Cross-Modal Scene Perception

Fiona N. Newell,
Department of Psychology,
Trinity College Dublin,
Ireland
fiona.newell@tcd.ie

Caroline W-Y. Ang
Department of Psychology,
Trinity College Dublin,
Ireland

Heinrich H. Bülthoff
Max-Planck Institute for
Biological Cybernetics,
Tübingen, Germany

Abstract

We report the effect of cross modal performance on the recognition of familiar objects arranged in a scene. In two separate experiments, participants had to learn the position of objects in a scene either by viewing the objects (visual learning) or by feeling the objects (haptic learning). After learning, the experimenter swapped two of seven object positions and the participant's recognition memory for the position of these objects was tested. The test was conducted in either the same modality as learning, or it was conducted in the other modality. Furthermore, participant's ability to recognise scenes across changes in orientation was also tested by rotating the scene by 60° relative to the observer on half the trials. In Experiment 1 we found a significant cost of transfer across modalities. We also found an overall effect of orientation change. In order to control for the possibility that a common verbal code was used to mediate between the modalities participants conducted a verbal interference task between learning and test in Experiment 2. Here we found that the cost of transfer was greater from vision to haptics than vice versa. Furthermore, effects of orientation were no longer present in the transfer conditions. Our finding are discussed with reference to separate perceptual codes for vision and haptics

1. Introduction

In order to recognise objects and scenes in the real world, the human visual system is faced with the problem of maintaining object constancy. Specifically, the problem is that despite changes in the retinal projection of an object (or scene) whenever the observer or objects in the environment move, the object representation must remain constant for recognition. In the past, several mechanisms have been proposed to allow for object constancy within the visual system. However, given that our exploration of our environment is generally multi-sensory, one proposal on how object constancy might be achieved is to assume a multi-modal representation of objects. In this way information that has changed, or is reduced, in one modality can be compensated by information from another modality. Thus, if a cat moves under your chair, it remains a cat because of the sound it makes, or by the way it feels against your legs. However, we know very little about how information from different modalities combines to form a single multi-modal representation of an object. It might be argued that in order for information to be shared across modalities, that these different modalities must deal with encoded information in a similar manner [1]. For example, if the visual system is sensitive to orientation, then we expect that the haptic system is also orientation sensitive, in order for corresponding surfaces of objects to be combined in an object's representation.

Recent research has shown that our visual memory for scenes is sensitive to changes in orientation with respect to the observer (2, 3, 4 and 5). We wondered whether haptic recognition of scenes is also sensitive to changes in orientation. In a previous study on cross-modal object recognition, Newell et al. found that single objects are recognised by both the visual system and the haptic system in a view-dependent manner [6]. Furthermore, this view specific representation promoted more efficient cross modal recognition with changes in orientation. In this study we tested observers' ability to recognise the spatial layout of objects in a scene both uni-modally (either haptically or visually) or across modalities. We predicted that, like single object recognition, the haptic and visual system would be sensitive to changes in orientations of scenes. However, we were also particularly interested in measuring recognition performance when there was a change of modality between learning and test. By investigating cross-modal recognition performance we may be able to provide a clearer understanding of the nature of the encoded information within each modality and how this information is shared in order to recognise scenes of objects.

2. Experiment 1

We tested observers ability to recognise a scene of objects when the objects were learned either within the same modality as learning (i.e. visual - visual or haptic-haptic), or in the other modality (i.e. visual-haptic, haptic-visual). Furthermore, the entire scene of objects was presented either in the same position to the observer as learning (0°) or rotated by 60° at test.

2.1. Method

16 participants from the MPI for Biological Cybernetics, Tübingen, Germany participated in this experiment for pay. The stimulus set of objects included 15 wooden shapes of familiar objects. All objects were positioned on a rotatable platform which had 19 position markers. Each position was equidistant from its neighbours. In each trial, 7 objects were randomly chosen from a full set and placed in random positions on the platform. The objects were placed in random orientations. The experiment was based on a 2x2x2 factorial design using repeated measures. The main factors were the learning modality (visual, haptic), transfer of modality at test (yes, no) and orientation of the scene (0°, 60°). The participant was required to learn the scene and the position of the objects in the scene. Each participant learned each scene either visually, for 10 seconds, or haptically for 1 minute. This timing were established in a pilot study and it allowed for equivalent performance within the visual and haptic modalities. Following learning, the position of two of the 7 objects was exchanged. In the test, the participant had to identify which 2 objects had swapped positions. Testing occurred either in the same modality as learning, or in the other modality. Furthermore, the position of the entire scene was either unchanged with respect to the observer, or rotated by 60°. The participant was unaware which trials included a rotation of the scene and which did not. The original orientation of the swapped objects was maintained after swapping. There was no time limit for responses. Performance was measured in terms of error rates, in that, if the two swapped objects were correctly identified this was recorded as 0 error. If only one object was correctly identified this was recorded as a 50% error and so on.

2.2. Results

The mean percentage error rates for each condition are plotted in Figure 1. A 3 way ANOVA was conducted on the error data using orientation (0°, 60°), learning modality (vision or haptics) and transfer (within or across modalities) as factors. A main effect of orientation was found [$F(1,15)=19.722$, $p<0.001$]. A main effect of transfer was also found [$F(1,15)=9.0548$, $p<0.001$]. There was no effect of learning modality [$F(1,15)=0.42677$, n.s.]. No other effects were found.

2.3. Discussion

In this experiment, we found that both within and cross modal scene perception was sensitive to the orientation of the scene with respect to the observer. Furthermore, recognition performance was worse when participants were required to conduct the task across modalities. Therefore, although cross-modal perception resulted in an overall cost in recognition performance, recognition was still better if the scene remained in the same position with respect to the observer between learning and test. This result suggests that vision and haptics share a common, orientation sensitive code.

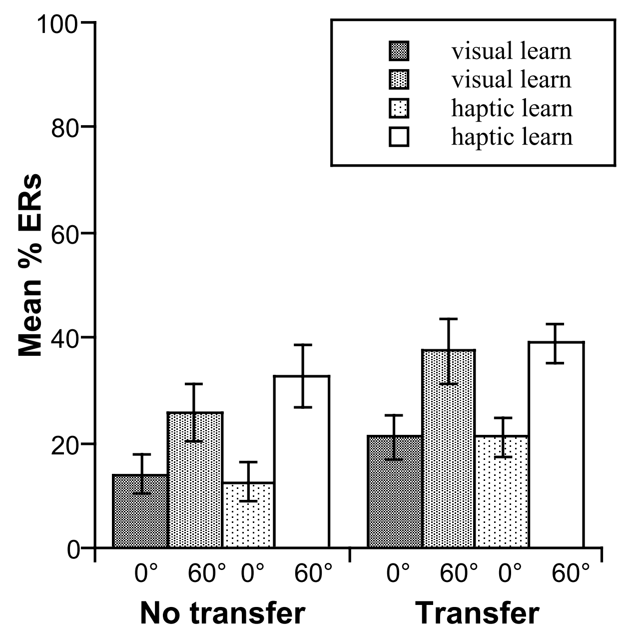

Figure 1: Plot showing mean percentage errors made across all conditions in Experiment 1. The error bars represent the standard error of the mean. 'No transfer' means that there was no change in modality between learning and test. "Transfer' means that a change in modality occurred between learning and test. The legend indicates the learning modality. The test involved either no rotation of the platform (0°) or the platform was rotated (60°).

3. Experiment 2

The purpose of Experiment 2 was to attempt to control for the possibility that participants were using a common verbal code to remember the objects in the scene, and rehearsing this information between learning and test. Here we replicated the procedure in Experiment 1 exactly, but participants were required to perform a verbal interpolation task between learning and test. If the visual or haptic encoding of the objects resulted in an abstract, verbal representation then a verbal interference task would reduce efficient recognition performance within either modality. If a verbal code was not used then we should expect no differences in performance between Experiments 1 and 2.

3.1. Method

16 undergraduate students from the Department of Psychology, Trinity College Dublin participated in this experiment for research credits. See Experiment 1 for a description of the procedure. During the gap between learning and test in each trial, participants were required to generate as many words as possible aloud, that began with a randomly selected letter of the alphabet. This task lasted 20 seconds which was the time needed for the experimenter to change the position of the objects between learning and test. The delay between learning and test was the same for both Experiments 1 and 2.

3.2. Results

The mean percentage error rates for each condition are plotted in Figure 2. A 3 way ANOVA was conducted using orientation (0°, 60°), learning modality (vision or haptics) and transfer (within or across modalities) as factors. A main effect of orientation was found [$F(1,15)=9.502$, $p<0.01$]. An effect of transfer was found [$F(1,15)=5.954$, $p<0.05$]. There was no effect of learning modality [$F(1,15)=2.632$, n.s.]. We found an interaction between transfer and learning [$F(1,15)=14.2913$, $p<0.002$]. A post-hoc Newman-Kuels analysis revealed that errors to the VH condition were significantly higher than errors to either the VV, HH or HV condition ($p<0.001$), with no differences between these three latter conditions. We also found an interaction between transfer and orientation [$F(1,15)=6.8182$, $p<0.02$]. Simple effects analyses revealed a significant effect of orientation within modalities ($p<0.001$) but no effect of orientation across modalities. No other effects were found.

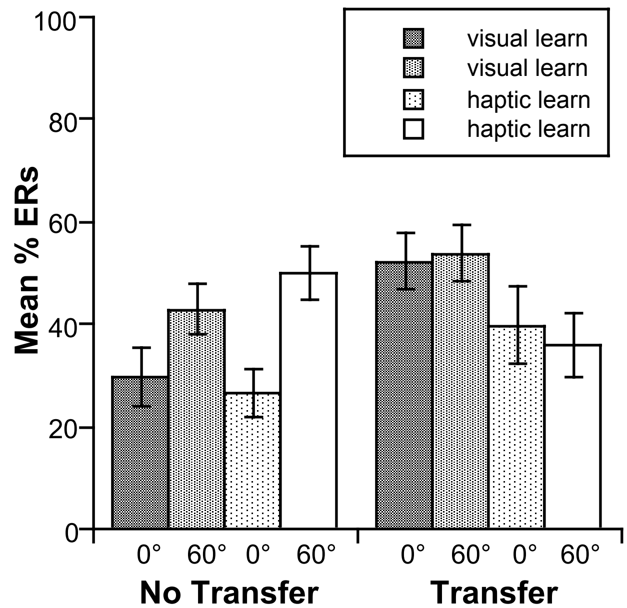

Figure 2. Plot showing mean proportion of errors made (N=16) across all conditions in Experiment 2.

3.3. Discussion

The verbal interpolation task had a clear effect on the performance. First, there were altogether more errors made by the participants in Experiment 2 than in Experiment 1, indicating that verbal interference made the task generally more difficult. Second, in the cross modal condition, we found that verbal interference produced more errors on transferring from visual learning to haptic testing (VH) than on transferring from haptic learning to visual testing (HV). Finally, the effect of orientation remained for the within modalities only, but was not found for cross modal recognition. The disappearance of the orientation effect in the transfer condition may be due to a lack of sensitivity in the experiment because of the relatively high error rate.

The finding that verbal interpolation resulted in more errors in the VH condition than on the HV condition suggests that the visual encoding of the objects was not the same as the haptic encoding. For example, the visual representation might not have been sufficiently shape specific to allow for efficient haptic recognition. These data might be explained by assuming a more abstract, non shape specific representation of the objects from visual sampling. If such an abstract representation had a verbal component to it then we might expect that any verbal interference task would disrupt recognition in the visual modality. Moreover, if haptic encoding involved a shape-specific representation, then this would be unaffected by a verbal secondary task and recognition would be efficient between haptics and vision. Clearly such a proposal requires further testing.

4. Overall conclusions

In general, the perception of scenes of objects is dependent on the orientation of the scene with respect to the observer [see also 1,2,3,and 4]. Both visual and haptic recognition performance was view dependent in that recognition was better when the scene was not rotated than when it was rotated by 60°. Furthermore, cross modal scene recognition was also dependent on orientation. Transferring between modalities from learning to test generally resulted in a cost in recognition performance, such that there were significantly more errors made in the transfer condition than in the no-transfer condition.

When the task involved verbal interference after learning, this resulted in selectively worse recognition performance from visual learning to haptic testing only. Performance of haptic learning to visual recognition was not any worse than within-modality performance across all orientation conditions.

Our findings suggest that a different perceptual code was used for vision and haptics. Under conditions of no interference between learning and test, then these codes can be shared between the modalities although they are highly sensitive to changes in orientation. Under conditions of verbal interference however, the visual code is disrupted such that transfer from vision to the haptic system is not as efficient. A verbal interference task does not affect the haptic code in the same way as it does the visual code, suggesting that there might be a verbal component to the visual representation. This suggestion is consistent with previous research where verbal interference was found to disrupt visual memory [7]. Further experimentation is planned to test for reasons why this effect might occur within the visual system only.

5. References

[1] Easton, R. D., Greene, A. J., & Srinivas, K. (1997). Transfer between vision and haptics: Memory for 2-D patterns and 3-D objects. *Psychonomic Bulletin and Review*, 4 (3), 403-410.

[2] Diwadkar V.A. and McNamara T.P. (1997). Viewpoint dependence in scene recognition. *Psychological Science*, 8 (4): 302-307

[3] Wang R.X.F. and Simons D.J. (1999). Active and passive scene recognition across views. *Cognition*, 70 (2): 191-210.

[4] Simons D.J. and Wang R.F. (1998). Perceiving real-world viewpoint changes. *Psychological Science*, 9 (4): 315-320.

[5] Christou C.G. and Bulthoff H.H. (1999). View dependence in scene recognition after active learning. *Memory & Cognition*, 27 (6): 996-1007.

[6] Newell F.N., Ernst M.O., Tjan B.S., and Bulthoff H.H. (2001). Viewpoint dependence in visual and haptic object recognition. *Psychological Science*, 12 (1): 37-42.

[7] Loftus, E.F., Miller, D.G. and Burns, H.J. (1978). Semantic intergration of verbal memory into a visual memory. *Journal of Experimental Psychology: Human Learning and Memory,* 4, 19-31.