The Role of 3D Immersive Environments in Assessment and Training Spatial Skills

Maria Kozhevnikov
National University of Singapore
Singapore
& Harvard Medical School
Cambridge, MA
psymaria@nus.edu.sg

ABSTRACT

Recently, increasingly realistic 3D visual displays have been designed to serve as new, more ecologically valid alternatives to conventional 2D visual displays. However, research has thus far provided inconsistent evidence regarding the effectiveness of 3D displays in facilitating training and task performance. We were interested in the contribution of "immersion" to individuals' ability to spatially transform 3D images; we compared subjects' performance on spatial transformation tasks in traditional 2D non-immersive (2DNI), 3D non-immersive (3DNI: stereo-glasses), and 3D immersive (3DI: head mounted display with position tracking) environments. Twenty-five participants completed a number of spatial transformation tasks, in which they were asked either to mentally rotate 3D objects along different planes (mental rotation task) or mentally rotate their imagined selves within the environment (perspective-taking task). While the patterns of subjects' responses were not significantly different between the 2DNI and 3DNI environments, we found a unique pattern of responses in the 3DI environment. Our findings suggest that 2DNI and 3DNI environments might encourage the use of more "artificial" encoding strategies, in which the 3D images are encoded with respect to a scene-based frame of reference (i.e. the computer screen). On the other hand, 3DI environments can provide the necessary feedback for an individual to use the same strategy and egocentric spatial frame of reference that he/she would use in a real-world situation. Overall, the results of this study suggest that immersivity might be one of the most important aspects to be considered for assessment and training in domains that rely on visual-spatial performance and require high spatial transformation skills (e.g., robotics, navigation, medical surgery).

ABOUT THE AUTHORS

Maria Kozhevnikov is an Associate Professor at the National University of Singapore and also a Visiting Associate Professor at Harvard Medical School. Maria Kozhevnikov received her Ph.D from Technion (Israel) jointly with UC Santa Barbara. Prior to joining the National University of Singapore, she has held faculty positions at Rutgers University (NJ) and George Mason University (VA). In the past two years, she has also served as a Program Director at the US National Science Foundation. Her research interests focus on examining the neural mechanisms of visual/spatial processing as well as in exploring ways to train visual/spatial skills using three-dimensional immersive technologies.

The Role of 3D Immersive Environments in Assessment and Training Spatial Skills

Maria Kozhevnikov
National University of Singapore
Singapore
& Harvard Medical School
Charlestown, MA
psymaria@nus.edu.sg

INTRODUCTION

The ability to translate information about one's environment from 2-dimensional into 3-dimensional form and the ability to perform complex 3D transformations are important for many military occupations such as air traffic controllers, space craft pilots and satellite imagery analysts. Visual-spatial cognition research has already extensively incorporated immersive 3D virtual reality (VR) technology (e.g., Chance et al., 1998; Darken & Sibert, 1996; Klatzky et al., 1998; Kozhevnikov, 2008; Richardson, Montello, & Hegarty, 1999) due to the ease with which one can both create a complex environment for participants to explore, and record their behavior (Loomis et al., 1999, Peruch & Gaunet, 1998). However, regarding assessment and training applications, although these 3D environments are both more appealing to the user and richer in spatial information, research thus far has not reached a strong conclusion regarding the effectiveness of 3D environments for promoting visual-spatial learning and task performance (e.g., Van Orden & Broyles 2000). For some tasks, such as collision avoidance, 3D displays have been found to facilitate learning and performance better than 2D displays (e.g., Van Orden & Broyles, 2000), but other researchers have reported that 3D displays as less efficient than 2D displays (e.g., Alexander & Wickens, 2005; Hollands et al., 1998). Currently, too little is known about the cognitive processes that underlie learning and training in 3D vs. 2D environments to fully justify using 3D immersive virtual reality displays. The focus of the current research is to understand how individuals process visual-spatial information in 3D immersive environments vs. 3D non-immersive (stereo-glasses) and conventional 2D non-immersive visual displays and how complex 3D immersive technology can facilitate assessment and training of spatial skills.

RESEARCH BACKGROUND

Neuroscience data suggest that even though there are many shared neural mechanisms behind processing information in 3D and 2D forms (primary visual cortex and some parts of ventral areas analyzing shapes), there are some differences, and especially, different roles for object-properties processing (ventral) and spatial-relation processing (dorsal) brain pathways. For instance, a part of the dorsal system that processes large-scale 3D spatial information, the caudal intraparietal sulcus, was found to respond to 3D surface orientation defined by various depth cues (stereoscopic, perspective and texture-based: Tsutsui & Taira, 2001 Tsutsui, Sakata, Naganuma, & Taira, 2002). In contrast, the ventral system seems to contribute largely to processing 3D shapes of specific small-sized objects (Connor, 2002). Overall, these studies suggest that while the ventral stream mostly processes smaller-scale information about the 3D shapes of individual objects, the dorsal system is responsible for processing large-scale 3D spatial information such as orientation and location.

The dorsal stream additionally deals with non-visual information (such as motion or vestibular cues) necessary for generating a 3D image of a large-scale environment, as well as for movement in 3D space, such as locomotion and navigation (e.g., Loomis & Beall, 1998). Visually controlled locomotion is often accomplished with the aid of supplementary non-visual information about the person's motion, such as signals provided by the vestibular and somatosensory systems that provide the operator of a vehicle with information about vehicle velocity and acceleration (Gillingham & Wolfe, 1986) and, in the case of flying at night, provide information about aircraft orientation (Loomis & Beall, 1998). Recent evidence (see Kozhevnikov et al, 2008) also suggests that spatial navigation may best be trained via paradigms that provide vestibular and proprioceptive feedback (e.g., immersive virtual environments with motion tracking or driving simulators mounted on rotating platforms).

Furthermore, there is evidence that involvement of vestibular and proprioceptive cues are crucial for performance on *egocentric* spatial transformations that require imagining taking different orientation in space (e.g., Easton & Sholl, 1995; Rieser, 1989). This is in contrast to *allocentric* spatial transformations, which require the mental manipulation of objects from a stationary point of view. Egocentric transformations, such as imagining a different orientation (perspective) involves movement of one’s egocentric frame of reference, which encodes objects’ locations with respect to the front/back, left/right and up/down axes of the observer’s body. The encoding of visual-spatial stimuli in relation to egocentric (body-centered) spatial frames of reference has been shown to be critical for successful performance in many real-world tasks, such as real-world navigation on land, air or water, scene encoding, remote operation, medical and dental surgery, weapon deployment, etc. (Kozhevnikov et al., 2007). In contrast, allocentric manipulation of objects or arrays of objects (e.g., mental rotation of cubes or other geometrical figures) involves imagining movement relative to an object-based frame of reference, which specifies the location of one object (or parts) with respect to other objects. Allocentric transformations are important in performance on small-scale spatial tasks such as manipulating graphs and diagrams, and success in mathematics and physical sciences. Although these might rely on motor-planning strategies, they do not require direct vestibular or proprioceptive feedback from the environment. Given that 2D and 3DNI displays do not provide the same level of non-visual cues for 3D image generation as would be experienced in a real-world environment, we suggest that 3DI environments might be particular efficient for assessment and training of large-scale spatial egocentric tasks.

METHOD AND RESULTS

Twenty-five undergraduate psychology students were administered the *Mental Rotation Task* and *Perspective-Taking Task* (described in the sections below) in each of the three different environments: 2D non-immersive (2DNI), 3D non-immersive (3DNI) and 3D immersive (3DI) environments. Environment order was counterbalanced across participants.

In the 2D environment, scenes and objects were presented to the participant on a standard computer screen (Figure 1c). In the 3DNI environment, scenes were presented to the participant on a computer screen Stereoscopic depth is provided by means of anaglyphic glasses (Figure 1b). In the 3DI virtual environment, scenes were presented through a stereo head-mounted display (HMD) with a visual field of 150 degrees (Figure 1a), which was used in conjunction with a computer a motion tracker.

The 3DI environments are interactive in that when the user turns his/her head, the image adjusts correspondingly. The position tracking system permits full 3D optical tracking of up to 4 wireless targets over large areas (more than 10 x 10 meters) with sub-millimeter precision. In conjunction with a gyroscopic orientation sensor, this position tracking system supports the real-time picture-to-position simulation in virtual reality, in which any movement of the user’s head immediately causes a corresponding change of the picture he/she sees in the head-mounted display.

A. 3D Immersive **B. 3D non-immersive** **C. 2D non-immersive**

3DI, which includes HMD with position tracking

3DNI with anaglyphic glasses to present a stereo picture of three-dimensional spatial forms

2D monocular viewing environment

Figure 1. Different Types of Testing Environments

\

Perspective-Taking Task

We designed a perspective-taking ability (PTA) test (Kozhevnikov & Hegarty, 2001; Kozhevnikov et al., 2006) as a measure of egocentric spatial transformation ability and compared subjects' performance on this test in 3DI, traditional 2DNI, and 3DNI (stereoscopic glasses) environments. In the 3DI condition, the test was presented via HMD (Figure 2), while in the traditional 2D and 3D non-immersive environments, spatial scenes were presented to the participant on a standard computer monitor.

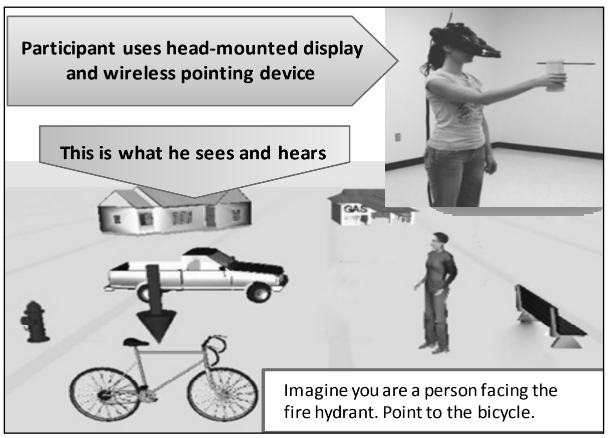

Figure 2. PTA Task administered in 3DI

On each trial, the participant was placed in a location inside the scene in 3DI environment, or shown the scene exocentrically in the 2D and 3D non-immersive environments. The participants were explicitly instructed to imagine taking the perspective of an avatar located within the array of objects, and then to point to a specific target from the new imagined perspective, using a pointing device. There were 16 practice and 54 test trials. Participants were not allowed to move their body or head during the test trials. The responses (pointing direction) were given by joystick in 2DNI and 3DNI environments and by a pointing device in the 3DI condition.

Results: Comparative analysis of subjects' responses in the three environments revealed that the 3DI environment best encourages the use of egocentric spatial encoding strategies. Specifically, while the participants were as accurate on performing the PTA task in 3DI as in 2DNI or 3DNI environments (see Figure 3), their errors were qualitatively different.

Particularly, in the 3DI environment, most errors were systematically due to confusion between "right-left" and "back-front" coding with respect to the participants' bodies, indicating that they indeed were relying more on body-centered frame of reference. In contrast, in the 2D and 3D desktop non-immersive environments, participants made more "allocentric" errors characterized by over-rotating or under-rotating the scene (see Figure 4).

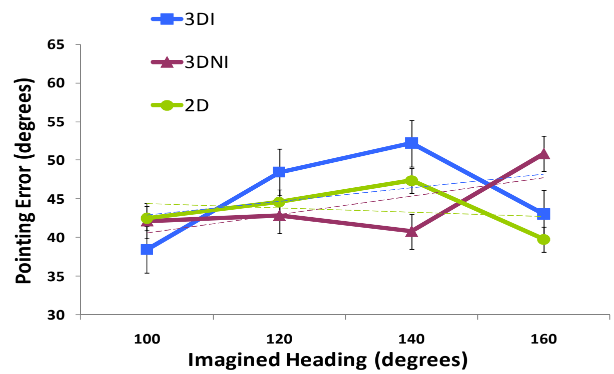

Figure 3. Pointing error on the Perspective-Taking Task in three different environments

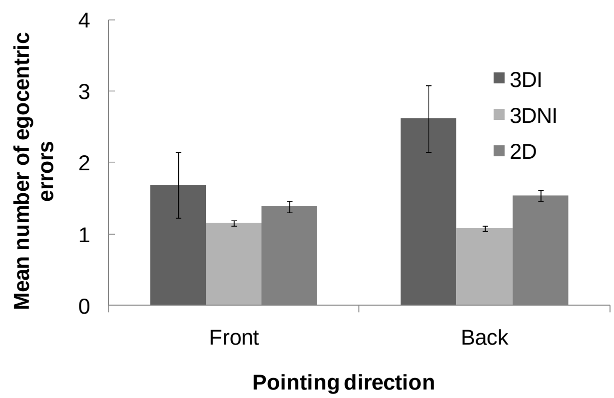

Figure 4. Mean number of egocentric error on the Perspective-Taking Task

Furthermore, our results showed that the 3DI condition of the PTA test had a significantly stronger training effect than the two other environments. The results revealed that the 3DI PTA test facilitated an increase in performance by 200% (i.e., the rate of error reduction), compared to the non-immersive 2D version of the test.

Mental Rotation Task

For the Mental Rotation Task, each participant completed a computerized adaptation of Shepard &

Metzler's (1971) Mental Rotation Test (MRT) in three different testing environments: 3DI, 3DNI, and 2DNI.

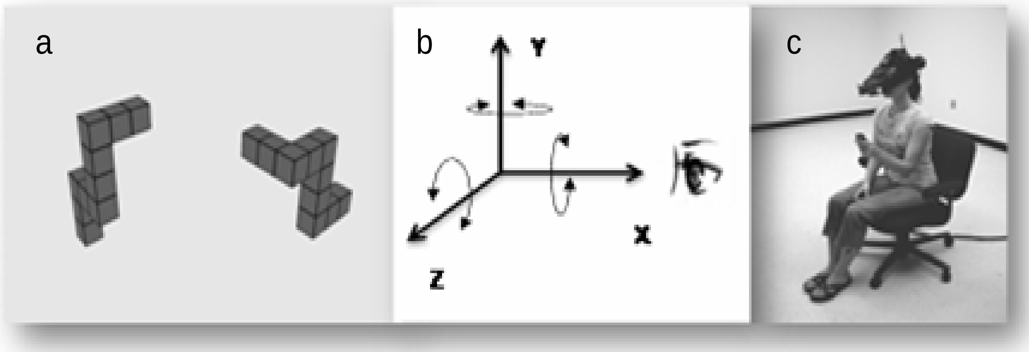

Figure 5. Mental Rotation Test: a) Example item, which includes two 3D shapes that have to be mentally rotated into alignment, b) Three principle axes of rotation, c) Test in 3DI environment, which includes HMD with position tracking.

There were 72 randomly ordered trials, in each of which participants viewed two 3D figures composed of cubes, one of which was rotated relative to the position of the other (see Figure 5a). Subjects were to imagine rotating one figure to determine whether or not it was identical to the other figure, and to indicate whether their response by pressing the the left (identical) or right (different) button on a remote control device. Participants were asked to respond as quickly and as accurately as possible. The figures were rotated around 3 spatial coordinate system axes including the picture (X), vertical depth (Y), and horizontal depth (Z) axes (Figure 5b).

Results: While the patterns of subjects' responses were not significantly different between the 2DNI and 3DNI desktop environments, we found a unique pattern of responses in the 3DI environment, suggesting that immersion triggered significantly greater use of an egocentric frame of reference (specifically a retinocentric frame) than the non-immersive environments. In particular, in 3DI, the rate of rotation around the depth axis (around Z axis) was significantly slower than the rate of rotation in the picture plane (around the X or Y axes) (see Figure 5).

This suggests that the subjects were in fact rotating 2D retina-based representations, since rotation in depth is more difficult than in the picture plane, due to foreshortening and occlusion. However, in 2D and 3D non-immersive environments, the rates of mental rotation around the X and Z axes were identical. This suggests that non-immersive displays encourage the use of more "artificial" encoding and transformation strategies, where the objects' components are encoded in terms of "vertical" and "horizontal" relations with regard to their own internal structure, as well as to the frame of the computer screen.

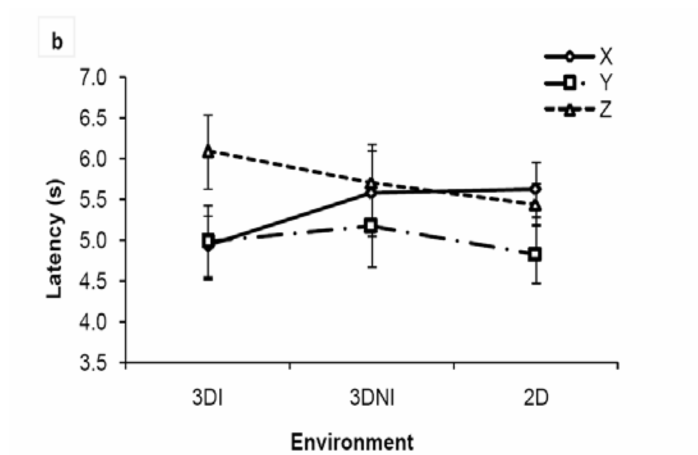

Figure 6. Latency as a function on axis of rotation and testing environment

CONCLUSIONS

Overall, the behavioral pattern of results was very similar for 2DNI and 3DNI environments, while response patterns in 3DI were unique. First, the findings suggest that immersive environments are different from 2D and 3D non-immersive environments, and that immersion is necessary to provide adequate information for building the spatial reference frame crucial for egocentric encoding. The fact that there was equivalent performance in 2D and 3D non-immersive environments suggests that the human visual system can extract the same information from binocular and monocular cues to the same degree of success. In contrast, the design of immersive environments might help to encourage encoding and transformation of an image with respect to the egocentric frame of reference, better approximating a real environment, as well as providing non-visual cues that are important for large-scale egocentric tasks (robotics, medical surgery, navigation). Thus, when designing an environment for assessing performance on large-scale egocentric spatial tasks, one might consider the use of 3DI environments as the most beneficial. However, for tasks that are more related to color and object shape processing (and depend more on the ventral pathway of the brain), such as recognizing objects in limited visibility, 3DI environments might not be necessary, and most likely would elicit similar patterns of responses as 3DNI and 2DNI environments.

Second, 3DI environments might be especially beneficial for training for real-world spatial tasks. Reliance on egocentric spatial encoding in immersive, but not non-immersive, environments would also explain why the results of the training studies show no transfer from training in 2D environments to performance in 3DI or to the real world. For instance, Pausch et al. (1997) reported that practice with conventional 2D displays in visual search tasks in fact impairs later performance on a similar task in a 3DI environment, but not vice versa. This implies that using desktop graphics to train users for real world visual tasks might not be effective, and may actually be counterproductive. The reason for this effect, we suggest, is that the encoding of spatial relations and the cognitive strategies applied to perform visual-spatial transformations in non-immersive versus immersive environments are different. We also suggest that 3DI environments with a variety of simulated 3D stimuli will provide the most efficient environment for training visual-spatial skills that will generalize and transfer to real-world tasks.

REFERENCES

Alexander, A.L. and Wickens, C.D. (2005). 3D Navigation and Integrated Hazard Display in Advanced Avionics: Performance, Situation Awareness, and Workload. Institute of Aviation, Aviation Human Factors Division, Savoy, IL, USA.

Chance, S, Gaunet, F, Beall, A, and Loomis, J. (1998). Locomotion Mode Affects the Updating of Objects Encountered During Travel: The Contribution of Vestibular and Proprioceptive Inputs to Path Integration, *Presence, 7*, 168-178.

Connor, C. E. (2002). Reconstructing a 3D world. *Science, 298*, 376-377.

Darken, R, and Sibert, J. (1996). Navigating large virtual spaces, *International Journal of Human-Computer Interaction, 8*, 49-71.

Easton, R. D., & Sholl, M. J. (1995). Object-array structure, frames of reference, and retrieval of spatial knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 483-500.

Gillingham, K. and Wolfe, J. (1986). Spatial Orientation in Flight (Technical Report USAFSAM-TR-85-31), Brooks Air Force Base, TX: USAF School of Aerospace Medicine, Aerospace Medical Division.

Hollands, J. G., Pierce, B. J., & Magee, L. E. (1998). Displaying information in two and three dimensions. *International Journal of Cognitive Ergonomics, 2*(4), 307-320.

Klatzky, R, Loomis, J, Beall, A, Chance, S, and Golledge, R. (1998). Spatial updating of self-position and orientation during real, imagined, and virtual locomotion, *Psychological Science, 9*, 293-298.

Kozhevnikov, M. (2008). The Role of Immersive 3-D Environments in Mental Rotation. Paper was be presented at the Human Factors and Ergonomics Society's 52nd Annual Meeting, New York, NY, Sept. 22-26.

Kozhevnikov, M., & Hegarty, M. (2001). A dissociation between object manipulation spatial ability and spatial orientation ability. *Memory & Cognition, 29*(5), 745-756.

Kozhevnikov, M., Motes, M. A., & Hegarty, M. (2007). Spatial visualization in physics problem solving. *Cognitive Science: A Multidisciplinary Journal, 31*, 549-579.

Kozhevnikov, M., Motes, M. A., Rasch, B., & Blajenkova, O. (2006). Perspective-Taking vs. Mental Rotation Transformations and How They Predict Spatial Navigation Performance. *Applied Cognitive Psychology, 20*, 397-417.

Loomis, J, Blascovich, J, and Beall, A. (1999). Immersive virtual environment technology as a basic research tool in psychology. *Behavior Research Methods, Instruments & Computers, 31*, 557-564.

Loomis, J. and Beall, A. (1998). Visually controlled locomotion: Its dependence on optic flow, three-dimensional space perception, and cognition, *Ecological Psychology, 10*, 271-285.

Pausch, R., Proffitt, D., & Williams, G. (1997). Quantifying immersion in virtual reality. In: *Proceedings of the 24th annual conference on Computer graphics and interactive techniques* (pp. 13-18). ACM Press/Addison-Wesley Publishing Co.

Péruch, P, and Gaunet, F. (1998). Virtual environments as a promising tool for investigating human spatial cognition, *Cahiers de Psychologie Cognitive/Current Psychology of Cognition, 17*, 881-899.

Richardson, A, Montello, D, and Hegarty, M. (1999). Spatial knowledge acquisition from maps and from navigation in real and virtual environments, *Memory & Cognition, 27*, 741-750.

Rieser, J: 1989, Access to knowledge of spatial structure at novel points of observation, *Journal of Experimental*

Psychology: Learning, Memory, and Cognition, *15*, 1157-1165.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*(3972), 701-703.

Tsutsui, K., Sakata, H., Naganuma, T., & Taira, M. (2002). Neural correlates for perception of 3D surface orientation from texture gradient. *Science, 298*, 409.

Tsutsui, K., & Taira, M. (2001). Monkeys perceive surface orientation from pictorial cues. *Japanese Journal of Psychonomic Science*, *20*, 57-58.

Van Orden, K, and Broyles, J. (2000). Visuospatial task performance as a function of two- and three-dimensional display presentation techniques, *Displays*, *21*, 17-24.