Scene Perception, Gaze Behavior, and Perceptual Learning in Virtual Environments

GUILLAUME ALBERT, B.SC.,1 PATRICE RENAUD, Ph.D.,1,2 SYLVAIN CHARTIER, Ph.D.,1,2 LISE RENAUD, Ph.D.,3 LOUISE SAUVÉ, Ph.D.,4 and STÉPHANE BOUCHARD, Ph.D.1

ABSTRACT

More and more immersive environments are developed to provide support for learning or training purposes. Ecological validity of such environments is usually based on learning performance comparisons between virtual environments and their genuine counterparts. Little is known about learning processes occurring in immersive environments. A new technique is proposed for testing perceptual learning during virtual immersion. This methodology relies upon eye-tracking technologies to analyze gaze behavior recorded in relation to virtual objects’ features and tasks’ requirements. It is proposed that perceptual learning mechanisms engaged could be detected through eye movements. In this study, nine subjects performed perceptual learning tasks in virtual immersion. Results obtained indicated that perceptual learning influences gaze behavior dynamics. More precisely, analysis revealed that fixation number and variability in fixation duration varied with perceptual learning level. Such findings could contribute in shedding light on learning mechanisms as well as providing additional support for validating virtual learning environments.

INTRODUCTION

The use of immersive technologies in research context enables a close control of variable setting and experimental situation altogether. Furthermore, immersive technologies are especially fit to merge spatial information into a visual environment close to reality. Immersive environments have been used in various field such as health,1 military,2 firefighter training,3 and clinical psychology.4,6 The bulk of these studies aims to design or validate virtual environments for treatment or resource training. In addition to being a promising tool for training purposes, immersive environments can also supply much information about motor behavior while virtual environments are interacted with. Assessing motor behavior in virtual reality may indeed provide additional information on how a subject assimilate or learn critical components of a new environment.

Perceptual learning

Perception involves the extraction of structures from the surrounding by means of the senses,7 and perceptual learning implies changes in the way an organism extracts information from its environment.8 These changes, induced by practice and experience, promote adaptation by tuning organisms’ perceptual systems to optimize extraction from
sensorial input. Perceptual learning differs from higher-level cognitive learning in the way changes are occurring at early stages of information processing.\(^9\) These adaptations carry profound implications due to their primary role in the control of behavior. Evidence suggests an association between these changes and modifications in the neural circuitry.\(^8\) Processes associated to perceptual learning fall in a twofold categorization: discovery processes and fluency processes. Discovery processes are concerned with the selection of information while modification in the ease of extraction is an expression of fluency. Stimulus imprinting is a perceptual learning process associated mainly with the second category. Stimulus imprinting is characterized by the development of perceptual receptors specialized for stimuli detection.\(^10\) Those internalized detectors are shaped by impinging stimuli inducing perceptual adaptation, which in return increases speed and accuracy for stimuli detection. Moreover, when stimulus exposition is repeated or prolonged, small deviation from the input stimulus is more easily detected.\(^10\) Imprinting on stimulus may occur for an entire stimulus or for selected parts from a stimulus. The latter may appear when selected parts from stimulus are isolated from others for varying independently, eliciting “compositional” processing. Another form of stimulus imprinting is topological imprinting. Topological imprinting occurs when position and orientation of visual patterns within space are learned as a result of training.\(^9\) In this case, internalized detectors are shaped by incident environmental variations in Cartesian and Eulerian parameters setting the surrounding layout. Indeed, adaptation results in the extraction of regularities distinctive to these spatial parameters.

In a visual task, this perceptual adaptation could be expressed through eye movements when learning would have taken place. The present research was designed to investigate if perceptual learning mechanisms engaged in visual searching tasks could be detected through eye movements.

**Eye movements**

Visual processing involves selecting a limited area in the visual field. This limitation is due to physiological characteristics of the fovea, a small region of the retina. The high acuity of this region forces its alignment with the selected area in order to ensure optimal processing. This alignment, called **foveation**, is performed approximately 230,000 times per day.\(^11\) Foveation implies the launching of a saccade on a selected region, followed by a fixation, a brief period of stillness where visual processing is allowed. Saccades are ballistic eye movements occurring very rapidly (100–300 msec) during which visual processing is upheld. Fixations are momentary lapse (150–600 msec) where the eyeball is stationary and visual processing enabled. Eye movements are realized given interplay of three pairs of agonist/antagonist ocular muscles. These muscles circumscribe eye movements within an abstract bidimensional plane on which eyeball positioning is determined by perceptual and higher demands.

**Eye movement research**

Numerous research in reading have stressed eye movement measurements as potential behavioral indices expressing ongoing visual and cognitive processing.\(^12\) Early studies in scene perception revealed organized fixation patterns clustering over informative scene regions.\(^13\) Moreover, these patterns varied according to task instructions.\(^14\) Subsequent studies stressed the influence of semantic and stimulus factors on fixation sites. Studies indicated furthermore that global scene gist and layout information acquired upon first fixation is used afterwards to cue target location in scene.\(^15,16\) While influence of stimulus salience on visual attention is obvious, attempts to predict eye fixation pattern relying solely on saliency maps are inconclusive.\(^17\) Saliency maps are representation of eye movements based on specific stimulus features. Yet, in scene perception ocular pattern conveys information about visual attention and cognitive processing.\(^15,16\)

Eye monitoring during execution of ordinary task also unveils the close coupling of visual attention and cognition. In that respect, two studies considered gaze behavior in ordinary activities such as making sandwiches\(^18\) and tea-making.\(^19\) Results strongly suggest that eye movements are embedded in general motor responses in order to provide information supporting each action during execution.\(^18,19\) Specific eye patterns have also been observed in more dynamically complex tasks such as driving\(^20\) and table tennis.\(^21\) In driving context, eye patterns have been found to vary accordingly with participants’ experience influencing visual search strategies. For example, in a real driving task, novice drivers tend to have longer fixation durations,\(^22\) and when context becomes more demanding, experienced drivers tend to expand their visual search array, producing shorter fixation durations.\(^22\) Hence, in a visual task, eye movements can reveal ease in information extraction that may contribute to differentiate experience
levels. Eye movement characteristics are also considered as a potential biological marker for attention-deficit/hyperactivity disorder (ADHD) diagnosis. For instance, in a simple target fixation task, saccades made by ADHD children were larger than those made by normal children. Authors indicated that saccades interrupting fixation were attributable to a primary failure in maintaining visual fixation. In this case, eye patterns could provide useful information for detecting learning difficulties.

Previous experiments

Studies in scene perception have addressed the influence of top-down and bottom-up effects on gaze behavior. However, to our knowledge interaction between perceptual learning and gaze behavior has been sparsely observed. Research on perceptual learning unveils results based mostly on time reaction analysis. Fluency in visual information extraction attributable to perceptual learning has not been directly observed. Also, few studies in scene perception have used virtual immersion technology as a research tool. This new trend in technology may provide a more realistic approach to scene perception enhancing ecological validity. The current study addresses these issues by investigating the effect of perceptual learning on eye movement behavior in virtual environment.

METHODS

Participants

Nine volunteers (six women, three men) participated in this experiment. Subjects ranged in age from 22 to 29 years. All subjects claimed to be free of serious vision problems and gave written informed consent according to institutional guidelines. The experimental procedure received local ethical review and approval.

Apparatus

The stimuli were displayed through a Virtual Research V8 binocular model head mounted display (HMD) with an image resolution of 640 × 480 pixels, a contrast ratio of 200:1 and a field of view of 60° diagonal. Display luminance was generally low and the room was not illuminated in order to avoid measure interference with the eye tracking device. The immersive system was supported by a 2.8-GHz computer equipped with an NVIDIA GeForce4 Ti 4200 video card (128 DDR). Eye tracking was done with an ASL 504 series video eye tracker integrated in the optics of the HMD and used to monitor position of the right eye (Fig. 1). This system relies on the corneal reflection of an infra-red source that is measured relative to the pupil center location. These particular corneal reflections, known as the first Purkinje images can be located with video-based eye trackers collecting infra-red reflections. A single eye tracker returns 2 DOF, i.e., variations in an x and y plane. Accuracy is of 0.5 degree. Data files generated with ASL software contained information on time and position relative to eye movements. Moreover, software analysis was used to extract information relative to fixation number and fixation duration.

Head position was monitored using an InterSense IS-900 tracking device. This tracking system renders the 6 degree-of-freedom (position : x, y and z / angular rotation : yaw, pitch, roll) of the sensor mounted on the HMD. The rendering is based on a hybrid technology of inertia and ultrasonic tracking. The accuracy of this system is of 3 mm RMS in translations, and 0.15 degree RMS in rotations.

Stimuli

Four virtual environments were created using 3D Studio Max software. VRML versions of these environments were used with our virtual reality custom software. The first step in creating these environments was to generate two learning environments for the low-level learning and high-level learning conditions (Fig. 2). These original environments would be subsequently modified to provide detection tasks. Luminosity was controlled to avoid interference with eye tracking equipment. Each original environment contained approximately...
35 objects. Objects in these environments differed from another on the basis of color, shape, and layout providing two unique learning environments.

Following from the original environments, two detection environments were created (Fig. 3). Each detection environment preserved the same general layout than its original counterpart but received transformations. Both environments displayed 24 transformations based on color modification, shape modification and object suppression/addition equally distributed between environments.

A fifth environment was generated for the control condition (Fig. 4). This environment displayed only walls, floor, and ceiling that were identical to experimental environments.

Design
This experiment used a repeated measures design. The within-subjects variable was gaze behavior characteristics manifested relatively to strength of perceptual learning experienced.

Procedures
Participants were brought in the experimental room where they received directives on how to install the equipment properly. Eye movement calibration took place when the participant obtained optimal viewing conditions. The calibration procedure varied in time from 5 to 20 min, depending on
the systems’ capacity to acquire gaze location. Afterwards, participant had a period to familiarize with the virtual reality system. Before every exposition, recorded directives were delivered through headphones. During the exposition white noise was sent through headphones in order to promote sensorial isolation along the task. Also, participants were informed that they could move their head freely in order to observe environments but had to remain seated. First, participants were exposed to a control environment for a two minutes period (Fig. 4). Directives given stated that the participant needed to stay calm all along and could explore the environment if wanted. No further directives were given. Following the control condition, the low-learning experimental block was initiated. This block contained two trials. In the first trial (low-level learning condition), the observers had 20 sec to learn and memorize the environment’s layout (Fig. 2), that is objects’ color, position, orientation, and form. Immediately after this first learning trial, participants had a 2-min, trial during which they had to detect any discrepancies between the layout they were observing (Fig. 3) and the one observed during the learning trial. Following detection trial, paper printed version of the simulated layout was presented upon which the subjects had to mark discrepancies detected during the detection trial. As stated earlier, the detection trial environments were designed like the learning trial environments except for 24 light modifications, those to be detected. At the end of the second block, participants removed the HMD and were asked to indicate discrepancies they observed on a depiction of the last seen environment. The rate of true identification served as measure of learning performance.

RESULTS

Results are presented in two subsections. The first section addressed the learning effect obtained across experimental conditions. The second section focused on eye movement analysis. In the latter section results were compared across control, low-level learning and high-level learning conditions. The data from one participant was removed from the following analyses because he failed to comply with the instructions that were given in not looking around the whole scene as required.

Learning effect

For learning effect (detection task), as expected the average number of discrepancies detected in the low-level learning condition was lower (3.1 detections) than in the high-level learning condition (9.6 detections). A GLM repeated measures analysis with pairwise comparisons revealed a significant experimental effect, based on detection performance ($F(1,7) = 53.773; p < 0.001$) attributable to perceptual learning. This effect ensured a valid behavioral basis for subsequent eye movement analysis regarding perceptual learning.

Eye movement

For eye movement data analysis, Table 1 shows average number of fixations for control, low-level learning and high level learning conditions. A repeated measures multivariate analysis of covariance (MANCOVA), with mean control fixation duration as a covariate, was performed on the number of fixations for each condition. There was a significant multivariate within-subject effect ($F(4,22) = 235.949; p < 0.0001$) with respect to control, low-level and high-level learning conditions. Pairwise comparisons (Table 2) revealed that subjects produced significantly more fixations in low-level condition than in high-level learning condition ($p < 0.05$). Significant differences, with regard to control condition, were observed for both high-level ($p < 0.05$) and low-level ($p < 0.01$) learning conditions. In low-level condition, subjects’
gaze behavior necessitated more visual anchorage to extract information needed. Hence, during discrepancy detection task, the number of fixations seemed to be modulated with respect to the learning effect due to perceptual learning.

Moreover, additional results unveiled differences regarding mean variability in fixation duration as indexed individually by the standard deviation of mean fixation duration recorded across conditions for each subject (Table 1). Results revealed (Table 3) that mean variability in fixation duration is lower in low-level condition than in high-level condition \( (p < 0.05) \). Again, significant differences, with regard to control condition, were reported for both high-level \( (p < 0.001) \) and low-level \( (p < 0.001) \) learning conditions. Such dissimilarity was attributable to lack of active visual search within control condition. Nevertheless, mean variability in fixation duration appeared to increase in relation with perceptual learning.

**DISCUSSION**

The aim of the experiment was to observe ocular movements as indices of perceptual learning. In this experiment, perceptual learning was accomplished through increased knowledge of environmental layout from virtual stimuli exposure. Focused has been given on this perceptual learning process known as topological imprinting which occurs when spatial attributes and objects’ relation within the space are learned as a result of experience.\(^9\) Results obtained from this preliminary study shed light upon perceptual learning as expressed by gazing behavior. Analysis revealed a decrease in fixation number during a detection task when participants had to detect discrepancies between slightly different virtual environments. This decrease was observed with longer stimuli exposure time, hence when perceptual learning was higher. Increasing perceptual learning seems to affect the perceptual system by promoting an optimal scanpath strategy. This interaction may be accounted for a propensity to extract invariants within the immersive environment due to a prolonged exposition. Therefore, the ease to extract invariants can explain a decrease in fixation number needed to accomplish the task. These findings are consequent with the results of Crundall and Underwood,\(^22\) where gazing properties are modulated with respect to driving experience. In their study, driving experience induced an ease in the picking up of information necessary to fulfill the driving task. In a real driving task, Crundall and Underwood\(^22\) observed that experienced drivers’ fixations across horizontal meridian were more widespread than for novice drivers when facing

<table>
<thead>
<tr>
<th>Measures</th>
<th>Control</th>
<th>Low-level learning</th>
<th>High-level learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of fixation</td>
<td>200.245</td>
<td>344.472</td>
<td>313.945</td>
</tr>
<tr>
<td>Mean variability (SD) fixation duration</td>
<td>2.224</td>
<td>0.253</td>
<td>0.310</td>
</tr>
</tbody>
</table>

**TABLE 2. PAIRWISE COMPARISONS OF FIXATION NUMBER ON CONTROL, LOW-LEVEL, AND HIGH-LEVEL LEARNING CONDITIONS**

<table>
<thead>
<tr>
<th>Measures</th>
<th>(I) Factor</th>
<th>(J) Factor</th>
<th>Mean Difference (I – J)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of fixation</td>
<td>Control</td>
<td>Low-level</td>
<td>(-144.181^{**})</td>
</tr>
<tr>
<td></td>
<td></td>
<td>High-level</td>
<td>(-113.699^{*})</td>
</tr>
<tr>
<td></td>
<td>Low-level</td>
<td>Control</td>
<td>(144.181^{**})</td>
</tr>
<tr>
<td></td>
<td></td>
<td>High-level</td>
<td>(30.482^{*})</td>
</tr>
<tr>
<td></td>
<td>High-level</td>
<td>Control</td>
<td>(113.699^{*})</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Low-level</td>
<td>(-30.482^{*})</td>
</tr>
</tbody>
</table>

\(^{*}p < 0.05.\)
\(^{**}p < 0.01.\)

**TABLE 3. PAIRWISE COMPARISONS OF MEAN VARIABILITY (SD) OF FIXATION DURATION ON CONTROL, LOW-LEVEL AND HIGH-LEVEL LEARNING CONDITIONS**

<table>
<thead>
<tr>
<th>Measures</th>
<th>(I) Factor</th>
<th>(J) Factor</th>
<th>Mean Difference (I – J)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of fixation</td>
<td>Control</td>
<td>Low-level</td>
<td>(1.972^{**})</td>
</tr>
<tr>
<td></td>
<td></td>
<td>High-level</td>
<td>(1.914^{**})</td>
</tr>
<tr>
<td></td>
<td>Low-level</td>
<td>Control</td>
<td>(-1.970^{**})</td>
</tr>
<tr>
<td></td>
<td></td>
<td>High-level</td>
<td>(-5.725E-02^{*})</td>
</tr>
<tr>
<td></td>
<td>High-level</td>
<td>Control</td>
<td>(-1.914^{**})</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Low-level</td>
<td>(-5.725E-02^{*})</td>
</tr>
</tbody>
</table>

\(^{*}p < 0.05.\)
\(^{**}p < 0.001.\)
more cognitively demanding context. Similar results are presented by Crundall et al.\textsuperscript{25} where novice drivers produced a narrower spread of search than experienced drivers during observation of police pursuit video clips (drivers’ view). In the same study, results indicated that novice tended to produce longer fixation than their experienced counterpart.\textsuperscript{25} Although, these studies do not reveal significant results related to fixation number, they nonetheless underline the effect of experience on visual search strategy in a visual task. This study brings results similar to those of researches previously discussed by revealing changes expressed in eye-movement measurements such as found in patterns of saccades and fixations (Fig. 5).

In addition, our results revealed that variability (SD) in fixation duration is affected by perceptual learning during detection tasks. Variability in eye fixation duration seems to increase with perceptual learning. Perceptual learning ensues from longer stimulus exposition, gaze behavior is more flexible regarding processing time granted to extract relevant information in the detection environment. Flexibility in visual search has also been observed in driving context. Crundall and Underwood\textsuperscript{22} demonstrated that experienced drivers were more able to adapt their visual search strategy to roadways’ cognitive demands while novice drivers tended to apply the same inflexible strategy. As observed here, this flexibility in strategy is reflected upon gaze behavior, particularly through fixation duration. Curiously, our results did not reveal significant differences regarding fixation duration as presented by Crundall and Underwood.\textsuperscript{22} Research using a larger sample may provide additional input on the subject.

**CONCLUSION**

As stated earlier, immersive technology embodies a promising tool for tailoring learning environments. Its flexibility and capability to consolidate spatial information in a veridical visual environment justify such expectations. This preliminary research provides promising results suggesting that immersive technologies, combined to eye tracking systems, may afford relevant input with regard to learning behavior unfolded in immersive environment. In this study, we provide data suggesting that perceptual learning impinges on standard eye movement measures such as fixations and saccades. To our knowledge, direct examination of gaze behavior in relation to perceptual learning is relatively unprecedented. Moreover, implementation of this experiment in immersive environments bares interesting implications in the virtual research and learning fields. By monitoring how the perceiver assimilates environmental features and relations within, it becomes possible to better understand how learning takes place in immersion as well as outside immersion. Such input could contribute to refine guidelines for creating virtual environment intended for learning and training at large. Assessment of behavioral response jointly with standard goal-attainment measurements would certainly enhance the ecological validation endeavors. Nevertheless, additional research is needed to further validate and deepen results obtained so far. For instance, in this research, standard eye movement measures such as fixations and saccades provide useful but limited input. Future studies of perceptual learning should be directed toward eye movement analysis in relation with virtual objects within immersive environments. By tracking eye

**FIG. 5.** Patterns of saccades and fixations as observed in a typical subject for low-level learning detection condition (A) and high-level learning detection condition (B); saccades are depicted as lines and fixation as dots whose dimension varies according to fixation duration; more fixations were observed in the low-level learning condition.
and head movements in immersion, it becomes possible to monitor the perceivers’ relation with virtual objects spatial characteristics.26,27

ACKNOWLEDGMENTS

Specific programming for this project was done by Christian Villemaire, computer scientist, and Serge Bernier, engineer. Thanks go to Canadian Foundation for Innovation Simulations and Advanced Gaming Environments for Learning (SAGE) project (funded by the Social Sciences and Humanities Research Council of Canada [SSHRC]); Université du Québec en Outaouais; Ministère du développement économique et régional et de la recherche du Québec.

REFERENCES


Address reprint requests to:
Guillaume Albert
Département de psychologie et de psychoéducation
Université du Québec en Outaouais
Laboratoire de Cyberpsychologie
C.P. 1250, succursale Gatineau
Gatineau, Qué., Canada J8X 3X7

E-mail: albgui@gmail.com