Serious Games: Customizing the Audio-Visual Interface

Bill Kapralos, Robert Shewaga, and Gary Ng

Faculty of Business and Information Technology,
University of Ontario Institute of Technology,
Oshawa, Ontario, Canada L1H 7K4
bill.kapralos@uoit.ca

Abstract. Serious games are gaining in popularity within a wide range of educational and training applications given their ability to engage and motivate learners in the educational process. Recent hardware and computational advancements are providing developers the opportunity to develop applications that employ a high level of fidelity (realism) and novel interaction techniques. However, despite these great advances in hardware and computational power, real-time high fidelity rendering of complex virtual environments (found in many serious games) across all modalities is still not feasible. Perceptual-based rendering exploits various aspects of the multi-modal perceptual system to reduce computational requirements without any resulting perceptual effects on the resulting scene. A series of human-based experiments demonstrated a potentially strong effect of sound on visual fidelity perception, and task performance. However, the resulting effects were subjective whereby the influence of sound was dependent on various individual factors including musical listening preferences. This suggests the importance of customizing (individualizing) a serious game's virtual environment with respect to audio-visual fidelity, background sounds, etc. In this paper details regarding this series of audio-visual experiments will be provided followed by a description of current work that is examining the customization of a serious game's virtual environment by each user through the use of a game-based calibration method.

Keywords: Serious games, virtual simulation, audio-visual interaction, audio-visual fidelity, calibration.

1 Introduction

The use of serious games within a wide range of educational and training applications, from military, health professions education, patient education, and business/corporate, amongst others, is becoming widespread particularly given the ubiquity of video game play by the current tech-savvy generation of learners. Recent hardware and computational advancements are providing designers and developers of serious games the opportunity to develop applications that employ

R. Shumaker and S. Lackey (Eds.): VAMR 2014, Part II, LNCS 8526, pp. 190–199, 2014.

a high level of fidelity/realism and novel interaction techniques using off-the-shelf consumer level hardware and devices. Devices such as the Microsoft Kinect motion sensing vision-based sensor allows users to interact with their application using a natural user interface that employs gestures thus eliminating the game controller and the typically non-natural and potentially limiting interaction it affords. For example, using the Kinect within a virtual operating room, surgery trainees are able to perform their required tasks in a more intuitive manner that is better representative of the real world (see [1]).

With respect to a simulation (including serious games), fidelity denotes the extent to which the appearance and/or the behavior of the simulation matches the appearance and behavior of the real system [2]. Despite the great computing hardware and computational advances we have experienced, real-time high fidelity rendering of complex environments (found in many serious games) across all modalities is still not feasible [3]. Designers and developers of serious games, and virtual simulations in general, typically strive for high fidelity environments, particularly with respect to the visual (graphical) scene. However, evidence suggests high fidelity simulation does not always lead to greater learning [4]), and striving for high fidelity can burden our computational resources (particularly when the simulation is intended to be used on portable computing devices), increase the probability of lag and subsequent discomfort and simulator sickness [5], and lead to increased development costs. Previous work has examined the perceptual aspects of multi-modal effects (including audio-visual), and numerous studies have demonstrated that multi-modal effects can be considerable, to the extent that large amounts of detail of one sense may be ignored in the presence of other sensory inputs. Perceptual-based rendering, whereby the rendering parameters are adjusted based on the perceptual system (typically vision), is often employed to limit computational processing. For example, it has been shown that sound can potentially attract part of the user's attention away from the visual stimuli and lead to a reduced cognitive processing of the visual cues [6]. Therefore, if the enhancement of visuals within a virtual environment is economically or technically limited, one may consider increasing the quality of the audio channels instead [7].

Motivated by these studies and the general lack of emphasis on audition in virtual environments and games (where historically the emphasis has been placed on the visual scene [8]), we have begun investigating multi-modal (audio-visual) interactions within virtual environments (serious games, virtual simulations, and games). So far, a series of experiments that examined the direct effect of sound on engagement, the perception of visual fidelity (the degree to which visual features in the virtual environment conform to visual features in the real environment [9]), 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 3D (S3D) and non-S3D viewing were conducted. Although this series of experiments have shown a strong influence of sound on visual fidelity, 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. This suggests the importance of individualizing (customizing) audio-visual fidelity, and the sounds employed within a virtual environment to take advantage of perceptual-based rendering. Building upon the results of these experiments, we are examining the customization of the serious game's virtual environment to each user via a novel game-based calibration technique that will allow users to customize the virtual environment before they begin using the serious game. The calibration process will be used to tailor the settings of various simulation parameters including S3D settings (interaxial settings), audio and visual fidelity, background sounds/sound effects, spatial sound settings (choosing head-related transfer functions from a pre-defined set, etc.), amongst others, to each user's preferences. Such customization provides the opportunity to increase user engagement and ultimately learning.

1.1 Paper Organization

The remainder of this paper is organized as follows. In Section 2 a brief discussion of previous work (with an emphasis on the series of our own previously conducted experiments), is provided. Details regarding the calibration game are provided in Section 3 while a discussion, concluding remarks, and plans for future research are provided in Section 4.

2 Background

Various studies have examined the perceptual aspects of audio-visual cue interaction, and it has been shown that the perception of visual fidelity can affect the perception of sound quality and vice versa [10]. For example, Mastoropoulou et al. [6] examined the influence of sound effects on the perception of motion smoothness within an animation and more specifically, on the perception of frame-rate, and infer that sound can attract part of the viewer's attention away from any visual defects inherent in low frame-rates [6]. Similarly, Hulusic et al. [11] showed that sound effects allowed slow animations to be perceived as smoother than fast animations and that the addition of footstep sound effects to walking (visual) animations increased the animation smoothness perception. Bonneel et al. [12] examined the influence of the level of detail of auditory and visual stimuli on the perception of audio-visual material rendering quality and observed that the visual level of detail was perceived to be higher as the auditory level of detail was increased. Although there are various other relevant studies, for the remainder of this section, emphasis will be placed on our own previous work that has examined visual fidelity perception in the presence of various auditory conditions. Greater details regarding the influence of sound over visual rendering and task performance is provided by Hulusic et al. [3] while an overview of "crossmodal influences on visual perception" is provided by Shams and Kim [13].

Our studies began with simple static environments that consisted of a single 2D image of a surgeon's head (a rendered 3D model). In the first study, visual fidelity was defined with respect to the 3D model's polygon count [14] while in the second study, polygon count was kept constant and visual fidelity was defined with respect to the 3D model's texture resolution [15]. A sample of the visual stimuli is provided in Fig. 1 where three renderings of the surgeon's head, each one with a constant polygon count but differing with respect to texture resolution, are shown. In both studies, participants were presented with the static visual (a total of six visuals were considered, each differing with respect to polygon count or texture resolution depending on the experiment), in conjunction 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 visuals, their task was to rate its fidelity on a scale from 1 to 7. With respect to polygon count, visual fidelity perception increased in the presence of classical music, particularly when considering images corresponding to higher polygon count. When considering texture resolution, sound consisting of white noise had very specific and detrimental effects on the perception of the quality of high-resolution images (i.e., the perception of visual quality of high fidelity visuals decreased in the presence of white noise). In contrast to the study that considered polygon count, sound consisting of music (classical or heavy metal) did not have any effect on the perception of visual quality when visual quality was defined with respect to texture resolution.

Fig. 1. Sample of the visual stimuli used in a previous experiment that examined the effect of sound visual fidelity perception [15]. Here, each model of the surgeon's head contained the same polygon count but the texture resolution differed.

These two experiments were repeated but now the visuals were presented in stereoscopic 3D [16]. When visual fidelity was defined with respect to polygon count, "classical music" led to an increase in visual fidelity perception while "white noise" had an attenuating effect on the perception of visual fidelity. However, both of these effects were evident for only the visual models whose polygon count was greater than 678 (i.e., auditory condition had no effect on the two smallest polygon count models), indicating that there is a polygon count threshold after which the visual distinction is not great enough to be negatively influenced by white noise. With visual fidelity defined with respect to texture resolution, both "classical music" and "heavy metal music" led to an increase in visual fidelity perception while "white noise" led to a decrease in visual fidelity perception.

Although the results of these four studies show that sound can affect our perception of visual fidelity, it is not known if this influence of sound is affected by the introduction of contextually specific sounds. The auditory conditions considered in our previous studies have been completely disjoint from the visuals. That is, there was no (direct) relationship between the auditory and visual cues (they were non-contextual). Two experiments were thus conducted to examine visual fidelity perception, defined with respect to texture resolution, in the presence of contextual sounds, that is, sounds that had a causal relationship to the visual cues [16, 17]. The visual stimuli consisted of six images of a surgeon holding a surgical drill, against a black background (similar to the visuals employed in the previous experiment shown in Fig. 1 but with the addition of the surgeon's upper body). The polygon count of the 3D model was kept constant but as with the previous experiment, the texture resolution of the surgeon and the drill was varied. The auditory conditions included the four non-contextual auditory conditions considered in the previous experiments in addition to the following three contextual sounds: i) operating room ambiance which included machines beeping, doctors and nurses talking, ii) drill sound, and iii) hospital operating room ambiance coupled (mixed) with the drill sound. The visuals remained static in both experiments but in the second experiment, stereoscopic 3D viewing was employed. With non-S3D viewing, results suggest that contextual auditory cues increase the perception of visual fidelity while non-contextual cues in the form of white noise leads to a decrease in visual fidelity perception particularly when considering the lower fidelity visuals [17]. However, the increase in visual fidelity perception was observed for only two of the three contextual auditory conditions and more specifically, for the operating room ambiance, and operating room ambiance + drill auditory conditions and not for the drill auditory condition despite the fact that the surgeon within the visual scene was holding a surgical drill. With respect to S3D viewing, "white noise" led to a decrease in visual fidelity perception across all of the visuals considered. However, none of the auditory conditions led to a statistically significant increase in visual fidelity perception [16]. That being said, none of the participants were surgeons or medical practitioners and may not have been familiar with an operating room and the sounds contained within an operating room. The notion of contextual auditory cues may also be subjective and may depend on prior experience and musical listening preferences.

The experiments described so far considered static visual environments where the visual scene (the 3D models presented to the participants), remained static. Two additional experiments were conducted to examine the effect of sound on visual fidelity perception, and task performance in dynamic virtual environments were the participants had to interact with the environment while completing a simple task. In both experiments, participants were presented with a virtual operating room and their task was to navigate through the virtual operating room from their starting position to a point in the room which contained a tray with surgical instruments (see Fig. 2). Once they reached the tray, they were required to pick up a surgical drill (they had to navigate around a bed and a non-player character nurse to reach the tray that contained the surgical instruments). In one of the experiments, visual fidelity was defined with respect to the level of (consistent) blurring of the entire screen (level of blurring of the scene was used to approximate varying texture resolution), and the auditory cues consisted of the three contextual cues considered in the previous experiments in addition to white-noise and no sound. Sound (contextual and non-contextual), did not influence the perception of visual fidelity irrespective of the level of blurring. However, sound did impact task performance (defined as the time to required to complete the task). More specifically, white noise led to a large decrease in performance (increase in task completion time) while contextual sound improved performance (decrease in task performance time), across all levels of visual fidelity considered. In the second experiment [18], visual cues consisted of: i) original (no effect), ii) cel-shading with three levels (i.e., color is divided into three discrete levels), and iii) cel-shading with six levels (i.e., color is divided into six discrete levels). The contextual auditory conditions consisted of: i) no sound (visuals only), ii) monaural (non-spatial) surgical drill sound, and iii) spatialized surgical drill sound. In contrast to the last study, in this experiment, spatial sound (acoustical occlusion and reverberation) was considered. Contrary to our previous work, the presence of sound (spatial and non-spatial) did not have any effect on either visual fidelity perception or task completion time. That being said, only six participants took part in the experiment (in contrast to 18 for each of our previous experiments), thus the results are preliminary.

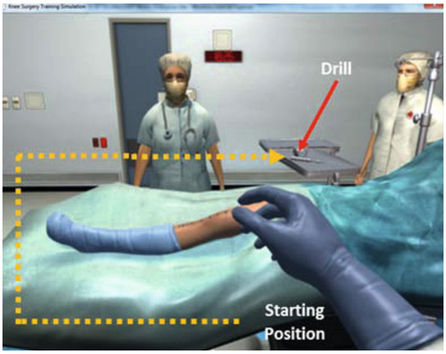

Fig. 2. View of the virtual operating room environment used in two previous experiments [18, 19]. The task of each participant was to navigate the environment from the starting position to the position of the surgical drill and then “choose” the drill.

2.1 Summary of Our Experimental Results

A total of eight experiments were conducted that examined the effect of sound on visual fidelity perception under a variety of conditions including static and dynamic environments, and stereoscopic 3D viewing. Results varied significantly across each of the experiments making it difficult to reach any firm consensus.

However, it is clear that white noise generally led to a decrease in the perception of visual fidelity and task performance, classical music led to an increase in visual fidelity perception (the majority of the time), and that the influence of sound on visual fidelity perception is very subjective. The visuals and many of the sounds considered in these experiments were medical in nature (e.g., surgeon, operating room, operating room ambiance sounds, drill sounds), yet many of the participants were students and although some were enrolled in Health Sciences-related programs, they had limited (if any), operating room exposure.

We hypothesize that the variation seen across the results of these experiments is due to subjective factors. Prior to the start of each experiment, participants were asked to complete a brief questionnaire regarding their video game play habits and game and musical preferences. A detailed analysis of the questionnaire that will examine whether any correlations exist between game/musical preferences and the experimental results is currently underway to confirm our hypothesis. However, informally, there does appear to be a relationship between musical genre preference and the influence of music on visual fidelity perception.

The variation observed across the results of all experiments and the potential consequences this variation may have on perceptual-based rendering and ultimately learning when considering serious games, motivated our work in the customization of audio-visual fidelity through a user-calibration method. This involves the use of a brief questionnaire that users complete prior to beginning the serious game followed by an interactive "calibration game" whereby the optimal audio-visual fidelity settings are determined dynamically by the player in the process of playing a game. How the questionnaire responses will be used will depend on the results of a meta-analysis that will be conducted on the results of our previous experiments but they may be used to drive the calibration game. Greater details regarding the calibration game are provided in the following section.

3 The Calibration Game: Calibration of Visual Fidelity

Although customizing the audio-visual interface using the results of a questionnaire presented to each user that may include visuals and audio clips, here, customization of the audio-visual interface is accomplished using a simple game-based approach, making the process interactive and far more engaging. Our approach is inspired by standard testing methodologies employed by optometrists to determine the optimum properties of corrective lenses in order to overcome a variety of visual deficiencies [20].

The calibration game presents the user with a split screen with the same game running in each window but under different fidelity/realism settings (see Fig. 3 for an example), with a single background sound. The player chooses the screen they prefer by clicking a button just above the corresponding window. Their choice will be registered and the audio-visual fidelity of the game running in the other window will change (increase or decrease). This process will be repeated over a number of cycles (the total number of cycles can be easily modified), until

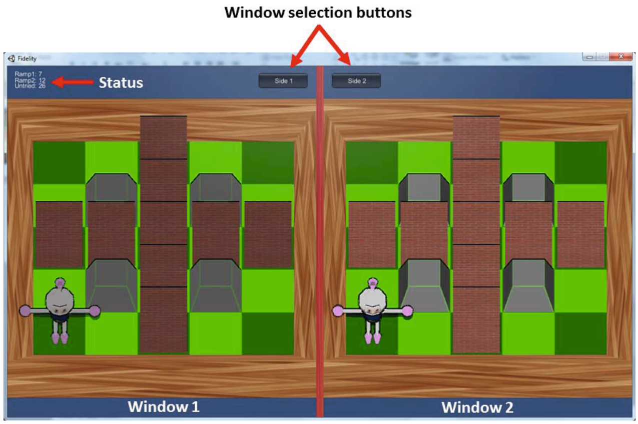

Fig. 3. Visual calibration game sample. Two versions of a game running in each window, each differing with respect to visual fidelity. The user then chooses which they prefer using one of the two selection buttons.

the optimal fidelity level is reached. Currently, the game used is a Bomberman strategic, maze-based video game where the player completes levels of the game by strategically placing bombs in order to kill enemies and destroy obstacles. The game is controlled using the 'W', 'A', 'S', 'D' keys to move the character (bomber), and bombs are placed by pressing the space bar. Both windows represent the same game-play (i.e., any actions to move the character or place a bomb will happen simultaneously in both windows). The calibration game was implemented using the Unity Game Engine and currently fidelity is defined by levels of cel-shading performed dynamically using a Unity shader. Although formal testing will follow, an informal test conducted with three participants revealed that the calibration game is easy to use and fun/enjoyable.

4 Discussion and Concluding Remarks

Prior work has demonstrated that the influence of sound on the perception of visual fidelity, and task performance within a virtual environment is complex and subjective, depending on a user's prior experience, and musical preference. However, this is rarely exploited as the vast majority of serious games take a "one-size-fits-all" approach with respect to audio-visual fidelity and the choice of background sounds and sound effects. Here, preliminary details of a novel "calibration game" being developed to custom-tailor the fidelity of the visuals within a serious game were provided. The game itself was inspired by standard optometrist testing and prior work that used a similar approach to determine the optimal interaxial distance of a stereoscopic S3D game and found the method to be effective [21]. Currently, fidelity was defined with respect to cel-shading implemented using the Unity Game Engine; this was done as a proof-of-concept to demonstrate the feasibility of such an approach and future work will examine other definitions of visual fidelity.

The work presented here is part of a larger initiative whose goal is to develop a greater understanding of visual fidelity, multi-modal interactions, perceptual-based rendering, user-specific factors, and their effect on learning. Although greater work remains, providing users the opportunity to customize the virtual environment of their serious game prior to using it will ultimately help us develop more effective serious games. Future work will see continued development and refinement of the calibration game. This will include experimenting with various other games aside from what was included here and conducting further experiments that examine audio-visual interactions and perceptual-based rendering. Furthermore, a meta analysis on the results of our previous experiments that examined audio-visual interactions (in addition to any subsequent experiments that will follow), will be conducted to identify any patterns or relationships among the study results and determine the most favorable fidelity settings that can be included in the calibration game. Future work will also further develop the calibration game to allow for additional definitions of visual fidelity, including polygon count and texture resolution, followed by a usability study to examine the effectiveness of the calibration game.

Acknowledgments. The financial support of the *Social Sciences and Humanities Research Council of Canada* (SSHRC), in support of the *Interactive and Multi-Modal Experience Research Syndicate* (IMMERSe) project, and the *Canadian Network of Centres of Excellence* (NCE) in support of the *Graphics, Animation, and New Media* (GRAND) initiative is gratefully acknowledged.

References

[1] Robison, R.A., Liu, C.Y., Apuzzo, M.L.J.: Man, mind, and machine: The past and future of virtual reality simulation in neurologic surgery. World Nuerosurgery 76(5), 419–430 (2011)
[2] Farmer, E., von Rooij, J., Riemersma, J., Joma, P., Morall, J.: Handbook of simulator based training. Ashgate Publishing Limited, Surrey (1999)
[3] Hulusic, V., Harvey, C., Debattista, K., Tsingos, N., Walker, S., Howard, D., Chalmers, A.: Acoustic rendering and auditory-visual cross-modal perception and interaction. Computer Graphics Forum 31(1), 102–131 (2012)
[4] Norman, G., Dore, K., Grierson, L.: The minimal relationship between simulation fidelity and transfer of learning. Medical Education 46(7), 636–647 (2012)
[5] Blascovich, J., Bailenson, J.: Infinite Reality. Harper Collins, New York (2011)
[6] Mastoropoulou, G., Debattista, K., Chalmers, A., Troscianco, T.: The influence of sound effects on the perceived smoothness of rendered animations. In: Proceedings of the Symposium on Applied Perception in Graphics and Visualization 2005, La Coruna, Spain, August 26-28, pp. 9–15 (2005)
[7] Larsson, P., Vstjll, D., Kleiner, M.: On the quality of experience: A multi-modal approach to perceptual ego-motion and sensed presence in virtual environments. In: Proceedings of the First International Speech Communications Association Tutorial and Research Workshop on Auditory Quality of Systems, Akademie Mont-Cenis, Germany (2003)

[8] Carlile, S.: Virtual auditory space: Generation and application. R. G. Landes, Austin TX (1996)
[9] Mania, K., Wooldridge, D., Coxon, M., Robinson, A.: 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 (2006)
[10] Storms, S.L., Zyda, M.J.: Interactions in perceived quality of auditory-visual displays. Presence: Teleoperators and Virtual Environments 9(6), 557–580 (2000)
[11] Hulusic, V., Debattista, K., Aggarwal, V., Chalmers, A.: Maintaining frame rate perception in interactive environments by exploiting audio-visual cross-modal interaction. The Visual Computer 27(1), 57–66 (2011)
[12] Bonneel, N., Suied, C., Viaud-Delmon, I., Drettakis, G.: Bimodal perception of audio-visual material properties for virtual environments. ACM Transactions on Applied Perception 7(1), 1–16 (2010)
[13] Shams, L., Kim, R.: Crossmodal influences on visual perception. Physics of Life Reviews 7(3), 295–298 (2010)
[14] Rojas, D., Kapralos, B., Cristancho, S., Collins, K., Conati, C., Dubrowski, A.: The effect of background sound on visual fidelity perception. In: Proceedings of ACM Audio Mostly 2011 (Extended Abstracts), Coimbra, Portugal, September 7-9, pp. 1–7 (2011)
[15] Rojas, D., Kapralos, B., Cristancho, S., Collins, K., Hogue, A., Conati, C., Dubrowski, A.: 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 (2013)
[16] Rojas, D., Kapralos, B., Hogue, A., Collins, K., Nacke, L., Cristancho, S., Conati, C., Dubrowski, A.: The effect of sound on visual fidelity perception in stereoscopic 3-d. IEEE Transactions on System Man and Cybernetics Part B 43(6), 1572–1583 (2013)
[17] Rojas, D., Kapralos, B., Collins, K., Dubrowski, A.: The effect of contextual sound cues on visual fidelity perception. Studies in Health Technology and Informatics (to appear, 2014)
[18] Cowan, B., Rojas, D., Kapralos, B., Collins, K., Dubrowski, A.: Spatial sound and its effect on visual quality perception and task performance within a virtual environment. In: Proceedings of the 21st International Congress on Acoustics, Montreal, Canada, June 2-7, pp. 1–7 (2013)
[19] Kapralos, B., Moussa, F., Dubrowski, A.: An overview of virtual simulations and serious games for surgical education and training. In: Brooks, A.L., Braham, S., Jain, L.C. (eds.) Technologies of Inclusive Well-Being. SCI, vol. 536, pp. 289–306. Springer, Heidelberg (2014)
[20] Kerr, D.S.: The refractor/phoropter an important tool in vision correction (December 7, 2010), http://dougkerr.net/pumpkin/articles/Refractor.pdf
[21] Tawadrous, M., Hogue, A., Kapralos, B., Collins, K.: An interactive in-game approach to user adjustment of stereoscopic 3D settings. In: Proceedings of Stereoscopic Displays and Applications XXIV, San Francisco, CA, USA, February 3-7, pp. 1–7 (2007)