Cross Modal Object Recognition is Viewpoint Independent

Final Project Report

Submitted by:

Hemangini Parmar (Y8214)

Abstract

In this paper the previous claims about cross modal object recognition being viewpoint dependent have been negated. It was observed that the recognition accuracy of the participants did not show any substantial change when objects were presented from different viewpoints cross modally but the accuracy was reduced a lot in the within modal case. However the hypothesis given by the authors in the parent work [1] claimed that cross modal object recognition was mediated by a higher-level representation, which could not be proved in this paper.

Methodology

The experiment was conducted on 20 female undergraduates residing in the Girls Hostel 1 at IIT Kanpur. Each participant was shown four sets of four objects in modes selected randomly out of the sixteen possible combinations of (modality, rotation) shown in table 1. This was done to ensure that the participants did not form a sequence of the four objects in their minds. The rotation was given as 180 degree clockwise about the axis.

Objects: The objects were of approximately 3-5cmx2cmx2cm in dimension and were made out of LEGO bricks. They were sixteen in number, one for each of the sixteen possible combinations. This was done to ensure that the memory of the participant was not reinforced with the repetition of objects. The objects were made very similar to avoid giving any kind of distinctive visual or haptic cues to the participants, though mirror images were avoided. However the parent paper had used 48 objects, each made out of six component wooden blocks measuring 1.6cmx3.6cmx2.2cm, resulting

into objects which were 9.5cm high each and the remaining dimensions varied depending on the arrangement of the constituent blocks. The objects used in the present experiment were numbered alphabetically (fig 1) and each object had a small dot to define the unrotated orientation of the object similar to the parent paper. Pilot testing was done to ensure that this dot was not noticed by the participants as it had a color dark enough to hide it from the background color of the object (black).

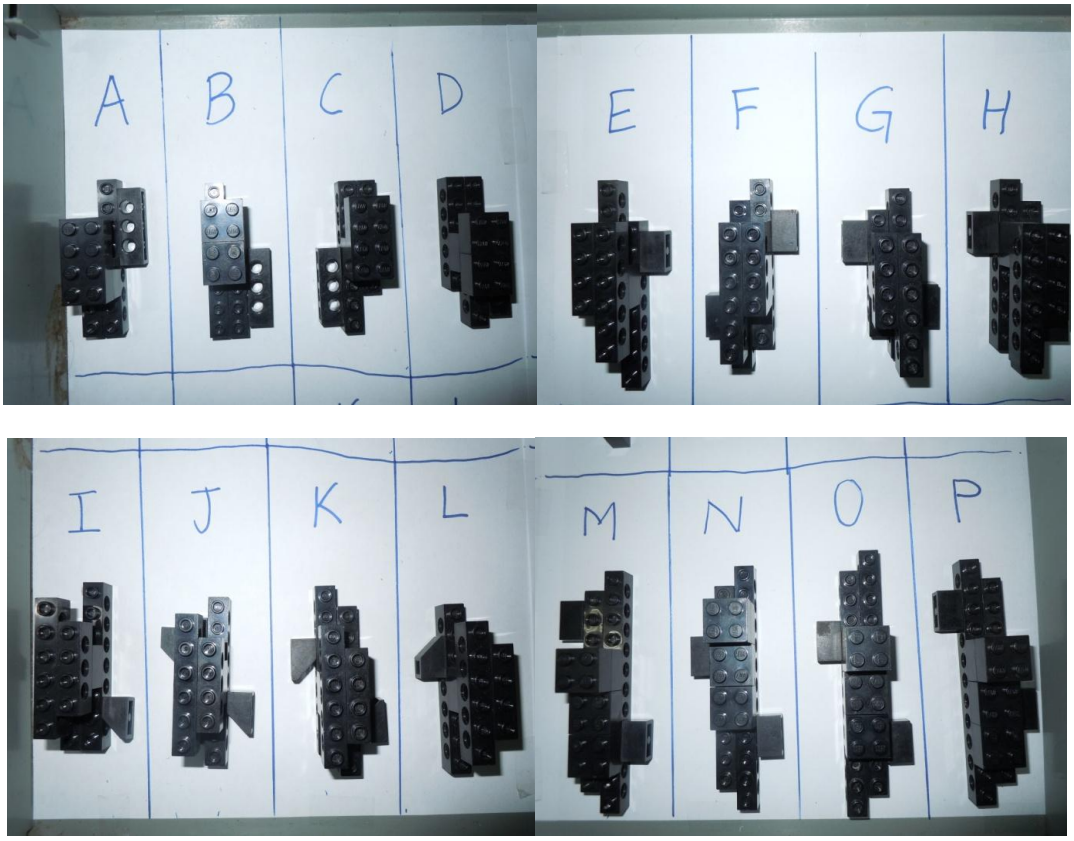

Fig 1: The sixteen objects used in the experiment

Modality → **Rotation ↓**	Visual-Visual	Visual-Haptic	Haptic-Visual	Haptic-Haptic
Unrotated				
Rotated about X				
Rotated about Y				
Rotated about Z				

Table 1: The modality and rotation combination in which the 16 objects were given to the participants

Experimental Setup : The participants were made to sit on a stool placed in front of a table (fig 2) such that the spacing between the eyes of the participant and the table was sufficient enough to allow her to view the object from all sides in case of visual mode. An additional setup (fig 2b) was permanently placed (removed in fig 2a to better explain the visual setup) which was high enough to block the view of the object completely from the participant but at an adequate distance to allow the participant to touch the object comfortably from all sides. As given in the parent paper, the visual mode prohibited the participant from touching the object or standing up and moving around the object but she was allowed to move her head to look all around the object, while in case of haptic mode the participants were restricted to keep the orientation of the object unchanged, i.e. the way it was given to them, while they were free to feel the object from all sides. The time limit which was adopted was the same as in the parent paper, i.e. 15 seconds for visual observation and 30 seconds for haptic observation [1].

Method: The sixteen possible combinations (table 1) were given randomly to the participants, to avoid formation of any kind of pattern in their minds. The 'learning level' consisted of showing four objects, identified by numbers (1-4), either visually or haptically to the participants, for the respective time durations per trial making a total of four trials (4x4 objects), and the 'recognition level' involved showing the same four objects in any four possible combinations out of the sixteen combinations and asking the participants to recognize the objects using their allotted numbers. In this way the sixteen objects were made to recognize in the sixteen possible combinations and the observation accuracy was analyzed (see Results).

Fig 2a: Visual Setup used in the experiment

Fig 2b: Haptic Setup used in the experiment

Results

The average observation accuracy (%) was plotted against the modality for both the rotated and the unrotated cases (fig 3). It was observed that the accuracy was marginally affected by rotation in the cross-modal case while it decreased substantially in the within-modal case. The results obtained agreed with the results obtained in the parent paper, thus validating the hypothesis that cross modal object recognition is viewpoint independent.

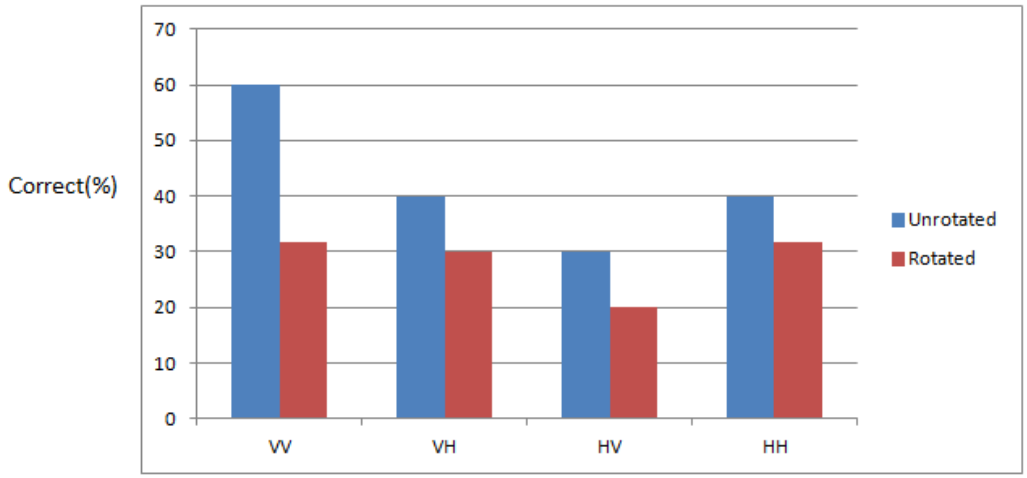

Fig 3: Average observation accuracy (%) versus the modality for both rotated and unrotated cases

Similar to the parent paper, an OSIQ questionnaire (shown in fig 4) was used here to assess the mental representations formed in the participants while recognizing the objects. 'OSIQ consists of two scales: **object imagery scale** assesses preferences for representing and processing colorful, pictorial, and high-resolution images of objects and a **spatial imagery scale** which assesses preferences for representing and processing schematic images, spatial relations amongst objects, and spatial transformations ', [2]. The resulting scores were correlated with the average observation accuracy for within-modal and cross-modal cases for both rotated and unrotated conditions, converted to a scale of {0,1} as shown in fig 5 and the correlation coefficient(r) and the probability of getting a correlation as large as the observed value by random chance when true correlation is zero (p) have been given in table 2 below. Each dot in the figure 5 denotes (observation accuracy, imagery score) for each participant (total six dots for six participants who were asked to give the questionnaire). There is an acceptable correlation between the cross-modal unrotated case and the spatial imagery score (r=0.5387) as expected but there is a discrepancy in the correlation between the cross-modal rotated case and spatial case.

	CM-R	CM-UR	WM-R	WM-UR
SS(r,p)	(-0.1161,0.8042)	(0.5387, 0.2121)	(-0.3702,0.4137)	(0.2929,0.5238)
OS(r,p)	(-0.3112,0.4970)	(0.3699,0.4141)	(0.0320,0.9458)	(0.6679,0.1011)

Table 2: Correlation coefficients (r) and probability (p) for correlation between imagery scores (SS and OS) and the modalities (CM and WM), for the rotated and unrotated conditions (R and UR)

		1	2	3	4	5
1	I was very good in 3-D geometry as a student.	1				
2	If I were asked to choose between engineering professions and visual arts, I would prefer engineering.	2				
3	Architecture interests me more than painting.	3				
4	My images are very colourful and bright.	4				
5	I prefer schematic diagrams and sketches when reading a textbook instead of colourful and pictorial illustrations.	5				
6	My images are more like schematic representations of things and events rather than detailed pictures.	6				
7	When reading fiction, I usually form a clear and detailed mental picture of a scene or room that has been described.	7				
8	I have a photographic memory.	8				
9	I can easily imagine and mentally rotate 3-dimensional geometric figures.	9				
10	When entering a familiar store to get a specific item, I can easily picture the exact location of the target item, the shelf it stands on, how it is arranged and the surrounding articles.	10				
11	I normally do not experience many spontaneous vivid images; I use my mental imagery mostly when attempting to solve some problems like the ones in mathematics.	11				
12	My images are very vivid and photographic.	12				
13	I can easily sketch a blueprint for a building that I am familiar with.	13				
14	I am a good Tetris player.	14				
15	If I were asked to choose between studying architecture and visual arts, I would choose visual arts.	15				
16	My mental images of different objects very much resemble the size, shape and colour of actual objects that I have seen.	16				
17	When I imagine the face of a friend, I have a perfectly clear and bright image.	17				
18	I have excellent abilities in technical graphics.	18				
19	I can easily remember a great deal of visual details that someone else might never notice. For example, I would just automatically take some things in, like what colour is a shirt someone wears or what colour are his/her shoes.	19				
20	In high school, I had less difficulty with geometry than with art.					

Fig 4: OSIQ Questionnaire used in the experiment (*Courtesy [2]*)

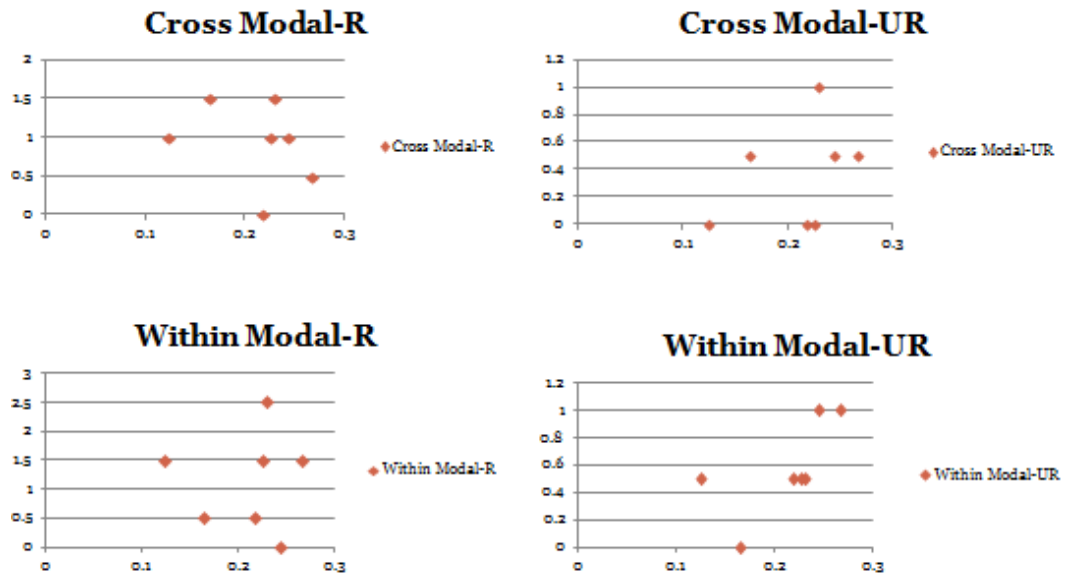

Fig 5a: Correlation between Spatial imagery scores and average observation accuracy for each participant; four graphs for both the modalities (within, cross) and orientations (R,UR)

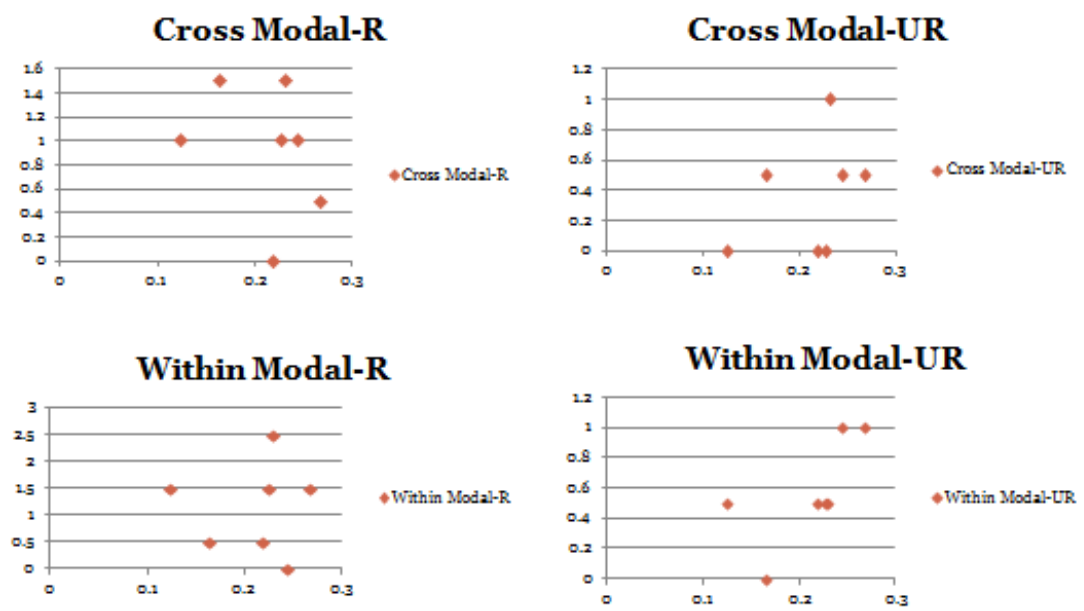

Fig 5b: Correlation between Object imagery scores and average observation accuracy for each participant; four graphs for both the modalities (within, cross) and orientations (R,UR)

Discussion

This experiment was able to validate the hypothesis that cross-modal object recognition is viewpoint independent. However no substantial evidence was collected in regard to the second hypothesis of cross-modal object recognition being mediated by a higher-level representation. As observed in table 2, acceptable correlation (r=0.5387) was observed between spatial imagery score and the cross-modal unrotated case validating the possibility that cross-modal object recognition could be mediated by an abstract spatial representation which is higher-level in nature. But the unexpected correlation coefficient value (r=-0.1161) for the cross-modal rotated case is an error and the possible reasons behind this discrepancy could be the less number of participants who took the questionnaire (six in number), the abstractness of the questionnaire itself, besides the correlation statistic is very sensitive to noise. Functional neuroimaging studies have observed overlapping in the regions related to visual and haptic shape processing in the brain. However the locus of the 'higher-level modality and viewpoint independent representation' [1] have not been located yet, hence obscuring the second hypothesis.

Acknowledgement

This report would not have been possible without the kind support and help of many individuals. I would like to extend my sincere thanks to all of them. First of all I am thankful to Professor Amitabha Mukerjee, Department of Computer Science and Engineering, IIT Kanpur, for giving me this opportunity to gain an indispensable experience and for constantly encouraging and monitoring me on every step. This project would not have been feasible without his inspiration and support. I would also like to thank the staff at the CSE Laboratory for providing me with the LEGO bricks.

References

[1] Lacey S, Peters A, Sathian K (2007): *Cross-Modal Object Recognition Is Viewpoint-Independent.* PLoS ONE 2(9): e890. doi:10.1371/journal.pone.0000890

[2]Olessia Blajenkova, Maria Kozhevnikov and Michaela Motes (2006): *Object-Spatial Imagery: A New Self-Report Imagery Questionnaire*, Appl. Cognit. Psychol. 20: 239–263

[3] Maximilian Riesenhuber and Tomaso Poggio (1999): *Hierarchical models of object recognition in cortex.* Nature Neuroscience, 2(11):1019–1025, November 1999.

[4]Lacey S, Campbell C (2006): *Mental representation in visual/haptic cross-modal memory: Evidence from interference effects.* QJ Exp Psychol, 59: 361-376