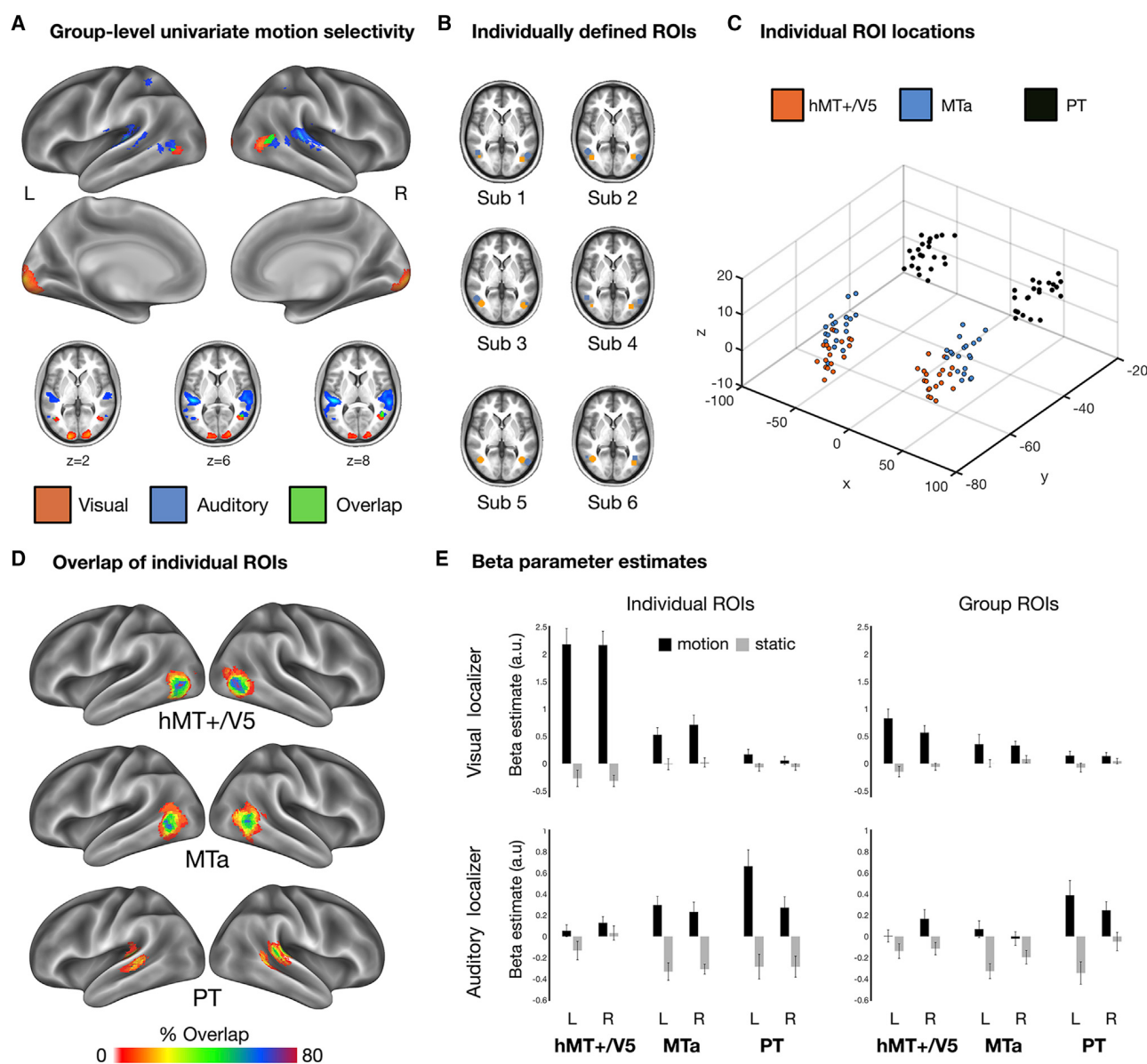


Figure 1. Univariate Results from the Visual and the Auditory Motion Localizers

(A) The univariate results of the contrast (visual motion > visual static) from the visual motion localizer (in warm colors) to functionally define hMT⁺/V5 and the contrast (auditory motion > auditory static) in the auditory motion localizer (in cold colors) to functionally define planum temporale (PT) and the auditory motion selective regions in the middle temporal cortex (MTa). The overlap between MTa and hMT⁺/V5 is highlighted in green. Results are FWE corrected and are displayed on an inflated brain [30].

(B) Examples of the individually defined ROIs in hMT⁺/V5 and MTa from 6 participants.

(C) A 3D representation of the Montreal Neurological Institute (MNI) coordinates of the individually defined hMT⁺/V5 (orange), MTa (blue), and PT (black).

(D) The overlap of each individually defined ROI across participants.

(E) Beta parameter estimates extracted from the group-level and individually defined hMT⁺/V5, MTa, and PT in the visual and the auditory motion localizer. No statistical analyses were performed for the beta estimates from hMT⁺/V5 in the visual motion localizer, nor from MTa and PT in the auditory motion localizer (for illustration purposes only). Error bars represent the standard error of the mean. See Tables S1 and S2 for more details.

cross-modal searchlight analysis that revealed a single cluster in the vicinity of the right hMT⁺/V5 ($p = 0.013$; corrected using a small volume correction [SVC]) where shared axis of motion information was present across modalities. This region overlapped with the posterior portion of the group-level visually defined hMT⁺/V5 from the visual motion localizer (Figure 4G).

Modality Decoding

Significant cross-modal decoding in right hMT⁺/V5 indicated the presence of shared motion-direction representation across modalities. It, however, did not mean that the activity patterns were identical across the sensory modalities. We reasoned that if a region is purely abstracted from the input modality conveying the

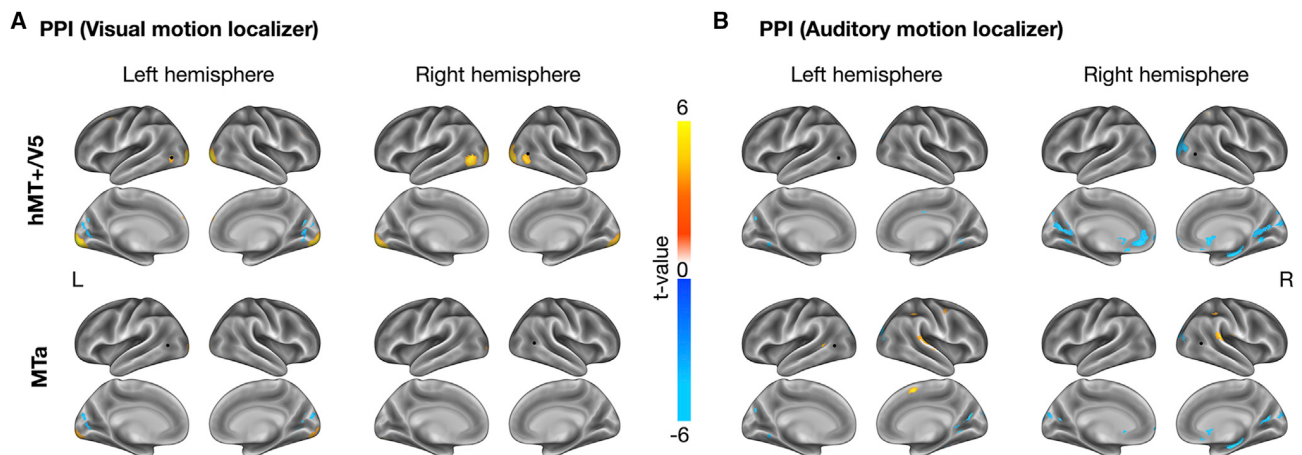


Figure 2. PPI Analysis

(A) PPI analysis in the visual motion localizer. Regions showing increased task-based functional connectivity with bilateral hMT⁺/V5 and MTa for the contrast (visual motion > visual static) in warm colors and for the contrast (visual static > visual motion) in cold colors.

(B) PPI analysis in the auditory motion localizer. Regions showing increased task-based functional connectivity with bilateral hMT⁺/V5 and MTa for the contrast (auditory motion > auditory static) in warm colors and for the contrast (auditory static > auditory motion) in cold colors. PPI results are displayed at $p < 0.001$, minimum cluster size of 20 voxels. See Table S3 for more details. L, left hemisphere; R, right hemisphere. See Table S3 for the PPI results.

directional motion information, decoding the modality (visual or auditory) of the presented motion stimuli would be difficult. The modality decoding was highly significant in all ROIs ($p < 0.0001$; see STAR Methods for analysis procedure), suggesting that the neural patterns associated with each direction are highly dissimilar across the senses (Figure 5A).

Representational Similarity Analysis

To further explore the differences in the representational format of auditory and visual motion directions in hMT⁺/V5, we tested the correlation between the neural dissimilarity matrix (DSM) of right hMT⁺/V5 in each subject with different theoretical models that included a pure modality-invariant model, a multimodal model, a series of intermediate gradients between them, a unimodal visual model, and a unimodal auditory model. The analysis revealed a negative correlation with the modality-invariant model (mean $r \pm SD = -0.3418 \pm 0.0209$) that increased gradually as the models progressed toward the multi-modal model. The model that correlated the most with our data was the multi-modal model (mean $r \pm SD = 0.9897 \pm 0.0194$), and this correlation was significantly higher than all other models (Figure 5B; Table S5). The gradual increase in correlation from the modality-invariant model to the multi-modal model added extra evidence that the representation of different motion directions in right hMT⁺/V5 was not abstracted from the input modality.

Pattern Correlations

To better understand the relation of the pattern geometries for the different motion directions across the senses, we performed a split-half correlation (SHC) analysis [33]. We observed that the motion-direction patterns were positively correlated in the odd and even runs in each individual modality (within-modality SHC) in each ROI (Figure 5C; Table S6). Yet, the cross-modal SHC showed a strong negative correlation in left ($p = 1.877 \times 10^{-5}$) and right hMT⁺/V5 ($p = 2.114 \times 10^{-6}$), no significant

correlation in bilateral MTa ($p > 0.05$), and a positive correlation in right PT ($p = 8.945 \times 10^{-4}$).

To compare the correlation values across the three different conditions (within audition, within vision, and across modalities) in each ROI, a repeated-measure ANOVA was implemented. Results showed that the correlation values were significantly different ($p < 0.001$) across the three conditions in each ROI. Post hoc pairwise comparisons revealed that in hMT⁺/V5, the visual SHC was not different from the auditory SHC. The cross-modal SHC was significantly lower than both the visual and the auditory within-modality SHC in both hemispheres and all ROIs (Table S6).

Differential Pattern Correlation

Despite the motion-direction patterns being negatively correlated across the two modalities in hMT⁺/V5, cross-modal decoding was successful in the right hMT⁺/V5 (see cross-modal decoding section). While this might appear surprising, we reasoned that a successful cross-modal decoding would be possible if the relative relation between the patterns of the different motion directions in vision is similar to that in audition. In other words, if the representation of the different motion directions in vision is similar to that in audition (despite the global negative correlation between vision and audition), we would be able to cross-decode motion directions across modalities. We tested this hypothesis by calculating the differential pattern from the mean for each motion direction in each modality independently. The activity patterns of the different motion directions were demeaned across directions in each modality separately, resulting in four differential patterns for the four motion directions in each modality. We then tested whether there was a significant correlation between the differential patterns of the four-motion direction across the two modalities (Figure 5E). The differential direction patterns were positively correlated ($r = 0.0827$, $p = 0.0047$, family-wise error [FWE]-corrected) only in the right

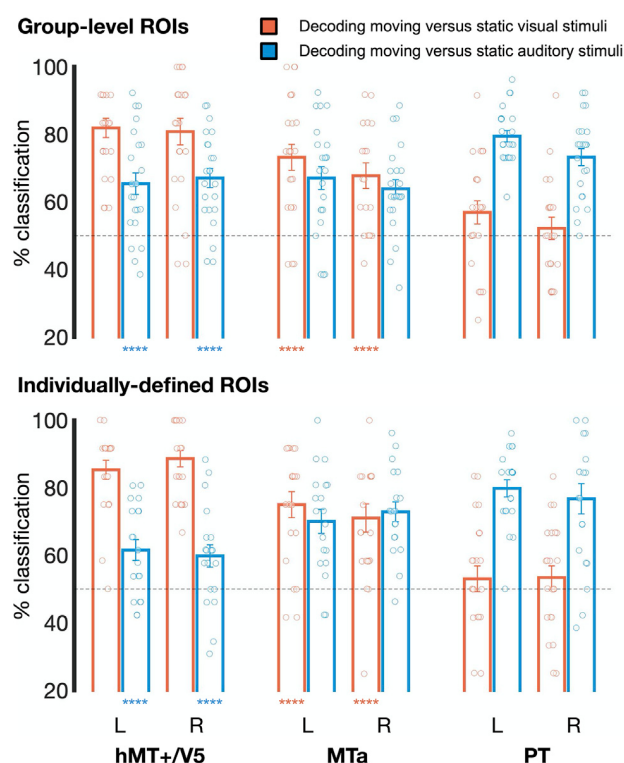


Figure 3. Motion versus Static Decoding

Decoding moving versus static stimuli from the visual and auditory motion localizers independently (chance level = 50%). Statistical analysis for decoding moving versus static sounds was implemented only in visually defined ROIs to avoid any circularity. Similarly, statistical analysis for decoding moving versus static visual dots was implemented only in auditorily defined ROIs. Individual points represent individual subjects' data (**** $p < 0.0001$). Error bars represent the standard error of the mean.

hMT⁺/V5. These results demonstrate that, while the motion-direction patterns are negatively correlated between vision and audition in the right hMT⁺/V5 at a global scale (across all directions), the relative differences between the directions in each modality are, however, similar across the senses. These findings explain the successful cross-modal decoding in the right hMT⁺/V5 despite the global negative correlation between visual and auditory patterns.

DISCUSSION

We demonstrate that a region that has long supported the idea of a “modular” visual system, hMT⁺/V5, is not impenetrable to information from the other senses and, in contrast, contain representation of auditory motion direction in a format that is partially aligned to the brain code this region uses to process the direction of visual motion. The presence of shared directional motion information across the senses in the right hMT⁺/V5 did not, however, indicate that motion processing is implemented using a fully abstracted response code since responses to visual and auditory motion were highly dissimilar at the global level. Our results therefore reveal a multifaceted representation of motion signals in hMT⁺/V5: while vision and

audition overall produce opposite voxel-wise responses in this region, the relationship between motion directions are, however, aligned across the senses.

We used separate visual and auditory motion localizers to define, in each subject individually, regions that show preferential recruitment for visual and auditory motion processing. As expected, hMT⁺/V5 showed prominent preferential response to visual motion [5, 34, 35], while PT showed prominent preferential response to auditory motion [36–40]. Interestingly, we also found preferential activity for auditory motion processing in the middle temporal cortex, in a region anterior to hMT⁺/V5, that we define here as MTa, as it was more anterior to hMT⁺/V5 and responded to auditory motion (Figure 1C). Earlier human studies already suggested that auditory and tactile motion selectively activate a region anterior to the visually defined hMT⁺/V5 cluster [10, 12, 20–22, 41]. Similarly, a recent study in the macaque demonstrated the presence of selective auditory motion responses in a region in the inferior bank of the superior temporal gyrus that was slightly anterior to the visual motion areas [38]. While we observed a degree of overlap between the visually and the auditory motion selective regions in the middle occipito-temporal cortex at the group level [10–12, 26], we did not observe a selective univariate preference for moving sounds in hMT⁺/V5 when avoiding smoothing or when using individual visual motion selective peaks to extract univariate beta parameters (Figure 1E), confirming previous univariate results [19–21]. However, this did not constitute evidence for the absence of auditory motion information in hMT⁺/V5 since MVPA revealed robust decoding between auditory motion and static conditions in visually defined hMT⁺/V5, highlighting that hMT⁺/V5 contained distributed information about the presence of motion information in sounds.

Using MVPA, we investigated the presence of motion-direction information across the senses in hMT⁺/V5. We demonstrate that auditory and visual motion directions could be reliably decoded in the right hMT⁺/V5. Moreover, we observed that decoding opposite directions within the same axis of motion (up versus down; left versus right) was less reliable than decoding across axes of motion (horizontal versus vertical) in both modalities. These results are reminiscent of animal studies showing that the MT cortex has an axis-of-motion columnar organization [4, 42–44] and high-field fMRI studies suggesting a large-scale axis-of-motion organization in hMT⁺/V5 in humans [6]. The multi-dimensional scaling outcome clearly demonstrate the axis-of-motion organization for both the visual and the auditory modalities, and the shared representation of motion directions across modalities (Figure 4F). Interestingly, a similar auditory “axis-of-motion” organization was recently disclosed in the human Planum Temporale, therefore suggesting that such representational structure for motion direction is shared across the senses and across brain regions involved in processing motion signals [40].

Our cross-modal decoding results not only demonstrate that hMT⁺/V5 implements an axis-of-motion representational structure in both vision and audition but further establish the presence of shared motion-direction information across the senses (Figure 4E). Previous neuroimaging studies also highlighted the interaction between the visual and the auditory modalities in hMT⁺/V5 during multisensory processing [45–48]. Our observation of shared motion-direction information between vision and

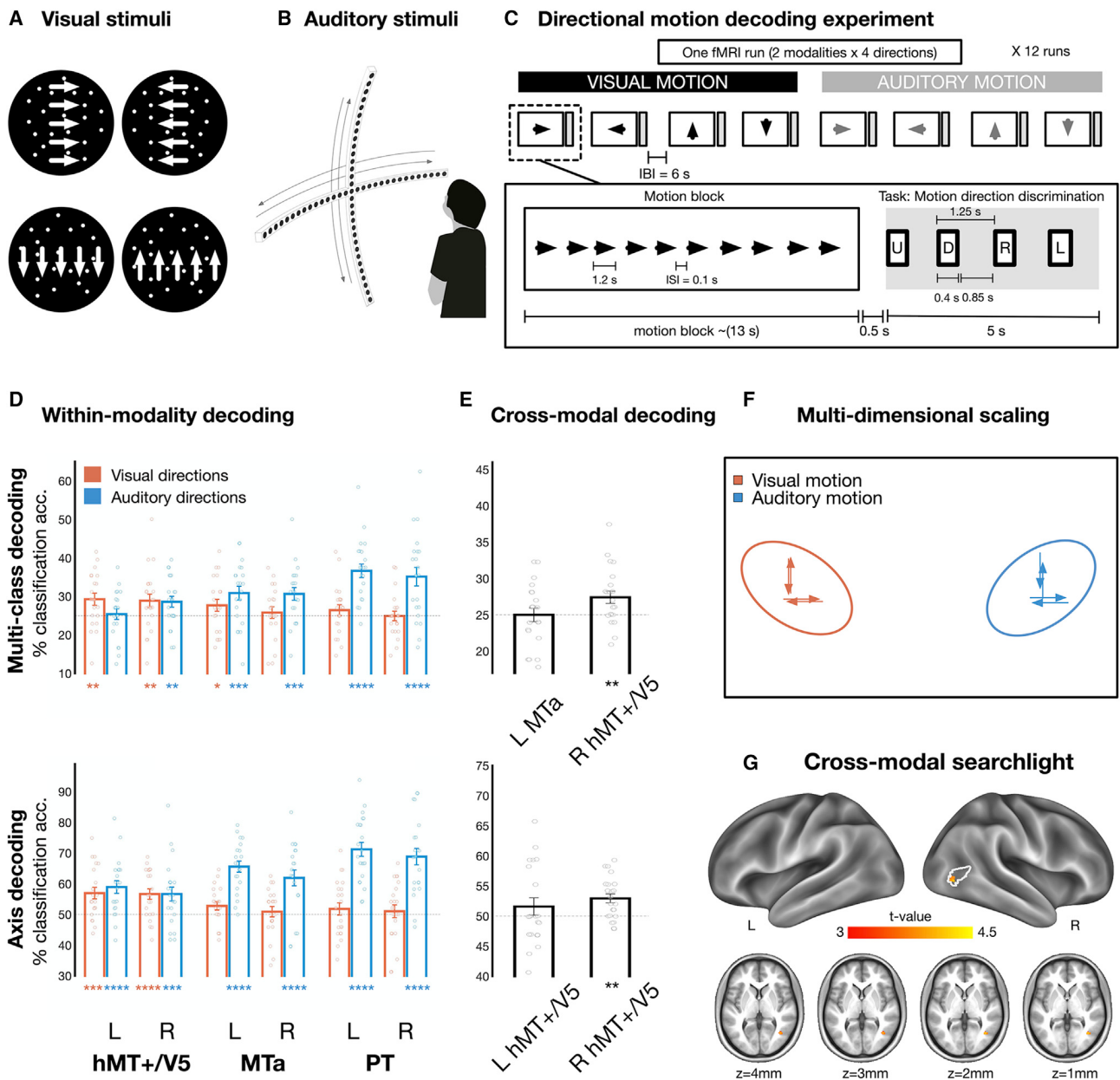


Figure 4. Directional Motion Decoding Experiment

(A and B) Exemplars from the (A) visual stimuli using random dot kinematograms and (B) the recorded auditory stimuli in 4 translational directions (upward, downward, rightward, and leftward motion).

(C) Experimental design of the directional motion decoding experiment (1 run). Each run had 8 motion blocks (2 modalities $\times$ 4 directions). The participants performed a motion-direction discrimination task following each block. The order of the blocks was randomized in each run.

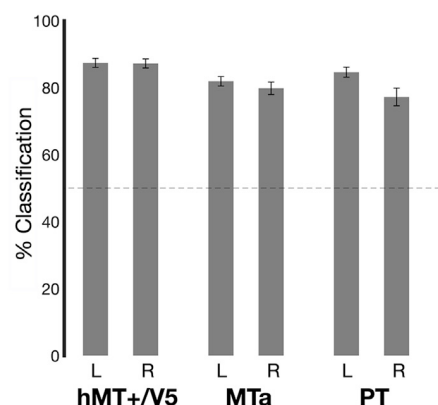
(D) Within modality multi-class classification of the 4 motion directions and the binary axis decoding (horizontal versus vertical) in the visual and auditory modality in individually defined hMT⁺/V5, MTa, and PT.

(E) Cross-modal decoding in regions that could successfully decode the motion directions (and axes) in both the auditory and the visual modality. Chance level (dotted line) is 25% for the multi-class direction decoding and 50% for the binary axis decoding. Error bars represent the standard error of the mean. Results are FDR corrected (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$). Individual points reflect individual subjects' results.

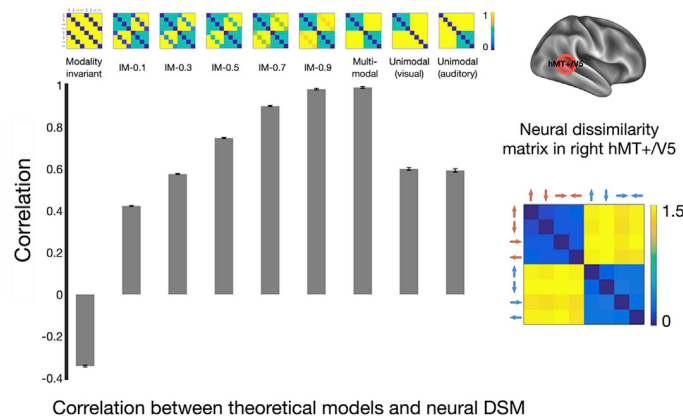
(F) Multi-dimensional scaling was used to visualize the distance between the patterns of the different motion directions in the visual and auditory modality using DISTATIS [32].

(G) Group-level searchlight analysis of the cross-modal axis of motion decoding. Only a region that overlapped with the right hMT⁺/V5 showed a successful cross-modal axis of motion decoding. L, left hemisphere; R, right hemisphere. See Figure S1 for the within-modality within-axis classification, Figure S2 for the control analysis for the cross-modal decoding, and Figure S3 for the control analyses for eye movements. See Table S4 for the within-modality decoding results.

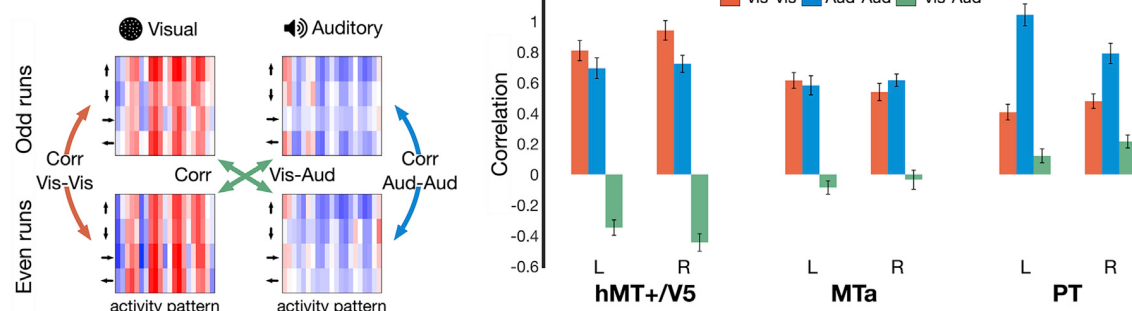
A Modality decoding



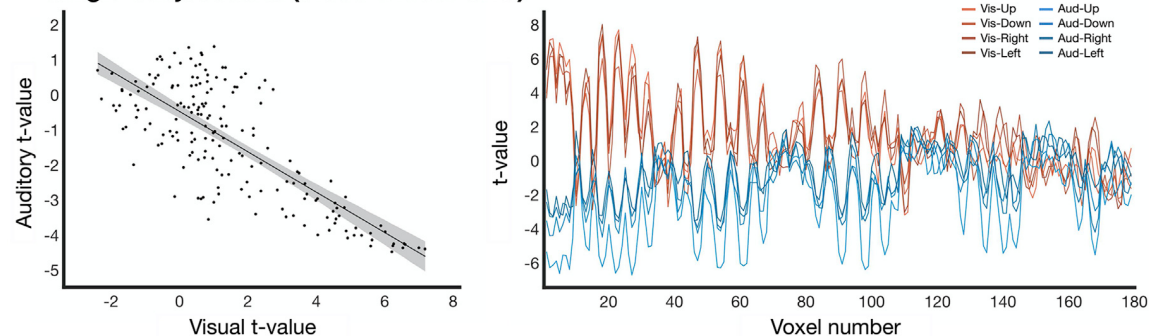
B Representational similarity analysis



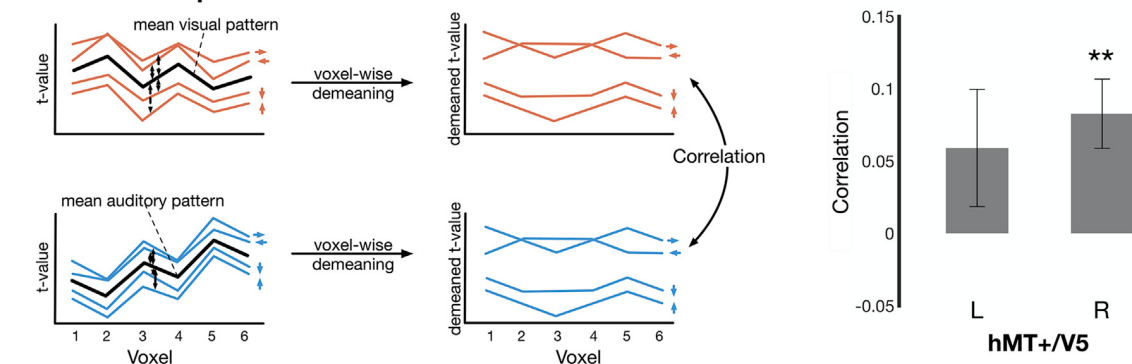
C Split-half correlation



D Single subject data (cross-modal SHC)



E Differential pattern correlation



(legend on next page)

audition in hMT⁺/V5 provide evidence that hMT⁺/V5 may play a crucial role in providing a common representational structure between the two modalities to link auditory and visual motion-direction information. The presence of a common brain code for directional motion in vision and audition might potentially relate to psychophysical studies showing cross-modal adaptation effects for motion directions [49–51].

The successful cross-modal decoding in right hMT⁺/V5 did not, however, indicate that the representation of motion direction was fully abstracted from sensory input. A series of complementary multivariate analyses demonstrated that the recorded activity in hMT⁺/V5 was highly dissimilar across the senses. For instance, motion-direction patterns from the visual modality were negatively correlated with those from the auditory modality (Figures 5C and 5D). The negative correlation between the two modalities suggest that there might exist different neural populations with a different level of preference for each modality in the voxels of hMT⁺/V5 and therefore voxels that have a higher density of visually responsive neurons have, de facto, a lower density of auditory responsive neurons. However, this hypothesis remains to be explored in future studies.

How does the presence of shared motion-direction information across the two modalities (as inferred by significant cross-modal decoding) relate to the fact that patterns observed in hMT⁺/V5 in vision and audition negatively correlate? This was explained by our findings that, while the motion-direction patterns were negatively correlated between vision and audition in right hMT⁺/V5 at a global scale (across all directions), the relation between the different motion directions in vision was similar to the one found in audition (see MDS, Figures 4F and 5E). These findings demonstrate that a successful cross-modal decoding indicates the presence of shared discriminating information between the different directions across modalities but does not, however, indicate that the region is immune to the sensory input modality. Our results therefore serve as a timely reminder that cross-modal decoding is not a proxy for the presence of fully abstracted representations in a brain region.

Interestingly, we observed that motion directions could only be decoded in the auditory modality but not with visual stimuli in PT. These results suggest that, whereas the right hMT⁺/V5

may contain information about auditory motion directions, the main auditory motion selective region (PT) does not reciprocally contain information about visual motion directions, at least in our dataset. Such a difference in the presence of cross-modal information in regions typically preferring visual or auditory motion signals may relate to the fact that vision typically plays a dominant role in discriminating motion information [49, 45, 52, 53]. It may therefore be speculated that the regions typically providing the most reliable sensory signal for discriminating motion directions (i.e., vision) are more likely to also represent cross-modal signals [54, 55].

It could be argued that the successful cross-modal decoding that we observed in right hMT⁺/V5 was attributed to the proximity to the MTa region. Our analyses demonstrate a clear anatomical and functional distinction between these two regions. The individually defined hMT⁺/V5 and MTa were non-overlapping in almost all participants (Figure S2A). Even when excluding the overlapping voxels between the right hMT⁺/V5 and the right MTa ROIs, we still replicated the significant cross-modal decoding results in the right hMT⁺/V5 (Figure S2B).

Could the decoded auditory motion-direction information in hMT⁺/V5 be related to eye movements? Unfortunately, we couldn't track the gaze location inside the scanner due to technical reasons. However, we tested additional participants outside the scanner and did not find any effect of the auditory motion direction on eye movements (Figures S3A and S3B). Eye movements can also be detected from the MRI signal from the eyes [56]. We tested whether the signal from the eyes varied across conditions and found no evidence for differences in eye movements profiles (Figure S3C). We are therefore convinced that the auditory moving information found in hMT⁺/V5 is unlikely to be explained by differences in eye movements.

Imagining visual motion can evoke a selective response in hMT⁺/V5 [57, 58], and the direction of visual motion can be decoded from visual imagery [59]. While the possibility of auditory-induced visual imagery cannot be fully excluded, we believe that it is unlikely. Previous studies demonstrating the recruitment of hMT⁺/V5 during visual imagery of motion needed to train the participants on performing a precise mental imagery task [58, 59], which was not the case in our study. In addition,

Figure 5. Multivariate Analyses from the Directional Motion Decoding Experiment

(A) Modality decoding results in hMT⁺/V5, MTa, and PT from the directional motion decoding experiment (chance level is 50%).

(B) Representational similarity analysis in the right hMT⁺/V5: the theoretical models ranged from a modality-invariant model to a multi-modal model, the intermediate gradients between them, and the two unimodal visual and auditory models. The neural DSM from the right hMT⁺/V5 showed a gradient increase in correlation with the theoretical models as the models progressed toward the multi-modal model and was higher than the unimodal models. The right panel shows the average DSM in the right hMT⁺/V5 across participants.

(C) Split-half correlation analysis (SHC) measured the correlation between the motion-direction patterns in the odd and the even runs in the directional motion decoding experiment. The SHC analysis was performed within modality (visual modality in orange, and auditory modality in blue) and across modalities (in green). The left-side panel is a graphical illustration of the SHC analysis. The right-side panel shows the results of the SHC in the individually defined ROIs.

(D) Single subject example: the visual motion patterns (averaged across directions) in the right hMT⁺/V5 of a representative subject shows a strong negative correlation with the auditory motion patterns. On the right panel, the plot shows the activity pattern of each motion direction in each modality in the right hMT⁺/V5 of the same subject. The single subject's raw data are for illustration purposes and are presented without demeaning. The figure illustrates the high correlation between the different directions within each modality, the negative correlation between the visual and the auditory motion patterns, and the stronger activation of hMT⁺/V5 for the visual motion stimuli than for the auditory motion stimuli.

(E) A graphical abstract illustrating the differential pattern correlation analysis. The analysis tested whether the relative differences between the different directions were similar across modalities. Note that the visual (in orange) and auditory motion patterns (in blue) are negatively correlated. The black line in each modality represents the mean pattern across the four directions. In each modality separately, the motion-direction patterns were demeaned across the different directions. The differential patterns from each modality were then correlated with each other. The right panel shows the results of the analysis and that the relative differences between the four motion-direction patterns were significantly correlated across modalities in the right hMT⁺/V5. Error bars represent the standard error of the mean. See Tables S5 and S6 for the results of the representational similarity analysis (RSA) and SHC analysis.

participants had to detect a change in color of the fixation cross during the auditory stimulation block in the directional motion decoding experiment, and performing this orthogonal visual task interferes with performing visual imagery [60, 61]. Maybe even more convincing is our functional connectivity analysis (Figure 2) revealing that, while hMT⁺/V5 showed enhanced connectivity with early visual regions during visual motion processing, this was not the case during auditory motion processing, suggesting that auditory motion did not re-instantiate the network-level activity triggered by visual motion [62]. Finally, the split-half correlation analysis demonstrated that the patterns linked to visual directions were, globally, negatively correlated with the patterns evoked by auditory motion (Figure 5C). Overall, these results support the notion that the information about auditory motion directions in hMT⁺/V5 was not a simple byproduct of visual motion imagery.

In conclusion, our study demonstrates that the right hMT⁺/V5, a region typically considered as purely sensitive to visual motion, also contains information about auditory motion directions. The representation of auditory motion directions in hMT⁺/V5 is aligned to the brain code this region uses to process the direction of visual motion. Such aligned representation between auditory and visual motion directions may allow efficient interaction between the senses and support fast and optimal representation of multisensory signals. However, the presence of shared motion-direction information did not indicate that hMT⁺/V5 was fully abstracted from the sensory input this region receives since the activity profile from the two modalities were highly distinguishable in this region. Altogether, our results have important implications for our understanding on how we classically consider the division of sensory labor between brain regions dedicated to a specific perceptual or cognitive function (here, motion). More precisely, our results suggest that regions classically considered exclusively visual may also contain information from the other senses in a format that is partially aligned to the brain code these regions use to process vision.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
 - Lead Contact
 - Materials Availability
 - Data and Code Availability
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
 - Participants
- **METHOD DETAILS**
 - Training
- **EXPERIMENTAL DESIGN**
 - Stimuli
 - Visual motion localizer
 - Auditory motion localizer
 - Directional motion decoding experiment
 - MRI acquisition
 - fMRI preprocessing
 - Eye tracker experiment

● QUANTIFICATION AND STATISTICAL ANALYSIS

- Behavioral Analysis
- Univariate analysis
- Beta parameter estimates
- Functional connectivity analysis (PPI)
- Multivariate pattern analysis (MVPA)
- Experiment 1 (visual and auditory motion localizer)
- Experiment 2 (Directional motion decoding experiment)
- Within-modality decoding
- Cross-modal decoding
- Representational similarity analysis (RSA)
- Split-half correlation analysis
- Differential pattern correlation

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.cub.2020.04.039>.

ACKNOWLEDGMENTS

We would like to express our gratitude to Roberto Bottini who helped with the data acquisition, Jorge Jovicich for helping setting up the fMRI acquisition parameters, Pietro Chiesa for continuing support with the auditory hardware, and Barbara K. Marebwa for her discussion and feedback on the manuscript. This work was supported by a European Research Council starting grant (MADVIS grant #337573) attributed to O.C. and by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 701250 awarded to V.O. M.R. is a PhD student and O.C. a research associate at the FRS-FNRS of Belgium.

AUTHOR CONTRIBUTIONS

M.R., S.C., C.B., S.M., and O.C. designed the research; S.C. and V.O. collected the data; S.C. and M.R. performed the research; M.R. and S.C. analyzed the data under the guidance of O.C.; M.R. and O.C. drafted the paper; all authors revised and edited the draft and agreed on the final version of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 3, 2019

Revised: March 3, 2020

Accepted: April 16, 2020

Published: May 21, 2020

REFERENCES

1. Baker, J.F., Petersen, S.E., Newsome, W.T., and Allman, J.M. (1981). Visual response properties of neurons in four extrastriate visual areas of the owl monkey (*Aotus trivirgatus*): a quantitative comparison of medial, dorsomedial, dorsolateral, and middle temporal areas. *J. Neurophysiol.* 45, 397–416.
2. Zeki, S.M. (1974). Functional organization of a visual area in the posterior bank of the superior temporal sulcus of the rhesus monkey. *J. Physiol.* 236, 549–573.
3. Zeki, S. (1980). The response properties of cells in the middle temporal area (area MT) of owl monkey visual cortex. *Proc. R. Soc. Lond. B Biol. Sci.* 207, 239–248.
4. Albright, T.D., Desimone, R., and Gross, C.G. (1984). Columnar organization of directionally selective cells in visual area MT of the macaque. *J. Neurophysiol.* 51, 16–31.

5. Zeki, S., Watson, J.D., Lueck, C.J., Friston, K.J., Kennard, C., and Frackowiak, R.S. (1991). A direct demonstration of functional specialization in human visual cortex. *J. Neurosci.* **11**, 641–649.
6. Zimmermann, J., Goebel, R., De Martino, F., van de Moortele, P.F., Feinberg, D., Adriani, G., Chaimow, D., Shmuel, A., Ugurbil, K., and Yacoub, E. (2011). Mapping the organization of axis of motion selective features in human area MT using high-field fMRI. *PLoS ONE* **6**, e28716.
7. Tootell, R.B.H., Reppas, J.B., Dale, A.M., Look, R.B., Sereno, M.I., Malach, R., Brady, T.J., and Rosen, B.R. (1995). Visual motion aftereffect in human cortical area MT revealed by functional magnetic resonance imaging. *Nature* **375**, 139–141.
8. Dormal, G., Rezk, M., Yakobov, E., Lepore, F., and Collignon, O. (2016). Auditory motion in the sighted and blind: Early visual deprivation triggers a large-scale imbalance between auditory and “visual” brain regions. *Neuroimage* **134**, 630–644.
9. Hagen, M.C., Franzén, O., McGlone, F., Essick, G., Dancer, C., and Pardo, J.V. (2002). Tactile motion activates the human middle temporal/V5 (MT/V5) complex. *Eur. J. Neurosci.* **16**, 957–964.
10. Poirier, C., Collignon, O., Devolder, A.G., Renier, L., Vanlierde, A., Tranduy, D., and 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.
11. Ricciardi, E., Vanello, N., Sani, L., Gentili, C., Scilingo, E.P., Landini, L., Guazzelli, M., Bicchi, A., Haxby, J.V., and Pietrini, P. (2007). The effect of visual experience on the development of functional architecture in hMT+. *Cereb. Cortex* **17**, 2933–2939.
12. van Kemenade, B.M., Seymour, K., Wacker, E., Spitzer, B., Blankenburg, F., and Sterzer, P. (2014). Tactile and visual motion direction processing in hMT+/V5. *Neuroimage* **84**, 420–427.
13. Wacker, E., Spitzer, B., Lützkendorf, R., Bernarding, J., and Blankenburg, F. (2011). Tactile motion and pattern processing assessed with high-field fMRI. *PLoS ONE* **6**, e24860.
14. Calvert, G.A., and Thesen, T. (2004). Multisensory integration: methodological approaches and emerging principles in the human brain. *J. Physiol. Paris* **98**, 191–205.
15. Schroeder, C.E., and Foxe, J. (2005). Multisensory contributions to low-level, ‘unisensory’ processing. *Curr. Opin. Neurobiol.* **15**, 454–458.
16. Driver, J., and Spence, C. (2000). Multisensory perception: beyond modularity and convergence. *Curr. Biol.* **10**, R731–R735.
17. Pascual-Leone, A., and Hamilton, R. (2001). The metamodal organization of the brain. *Prog. Brain Res.* **134**, 427–445.
18. Ricciardi, E., Handjaras, G., Bonino, D., Vecchi, T., Fadiga, L., and Pietrini, P. (2013). Beyond motor scheme: a supramodal distributed representation in the action-observation network. *PLoS ONE* **8**, e58632.
19. Jiang, F., Stecker, G.C., and Fine, I. (2014). Auditory motion processing after early blindness. *J. Vis.* **14**, 4.
20. Jiang, F., Beauchamp, M.S., and Fine, I. (2015). Re-examining overlap between tactile and visual motion responses within hMT+ and STS. *Neuroimage* **119**, 187–196.
21. Beauchamp, M.S., Yasar, N.E., Kishan, N., and Ro, T. (2007). Human MST but not MT responds to tactile stimulation. *J. Neurosci.* **27**, 8261–8267.
22. Saenz, M., Lewis, L.B., Huth, A.G., Fine, I., and Koch, C. (2008). Visual Motion Area MT+/V5 Responds to Auditory Motion in Human Sight-Recovery Subjects. *J. Neurosci.* **28**, 5141–5148.
23. Chaplin, T.A., Allitt, B.J., Hagan, M.A., Rosa, M.G.P., Rajan, R., and Lui, L.L. (2018). Auditory motion does not modulate spiking activity in the middle temporal and medial superior temporal visual areas. *Eur. J. Neurosci.* **48**, 2013–2029.
24. Beckett, A., Peirce, J.W., Sanchez-Panchuelo, R.M., Francis, S., and Schluppeck, D. (2012). Contribution of large scale biases in decoding of direction-of-motion from high-resolution fMRI data in human early visual cortex. *Neuroimage* **63**, 1623–1632.
25. Kamitani, Y., and Tong, F. (2006). Decoding seen and attended motion directions from activity in the human visual cortex. *Curr. Biol.* **16**, 1096–1102.
26. Battal, C., Rezk, M., Mattioni, S., Bottini, R., Bertonati, G., Occelli, V., Targher, S., and Collignon, O. (2018). Decoding auditory motion direction and location in hMT+/V5 and Planum Temporale of sighted and blind individuals. *J. Vis.* **18**, 436.
27. Kaplan, J.T., Man, K., and Greening, S.G. (2015). Multivariate cross-classification: applying machine learning techniques to characterize abstraction in neural representations. *Front. Hum. Neurosci.* **9**, 151.
28. Dumoulin, S.O., Bittar, R.G., Kabani, N.J., Baker, C.L., Jr., Le Goualher, G., Bruce Pike, G., and Evans, A.C. (2000). A new anatomical landmark for reliable identification of human area V5/MT: a quantitative analysis of sulcal patterning. *Cereb. Cortex* **10**, 454–463.
29. Huk, A.C., Dougherty, R.F., and Heeger, D.J. (2002). Retinotopy and functional subdivision of human areas MT and MST. *J. Neurosci.* **22**, 7195–7205.
30. Marcus, D.S., Harms, M.P., Snyder, A.Z., Jenkinson, M., Wilson, J.A., Glasser, M.F., Barch, D.M., Archie, K.A., Burgess, G.C., Ramaratnam, M., et al.; WU-Minn HCP Consortium (2013). Human Connectome Project informatics: quality control, database services, and data visualization. *Neuroimage* **80**, 202–219.
31. Gitelman, D.R., Penny, W.D., Ashburner, J., and Friston, K.J. (2003). Modeling regional and psychophysiological interactions in fMRI: the importance of hemodynamic deconvolution. *Neuroimage* **19**, 200–207.
32. Abdi, H., O’Toole, A.J., Valentin, D., and Edelman, B. (2005). Dros. Inf. Serv.TATIS: The analysis of multiple distance matrices. In *Computer Vision and Pattern Recognition-Workshops, 2005. CVPR Workshops (IEEE Computer Society Conference on (IEEE))*, p. 42.
33. Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., and Pietrini, P. (2001). Distributed and Overlapping Representations of Face and Objects in Ventral Temporal Cortex. *Science* **293**, 2425–2430.
34. Watson, J.D.G., Myers, R., Frackowiak, R.S.J., Hajnal, J.V., Woods, R.P., Mazziotta, J.C., Shipp, S., and Zeki, S. (1993). Area V5 of the human brain: evidence from a combined study using positron emission tomography and magnetic resonance imaging. *Cereb. Cortex* **3**, 79–94.
35. Tootell, R.B.H., Reppas, J.B., Kwong, K.K., Malach, R., Born, R.T., Brady, T.J., Rosen, B.R., and Belliveau, J.W. (1995). Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *J. Neurosci.* **15**, 3215–3230.
36. Warren, J.D., Zielinski, B.A., Green, G.G.R., Rauschecker, J.P., and Griffiths, T.D. (2002). Perception of sound-source motion by the human brain. *Neuron* **34**, 139–148.
37. Pavani, F., Macaluso, E., Warren, J.D., Driver, J., and Griffiths, T.D. (2002). A common cortical substrate activated by horizontal and vertical sound movement in the human brain. *Curr. Biol.* **12**, 1584–1590.
38. Poirier, C., Baumann, S., Dheerendra, P., Joly, O., Hunter, D., Balezeau, F., Sun, L., Rees, A., Petkov, C.I., Thiele, A., and Griffiths, T.D. (2017). Auditory motion-specific mechanisms in the primate brain. *PLoS Biol.* **15**, e2001379.
39. Baumgart, F., Gaschler-Markefski, B., Woldorff, M.G., Heinze, H.-J., and Scheich, H. (1999). A movement-sensitive area in auditory cortex. *Nature* **400**, 724–726.
40. Battal, C., Rezk, M., Mattioni, S., Vadlamudi, J., and Collignon, O. (2019). Representation of auditory motion directions and sound source locations in the human planum temporale. *J. Neurosci.* **39**, 2208–2220.
41. Sani, L., Ricciardi, E., Gentili, C., Vanello, N., Haxby, J.V., and Pietrini, P. (2010). Effects of Visual Experience on the Human MT+ Functional Connectivity Networks: An fMRI Study of Motion Perception in Sighted and Congenitally Blind Individuals. *Front. Syst. Neurosci.* **4**, 159.
42. Malonek, D., Tootell, R.B., and Grinvald, A. (1994). Optical imaging reveals the functional architecture of neurons processing shape and motion in owl monkey area MT. *Proc. Biol. Sci.* **258**, 109–119.
43. Shmuel, A., and Grinvald, A. (1996). Functional organization for direction of motion and its relationship to orientation maps in cat area 18. *J. Neurosci.* **16**, 6945–6964.

44. Diogo, A.C.M., Soares, J.G.M., Koulakov, A., Albright, T.D., and Gattass, R. (2003). Electrophysiological imaging of functional architecture in the cortical middle temporal visual area of Cebus apella monkey. *J. Neurosci.* 23, 3881–3898.
45. Alink, A., Singer, W., and Muckli, L. (2008). Capture of auditory motion by vision is represented by an activation shift from auditory to visual motion cortex. *J. Neurosci.* 28, 2690–2697.
46. Lewis, R., and Noppeney, U. (2010). Audiovisual synchrony improves motion discrimination via enhanced connectivity between early visual and auditory areas. *J. Neurosci.* 30, 12329–12339.
47. Scheef, L., Boecker, H., Daamen, M., Fehse, U., Landsberg, M.W., Granath, D.O., Mechling, H., and Effenberg, A.O. (2009). Multimodal motion processing in area V5/MT: evidence from an artificial class of audio-visual events. *Brain Res.* 1252, 94–104.
48. Sadaghiani, S., Maier, J.X., and Noppeney, U. (2009). Natural, metaphoric, and linguistic auditory direction signals have distinct influences on visual motion processing. *J. Neurosci.* 29, 6490–6499.
49. Berger, C.C., and Ehrsson, H.H. (2016). Auditory Motion Elicits a Visual Motion Aftereffect. *Front. Neurosci.* 10, 559.
50. Kitagawa, N., and Ichihara, S. (2002). Hearing visual motion in depth. *Nature* 416, 172–174.
51. Konkle, T., Wang, Q., Hayward, V., and Moore, C.I. (2009). Motion aftereffects transfer between touch and vision. *Curr. Biol.* 19, 745–750.
52. Charbonneau, G., Véronneau, M., Boudrias-Fournier, C., Lepore, F., and Collignon, O. (2013). The ventriloquist in periphery: impact of eccentricity-related reliability on audio-visual localization. *J. Vis.* 13, 20.
53. Hidaka, S., Teramoto, W., Sugita, Y., Manaka, Y., Sakamoto, S., and Suzuki, Y. (2011). Auditory motion information drives visual motion perception. *PLoS ONE* 6, e17499.
54. Rohe, T., and Noppeney, U. (2015). Cortical hierarchies perform Bayesian causal inference in multisensory perception. *PLoS Biol.* 13, e1002073.
55. Ward, L.M., McDonald, J.J., and Lin, D. (2000). On asymmetries in cross-modal spatial attention orienting. *Percept. Psychophys.* 62, 1258–1264.
56. Beauchamp, M.S. (2003). Detection of eye movements from fMRI data. *Magn. Reson. Med.* 49, 376–380.
57. Goebel, R., Khorram-Sefat, D., Muckli, L., Hacker, H., and Singer, W. (1998). The constructive nature of vision: direct evidence from functional magnetic resonance imaging studies of apparent motion and motion imagery. *Eur. J. Neurosci.* 10, 1563–1573.
58. Kaas, A., Weigelt, S., Roebroek, A., Kohler, A., and Muckli, L. (2010). Imagery of a moving object: the role of occipital cortex and human MT/V5+. *Neuroimage* 49, 794–804.
59. Emmerling, T.C., Zimmermann, J., Sorger, B., Frost, M.A., and Goebel, R. (2016). Decoding the direction of imagined visual motion using 7T ultra-high field fMRI. *Neuroimage* 125, 61–73.
60. Johnson, P. (1982). The functional equivalence of imagery and movement. *Q. J. Exp. Psychol. A* 34, 349–365.
61. Schorr, D., Balzano, G., and Smith, E.E. (1978). Selective interference between imagery and perception: Is it modality specific or relation specific? *Bull. Psychon. Soc.* 12, 419–422.
62. Wolbers, T., Klatzky, R.L., Loomis, J.M., Wutte, M.G., and Giudice, N.A. (2011). Modality-independent coding of spatial layout in the human brain. *Curr. Biol.* 21, 984–989.
63. Oosterhof, N.N., Connolly, A.C., and Haxby, J.V. (2016). CoSMoMVA: Multi-Modal Multivariate Pattern Analysis of Neuroimaging Data in Matlab/GNU Octave. *Front. Neuroinform.* 10, 27.
64. Bex, P.J., Dakin, S.C., and Simmers, A.J. (2003). The shape and size of crowding for moving targets. *Vision Res.* 43, 2895–2904.
65. Møller, H. (1992). Fundamentals of binaural technology. *Appl. Acoust.* 36, 171–218.
66. Corbetta, M., Akbudak, E., Conturo, T.E., Snyder, A.Z., Ollinger, J.M., Drury, H.A., Linenweber, M.R., Petersen, S.E., Raichle, M.E., Van Essen, D.C., and Shulman, G.L. (1998). A common network of functional areas for attention and eye movements. *Neuron* 21, 761–773.
67. Petit, L., and Haxby, J.V. (1999). Functional anatomy of pursuit eye movements in humans as revealed by fMRI. *J. Neurophysiol.* 82, 463–471.
68. Griffiths, T.D., and Green, G.G.R. (1999). Cortical activation during perception of a rotating wide-field acoustic stimulus. *Neuroimage* 10, 84–90.
69. von Saldern, S., and Noppeney, U. (2013). Sensory and striatal areas integrate auditory and visual signals into behavioral benefits during motion discrimination. *J. Neurosci.* 33, 8841–8849.
70. Alink, A., Euler, F., Kriegeskorte, N., Singer, W., and Kohler, A. (2012). Auditory motion direction encoding in auditory cortex and high-level visual cortex. *Hum. Brain Mapp.* 33, 969–978.
71. Zaitsev, M., Hennig, J., and Speck, O. (2004). Point spread function mapping with parallel imaging techniques and high acceleration factors: fast, robust, and flexible method for echo-planar imaging distortion correction. *Magn. Reson. Med.* 52, 1156–1166.
72. Krumbholz, K., Schönwiesner, M., Rübsem, R., Zilles, K., Fink, G.R., and von Cramon, D.Y. (2005). Hierarchical processing of sound location and motion in the human brainstem and planum temporale. *Eur. J. Neurosci.* 21, 230–238.
73. Friston, K.J., Buechel, C., Fink, G.R., Morris, J., Rolls, E., and Dolan, R.J. (1997). Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage* 6, 218–229.
74. Chang, C.-C., and Lin, C.-J. (2011). Libsvm. *ACM Trans. Intell. Syst. Technol.* 2, 1–27.
75. Stelzer, J., Chen, Y., and Turner, R. (2013). Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): random permutations and cluster size control. *Neuroimage* 65, 69–82.
76. Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc.* 57, 289–300.
77. Kriegeskorte, N., Goebel, R., and Bandettini, P. (2006). Information-based functional brain mapping. *Proc. Natl. Acad. Sci. USA* 103, 3863–3868.
78. Albright, T.D. (1984). Direction and orientation selectivity of neurons in visual area MT of the macaque. *J. Neurophysiol.* 52, 1106–1130.
79. Kriegeskorte, N., Mur, M., and Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* 2, 4.
80. Kriegeskorte, N., Mur, M., Ruff, D.A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., and Bandettini, P.A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron* 60, 1126–1141.
81. Zabicki, A., de Haas, B., Zentgraf, K., Stark, R., Munzert, J., and Krüger, B. (2017). Imagined and executed actions in the human motor system: Testing neural similarity between execution and imagery of actions with a multivariate approach. *Cereb. Cortex* 27, 4523–4536.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
fMRI data	This paper	Available on request
Software and Algorithms		
MATLAB 2012b	Mathworks	RRID:SCR_001622
SPM8	Wellcome Trust Centre for Neuroimaging	https://www.fil.ion.ucl.ac.uk/spm/software/spm8/
CoSMoMvPA	[63]	http://www.cosmomvpa.org/
Connectome workbench	[30]	https://humanconnectome.org/software/connectome-workbench

RESOURCE AVAILABILITY

Lead Contact

Further information and requests should be directed to and will be fulfilled by the Lead Contact, Mohamed Rezk (mohamed.rezk@uclouvain.be).

Materials Availability

This study did not generate new unique reagents.

Data and Code Availability

The dataset supporting the current study has not been deposited in a public repository due to restrictions on data sharing in our ethical approval, and as the participants did not consent to sharing their data publicly. However, fully anonymized data can be shared upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Participants

Twenty-four individuals participated in the experiment (12 males, mean age $\pm$ SD = 27.68 $\pm$ 5.11 years, age range = 20 - 41 years). All participants were right-handed and had normal or corrected-to-normal vision, and self-reported normal audition. Participants had no reported history of psychiatric or neurological problems. One subject was excluded from the analysis due to excessive head motion and very low behavioral performance in discriminating visual and auditory motion directions during fMRI acquisition suggesting no attention to the task. All participants were naive to the purpose of the experiment. The protocol was approved by the research ethics committee in University of Trento and the Center for Mind/Brain Sciences (CIMEC). A written consent was acquired from all participants. Participants were monetarily compensated for their participation.

METHOD DETAILS

Training

Prior to testing, participants underwent a two-parts training session. First, the participants were trained in MRI-like conditions for the auditory motion direction discrimination task to get familiarized with the procedure and to ensure the best sound perception inside the scanner. Participants were asked to lie down during the training session to get familiarized with the space shift of the motion sound when they are in the scanner. They were trained using in-ear-plugs under the headphones while loud fMRI acquisition noise was played from a speaker behind their head. The training for the auditory motion direction discrimination task in the decoding experiment lasted for a maximum of 30 min or until the participants performed two complete training runs without any mistakes. Each training run had 8 blocks with 4 auditory motion directions and 2 repetitions per direction. The percentage of correct responses and feedback was provided at the end of each run. Training for visual motion direction discrimination with 100% coherence in MRI-like conditions was not necessary due to the easiness of the task (all participants reach ceiling effect without training). After the auditory training in MRI-like conditions was over, the participants performed a 1-run training session containing both visual and auditory blocks (2 modalities $\times$ 4 directions $\times$ 1 repetition) while sitting in front of the computer with MRI noise played in the background.

EXPERIMENTAL DESIGN

Stimuli

Visual Stimuli

Visual stimuli were projected on a screen (frame rate: 60 Hz; screen resolution: 1024 × 768 pixels; mean luminance: 109 cd/m²) behind the scanner via an EMP 7900 Epson projector. Participants viewed the screen (distance from head = 134 cm) through a mirror mounted on the MRI head coil. Participants performed a task in each experiment by responding with a Lumina MR-compatible response button. Visual stimuli consisted of random dot kinematograms (RDK) within an invisible circular aperture of 8 visual degrees centered around a white fixation cross. Each visual event was composed of 300 white dots (diameter = 0.1°) on a black background. Motion dots had a speed of 4°/s and a limited dot lifetime of 200 ms. The use of the limited lifetime ensures that motion direction discrimination could be achieved by relying on the global motion perception rather than focusing on a single dot [64]. Dots moved in one of four possible motion directions [upward, downward, rightward, and leftward] with 100% coherence level (Figure 4A). Each visual motion event lasted 1.2 s. In the visual static condition, each static event had RDKs of 300 dots that remained static. The location of the dots was randomized in each event.

Auditory Stimuli

To provide an externalized and ecological sensation of auditory motion and accurate motion localization in the MRI, we recorded the auditory stimuli using in-ear binaural microphones [65] in a semi-anechoic room for each subject individually prior to the scanning session (Zoom H4n digital wave recorder – 200 m, with microphones Master Series – Sound Professionals-TFB-2). Participants were positioned at the center of the sound setup with their head on a chin-rest, facing one vertical and one horizontal semi-circular sound bar of 31 speakers each. The sound bars had a radius of 1.1 m and provided a fixed distance (1.1 m) between each speaker and the participant's head (Figure 4B). The position of the horizontal bar was at the same level of the subject's ear level, and the position of the vertical bar was at the participant's mid-sagittal plane. Pink noise (duration 1.2 s, fade in/out of 50 ms each) was divided in 31 equal segments and was replayed sequentially in the 31 corresponding speakers, to create smooth sensation of motion (no gap or overlap between the different segments). The sound's level was constant at 65 dB from the subject's head position. Four translational motion directions were recorded [upward, downward, rightward, and leftward]. Motion speed was 2 m/s and covered 120° of the subject's peripheral space. For the target detection task, similar motion stimuli with faster speed (4 m/s) and shorter duration (0.6 s) were recorded. An additional static condition was recorded at the central speaker, located at the intersection of the horizontal and vertical planes of the sound bars. Static events had a duration of 1.2 s for the normal event, and 0.6 s for the target event. The recordings from each participant were re-played inside the MRI scanner for the auditory motion localizer and the directional motion decoding experiment. By using such sound system and in-ear recording in each subject, the auditory stimuli are convolved with each individuals' own pinna and head related transfer function producing a vivid auditory perception of the external space.

Visual motion localizer

We implemented a traditional visual motion localizer to localize hMT+/V5 both at the group-level and in each subject individually [29]. Visual motion and static conditions were generated using white random dot kinematograms (RDK) on a black background. Since the visual stimuli were presented at fixation, our localizer defines the whole hMT+/V5 complex including both MT and MST [29]. We used dots moving in one of four possible translational motion directions [upward, downward, rightward, and leftward]. The visual motion localizer started with an initial 5 s of blank screen and ended with 13 s of blank screen. The run had 6 blocks of motion and static conditions. Blocks were separated by each other with an inter-block interval (IBI) of 8 s. Each block (~15.6 s) had 12 stimuli of 1.2 s each, and an inter-stimulus interval (ISI) of 0.1 s. Each motion block had 3 repetitions of the 4 motion directions. The order in which the motion directions were presented was randomized in each block and balanced across the different motion blocks. In the static condition, the location of the dots was randomized for each event (inducing 12 changes within one block). The fixation cross was presented during the whole duration of the visual localizer. To minimize eye movements and saccadic shifts [66, 67], the participants were asked to detect a brief change (150 ms) in the fixation cross color. The number of targets (range: 0–2 in each block) was randomized and balanced across conditions. The participants performed the task while the fMRI data were acquired with an accuracy (mean ± SD) of 98.62% ± 2.36%.

Auditory motion localizer

A localizer was implemented to define regions responding preferentially to auditory motion sounds. Previous studies demonstrated that regions within the middle temporal cortex and PT are selectively recruited during auditory motion processing compared with static sounds [10, 22, 36, 68–70]. We used an experimental design matching the one implemented for the visual motion localizer. The participants were blindfolded during the auditory motion localizer. The run started with an initial 5 s of silence and ended with 11 s of silence. The localizer run had 13 blocks of auditory motion and 13 blocks of auditory static conditions. Auditory blocks were separated by an interval of 6 s. Each block had 12 events of 1.2 s, and an ISI of 0.1 s. In the motion blocks, auditory motion stimuli were presented in one of four directions [upward, downward, rightward, leftward] (Figure S2). Each motion block had 3 repetitions of each motion directions. The presentation order of the different auditory motion directions within each block was randomized and balanced across blocks. The static blocks had 12 stimuli of static events separated by the 0.1 s inter-trial interval. Participants were asked to detect target events that were faster in speed and shorter in duration (0.6 s [8]). The number of targets ranged between 1 and 3 targets in each block and was balanced across conditions. The duration of each block varied depending on the

number of targets present (range 14.4 – 15.6 s). Auditory stimuli were delivered through a SereneSound MR-compatible in-ear headphones inside the scanner. The participants performed the task while the fMRI data were acquired with an accuracy (mean $\pm$ SD) of $86.96\% \pm 14.71\%$.

Directional motion decoding experiment

The directional motion decoding experiment was composed of 12 runs, each run comprising 8 motion blocks (4 directions $\times$ 2 modalities). Each motion block had either visual or auditory stimuli moving in one of four possible directions [upward, downward, rightward, or leftward]. Each run started with an initial 5 s of blank screen and ended with 11 s of blank black screen. Each block had 10 events of a 1.2 s, all moving in the same direction. Events were separated by a 100 ms inter-trial interval resulting in a block duration of 13 s. To ensure fixation and to minimize eye movements during the motion blocks, the participants were asked to press a button whenever the fixation cross changed color (duration = 150ms). Performing a visual detection task during motion processing also decreases the possibility of visual imagery [60, 61]. Following each block, there was a response collection period of 5 s where the participants heard a voice saying “up,” “down,” “right,” and “left” in a pseudo-randomized order. The participants were instructed to press a button with their index finger when they heard the cue that matched the block’s motion direction in vision or in audition. The number and the order of the correct button presses were counterbalanced across the different conditions. The task was identical during the visual and the auditory motion blocks in the directional motion decoding experiment. After the response collection, there was a silent inter-block interval of 6 s. The order of the blocks was randomized for the motion direction and the presented modality in each run (Figure 4C).

MRI acquisition

Structural and functional data were acquired at the Center for Mind/Brain Sciences (CIMEC - University of Trento, Italy) using a Bruker MedSpec BioSpin 4-Tesla MR scanner with a standard eight-channel head coil. A whole-brain structural scan was acquired with a high-resolution T1-weighted MPRAGE sequence (176 sagittal slices, repetition time (TR) = 2700 ms, echo time = 4.18 ms, voxel size = $1 \times 1 \times 1$ mm, flip angle = 7° , field of view = 256×224 mm).

Functional data were acquired with a T2* weighted gradient echo-planar imaging sequence (EPI). A total of 39 slices were acquired in an ascending interleaved order. Acquisition parameters included a TR of 2500 ms, echo time of 26 ms, flip angle = 73° , field of view = 192×117 mm, gap = 0.3 mm, and voxel size of $3 \times 3 \times 3$ mm. The number of volumes acquired in the visual motion localizer and the auditory motion localizer were 126 and 235 volumes, respectively. For the directional motion decoding experiment, 1068 functional volumes were acquired (89 volumes $\times$ 12 runs). The first four initial scans of each run were discarded to allow for equilibrium magnetization. To apply distortion correction to the acquired functional images, point spread function (PSF) runs were acquired [71] before every three consecutive runs. Participants entered the scanner blindfolded for the MPRAGE and the auditory motion localizer. The blindfold was removed afterward for the visual motion localizer and the directional motion decoding experiment.

fMRI preprocessing

fMRI data were preprocessed in statistical parametric mapping (SPM8 – Wellcome Department of Imaging Neuroscience, University College London, UK) implemented in MATLAB R2012b (Mathworks, Inc.). Preprocessing steps included slice time correction, EPI alignment to the mean functional image with a 2nd degree B-spline interpolation, coregistration of the functional volumes to the structural image and normalization to the Montreal Neurological Institute (MNI) template using a resampling of the structural and functional data to an isotropic 2 mm resolution.

Eye tracker experiment

To investigate whether different auditory motion directions elicited systemic eye movement patterns, we tested 7 participants in an eye tracking experiment outside the fMRI scanner. We used an Eyelink 1000 eye tracker (SR Research, Mississauga, Canada) to record the gaze location. The eye tracker had a sampling rate of 1000 Hz, and gaze tracking range of 32° angle horizontally and 25° angle vertically. The eye tracker was calibrated using a built-in 9-point protocol. The participants performed 8 runs of the same experiment that was done inside the scanner for the auditory modality and performed the same task. We tested in each participant if the eye position was different when perceiving different auditory motion directions. First, we performed a within-subject per-time point analysis where we compared the gaze location in the horizontal and vertical plane at each time point (1 ms resolution) of the auditory motion blocks. Since the analysis was done within-subject, the variance of the gaze location for each motion direction was estimated from the different runs. Additionally, we performed a within-subject analysis where we averaged the gaze position across time in each run for the horizontal and the vertical plane for each motion direction (See Figure S4A and S4B for the results).

QUANTIFICATION AND STATISTICAL ANALYSIS

Behavioral Analysis

The behavioral accuracy for discriminating the different motion directions in the directional motion decoding experiment was assessed using a 2 (modalities) $\times$ 4 (directions) repeated-measures ANOVA. A Greenhouse–Geisser correction was applied to the significance levels and degrees of freedom whenever the assumption of sphericity was violated. Further pairwise comparisons were performed and corrected for multiple comparisons using Family-wise error (FWE) correction. See Data S1 for the behavioral data.

Univariate analysis

Functional data were spatially smoothed with a 6-mm full-width half-maximum (FWHM) smoothing kernel for the univariate analysis. For the visual motion localizer, a general linear model (GLM) was fitted for every voxel with the visual motion and the visual static conditions as regressors of interest. Each regressor was modeled with a boxcar function and convolved with the canonical hemodynamic response function (HRF). The 6 head motion parameters (3 translation and 3 rotation parameters) were included in the model as regressors of no interest. A high-pass filter of 128 s was used to remove low-frequency signal drifts. The contrast [Motion > Static] was computed for every subject individually [29, 35]. A standard group-level analysis was performed. The individual contrasts were further smoothed with a 6-mm FWHM kernel and entered into a random effects model for the second-level analysis. A one-sample *t* test was performed, and group-level statistical inferences were made at a voxel-level threshold of $p < 0.05$, FWE-corrected. The group-level analysis revealed significant activations in V1, the inferior occipital gyrus, and the middle temporal gyrus (hMT+/V5) bilaterally [MNI peak coordinates: $-42 -66 2$ and $46 -62 6$] (Figure 1A; Table S1).

To define subject-specific ROIs that will be used for the MVPA on a separate dataset (directional motion decoding experiment), a lenient threshold of $p < 0.01$ uncorrected was used to locate hMT+/V5 bilaterally. The individually defined hMT+/V5 was identified in each subject as the area that responded significantly to motion more than static in the occipito-temporal cortex, and was constrained by the anatomical location of hMT+/V5 and an Euclidean distance not more than 12 mm away from the group-level hMT+/V5 peak coordinate. The peak coordinate of hMT+/V5 in each subject was used to create a 7-mm radius sphere to be used as individually defined ROIs in the multivariate analysis. Out of the 23 participants, the left hMT+/V5 was identified in 21 participants, and the right hMT+/V5 was identified in 22 participants (Figure 1B and 1C, see Table S2 for hMT+/V5 MNI coordinates in each subject). For each ROI, the participants whose individually defined regions couldn't be identified were not included in any analysis relying on individually defined ROIs.

For the auditory motion localizer, the same procedure was applied to defined regions that show enhanced recruitment during processing auditory motion compared with auditory static stimuli. Previous studies demonstrated that the PT and regions within the middle temporal cortex are recruited during auditory motion processing [10, 22, 36, 72]. PT and MTa were defined both at the individual-level as individually defined ROIs for the MVPA, and at the group-level using the same procedure previously described for the visual motion localizer experiment. At the group-level, we observed an enhanced recruitment of the planum temporale (PT) (MNI peak coordinates: $[-46 -30 8]$ and $[8, 30, 35]$), the superior temporal gyri and the middle temporal gyrus (hereafter MTa, MNI coordinates: $[-46 -64 6]$ and $[8, 46-48, 52-55]$) bilaterally (Figure 1A; Table S1). In addition to the group-level results, the peak coordinate for PT and MTa were individually defined from the contrast [auditory motion > auditory static] which was constrained by anatomical location, and less than 12 mm away from the group-level PT and MTa peak coordinates. We were able to define bilateral MTa and PT in 22 out of the 23 participants (Figure 1C; Table S2).

By overlaying the group-level maps from the visual and auditory motion localizer, we observed an overlap between the regions processing visual and auditory motion. However, regions preferential to auditory motion were more anterior and dorsal than the visually defined hMT+/V5 (Figure 1A).

Beta parameter estimates

We used the unsmoothed data from the visual and the auditory motion localizer for the extraction of the beta parameter estimates. The use of the unsmoothed data prevents spill-over effects that may arise due to the close proximity between the visually defined hMT+/V5 and the auditory-defined MTa. Beta extraction was performed on both individually defined and group-level ROIs [8, 19, 20]. For every ROI in each subject, we created a sphere of 3-mm radius around the subject-defined coordinates (and the group-level coordinates) and extracted the mean beta parameter estimates. To test if visually defined hMT+/V5 showed enhanced recruitment during auditory motion processing, we extracted the beta parameter estimates from the individually defined and the group-level hMT+/V5 for the auditory motion and static conditions in the auditory motion localizer. We implemented a paired-sample *t* test to compare the beta parameter estimates of the auditory motion and auditory static conditions in hMT+/V5 bilaterally. To test if auditory-defined ROIs (MTa and PT) showed enhanced recruitment during visual motion processing, the beta parameter estimates from the individually defined and group-level MTa and PT were extracted for the visual motion and visual static conditions in the visual motion localizer. For each ROI in each hemisphere, a paired-sample *t* test was implemented to compare the beta parameter estimates from the visual motion and the visual static condition. Results were corrected for multiple comparisons using Bonferroni correction. For illustration purposes only, the beta estimates for visual motion and visual static conditions were extracted from the individually and group-level defined hMT+/V5 in the visual motion localizer. Similarly, the beta estimates for the auditory motion and static conditions were extracted from the individually and group-level defined MTa and PT in the auditory motion localizer. These beta estimates were not analyzed statistically since there would be clear circularity between the ROI definition and beta extraction.

Functional connectivity analysis (PPI)

Psychophysiological Interaction (PPI) analysis [31, 73] identifies voxels in which the activity is more related to the activity in a seed region of interest (seed ROI) in a given psychological context. In the present study, PPI analyses were performed in order to identify brain regions that were more functionally coupled with MTa or hMT+/V5 while processing moving versus static auditory or visual information. The locations of the seed regions were defined from the group-level univariate analysis from the visual and auditory motion localizers. In each subject, the time series (first eigenvariate) from each seed region was extracted from a sphere (radius = 8 mm) around the local peak: left hMT+/V5 $[-42 -66 2]$, right hMT+/V5 $[6, 30, 45-48, 51-62]$, left MTa $[-46 -64 6]$, and right MTa

[8, 46–48, 52–55]. A first-level GLM analysis was performed in each subject where the model included the following regressors [1]: psychological regressor [motion > static] [2], the extracted activity from the seed region (physiological regressor) [3], and the interaction term (PPI) between the previous two regressors. The PPI term was estimated by first deconvolving the BOLD activity to estimate the neuronal response in the ROIs [31], then the interaction term was calculated as the element-wise product of the estimated neuronal activity and the psychological regressor. The interaction term was further convolved with the canonical HRF. The six motion parameters were included in the model as regressors of no interest. After estimating the PPI models for each subject, a random effects model was constructed and a one-sample *t* test was performed. In the second level analysis, we tested for regions that showed enhanced (or decreased) coupling with each of the seed regions in the motion condition compared with the static condition in each of the visual and auditory motion localizers separately.

Multivariate pattern analysis (MVPA)

ROI definition for MVPA

Group-level ROIs: Group-level hMT+/V5 was defined by creating a 7mm sphere (179 voxels) around the peak coordinates from the contrast [visual motion > visual static] obtained from the group-level univariate analysis in the visual motion localizer. MTa and PT were defined from the peak coordinates from the contrast [auditory motion > auditory static] obtained from the group-level univariate analysis in the auditory motion localizer. The group-level ROIs were used only in the motion versus static decoding in experiment 1. Group-level ROIs were not used in the motion direction decoding in experiment 2.

Individually defined ROIs

hMT+/V5 was defined in each subject by creating a 7mm sphere (179 voxels) around the peak coordinates from the first-level (individual) contrast [visual motion > visual static] obtained from the univariate analysis of the visual motion localizer. Similarly, MTa and PT was defined in each subject by creating a 7mm sphere around the peak coordinates from the first-level contrast [auditory motion > auditory static] obtained from the univariate analysis in the auditory motion localizer.

Experiment 1 (visual and auditory motion localizer)

Motion versus static decoding

For the visual and the auditory motion localizer separately, the preprocessing steps were similar to the ones implemented in the univariate analysis without applying smoothing to the functional data. A general linear model (GLM) analysis was performed to estimate the BOLD signal for every voxel. The GLM included a separate regressor for each motion and static block. The six head motion parameters (translation and rotation) were included in the model as regressors of no interest. A *t*-map was estimated for each motion and static block using the contrast [motion > baseline] and [static > baseline] respectively. The *t*-maps were used for further MVPA.

Multivariate analysis was performed using CoSMoMVPA (<http://www.cosmomvpa.org> [63]; implemented in MATLAB. Classification was performed using support vector machine (SVM) classifier as implemented in LIBSVM [74]. To test if the representation of the motion condition was different from the representation of the static condition, a [motion versus static] decoding analysis was implemented in the group-level and the individually defined ROIs (hMT+/V5, MTa, and PT bilaterally). Before running the decoding analysis in each ROI, the patterns of each block and each condition (motion and static) was demeaned individually to minimize the univariate activation level differences and to mathematically equate the mean activity of each condition and each block. For each localizer, a linear SVM classifier (*C* parameter = 1) was trained and tested to discriminate the motion and the static blocks in each subject individually. A leave-one block out cross-validation scheme was implemented. In each cross-validation fold, the training data (5 blocks per condition for the visual motion localizer, 12 blocks per condition for the auditory localizer) were normalized (*z*-scored) across conditions and the classifier was trained to discriminate the motion and static conditions. An ANOVA-based feature selection was implemented on the training data in each cross-validation fold to identify the most informative 150 voxels. The normalization parameters from the training data were applied to the test data, and the performance of the SVM classifier was evaluated on the test data (1 block). The previous step was repeated *n* times (*n* = 6 for the visual motion localizer, *n* = 13 for the auditory motion localizer) where in each fold the classifier was tested on a different block. A single classification accuracy was obtained by averaging the classification accuracies for all the cross-validation folds. If the classifier could successfully discriminate the two conditions above chance (50%), this would indicate the presence of motion information in the tested modality that the classifier was able to learn during the training. No statistical analysis or inferences were made for the [visual motion versus visual static] decoding in hMT+/V5 to avoid circularity as this region was defined from the univariate analysis by contrasting visual motion versus visual static. Similarly, no statistical analysis was performed for the [auditory motion versus auditory static] decoding in MTa and PT for the same reason.

Statistical significance of the classification results was assessed using a non-parametric technique by combining permutations and bootstrapping [75]. For each subject, the labels of the different conditions (motion and static) were permuted, and the same decoding analysis was performed. The previous step was repeated 100 times for each subject. A bootstrap procedure was applied in order to obtain a group-level null distribution that was representative of the whole group. From each subject's null distribution, one value was randomly chosen (with replacement) and averaged across all participants. This step was repeated 100,000 times resulting in a group-level null distribution of 100,000 values. The statistical significance of the MVPA results was estimated by comparing the observed result to the group-level null distribution. This was done by calculating the proportion of observations in the null distribution that had a classification accuracy higher than the one obtained in the real test. To account for the multiple comparisons, all *p* values were corrected using false discovery rate (FDR) correction [76].

Experiment 2 (Directional motion decoding experiment)

For the directional motion decoding experiment, the preprocessing steps were similar to the ones implemented in the univariate analysis except no smoothing was applied to the functional data. A general linear model (GLM) analysis was performed to estimate the BOLD signal for every voxel. The GLM included a separate regressor for each direction in each modality in each run. Each regressor was convolved with the canonical HRF. The six head motion parameters (translation and rotation) for each run were included in the model as regressors of no interest. An additional regressor of no interest was included in each run to remove the effects related to the named directions after each block and the participants' responses. For each run, t-maps were estimated for each motion direction in each modality using the contrast [motion direction > baseline]. The t-maps were used for further MVPA.

Within-modality decoding

The MVPA was performed using SVM classifier as implemented in LIBSVM [74]. To test for the presence of motion direction information, a within-modality decoding was implemented in each of the individually defined ROIs (hMT+/V5, MTa, and PT bilaterally). For each modality, a linear SVM classifier (C parameter = 1) was trained and tested to discriminate the 4 motion directions in each subject individually. A leave-one run out cross-validation scheme was implemented. In each cross-validation fold, the training data (11 runs) were normalized (z-scored) across conditions and the classifier was trained to discriminate the patterns of the 4 motion directions. We implemented an ANOVA-based feature selection on the training data in each cross-validation fold to use the most informative 150 voxels in each subject and each ROI. The normalization parameters from the training data were applied to the test data, and the performance of the SVM classifier was evaluated on the test data (1 run). The previous step was repeated 12 times (N-fold cross-validation) where in each fold the classifier was tested on a different run. A single classification accuracy was obtained by averaging the classification accuracies for all the cross-validation folds. If the classifier could successfully discriminate the different motion directions above chance (25%), this would indicate the presence of motion direction information in the tested modality that the classifier was able to learn during the training. Statistical significance was assessed non-parametrically as previously described [75] in the motion versus static decoding section. To account for the multiple comparisons, all p values were corrected using false discovery rate (FDR) correction [76].

In addition to the 4-way multi-class decoding, we performed a within-axis, and axis of motion direction decoding. In the within-axis binary decoding, we tried to discriminate motion patterns from the opposite directions within the same axis (upward versus downward, and rightward versus leftward motion). In the axis of motion decoding, we tried to discriminate the horizontal and vertical axes of motion by treating rightward and leftward motion as horizontal motion, and upward and downward motion as vertical motion (chance level = 50%). The analysis procedure was identical to one described in the multi-class decoding.

Cross-modal decoding

To test whether there was shared motion direction information between the visual and auditory modalities, cross-modal classification was performed in regions that could successfully decode motion directions in both the visual and auditory modalities. Before the cross-modal decoding analysis in each ROI, the patterns of each direction in each modality were demeaned individually to mathematically equate the mean activity of the visual and the auditory modalities. The SVM classifier was trained on one modality (e.g., vision) and tested on the other modality (e.g., audition), and vice versa. This resulted in a cross-validation scheme of 24 folds. In every cross-validation fold, the classifier was trained on N-1 runs (11 runs) from one modality and tested on 1 run from the other modality. An ANOVA-based feature selection was implemented on the training data in each cross-validation fold to use the most informative 150 voxels. The cross-modal decoding accuracy was obtained by averaging the classifications of all the cross-validation folds. Successful classification above the chance level (25%) would suggest the presence of shared motion direction information across the different modalities. Statistical significance was assessed non-parametrically as previously described [75].

In addition to the cross-modal decoding in the ROIs, we implemented a whole-brain searchlight approach [77] to test the regional specificity at the whole-brain level of the shared motion direction information across different modalities. A sphere of 3-voxels radius moved across the whole brain, where in each step, each voxel became the center of the searchlight sphere. Within each sphere, cross-modal decoding was implemented in the same manner described in the ROI analysis. The searchlight analysis results in a whole brain cross-modal decoding accuracy map for each subject. The individual accuracy maps were smoothed with 6 mm FWHM smoothing kernel and entered in a second-level model. A one-sample t test was performed to test if the cross-modal decoding accuracy was above the chance level (25%). Statistical maps were corrected using a SVC of 10-mm radius at the location of hMT+/V5 that was defined independently from the visual motion localizer.

Previous studies suggested that MT in the macaques and hMT+/V5 in humans have an axis-of-motion organization [4, 6, 78]. To test if the representation of auditory motion shows a similar axis-of-motion organization to the one in vision, we performed the same ROI decoding and searchlight analysis to test if we can classify the horizontal and the vertical axes (chance level is 50%) in each of the visual and the auditory modality. We aggregated the rightward and leftward motion conditions to the horizontal axis, and the upward and downward motion to the vertical axis. In the ROI cross-modal decoding analysis, first we identified regions where we could classify the different axes of motion in both the visual and auditory modality. Then an SVM classifier was trained to discriminate the patterns from the horizontal and the vertical axes from one modality (vision) and tested on the other modality (audition), and vice versa. Additionally, the same analysis was done across the whole brain using a whole-brain searchlight approach [77].

Modality decoding

To test whether the representation of motion directions across the visual and auditory modalities could be differentiated based on their activity patterns in each of the motion responsive areas, each motion direction pattern in each modality was demeaned separately to equate the mean activity of the visual and auditory modalities. For each direction separately, an SVM classifier was trained and tested to discriminate the two modalities (chance level = 50%), and a decoding accuracy was obtained. The overall modality decoding accuracy was then obtained by averaging the results from the four directions. The decoding procedure and the non-parametric significance testing [75] were identical to the previously described ROI decoding analysis.

Representational similarity analysis (RSA)

Using representational similarity analysis (RSA) [79, 80], we further investigated the differences in the representational geometry of the four motion directions within and across modalities. The advantage of using RSA is that it captures the (dis)similarity structure between different categories (e.g., motion directions in this experiment) and different modalities [79]. While cross-modal decoding analysis highlights the presence of common/shared information across modalities, it doesn't inform us about how (dis)similar the patterns are across directions and across modalities. Additionally, by using RSA we can construct different theoretical computational models and test which model best explains our observed data. By constructing a neural dissimilarity matrix from our fMRI data, we obtained an estimate of the (dis)similarity between the elicited patterns from the different motion directions and the different modalities in the right hMT+/V5. Pearson's correlation was used as a measure of dissimilarity (1 - Pearson's correlation) resulting in an 8×8 dissimilarity matrix in each subject. To investigate the multi-modal or the modality-abstracted nature of hMT/V5, we created multiple theoretical models in the form of DSMs ranging from a pure modality-invariant model to a pure multi-modal model with intermediate gradients in between them, and two unimodal (visual and auditory) models [81].

Theoretical models: In the modality-invariant model, the dissimilarity between the same motion direction across the different modalities were set to zero (hence, motion direction processing is abstracted from the input modality), and the dissimilarity between different motion directions was set to 1 regardless of the modality. Diagonal values were set to 0. On the other hand, the multi-modal model suggests that while motion directions within one modality are different from each other, they are more dissimilar than motion directions from other modalities. In the multi-modal model, the dissimilarity between the different motion directions within the same modality was set to 0.5, while the dissimilarity to motion directions from the different modality was set to 1. Diagonal values were set to 0.

In the intermediate models, we created a gradient from the modality-invariant to the multi-modal model. The dissimilarity between directions within the same modality was set to 0.5. The dissimilarity between the different directions across modalities was set to 1. While for the dissimilarity between the same directions across modalities were set to a gradient of values ranging from 0.1 to 0.9 with steps of 0.2.

In the unimodal visual and auditory models, the dissimilarity between the different motion directions within the dominant unimodal modality was set to 0.5. The dissimilarity between the different motion directions within the other non-dominant modality was set to 1 as this model assumes that there is no motion direction information in the non-dominant modality. The dissimilarity between the different directions across modalities was set to 1.

We correlated the neural DSM from the right hMT+/V5 with each of the theoretical models (Figure 5B). If hMT+/V5 is a multi-modal region, we would expect to see a low correlation with the modality-invariant model and a gradual increase in correlation as we progress in the direction of the multi-modal model. Oppositely, if hMT+/V5 is a supramodal region, we would expect to see a low correlation with the multi-modal model, and a gradual increase in correlation as we progress toward the modality-invariant model. If hMT+/V5 is a unimodal region, we would expect a higher correlation with one of the unimodal models than with the other models. We performed a 2-sided Wilcoxon signed rank test to investigate whether there was a difference between the correlation values of each two consecutive theoretical models and the neural DSM. Results were corrected for multiple comparisons using FWE-correction.

Split-half correlation analysis

In order to have a more in-depth understanding of the similarity between the motion directions' patterns within-modalities and across-modalities, we implemented a split-half correlation analysis [33] in the individually defined ROIs. We first demeaned every pattern for every direction and for every modality separately to equate the mean activity of the visual and the auditory modality. For each condition in each modality, the t-maps of the odd and even runs were averaged separately resulting in 1 t-map per condition for the odd runs, and 1 t-map per condition for the even runs. Using Pearson's correlation as a measure of similarity, we estimated the within-modality split-half correlation of the pattern of each motion direction from the odd runs with the pattern of the same motion direction from the even runs (within the same modality). The correlations of the four directions within the same modality were averaged, resulting in one value which was considered to be the estimate of the within-modality split-half correlation. This was performed for both the visual and the auditory modalities separately. To estimate the cross-modal split-half correlation, we correlated the pattern of each motion direction from the odd runs in one modality (e.g., visual leftward motion pattern) with the pattern of the same motion direction from the even runs in the other modality (e.g., auditory leftward motion pattern), and the other way around. The correlation results from the four motion directions, and from both conditions ([odd: Vision - even: Audition] and [odd: Audition - even: Vision]) were averaged resulting in one value for each participant which was considered to be the estimate for the cross-modal split-half correlation (Figure 5C). The correlation values were Fisher-transformed and tested against zero using a one-sample t test. Results were corrected for multiple comparisons using FWE-correction. To test whether there are differences in the correlation values

across the three different conditions (within-modality visual, within-modality auditory, and crossmodal correlation) in each ROI, we implemented a repeated-measures ANOVA in each ROI separately. A Greenhouse–Geisser correction was applied to the significance levels and degrees of freedom whenever the assumption of sphericity was violated. Further pairwise comparisons were performed and corrected for multiple comparisons using FWE-correction.

Differential pattern correlation

To investigate how cross-modal decoding was successful in the right hMT+/V5 despite the negative correlation across modalities, we tested whether the relative relation between the different motion directions was similar across modalities. For instance, the relation between the pattern of the upward motion direction and that of the other three motion directions in vision was similar to that in audition. First, we demeaned the motion direction patterns across the different directions in each modality separately. Based on the obtained differential motion direction patterns from the mean, we correlated the differential pattern of each motion direction across the two modalities using Pearson's correlation (Figure 5E). We obtained four correlation values per subject in the left and the right hMT+/V5 which correspond to the four motion directions. The average of the four correlation values was used as an estimate of the differential pattern correlation in each ROI. Correlation values were Fisher-transformed and tested against zero using a one-sample *t* test to determine if there is a significant correlation of the differences between motion directions across modalities (Figure 5E). Results were corrected for multiple comparisons using FWE-correction.