The Influence of Sound Effects on the Perceived Smoothness of Rendered Animations

Georgia Mastoropoulou, Kurt Debattista, Alan Chalmers*
Department of Computer Science,
University of Bristol, UK

Tom Troscianko
Department of Experimental Psychology,
University of Bristol, UK

Abstract

The developers and users of interactive computer graphics (CG), such as 3D games and virtual reality, are demanding ever more realistic computer generated imagery delivered at high frame rates, to enable a greater perceptual experience for the user. As more computational power and/or transmission bandwidth are not always available, special techniques are applied that trade off fidelity in order to reduce computational complexity, while trying to minimise the perceptibility of the resulting visual defects. Research on human visual perception has promoted the development of perception driven CG techniques, where knowledge of the human visual system and its weaknesses are exploited when rendering/displaying 3D graphics. It is well known in the human perception community that many factors, including audio stimuli, may influence the amount of cognitive resources available to perform a visual task. In this paper we investigate the influence sound effects have on the perceptibility of motion smoothness in an animation (i.e. on the perception of delivered frame rate). Forty participants viewed pairs of computer-generated walkthrough animations (with the same visual content within the pair) displayed at five different frame rates, in all possible combinations. Both walkthroughs in each test pair were either silent or accompanied by sound effects and the participant had to decide which one had a smoother motion. A significant effect of sound effects on the perceived smoothness was revealed. The participants who watched the audiovisual walkthroughs gave more erroneous answers while performing their task compared to the subjects in the "No Sound" group, regardless of their familiarity with animated CG. Especially the unfamiliar participants failed to notice motion smoothness variations which were apparent to them in the absence of sound. The effect of the type of camera movement in the scene (translation or rotation) on the viewers' perception of the motion smoothness/jerkiness was also investigated, but no significant association between them was found. Our results should lead to new insights in 3D graphics regarding the requirements for the delivered frame rate in a wide range of applications.

CR Categories: I.3.3 [Computer Graphics]: Picture/Image Generation— [I.3.7]: Computer Graphics—Three-Dimensional Graphics and Realism - Animation

Keywords: 3D graphics, perceptually-adaptive techniques, cross-modal interactions, auditory-visual interactions, frame rate, perceived smoothness, animations

*e-mail: mastorop@cs.bris.ac.uk

1 Introduction

Despite the performance of modern computer graphics hardware, it is still not possible to achieve the rendering/display of a complex 3D scene on a single computer at reasonable frame rates, let alone the real-time performance of 25 frames per second and above as demanded by games and virtual reality systems. A trade-off thus exists between spatial image quality and temporal quality (frame rate).

In the real world there is an intimate linkage between sound and visual stimuli and this has started to be taken into consideration by the developers of modern computer graphics applications, although most emphasis is still put on the visual domain. A number of researchers have investigated cross-modal interactions between vision and audition. The general conclusion from the research in auditory-visual perception is that audio stimuli can potentially attract a part of the user's attention away from the visual stimuli, resulting in the reduced cognitive processing of the latter (see, for example, [Massaro and Warner 1977; Tellinghuisen and Nowak 2003]. A common example of this is turning down the radio in a car while looking for a particular street sign. Furthermore, Welch et. al [Welch et al. 1986] and Recanzone [Recanzone 2003] found that the perceived rate of an audiovisual stimulus is determined primarily by audition.

The study presented in this paper was inspired by the research findings on crossmodal perception. The results of the experiment that we conducted confirm that in the presence of audio stimuli, and more specifically sound effects, viewers fail to notice variations in the motion smoothness between walkthrough animations displayed at different rates, which are apparent in the absence of sound (Figure 1). This is probably due to the fact that the auditory stimuli attract part of viewer's attention to the sound and away from the visual defects, such as jerky motion, which result from low frame rates.

These findings may have major implications for the developers of applications where the delivered frame rate is one of the QoS (Quality of Service) parameters. These applications include, but are not limited to, desktop VR systems, 3D games and simulation applications, especially the multiuser networked ones, as the savings in the computational time and the transmission resources by exploiting this observed auditory bias of visual attention can be dramatic. We would not suggest the application of these findings to full-scale virtual environments without further detailed investigation, as variations in the frame rate can cause motion sickness. In addition, video compression algorithms, which have started to take into account perceptual issues for low bit-rate applications, would

also benefit from our findings. By encoding frame-rate control, the sudden frame skipping which results from existing techniques and degrades motion smoothness significantly, could be reduced.

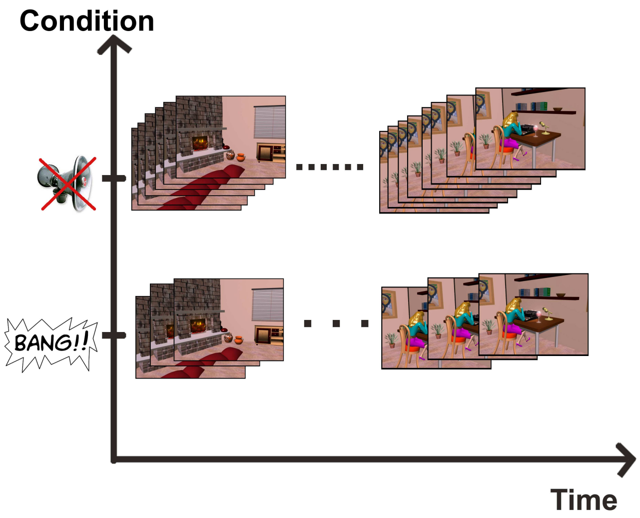

Figure 1: With the use of sound effects, fewer frames may be displayed per second compared to a silent animation, without any noticeable difference in the motion smoothness.

2 Previous Work

Perception, and in particular visual perception, is becoming increasingly important in computer graphics. Knowledge of the human visual system has been used to improve the quality of the displayed image, for example [Ramasubramanian et al. 1999; Pattanaik et al. 2000; Myszkowski et al. 2001]. Other research has shown how images can be selectively rendered without perceptual difference to the user, using level-of-detail, peripheral vision, saliency and visual tasks, for example [Luebke and Hallen 2001; Watson et al. 2001; Yee et al. 2001; Cater et al. 2003]. A review on the latest advances in perceptually adaptive computer graphics can be found in O'Sullivan et. al [O'Sullivan et al. 2004].

Despite the recent growth in the development of perceptually adaptive graphics techniques which employ perceptual criteria to reduce the computational complexity, researchers have restricted their focus on the visual stimuli and have not as yet taken into consideration the auditory-visual crossmodal interactions, although it is well known that stimuli reaching the various senses are, in general, not processed independently. For instance, Storms investigated crossmodal interactions between auditory-visual stimuli in VR and found that high-quality sounds coupled with high-quality visual stimuli increase the perceived quality of the visual displays [Storms 1998].

Winkler and Faller (2005), as well, investigated the factors which affect the evaluation of audiovisual quality and found that both audio and video quality contribute significantly to the perceived audiovisual quality [Winkler and Faller 2005]. ??? ADD PREVIOUS WORK IN AUDIO FOR GRAPHICS HERE ???

The spatial and temporal compression artifacts of coded video have recently been studied intensively. It is difficult to support both good spatial and temporal quality at very low bit rates. Up until now, developers have mainly opted for the degradation of the spatial quality under a fixed frame rate. Apteker et. al (1995) examined the effects that degrading frame rates have on user perception of a video application and showed that the perceived differences depend on the nature of the application [Apteker et al. 1995]. According to Song et. al (2001), "more flexible and robust rate control is needed under time-varying communication channels, such as the Internet, and under these environments, variable-encoding frame-rate control can provide a satisfactory solution" [Song and Kuo 2001]. They developed a variable-encoding frame rate control scheme which pursues an efficient tradeoff between spatial and temporal qualities.

2.1 Auditory-Visual Interactions on Attention and Processing Resources

A human's senses are stimulated by external factors, such as visual and auditory stimuli, and that stimulation results in the perception of the environment. As we cannot attend to all stimuli at once, attention filters and brings part of the information about the world around us into awareness. Broadbent concluded that although a large amount of sensory information can be absorbed simultaneously, a selective filter (attention) reduces the input from one source while that from another source is analysed by the brain [Broadbent 1958]. Attention, thus, determines which sensory stimuli will be further mentally processed, and which will be ignored [Carlson 1993].

In the area of cross-modal interaction between vision and audition in attention, Massaro and Warner (1977) showed there is a small but significant limitation of attentional capacity during visual and auditory perception [Massaro and Warner 1977]. Tellinghuisen and Nowak (2003) investigated the ability to ignore visual and auditory distractors presented during a visual search task and concluded that auditory distractors are processed regardless of visual perceptual load and also that the ability to inhibit cross-modal influence from auditory distractors is reduced under high visual load [Tellinghuisen and Nowak 2003].

In addition, Welch et. al [Welch et al. 1986] found that the perceived rate of an audiovisual stimulus is determined primarily by audition. Recanzone (2003) replicated this finding and concluded that it provides support for a *modality appropriateness* hypothesis, according to which the modality that is most appropriate with respect to a given task is the modality that influences most the perception in the context of that task. Audition has a higher temporal resolution and therefore, during temporal rate judgements (for instance, while watching an animation) the observer will put more weight on the auditory than the visual stimuli, without realising it [Recanzone 2003].

3 The Experiment

From the research findings in multimodal perception we can infer that the redirection of attention and the allocation of cognitive resources to the processing of sound effects while watching rendered animations, may reduce the viewer's cognitive resources allocated to the processing of the visual cues employed in judging the motion smoothness / jerkiness.

We, hence, hypothesized that it would be more difficult for subjects to distinguish differences in the motion smoothness between audiovisual composites than between silent animations. Two conditions were considered: "Sound Effect" and the control, "No Sound", condition. We also hypothesized that familiarity with animated computer graphics would help the corresponding subjects perform the

experimental task more efficiently than participants without any prior experience.

3.1 Participants

Forty participants, of ages ranging from 21 to 32, from the undergraduate and postgraduate student population volunteered to participate in this study. The participants were initially divided into two groups according to their familiarity with animated computer graphics, as we also wanted to investigate whether the user's experience would affect the degree of the "distractive" influence of the sound. The first group included the subjects who had attended a Computer Graphics course and were moderately/very familiar and the second group included participants whose studies were irrelevant to Computer Science and had no or little experience in animated computer graphics. The members of each of these two groups were randomly subdivided across the two conditions. Participants were informed that they could withdraw at any time during the experiment and they were naive as to its purpose. They all had either normal or corrected-to-normal vision and they did not report any hearing impairment.

3.2 Design

For our experiment we used an independent samples design. The dependent variable was the perceived relative motion smoothness of the two animated sequences in each test pair. The independent variable was the auditory background of the movie clips (sound effects or silence). The conditions tested are shown in Figure 2.

During each session the subjects watched pairs of clips depicting walkthroughs, one after the other, and they had to judge in which of the two the motion was smoother. The two clips in each test pair had the same visual content, but were rendered at varying frame rates (either different or the same within each pair).

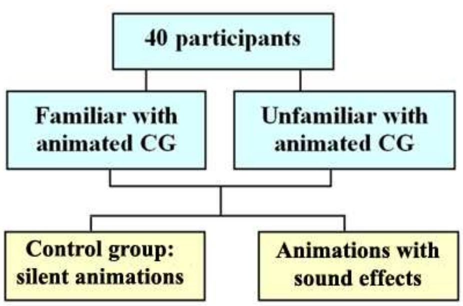

Figure 2: The Conditions tested

3.3 Equipment and materials

The test environment comprised a PC placed on a desk in an empty room, so that the subjects would not be distracted by surrounding objects. The subjects watched the walkthroughs full-screen on the 19" CRT monitor of the PC (resolution: 1280×1024 pixels). They were seated at normal viewing distance from the monitor (≈ 60 cm). Auditory stimuli were presented through quality headphones, with frequency response 18 - 22 KHz, isolated from outside noise. The volume of the sound remained the same for all subjects of the "Sound Effect" condition.

The experiment was preceded by a series of exploratory studies that would help us decide on the selection of the audio and visual stimuli materials used in our main study.

All the clips used in our experiment were based on six animated sequences of images (rendered at 640 x 480 pixels resolution), depicting 6 distinct parts of a walkthrough in a 3D interior scene (see example frames in Figure 3). We used six different sequences in order to reduce the boredom and fatigue which may result from watching the same set of visual stimuli over and over again.

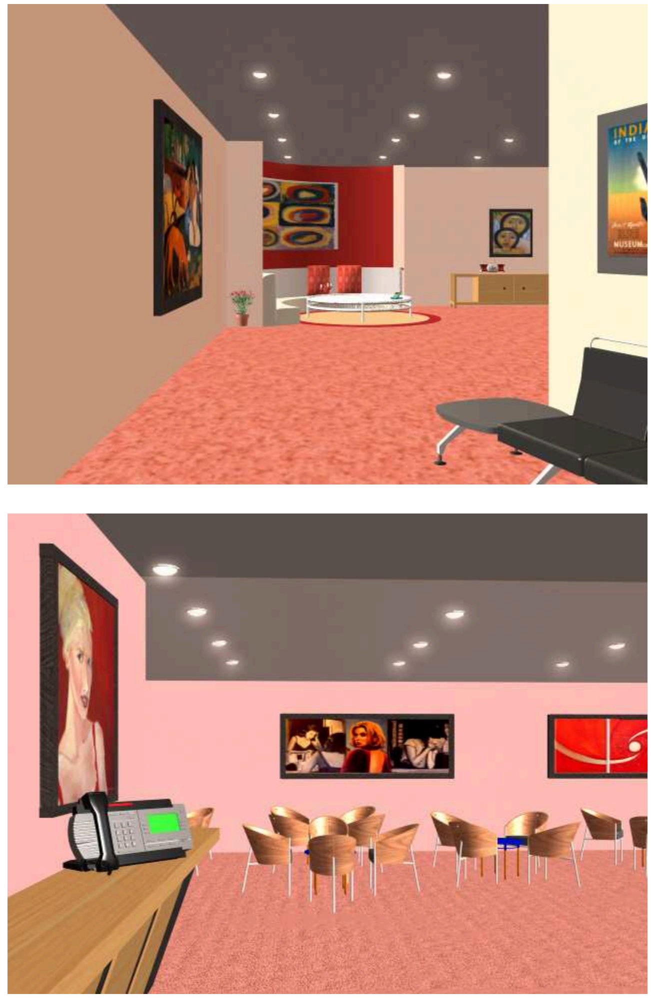

Figure 3: Example frames from the animated sequences

In each of the six animated sequences we used, the camera employs one of the following two kinds of camera motion: a) translation along the x-axis (walk through, forward type of movement) or b) rotation around the y-axis. In half of the sequences we used the first type of camera motion and in the rest the second type of camera movement was employed. We decided to further investigate whether the type of camera motion would make any difference to the results, as we expected the second type of motion to accentuate the motion jerkiness caused by reduced frame rates more than a "walk through" type of motion.

Each animated sequence was rendered at 5 different frame rates to produce the movies used in the experiment: 24 fps, 20 fps, 15 fps, 12 fps and 10 fps, giving 25 frame rate combinations. We did not include more frame rates, as the additional test pairs would prolong significantly each experimental session. The 25 rate pairs coupled

with the 6 animated sequences gave a total number of 150 test pairs. In 30 of the pairs the two clips were identical and they were included as a means of gauging each participant's performance. All paired comparisons were randomly generated. Each movie clip lasted for 3 seconds, in order to keep the duration of the experimental session to a minimum. A pilot study confirmed that 3 seconds were enough for a viewer to judge the motion smoothness of a test animation.

Three sound effects were selected to compose the auditory background for the "Sound Effect" condition: phone ringing, cell phone beeping and thunder. The first two sound effects were related with the visual content (phone or cell phone present in the scene) and the third was unrelated (ambient sound effect), as we wanted also to investigate whether the distracting influence of a sound depends on the visibility of the object emitting it or not. The sound files were synthesised by manipulating (multiplying and stretching) relevant freeware sounds. The produced sound effects were equated for peak amplitude and their properties were: 44100Hz, 16 bit, Stereo. The movie clips in each test pair had the same auditory background. For the audiovisual composites, each sound effect was assigned to two of the six rendered image sequences: one for each type of camera motion, in order to counterbalance for possible interactions between one of the sound effects and a specific type of camera movement.

Due to the prohibitive memory requirements of loading 150 uncompressed clips (together with the countdown sequences), we had to compress the video clips. Cinepac Codec by Radius was used for the compression, because it gave the best results, compared to other codecs, for our sequences. It was not possible to avoid completely the blurriness and flickering which results from the video compression, but the same artefacts were present in all conditions and therefore they should not affect our results.

3.4 Procedure

Each participant was tested individually. Participants were informed that they should watch carefully pairs of computer-generated walkthroughs with varying audiovisual content. The subjects in the "Sound Effect" group were told that the animations would be accompanied by sound effects and the sound would be delivered to them through headphones. Even the participants of the "No Sound" group had the headphones on during their experimental task, so as to be better isolated from outside noise. When each pair finished they would have five seconds before the next pair loaded to answer the question: "Which of the two movies in the test pair you just watched do you think had a better visual quality taking into consideration only the motion smoothness or on the contrary jerkiness?"

For each pair, the participants could select one of: "The first seemed better", "The second seemed better". They were instructed that they should pick one of the options even when they could not perceive any difference between the motion smoothness of the two movies (2-Alternative Forced-Choice method, 2AFC) [Abbey et al. 1999]. Visual signals indicated the beginning of the two clips within each check trial and a count down was displayed between consecutive trials so the participants knew exactly how much time was left for them to give their answer. During the five-second countdown there was silence for all conditions.

Before the actual experimental task, all participants received a familiarization phase, during which they watched a training sample that consisted of sample pairs of movies (divided by the countdown periods) of varying frame rate difference within pair. The training sample was played as many times as each participant wished and during its playback he/she received instructions from the experimenter about the visual cues that would help him/her perform the task.

Each experimental session lasted 35-40 minutes. To minimise the effect of fatigue/boredom (due to the repeated watching of the same visual stimuli) on the results, the subjects were instructed that they could pause the display during any countdown interval and continue when they felt ready again. Less than five participants chose to pause the experiment, and this happened only once during their experimental task.

Each questionnaire concluded with two questions about how relaxed/comfortable the participant felt and how focused he/she was while watching the animations: a) at the beginning and b) towards the end. They could select one of the following options: Not at all / A little bit / Moderately / Very much. These questions were used to check whether any change in the focus or comfort levels would affect the subjects' performance.

4 Results

Measure of performance in our experimental task was the percentage of times each subject correctly identified the smoother animation (i.e higher frame rate) within a pair of displayed animations. The performance was averaged for each pair of frame rates across all subjects within each group. For example, a performance of 100% for a pair of frame rates within a group indicates that all subjects from this group correctly identified the clip with the smoother motion whenever they came across the corresponding pair of rates during the experimental task. Figures 5 and 6 illustrate the performance of Familiar versus Unfamiliar participants within each condition and Figure 7 compares the performances measured across the two conditions.

We decided to present and analyse the results for the actual frame rate combinations (e.g 10 vs 15 fps) and not to consider only the difference between a pair of rates (e.g. 5 fps difference), because the perceptibility of visual defects is reduced when the absolute frame rate values are close to 20-24 fps, compared to lower values.

As we have already mentioned, in 30 of the test pairs the two animations were identical and they were included as a means of gauging each participant's performance. If the experiment was designed and conducted properly, the participants should pick one of the two options randomly for these pairs and, therefore, in our results we should get each of the two possible answers in approximately 50% of the cases. The actual percentage in our results was 50.5%, proving that our experimental results were valid and could be statistically analysed further.

From Figures 4 and 5 it is clear that the performance of subjects who were familiar with computer graphics in detecting the animation that was displayed at the higher frame in each pair of animations, was generally better than the performance of the unfamiliar participants across both conditions. This is in accord with the second part of our hypothesis that familiarity would be an affecting factor on the participants' performance for the specific experimental task.

As we can see in Figure 7, the control, "No Sound", group consistently gave more correct answers than the other group. Figure 6 reveals that both Familiar and Unfamiliar subjects contributed to the performance drop of the "Sound Effect" group. Therefore, not even the viewers who were very familiar with animated computer graphics escaped the influence of the sound effects. Further statistical analysis of the results was nevertheless necessary, in order to

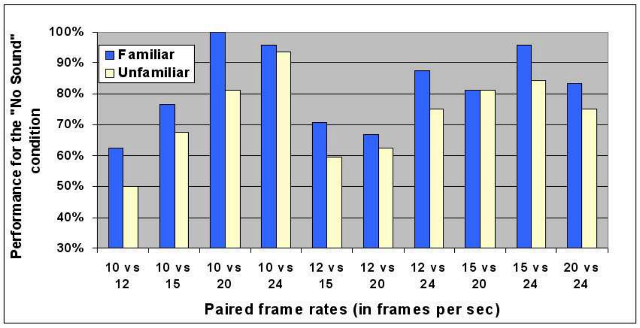

Figure 4: The performance of Familiar vs. Unfamiliar Subjects for the control ("No Sound") condition across the test frame rate pairs.

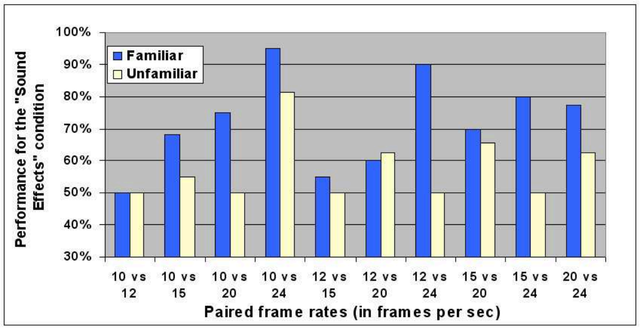

Figure 5: The performance of Familiar vs. Unfamiliar Subjects for the "Sound Effect" condition across the test frame rate pairs.

find out whether this drop in the performance of the "Sound Effect" group was significant and thus whether we should accept our initial hypothesis that sound effects in an animation make it more difficult for the viewer to detect frame rate variations.

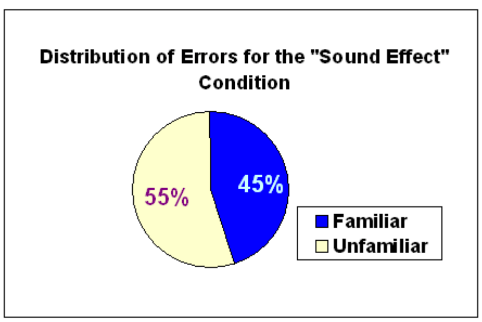

Figure 6: Distribution of erroneous answers between Familiar and Unfamiliar subjects for the "Sound Effect" condition.

We first had to decide whether we should use parametric or non-parametric tests for the statistical analysis of the results. We were not certain whether the distribution of our population was Gaussian, but according to the *Central Limit Theorem* if the samples are large enough, the distribution of means will follow a Gaussian distribution even if the population is not Gaussian. Assuming the population doesn't have a really peculiar distribution, a sample size of 10 is generally enough to invoke the Central Limit Theorem. Since most parametric tests, such as the t-test and ANOVA, are concerned only with differences between means, the Central Limit Theorem lets these tests work well even when the populations are not Gaussian.

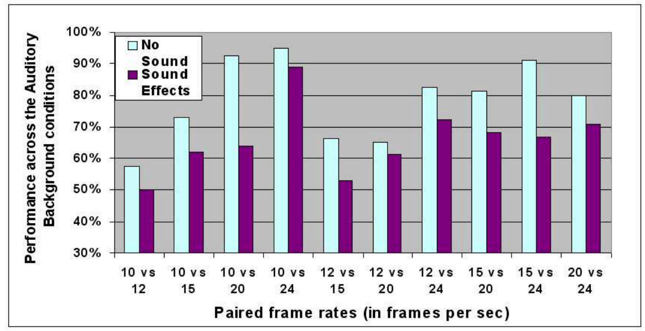

Figure 7: The performance of all Subjects across the "No Sound" and "Sound Effect" conditions, separately for each frame rate combination.

A two-way ANOVA (ANalysis Of VAriance) for independent samples was first carried out in order to investigate if the performance in the task was jointly influenced by the auditory background and the familiarity of the subjects with animated computer graphics (i.e. whether there is an interaction between these two independent variables). A two-factor analysis of variance consists of three significance tests: a test of each of the main effects of the two independent variables and a test of the interaction between them. The main effect of the auditory background was found to be very significant, even at the 0.0001 level of risk (F=26.51, df=1 and thus P=0.00001 $<$0.0001). The main effect of the familiarity with animated CG was also very significant, even at the 0.00001 level of risk (F=29.41 df=1, P=0.000004 $<$0.00001), revealing that prior experience with computer graphics has a major influence on the ability of the viewer to perceive temporal defects which result from low frame rates. Furthermore, a significant interaction between these two independent variables (auditory background and degree of familiarity) was revealed at the 0.05 level of risk (F=5.13 df=1, P_{rxc}=0.0296 $<$0.05).

The data were then analysed by carrying out an unpaired t-test for 2 Independent Samples [Box et al. 1978] between the means of our independent "No Sound" and "Sound Effect" conditions, separately for Familiar and Unfamiliar subjects, in order to determine whether or not there was a significant between-subjects performance difference as a function of the auditory background only.

The t-test gave a statistically significant result for both Unfamiliar and Familiar participants at the 0.001 and 0.01 levels of risk, respectively. More specifically, for Familiar subjects t=2.9469 df=18 ($N_{Familiar}$=20) and $P_{one-tailed}$=0.004312. For Unfamiliar subjects ($N_{Unfamiliar}$=20) t=4.4764, df=18 and $P_{one-tailed}$=0.000146. Therefore, the null hypothesis that there is no difference between silence and sound effects regarding the ability of the viewer to detect smoothness/jerkiness variations, could be rejected.

We further analysed the results from the "Sound Effect" group to examine whether the distracting influence of a sound depends on the presence in the scene of the object emitting it, but no statistically significant effect of the type of sound, scene-related or unrelated, on our subjects' performance was found.

In Figure 8, we can see the performance of all subjects from both conditions for the two types of camera motion, translation and rotation. Another two-way ANOVA for independent samples was carried out, in order to investigate the main effect of the type of camera movement on the subjects' in the task performance and also if there was a significant interaction between the type of camera motion and the auditory background. The main effect of type of camera movement was not significant (F=0.51 df=1, P=0.4743), and therefore the null hypothesis, that there is no difference between the "trans-

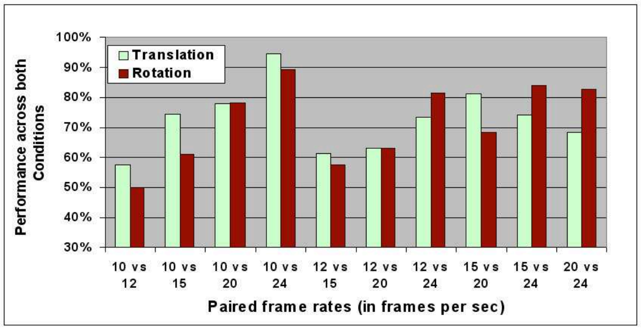

Figure 8: The performance of all subjects across the 2 types of camera motion (translation and rotation).

lation" and "rotation" types of movement in the scene regarding the ability of the viewer to detect smoothness/jerkiness variations, could not be rejected. This was a bit surprising as we had expected that participants would detect temporal artefacts more easily for the animations where the camera was rotating. Furthermore, no interaction between the auditory background of the scene and the type of camera movement was revealed (F=1.52 df=1, P_{rxc}=0.2214). The main effect of the auditory background was again found to be very significant, even at the 0.0001 level of risk (F=23.71, df=1 and thus P=0.000006 $<$0.0001).

During the design of our experiment we were concerned about the influence of the familiarisation to the three sound effects on their distracting influence, and more specifically that after a number of animation pairs the viewers would get used to the sounds and their attention would not be "attracted" by them. Nevertheless, the frequency of the correct answers did not increase during the course of the experimental task for the vast majority of the participants, revealing that the influence of the sound effects remained significant throughout the task, despite the fact that they were no longer novel to the viewers.

Some participants had reported reduced focus towards the end of their experimental session, but the pattern of their answers revealed no affect of focus on correctness, and therefore we cannot attribute reduced performance to fatigue and/or boredom.

We also analysed our results using frequencies (percentages) because Frequency Analysis can provide informative data on the levels of observed differences between participants. The results in percentages indicate, amongst others, the following.

Unfamiliar subjects in the "Sound Effect" condition could identify with ease only the difference between 10 and 24 frames per second, in 81.25% of the cases. Their performance for almost all the other frame rate pairs (10 vs. 12 fps, 12 vs. 15 fps, 10 vs. 15 fps, 15 vs. 24 fps, 10 vs. 20 fps, 12 vs. 24 fps) was just 50%-55%, which indicates that they gave random answers when they encountered those pairs while performing their experimental task. On the other hand, Unfamiliar subjects in the "No Sound" condition had difficulty in "identifying" the frame rate difference only for the pairs "10 vs. 12 fps" and "12 vs. 15 fps".

Familiar subjects in the "Sound Effect" group could not distinguish between animations displayed at frames rates with difference 2-5 frames per second, while the Familiar subjects in the "No sound" group could generally distinguish between clips with difference greater than 2 frames per second.

The results presented above confirm our hypothesis that sound effects can significantly affect the viewers' ability to perceive motion smoothness/jerkiness variations which result from changes to the frame rate, regardless of the type of camera movement in the scene (translation or rotation). The observed influence of the sound does not depend on whether the sound affects are ambient sounds (unrelated to the scene content) or they are related to specific objects in the 3D scene. Users who are unfamiliar with computer graphics are affected more than the familiar viewers by the presence of sound effects and they do not detect smoothness/jerkiness variations which are obvious when there is no sound.

5 Conclusions

The delivery of high-fidelity graphics for interactive graphics and multimedia applications, such as simulations, 3D games and VR environments, remains one of the major challenges for computer graphics practitioners, despite the huge progress in the related hardware and software during the past few years. For such applications, limitations of the human visual perception and cross-modal interactions on visual attention can be exploited in order to reduce the computational complexity and make more efficient use of the available resources, while trying to minimise the perceptibility of the resulting visual defects.

In this paper we showed that, audio stimuli do indeed attract a part of the viewers' attention away from the visuals and as a result, the observers find it more difficult to distinguish smoothness variations between audiovisual composites displayed at different rates, than between silent animations. We also demonstrated that viewers who are not familiar with animated computer graphics, can much harder notice variations in the motion smoothness between two audiovisual animations, compared to people with prior experience. Our findings could have major implications for the applications where the delivered frame rate is one of the QoS (Quality of Service) parameters.

References

ABBEY, C. K., ECKSTEIN, M. P., AND BOCHUD, F. O. 1999. Estimation of human-observer templates for 2 alternative forced choice tasks. In *Proceedings of SPIE*, vol. 3663, 284–295.

APTEKER, R. T., FISHER, J. A., KISIMOV, V. S., AND NEISHLOS, H. 1995. Video acceptability and frame rate. *IEEE Multimedia 2*, 3 (Fall), 32–40.

BOX, G. E. P., HUNTER, W. G., AND HUNTER, J. S. 1978. *Statistics for Experimenters*. New York: Wiley.

BROADBENT, D. 1958. *Perception and Communication*. London: Pergamon Press.

CARLSON, N. R. 1993. *Psychology, The Science of Behavior*. Needham Heights, MA: Allyn and Bacon.

CATER, K., CHALMERS, A., AND WARD, G. 2003. Detail to attention: Exploiting visual tasks for selective rendering. In *Proceedings Eurographics Rendering Symposium, Leuven.*

LUEBKE, D., AND HALLEN, B. 2001. Perceptually driven simplification for interactive rendering. In *Proceedings of 12th Eurographics Workshop on Rendering*, 221–223.

MASSARO, D. W., AND WARNER, D. S. 1977. Dividing attention between auditory and visual perception. *Perception & Psychophysics 21*, 569–574.

MYSZKOWSKI, K., TAWARA, T., AKAMINE, H., AND SEIDEL, H.-P. 2001. Perception-guided global illumination solution for animation rendering. In *Proceedings of SIGGRAPH 2001*, E. F. (Ed.), Ed., Computer Graphics Proceedings, Annual Conference Series. ACM Press / ACM SIGGRAPH, New York, 221–230.

O'SULLIVAN, C., HOWLETT, S., MORVAN, Y., MCDONNELL, R., AND O'CONOR, K. 2004. Star report on perceptually adaptive graphics. In *Eurographics 2004, Eurographics Association*.

PATTANAIK, S., TUMBLIN, J., YEE, H., AND GREENBERG, D. 2000. Time-dependent visual adaptation for realistic real-time image display. In *Proceedings of SIGGRAPH 2000*, E. Fiume, Ed., Computer Graphics Proceedings, Annual Conference Series. ACM Press / ACM SIGGRAPH, New York, 47–54.

RAMASUBRAMANIAN, M., PATTANAIK, S., AND GREENBERG, D. 1999. A perceptually based physical error metric for realistic image synthesis. In *Proceedings of SIGGRAPH 1999*, A. Rockwood, Ed., Computer Graphics Proceedings, Annual Conference Series. ACM Press / ACM SIGGRAPH, New York, 73–82.

RECANZONE, G. H. 2003. Auditory influences on visual temporal rate perception. *Journal of Neurophysiology 89*, 1078–1093.

SONG, H., AND KUO, C. C. J. 2001. Rate control for low-bit-rate video via variable-encoding frame rates. *IEEE Transactions on Circuits and Systems for Video Technology 11*, 4, 512–521.

STORMS, R. L. 1998. *Auditory-Visual Cross-Modal Perception Phenomena*. PhD thesis, Naval Postgraduate School, Monterey, California.

TELLINGHUISEN, D. J., AND NOWAK, E. J. 2003. The inability to ignore auditory distractors as a function of visual task perceptual load. *Perception & Psychophysics 65*, 5, 817–828.

WATSON, B., FRIEDMAN, A., AND MCGAFFEY, A. 2001. Measuring and predicting visual fidelity. In *Proceedings of SIGGRAPH 2001*, E. Fiume, Ed., Computer Graphics Proceedings, Annual Conference Series. ACM Press / ACM SIGGRAPH, New York, 213 – 220.

WELCH, R. B., DUTTONHURT, L. D., AND WARREN, D. H. 1986. Contributions of audition and vision to temporal rate perception. *Perceptual Psychophysiology 39*, 294–300.

WINKLER, S., AND FALLER, C. 2005. Audiovisual quality evaluation of low-bitrate video. In *SPIE/IS&T Human Vision and Electronic Imaging*, vol. 5666, 139–148.

YEE, H., PATTANAIK, S., AND GREENBERG, D. 2001. Spatiotemporal sensitivity and visual attention for efficient rendering of dynamic environments. In *ACM Transactions on Computer Graphics*, M. Cohen, Ed., Computer Graphics Proceedings, Annual Conference Series, vol. 20, 39–65.