The Effect of Sound on Visual Realism Perception and Task Completion Time in a Cel-Shaded Serious Gaming Virtual Environment

David Rojas[2], Brent Cowan[1], Bill Kapralos[1], Karen Colllins[3], and Adam Dubrowski[4]

[1]Faculty of Business and IT, University of Ontario Institute of Technology, Oshawa, Ontario, Canada.
[2]Institute of Medical Science, University of Toronto, Toronto, Canada.
[3]The Games Institute, University of Waterloo, Waterloo, Ontario, Canada
[4]Faculty of Medicine, Memorial University. St. John's, Newfoundland, Canada.

Email: bill.kapralos@uoit.ca

Abstract—**Here we investigate the effect of sound on the perception of visual realism and the time required to complete a simple navigation-based task within a serious gaming (virtual) environment under various sound and visual conditions. Results indicate that the perception of visual realism and task completion time can be affected by sound. Designers and developers of serious games (and virtual environments in general) should be aware of the effects of sound on a user's perception of the visual scene and on task completion time, and they should thus ensure that sound is carefully considered when creating such environments.**

Keywords—Serious games; virtual simulation; multi-modal interactions; audio-visual interaction; realism.

I. Introduction

Serious games, that is video games whose primary purpose is education/training [2] take advantage of the motivating, engaging, and fun factors inherent in video games to engage players/learners for a specific purpose such as to develop new knowledge or skills [6]. Serious games provide a rich and engaging environment for learning and for training, making them an attractive alternative to existing methods or training [13]. The rising popularity of serious games has seen a recent push towards the application of serious games to medical education and training. With their ability to engage players/learners for a specific purpose, serious games provide an opportunity to acquire cognitive and technical surgical skills outside the medical theater (e.g., emergency room, operating room, etc.), thereby optimizing the limited exposure time trainees have with live patients. However, before the application of serious games for medical education and training becomes more widespread, there are a number of open questions and issues that must be addressed. More specifically, there is a lack of research that investigates the technological conditions under which virtual learning can be maximized, and a lack of research that links virtual learning with proven educational theory and practice. With respect to medical education, Tashiro and Dunlap [22] identified seven areas that require further research and improvements for the effective development of serious games: i) disposition to engage in learning, ii) impact of fidelity on learning, iii) threshold for learning, iv) process of cognitive development during knowledge gain, v) stability of knowledge gain (retention), vi) capacity for knowledge transfer to related problems, and vii) disposition toward sensible action within clinical settings. In the context of a virtual learning environment including serious games, fidelity denotes the degree of similarity between the training situation and the operational situation which is simulated [8]. In other words, fidelity describes the extent to which the virtual environment emulates the real world [1]. However, as described by Cook et al. [5], in the context of a simulation, fidelity can encompass a wide variety of aspects related to the simulation. The term can therefore convey various meanings thus "losing its usefulness" [5]. Therefore, to avoid any confusion with the term, for the remainder of this paper, we replace the term "fidelity" with "realism" which, in the context of a virtual reality, refers to the degree that a user perceives the virtual environment as being "plausible" for a real one.

We have recently begun investigating the perception of visual realism of 3D rendered (virtual) scenes under various auditory conditions. This is part of our larger-scale effort that examines virtual environment realism, multi-modal interactions, user-specific factors (e.g., personality, learning style, existing knowledge, level of attention, motivation etc.) and their effect on knowledge transfer and retention within a medical education context. Our work seeks to answer the questions related to "how much realism is actually needed to maximize transfer?" and "what effect do multi-modal cues have on knowledge transfer and retention?" These questions may have a number of implications, particularly when considering that any training device—be it a virtual or physical simulator—will likely never be able to faithfully completely replicate the real-world with all its sensory stimuli; in other words, complete (perfect) realism appears to be impossible to achieve, at least with our current technology [10]. In addition, striving to reach full realism can also lead to increased development costs, may lead to lag and motion sickness, and it remains unclear if such high levels of realism are needed for either enjoyment or knowledge transfer (recent work suggests it may not be necessary [5,7]).

To date, we have conducted a series of experiments that examined the direct effect of sound on engagement, and on the perception of visual realism (the degree to which visual features within a virtual environment conform to visual features in the real environment [11]), and task performance (the time required to complete a task within a virtual environment), of both static and dynamic 3D rendered (virtual) scenes in both stereoscopic

978-1-4799-8958-4/15/$31.00 ©2015 IEEE QoMEX 2015

3D (S3D) and non-S3D viewing were conducted. Although this series of experiments have shown a strong influence of sound on visual realism perception, engagement, and task performance, results have also shown strong subjective effects, whereby the influence of sound is dependent on various individual factors including musical listening preferences.

In our previous studies, visual realism was defined with respect to texture resolution, polygon count, or blurring of the scene (to approximate texture resolution). However, there are many ways to define a virtual visual scene and many ways to define visual realism. Here, we have chosen to define visual realism with respect to levels of *cel-shading* (also known as *toon-shading*), a popular 3D rendering technique used in games that attempts to recreate the look of traditional 2D animation with the use of flat colors to light 3D surfaces in an unrealistic way. Using cel-shading instead of striving for "photorealism" can save developers both time and money. More specifically, cartoon style environments and characters generally require less surface detail than photorealistic settings, thus leading to significant reductions in development time and rendering (computational) game resources (cel-shading is relatively simple to implement for real-time rendering [12]). Animated cartoons are often shaded using a discrete number of color values or brightness levels as sharp lines are easier to animate than smooth gradients.

II. Background

Various studies have examined the perceptual aspects of audio-visual cue interaction, and have shown that the perception of visual realism can affect the perception of auditory quality and vice versa [20]. Here, we will focus on our own prior work as it relates to the work described in this paper. That being said, greater details regarding the influence of sound on visual rendering is provided by Hulusic et al. [9] while an overview of "crossmodal influences on visual perception" is provided by Shams and Kim [19].

Our prior work has so far examined visual realism perception in the presence of various auditory conditions. We have begun by conducting a number of studies that examine the perception of visual realism of static 3D rendered (virtual) scenes under various auditory conditions. The purpose of these studies was to determine whether sound can lead to an increase in visual realism perception and allow us to employ lower realism visuals and therefore reduce computational load. Beginning with simple, static environments provides us the opportunity to build upon our results in a methodical manner and this will ultimately aid in our long-term goal of developing an understanding of simulation (serious games in particular) realism, multimodal interactions, user-specific factors and their effects on learning and ultimately develop more effective virtual training environments and serious games.

Our initial studies included simple static environments comprised of a single 2D image of a surgeon's head (a rendered 3D model). Visual realism was defined with respect to either the 3D model's polygon count [15] or, with respect to the 3D model's texture resolution while polygon count remained constant [17]. Participants were presented with the static visual (the rendering of a surgeon's head), and each visual was presented along with one of four auditory conditions: i) no sound at all (silence), ii) white noise, iii) classical music (Mozart), and iv) heavy metal music (Megadeth). For each of the audio-visual presentations, the participants' task was to rate the realism of the visual on a Likert scale from 1 to 7. With respect to polygon count, visual realism perception increased in the presence of classical music, and this increase was more pronounced with the renderings corresponding to higher polygon count (higher realism). With respect to texture resolution, white noise sound had very specific and detrimental effects on the perception of visual realism. In a follow-up study, this was repeated but the visuals were presented in stereoscopic 3D [18]. When visual realism was defined with respect to polygon count, "classical music" led to an increase in visual realism perception while "white noise" had an attenuating effect on the perception of visual realism. However, both of these effects were evident for only the visual models whose polygon count was greater than 678, indicating that there is a polygon count threshold after which the visual distinction is not great enough to be negatively influenced by white noise. When visual realism was defined with respect to texture resolution, both "classical music" and "heavy metal music" led to an increase in visual realism perception while "white noise" caused a decrease in visual realism perception.

The results of these initial experiments showed that sound can affect visual realism perception, and at times, the resulting effect can be substantial. However, the auditory conditions considered in those studies were non-contextual (no direct relationship) to the visual scene. Two experiments were thus conducted to examine visual realism perception in the presence of contextual sounds, that is, sounds that had a causal relationship to the visual cues which once again, remained static [16, 18]. With non-S3D viewing, contextual auditory cues led to an increase in the perception of visual realism for the majority of the contextual auditory cues, while non-contextual cues in the form of white noise led to a decrease in visual realism perception, particularly when considering the lower realism visuals [16]. The notion of contextual auditory cues may be subjective and may depend on various factors including prior experience and musical listening preferences. An additional experiment were conducted to examine the effect of sound on visual realism perception, and task performance (defined as the time required to complete the task) in dynamic virtual environments were the participants had to interact with the environment while completing a simple task (navigating through a virtual surgical operating room to pick up a surgical drill). Visual realism was defined with respect to the level of (consistent) blurring of the entire screen and the auditory cues consisted of three contextual cues in addition to white-noise, and no sound. Sound (contextual and non-contextual), did not influence the perception of visual realism irrespective of the level of blurring. However, sound did impact task performance. More specifically, white noise led to a large decrease in performance (increase in task completion time, defined as the time between the start of a trial to the time at which the participant chose the surgical drill), while contextual sound improved performance (decrease in task performance time), across all levels of visual realism considered.

The study presented in this paper follows up on our last study that considered a dynamic virtual environment where the participants were required to perform a task. However, here, rather than define visual realism with respect to blurring of the scene we have chosen to consider cel-shading and gray-scale as a more realistic representation of visual effects that may be encountered (e.g., cel-shading is a popular graphical technique applied to a number of games).

III. Methods and Materials

A. Participants

Participants consisted of unpaid volunteers students from the University of Ontario Institute of Technology (UOIT). Eight female, and seven male (total of 15) volunteers participated in the experiment (average age was 23 years). None of the participants reported any hearing or visual defects. The authors did not participate in the experiment. The experiment abided by the UOIT Research Ethics Review process.

B. Visual Stimuli

The visual scene consisted of five rendered (3D) versions of an operating room with various tools, and equipment (see Fig. 1(a)). Within the operating room were three non-player characters (nurses) which, for the purposes of this experiment, remained static and did not afford any interaction with the participants. The five conditions included the following (see Fig. 1): a) original, no effect, b) gray-scale (all colors removed), c) outline (i.e., normal color but with highlighted edges), d) 3-level cel-shading (i.e., color is divided into three discrete levels), and e) 6-level cel-shading (i.e., color is divided into six discrete levels). The experiment was carried out (and all visuals were thus displayed), on an Acer Aspire laptop with a 15.6" screen size and a screen resolution of 1366 × 768. Participants were seated approximately 0.7 m in front of the laptop and the operating room environment was viewed in "full screen" mode.

Here, cel-shading was approximated using *posterization* (also known as banding). Posterization is similar to cel-shading in that the value or brightness of the image is discretized causing smooth gradients to be replaced by abrupt changes in brightness. This has the effect of segmenting an image based on the brightness and hue. The effect of posterization is generally indistinguishable from cel-shading when the source image has been rendered with smooth or Gouraud shading using 3D models as done in this work. Our posterization is performed as a post-processes which converts a realistically rendered 3D scene to a non-photorealistic rendering of a scene with a resulting appearance that is similar to cel-shading. The red, blue, and green values together can be thought of as a 3D vector in Cartesian space. In this representation, the length of the vector represents the brightness or value of the color, while the direction of the vector represents the hue and saturation. The shader program resizes the vector (color) to the nearest discrete length without changing the direction. The result is a scene which contains many colors but only a discrete number of brightness levels leading to an image that appears to have been rendered with a discrete number of colors. In reality, only the brightness level has been discretized. The outline visual condition was implemented using an outline shader program that renders thick black outlines around 3D objects in the scene. Unlike the other effects, the input texture for this shader program is a black and white depth map. Every surface in the scene is assigned a value between 0 and 1 based on its distance from the camera. To determine if a given pixel is on a line, the depth is sampled and compared to that of the neighboring pixels. If the pixel's depth is different from that of any of its neighbors, than it is close to an edge and is therefore rendered black.

C. Auditory Stimuli

Five ambient (background) auditory conditions were examined: i) no sound, ii) white noise, iii) operating room ambiance mixed with a surgical drill sound, iv) operating room ambiance without the drill sound, and v) classical music (Mozart). The operating room ambiance sound included typical operating room type sounds (e.g., machines beeping, doctors and nurses talking, etc.) and was purchased from AudioSparx.com. The operating room ambiance with the drill sound was made by mixing the operating room ambiance sound with a recording of an actual drill sound. The recording was made in an Eckel audiometric room to limit any external noise (air condition "hums", etc.) and reverberation of the generated sounds within the environment, at a sampling rate of 44.1 kHz. The white noise sound was sampled at a rate of 44.1 kHz and band-pass filtered using a 256-point Hamming windowed finite impulse response (FIR) filter with low and high frequency cut-off frequencies of 200 Hz and 10 kHz respectively. All auditory stimuli were presented to the participants with a pair of Sony MDR 110LP headphones. It should be noted that we did not ask participants to provide any information regarding their perception of the auditory stimuli. Furthermore, we did not explicitly test whether the surgical drill sound was in fact contextualized to the surgical drill by the participants.

D. Experimental Method

Participants were seated in front of the laptop computer and were provided with an overview of the experiment followed by a description of their required task by one of the experimenters. In each trial, participants were presented with one of the five versions of the rendered operating room in conjunction with one of the five ambient sound conditions previously described (see Fig. 1). Their task was to navigate through the operating room from their starting position to a point in the room that contained a tray with surgical instruments and pick up a surgical drill (they had to navigate around the bed and one of the NPC nurses to reach the tray).

Navigation through the environment was accomplished in a first-person perspective (taking on the role of the surgeon) using the standard arrow keys (to move the "player"), and mouse (to move the "camera"). Within this first-person view, the hand and lower arm of the participant's avatar was displayed. Choosing the surgical drill involved moving their avatar's hand over the drill and clicking the left mouse button. Once the drill was chosen, it appeared in the hand of the user's avatar and the participant was prompted to rank the visual scene with respect to their perceived visual realism on a scale from 1 (lowest perceived realism), to 7 (highest perceived realism). Aside from interacting with the surgical instruments, there were no other interactions permitted (e.g., the participants could not interact with any of the NPCs or other objects in the room). Entering their choice signaled the end of the trial; the following trial began after the user clicked the "Continue" button that appeared on the visual realism-ranking screen after the participant entered their choice. Each of the five rendered versions of the operating room and each of the five ambient sound combinations (25 combinations in total) were repeated three times for a total of 75 trials, all of which were presented in a randomized ordering. The experiment took approximately 30 minutes to complete and all of the participants completed the experiment in a single session.

IV. Results

The Analysis of Variance (ANOVA) was used as the statistical method with "image" and "sound" as factors. All effects with an alpha value lower than .05 were further analyzed using Tukey's Honestly Significant Difference post-hoc test. The two dependent variables analyzed for this experiment were: i) *visual realism perception*, and ii) *task completion time* (the time between the start of a trial to the time at which the participant chose the surgical drill).

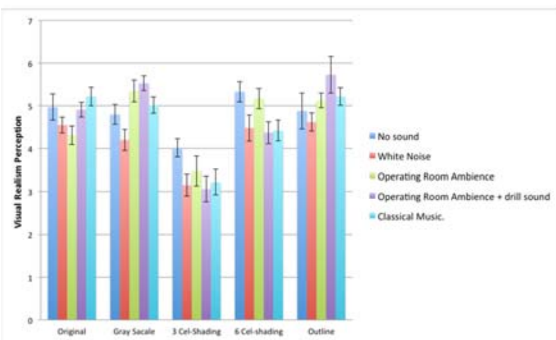

Fig 2. Visual realism perception scores.

A graphical summary of the visual realism perception results are shown in Fig. 2. The analysis of visual realism perception showed that there were significant effects for image [$F(1,4)=9.964$, $p<0.001$)] and sound [$F(1,4)=3.8$, $p=0.005$)] and the interaction between these two factors [$F(1,16)=2.384$, $p=0.002$)]. A subsequent post-hoc comparison (Tukey's HSD) revealed a significant difference in participants' perception of visual realism for the 3-level cel-shading condition when compared to all of the other visual conditions. More specifically, participants rated this visual condition with lower realism than the other ones. There was no significant difference with respect to the visual realism perception amongst the other visual conditions: original, grey scale, outline (highlighted edges), and 6-level cel-shading. This implies that the participants' perception of visual realism varied significantly (i.e., decreased) only when the visual condition presented was 3-level cel-shading. This detriment to the participants' visual realism perception was not influenced by the sound conditions that they were presented with.

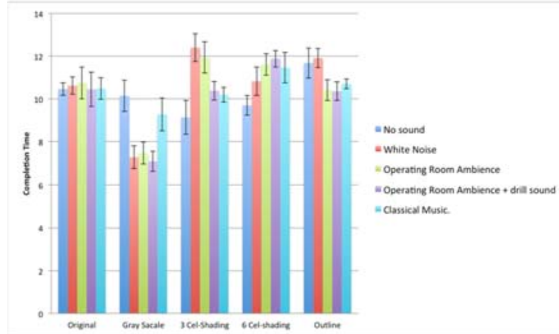

Fig 3. Completion time results.

With respect to sound, (Tukey's HSD) post-hoc comparisons revealed that the white noise condition had an attenuating effect on the participants' visual realism perception. Significant differences were found when the white noise condition was compared against the no sound condition, operating room ambiance without the drill, and operating room ambiance mixed with surgical drill conditions. Interestingly, no significant difference was found when compared to the classical music condition. This could be seen as the classical musical condition leading to a decrease in the participants' visual realism perception. However, that was not the case as there were no significant differences between the classical musical condition and the other auditory conditions. In summary, aside from the white noise condition, sound did not significantly influence participants' visual realism perception (white noise had an attenuating effect in participant's visual realism perception).

Task completion time was analyzed separately and a graphical summary of the task completion time results is provided in Fig. 3. The analysis showed that there were significant effects for image [$F(1,4)=21.355$, $p<0.001$)] and the interaction between image and sound [$F(1,16)=3.577$, $p<0.001$)], but not for sound [$F(1,4)=0.742$, $p=0.564$)]. A subsequent post-hoc comparison for the visual conditions (Tukey's HSD) revealed that the participants' completion time in the gray scale visual condition was significantly lower (i.e., the task was performed faster) when compared to the other four visual conditions: original, outline (highlighted edges), cel- 3-level cel-shading, and 6-level cel-shading.

V. Discussion

In this paper we have described a study that examined the influence of ambient sound on visual realism perception where visual realism was defined with respect to the levels of a cel-shaded virtual environment. In order to accomplish this task, an experiment was conducted whereby participants were presented with five different visual conditions as shown in Fig. 1: a) original, b) gray scale, c) outline (highlighted edges), d) 3-level cel-shading, and e) 6-level cel-shading, and v) under five ambient auditory conditions: i) no sound, ii) white noise, iii) operating room ambiance mixed with a surgical drill sound, iv) operating room ambiance without the drill sound, and v) classical music (Mozart). Our previous studies have shown that sound can influence visual realism perception, particularly when the sound presented to the participants was contextually related with the environment presented [16]. In our prior studies, visual realism was defined with respect to texture resolution, polygon count, and blurring of the scene, amongst others. However, no previous studies have considered a cel-shaded or highlighted edge visual environment as resolution-dependent, despite the fact that cel-shading is a popular visual effect common in a variety of applications (e.g., cartoons such as *Futurama* or *Appleseed*, or popular games including the *Legend of Zelda: Wind Waker* and *Borderlands*).

The results presented here, and similarly to our previous work [15-18], show that sound (and white noise in particular), can influence participants' visual realism perception. More specifically, the white noise auditory condition had a detrimental effect on visual realism perception when compared to the no sound, operating room ambiance sound, and the operating room ambiance with surgical drill sound auditory conditions. In contrast to our prior work [16], here, contextual sounds did not have a significant effect on visual realism perception. That being said, the auditory conditions that led to higher ratings of visual realism perception (albeit, not significant), were those in which the content (e.g., operating room ambiance and the operating room ambiance with surgical drill sound) was related to the context of the visual scenario that they were being presented with (e.g., a graphically rendered operating room). On the other hand, the auditory condition that led to the lowest visual realism ratings (with significance) was the white noise condition. This condition could have led to a disruption to the visual realism analysis process and therefore to the participants' lower visual realism perception ratings. This hypothesis is in agreement with Chang and Thompson [3] who demonstrated that whines, cries, and "child-directed speech" distracted listeners completing simple mathematical (subtraction) problems. Similarly, Conrad et al. [4] discovered that stressful music (e.g., heavy metal music) had a negative impact on the time required to complete a laparoscopic surgery task but did not impact task accuracy. It is important to highlight that the previously described auditory effects on participant's visual realism perception were not seen for the visual condition that was rated as the one with lowest visual realism (e.g., 3 cel-shading levels). This implies that the influence of sound on visual realism perception is secondary to the main/primary visual judgment; when a condition is catalogued as low in visual realism, auditory cues have no influence in the further judgment of visual realism perception.

With respect to task completion time, the analysis was conducted separately as this variable measured performance and not visual realism perception. Although, it could be inferred that visual realism would play a role in performance, results indicate that participants performed the task faster under the gray scale visual condition when compared to the other four visual conditions. Sound did not have any effect on task completion time.

These two findings, and more specifically, the effect of sound (white noise in particular) on visual realism perception, and the reduction in task completion time (increase in performance) in the presence of the gray-scale visual condition could be explained by cognitive load theory (CLT) [14,21]. CLT suggests that there are limited cognitive resources available when someone is trying to perform a task. When cognitive load is low, the participant has a greater amount of cognitive resources available to allocate to a task. In such a situation, it is therefore possible that the availability of these cognitive resources leads to better performance. In contrast, when cognitive load is high, it is possible that the cognitive capacity becomes overloaded. In this case, participants will have a harder time to perform a task or to pay attention to the different cues being presented to them at the time and as a result, details could be missed. It is interesting to note that when considering the visual realism perception results presented here, the white noise auditory condition led to a decrease in visual realism perception. Considering this result from a CLT perspective, this suggests that a "disturbing" auditory cue could overload the cognitive resources available to perform the particular task (i.e., visual realism analysis), thus leading to the decrease in visual realism perception. The decrease in the perceived realism ratings is a consequence of a lack of resources to properly perform the task (i.e., visual realism perception analysis). It is important to clarify that an overload of cognitive capacities does not lead to un-attendance or a distracted participant. Rather, the participant is still fully engaged in completing the task at hand but not enough cognitive resources are available to properly achieve the expected goal, thus the decrease in performance (increase in the time required to complete the task). With respect to the task completion time, a similar explanation could be applied to the fact that the gray scale visual condition led to greater performance. When participants were presented with the gray scale visual condition (i.e., no color), less cues were being presented to them (i.e., no color information). CLT would suggest that participants' cognitive load had more resources available due to the reduced amount of information (i.e., color cues) provided to them. The availability of cognitive resources led to a decrease in the task completion time (i.e., an increase in performance). In other words, color can be viewed as an external/additional cue requiring cognitive resources in order for it to be processed.

VI. Conclusions and Future Work

Given the popularity and ubiquitous use of video games, serious games, that is, video games whose primary purpose is education and/or training, have been applied to a wide variety of applications and are gaining in popularity. However, before their use becomes more widespread, there are a number of open problems that must be addressed. This includes questions related to realism and multi-modal interactions, which can have severe consequences when considering that with our current technology we cannot faithfully recreate a real environment in the virtual domain. Our ongoing work is examining issues related to audio-visual realism and multimodal interactions in order to determine how such factors interact and ultimately to determine what effect they have on learning.

As part of our ongoing effort, here we have presented the results of an experiment that examined the influence of sound cues on visual realism perception (when the visual conditions were varied in levels of cel-shading; a popular graphical technique), and on the time required to complete a simple task (task completion time) within a virtual environment. Results showed that sound does influence visual realism perception in some cases and more specifically, white noise (a distracting sound) leads to a decrease in visual realism perception. Results also showed that the gray scale visual condition (i.e., no color cues) increased performance (decreased task completion time); participants performed the task faster when compared to the other visual conditions that included color. Whether or not this reduction in completion time impacts learning has yet to be shown. Noise can be distracting in both simulation and real-world operating theatres thus, perhaps such distractions can be employed in a virtual learning environment (e.g., serious game) to train medical trainees to focus during distracting moments.

These findings are important to serious game designers and developers (and to the gaming industry in general). They must carefully consider the use of sounds within their games and understand that sound can have an effect on the visuals and to any intended visual effects. Furthermore, designers and developers of serious games and virtual learning environments should be careful not to overload the user's cognitive resources/capacity at any time during the learning experience. Of course, further work remains and future work will consider examining what consequences the results provided by this study have on learning. In other words, although the results of this study indicate that non-contextual sounds will lead to a burdening of cognitive resources, will this transfer negatively to learning and what influence and how much of a role will the learner's prior knowledge level play in this? These, amongst others, are questions for future work.

ACKNOWLEDGMENT

The financial support of the *Natural Sciences and Engineering Research Council of Canada* (NSERC), the *Social Sciences and Humanities Research Council of Canada* (SSHRC), and *Canadian Network of Centres of Excellence* (NCE), *Graphics, Animation, and New Media* (GRAND) initiative, and the in the is gratefully acknowledged.

REFERENCES

[1] Andrews, D. H., Carroll, L. A., and Bell, H. (1996). The future of selected fidelity in training devices. *Educational Technology*, 35(6):32-36.

[2] Bergeron, B. (2006). *Developing serious games*. Hingham, MA. USA: Thomson Delmar Learning.

[3] Chang, R. S., and Thompson, N. S. (2011). Whines, cries, and motherese: their relative power to distract. *Journal of Social, Evolutionary, and Cultural Psychology* 5(2):131-141.

[4] Conrad, C., Konuk, Y., Werner, P., Cao, C. G., Warshaw, A., Rattner, D., Jones, D. B., and Gee, D. (2010). The effect of define auditory conditions versus mental loading on the laparoscopic motor skill performance of experts. *Surgical Endoscopy*, 24:1347–1352.

[5] Cook D. A., Hamstra, S. J., Brydges, R., Zendejas, B., Szostek, J. H., Wang, A. T., Erwin, P. J., Hatala, R. (2013). Comparative effectiveness of instructional design features in simulation-based education: Systematic review and meta-analysis. *Medical Teacher*, 35(1), e867-898.

[6] Corti, K. (2006). Game-based learning; a serious business application, PIXELearning. Coventry, UK.

[7] Hamstra, S. J., Brydges, R., Hatala, R., Zendejas, B., and Cook, D. A. (2014). Reconsidering fidelity in simulation-based training, *Academic Medicine*, 89(3):1-6.

[8] Hays, R. T., and Singer, M. (1989). Simulation fidelity in training system design. New York: Springer-Verlag.

[9] Hulusic, V., Harvey, C., Debattista, K., Tsingos, N., Walker, S., Howard, D., and Chalmers, A. (2012). Acoustic rendering and auditory-visual cross-modal perception and interaction. *Computer Graphics Forum*, 31(1):102-131.

[10] Hulusic, V., Aranha, M., and Chalmers, A. (2008). The influence of cross-modal interaction on perceived rendering quality thresholds. In *Proceedings of the. Computer Graphics, Visualization & Vision Conference 200*8, Plzen - Bory, Czech Republic, pp. 41-48.

[11] Mania, K., Wooldridge, D., Coxon, M., and Robinson, A.(2006). The effect of visual and interaction fidelity on spatial cognition in immersive virtual environments? *IEEE Transactions on Visualization and Computer Graphics*, 12(3):396-404.

[12] Masuch, M., and Röber, N. (2004). Game graphics beyond realism: Then, now and tomorrow. In *Proceedings of the 2005 DiGRA Conference: Changing Views: Worlds in Play*, Vancouver, Canada, June 16-20, 2005.

[13] Mayo, M. J. (2007). Games for science and engineering. Communications of the ACM, 50(7):31-35.

[14] Marois, R., and Ivanoff, J. (2005). Capacity limits of information processing in the brain. *Trends in Cognitive Sciences*, 9(6):296-305.

[15] Rojas, D., Kapralos, B., Crsitancho, S., Collins, K., Conati, C., and Dubrowski, A. (2011). The effect of background sound on visual fidelity perception. In: *Proceedings of the ACM Audio Mostly Conference 2011*, Coimbra, Portugal, September 7-9, 2011.

[16] Rojas, D., Kapralos, B., Collins, K., and Dubrowski, A. (2014). The effect of contextual ound cues on visual fidelity perception. *Studies in Health Technology and Informatics*, 196:346-352, 2014.

[17] Rojas, D., Kapralos, B., Crsitancho, S., Collins, K., Hogue, A., Conati, C., and Dubrowski, A. (2012) Developing effective serious games: The effect of background sound on visual fidelity perception with varying texture resolution. *Studies in Health Technology and Informatics,* 173:386-392.

[18] Rojas, D., Kapralos, B., Hogue, A., Collins, K., Nacke, L., Cristancho, S., Conati, C., and Dubrowski A. (2013). The eect of sound on visual delity perception in stereoscopic 3D. *IEEE Transactions on System Man and Cybernetics part B*, 43(6):1572-1583.

[19] Shams, L., and Kim, R. (2010). Crossmodal infuences on visual perception. Physics of Life Reviews, 7(3):295-298.

[20] Storms, S.. L., and Zyda, M. J. (2000). Interactions in perceived quality of auditory-visual displays. *Presence: Teleoperators and Virtual Environments*, 9(6):557-580.

[21] Sweller, J., van Merrienboer, J., & Paas, F. 1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10, 251-296.

[22] Tashiro, J., and Dunlap, D. (2007). "The impact of realism on learning engagement in educational games," In *Proceedings of ACM Future Play 2007.* Toronto, Ontario, Canada, November 15-17.

(a) (b) (c) (d) (e)

Fig. 1. Visual stimuli considered in this experiment. (a) Original (non-filtered), (b) gray-scale (removal of all color information), (c) outline (normal color but highlighting of edges, (d) cel-shading with three levels, and (e) cel-shading with six levels (i.e., color is divided into six discrete levels).