

Maintaining Frame Rate Perception in Interactive Environments by Exploiting Audio-Visual Cross-Modal Interaction

Vedad Hulusić · Kurt Debattista · Vibhor Aggarwal · Alan Chalmers

Received: date / Accepted: date

Abstract The entertainment industry, primarily the video games industry, continues to dictate the development and performance requirements of graphics hardware and computer graphics algorithms. However, despite the enormous progress in the last few years it is still not possible to achieve some of industry’s demands, in particular high-fidelity rendering of complex scenes in real-time, on a single desktop machine. A realisation that sound/music and other senses are important to entertainment, led to an investigation of alternative methods, such as cross-modal interaction in order to try and achieve the goal of “realism in real-time”. In this paper we investigate the cross-modal interaction between vision and audition for reducing the amount of computation required to compute visuals by introducing movement related sound effects. Additionally, we look at the effect of camera movement speed on temporal visual perception. Our results indicate that slow animations are perceived as smoother than fast animations. Furthermore, introducing the sound effect of footsteps to walking animations further increased the animation smoothness perception. This has the consequence that for certain conditions the number of frames that need to be rendered each second can be reduced, saving valuable computation time, without the viewer being aware of this reduction. The results presented are another step towards the full understanding of the auditory-visual cross-modal interaction and its importance for helping achieve “realism in real-time”.

Keywords cross-modal · perception · high-fidelity rendering

V. Hulusić
International Digital Laboratory, WMG, University of Warwick,
Coventry, UK
E-mail: v.hulusic@warwick.ac.uk

1 Introduction

Rendering realistic high-fidelity graphics in real-time is a computationally expensive process, that is still unachievable for complex scenes even on high-end desktop machines. Furthermore, in virtual environments, such as video games, stimulation of auditory, and possibly some other senses, is often also required. This, however, does not necessarily need to be considered as an additional work load, but instead, can be exploited, so that the overall work load is balanced or even reduced, without any perceivable loss in quality. This is possible due to various limitations of the Human Sensory System (HSS). One such limitation is the influence of one sensory input on another, commonly termed cross-modal interaction. One particular cross-modal effect, which has been successfully exploited in the field of computer graphics is that of vision and audition [31, 17–19]. Mastoropoulou et al. demonstrated how cross-modal effects can be combined with selective rendering techniques to speed up the rendering process [33]. Subsequent work demonstrated how sound effects could be used as a distracter to reduce the computed frame rate of an animation without the participants perceiving any difference in the visual quality of the animation [34]. Hulusic et al. achieved similar results using rhythmically significant audio, played at different beat rates [18].

Although it has been shown that cross-modal interaction exists and it can be utilised in computer graphics to speed up rendering, the influence of movement related sound effects on perceived temporal visual quality of animations has only recently been considered. Hulusic et al. [19] showed that camera movement speed in a virtual environment influences visual perception in relation to the audio condition, and that a direct rela-

tionship between movement related sound effects and frame rate perception could exist. In this paper, we extend this early work by investigating the efficacy of such cross-modal methods using an effect, strongly associated with the movement being performed. We conducted a detailed psychophysical experiment to determine whether the sound effect of footsteps affects the perception of the smoothness of an animation, under two movement conditions: running and walking. Our results show a strong indication of the effect in walking animations. The difference between discrepant frame rates was less when lower frame rates were accompanied by the sound effect of footsteps. These results have implications for interactive virtual environments, such as video games. Once understood and harnessed correctly, the rendering engine could introduce movement related sound effects in order to decrease or balance its workload as necessary.

The rest of the paper is organised as follows. Section 2 gives an overview of the previous work done on this topic both in psychology and computer graphics. In Section 3, we explain the design, methodology, procedure and study question of the psychophysical study. The results are presented in Section 4. Finally, in Section 5 we discuss and conclude the paper and provide some directions for future work.

2 Related work

Human perception of individual senses has been studied for more than a hundred years [21]. A number of limitations of the Human Visual System (HVS) have been identified in such studies. These include Inattentional Blindness [29,47] in which objects, although being in plain sight, may remain unperceived when not the focus of attention. Similarly, the Human Auditory System (HAS) has various limitations, such as the Continuity Illusion phenomenon [26]. In this case, a discontinued audio is perceived as it has not been interrupted. Another phenomenon related to the HAS is auditory masking [37], also known as the *cocktail party effect*. This is a person’s ability to isolate the voice of a single talker while masking all the noise coming from the environment. Additionally, there are factors which effect both sound and vision such as angular sensitivity and the internal spotlight phenomenon [20,21].

In the last few decades, another area of interest in psychology has been the cross-modal interaction between vision and audition [13,16,35,43,38,41]. To date there has been little work to exploit the potential such cross-modal interactions may have in computer graphics.

2.1 Auditory-visual cross-modal interaction research in psychology

Although it has been shown that cross-modal interaction exists between different modalities such as smell and vision [40], or touch and vision [25], in this paper we focus solely on the auditory-visual cross-modal interaction. A highly important feature of the cross-modal interaction is that it works in both “directions”. Audition influences visual perception and vision can also have an effect on auditory perception. The foundation in this area of research was presented by Welch and Warren [53], who introduced the *modality appropriateness hypothesis*. This hypothesis states that the modality that is more appropriate for a certain task dominates the perception in that particular task. In other words, vision dominates perception in the spatial domain because of its higher acuity, while in the temporal domain audition is the dominant modality.

A common example of the visual dominance over audition is the ventriloquism effect [16,9,50,51]. This effect shows that humans, while watching TV or a puppet show, associate a sound source to a speaking person/puppets mouth even though it originates from the speaker/ventriloquist positioned at a different location. Another example is the McGurk effect [35], where a sound of /ba/ is perceived as /da/ when accompanied with lip movement corresponding to the pronunciation of /ga/.

The effect of audio on visual perception is of more interest for the work presented in this paper. As a direct consequence of the modality appropriateness hypothesis, our focus has been on the temporal domain. For example, the auditory driving effect shows that, when presented simultaneously, sound *drives* vision in the temporal domain [13,46,52]. Another phenomenon investigated by Shams et al. [43,44] is the illusory flash induced by sound. This effect illustrates how an illusory flash can be induced by a sound beep, where, for example, we observe two flashes accompanied by three beeps. Analogous to the ventriloquism effect, Morein-Zamir et al. [38] investigated temporal ventriloquism. This effect shows that the time between two light flashes is perceived longer when a sound is presented before the first and after the second light, but shorter when there are two sound beeps between the lights. Some other examples can be found in [42,23,41,45,14].

The limitations observed in cross-modal interaction might be attributed to limited attentional capacity. Two alternative theories, which look at how our attentional capacity can affect perception, have been developed: the central and divided attentional resources theories [2,30,11,6,27,1,7]. Some models of attention claim that

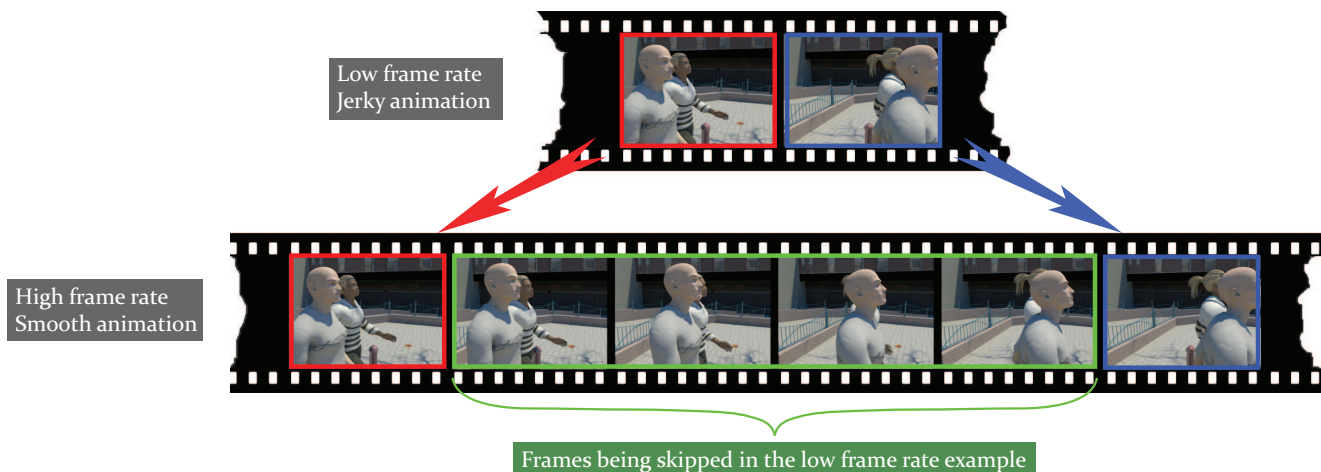


Fig. 1 Illustration of the difference between the low and the high frame rate.

our attention operates on a global level and that is divided across multiple senses. That means that the performance of a task requiring attention for one modality will be affected by concurrent task in some other modality. However, another model suggests separate attentional resources for vision and audition. It shows that, at least for low-level tasks (discrimination of pitch and contrast), there is no attentional dependencies between modalities.

2.2 Auditory-visual cross-modal interaction research in computer graphics

Perceptual methods, such as perceptually-based rendering [54,8], have over the past few years frequently been used within computer graphics, in order to speed up rendering process while keeping the same perceptual quality. Similarly, perceptual factors were considered in some research on audio rendering [24,36,49]. However, these methods focus on a single modality and the factors influencing that modality. Mastoropoulou, in her PhD thesis, was the first to investigate the cross-modal interaction between vision and audition in computer graphics field [31].

The auditory influence on visual perception can be investigated from two perspectives: the spatial and the temporal. In the former, focus is on the spatial image quality while the latter focuses on temporal quality such as frame rate perception.

Mastoropoulou et al. [33] demonstrated how sound effects can be efficiently used with selective rendering so that only sound emitting objects are rendered in high quality and the remainder of the scene in much lower quality, without any significant difference in perceived visual quality. Hulusic et al. [17] examined how related

and unrelated audio influences visual perception and showed that unrelated sound can be used for increasing the perceptual quality of static images.

Mastoropoulou and Chalmers [32] investigated how music can influence perception of frame rate and the perceived duration of a video animation. Subsequently, they investigated how sound effects, e.g. a phone ringing or a thunder clap can be used as a distracter to vision [34]. This work showed that when accompanied by these scene-unrelated sound effects it is possible to decrease the frame rate of a video animation without perceivable difference in visual quality. Most recently, Hulusic et al. [18] focused particularly on the effect of beat rate, scene and familiarity on the perception of frame rate. They showed that in case of static scenes lower beat rates have a significant effect on perception of low frame rates.

More recent work using cross-modal interaction can be found in [15,4,48,5].

3 Experiments

In this study, the effect of two factors on our visual perception were investigated, by conducting two sets of experiments, see Table 1. Two tests were performed with the data from the first experiment and one test with the data from the second experiment. In the first test (*Test 1*) we looked at the effect of camera movement speed (walking and running) on temporal visual perception. Our research hypothesis was that the speed of the camera movement will affect smoothness perception. Additionally, in *Test 2*, the perceived smoothness threshold for the animations accompanied by the audio effects was investigated. Our research hypothesis was that there will be difference between the preference of

Experiment	Test	Observed effect	Compared frame rates	
Experiment 1	Test 1	Camera movement speed	Audio 60r vs 60w	NoAudio 60r vs 60w
	Test 2	Perceived smoothness threshold	Audio 10r-60w, 20r-60w, 30r-60w vs 60r-60w	
Experiment 2	Test 3	Movement related sound effect	Audio vs NoAudio	
			Run 10-20, 10-30, 10-60 20-30, 20-60 30-60	Walk 10-20, 10-30, 10-60 20-30, 20-60, 30-60

Table 1 The details of the experimental design for each test. Numbers represent frame rate, while "r" and "w" stand for *running* and *walking* animations respectively.

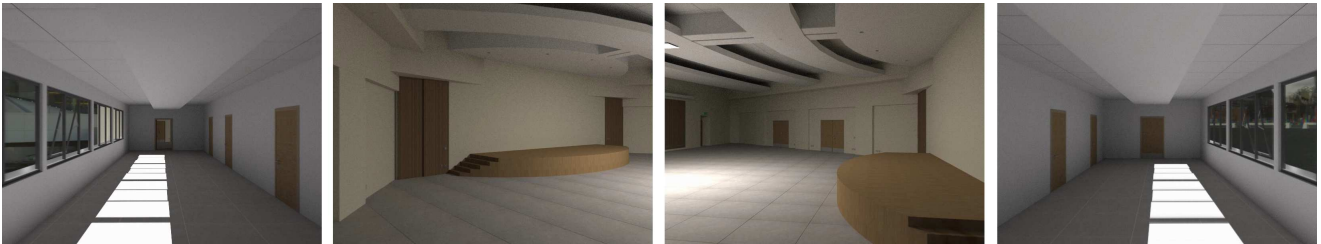


Fig. 2 Four frames taken from the walk-through animations. The first two frames are from the animations with camera moving from the corridor to the conference hall, and last two from the animations where the camera is moving from the conference hall to the corridor.

the discrepant frame rates and preference of the control condition (60r-60w). In the second experiment (*Test 3*), the movement related sound effect on the running and walking animations were investigated separately. The research hypothesis was that scene related audio effects will increase visual smoothness perception.

3.1 Design

Both experiments used a within-participant design with three independent variables: camera movement, frame rate and auditory condition. We used two camera conditions: *walking* (slow) and *running* (fast), and four different frame rates: 10, 20, 30 and 60 frames per second (fps) in different combinations, see Table 1. Audio conditions were: *Audio* (foot steps sound effect) and *NoAudio* (silent animation). The dependent variable was the perceived smoothness of the animations. This was measured using the Two Alternative Force Choice (2AFC) method in a complete randomised design. To control for fatigue and familiarity, 10 different animations of the same scene were used. For the Audio condition the foot steps sound effects were always synchronised with the visual stimulus.

3.2 Participants

In the experiments 86 people volunteered, 71 of whom were university students studying a variety of subjects, and the rest from university staff. The participants' age varied from 17 to 58 with an average age of 26. Out of 86 participants, 61 were male and 25 female. All of them had normal or corrected to normal vision. None of the participants reported any hearing impairments.

3.3 Apparatus

The experiment was conducted in a dark, quiet room with no distracters. In the first experiment the visual stimuli were presented on a calibrated 17 inch Philips 170B6 monitor with 1280×1024 pixel resolution and a refresh rate of 60 Hz. In the second set of experiments we used an LG W2234S 22 inch monitor with a refresh rate of 60 Hz and resolution 1680×1050. The stimuli were positioned at eye level, 60-70 cm from the participants eyes. For auditory stimuli an LTB Magnum 5.1 AC97 Headphone set was used.



Fig. 3 The experimental procedure. From left to right: grey box, first animation, grey box, second animation and A/B evaluation screen.

3.4 Stimuli

The visual stimuli were based only on one scene, see Figure 2. To control for familiarity, two animations, at a resolution of 800×600 , were rendered along the same path but in opposite directions, see Figure 4. The animations were rendered using our own implementation of path tracing [22]. All scenes were static with only frontal camera movement and no rotation relative to the motion path. For each of them a curved motion path with the oscillating motion of the camera along the vertical axis was used, see Figure 5. The oscillating motion was used to improve the sensation of walking in the experiment [28]. The strides in walking and running animations were 0.8m and 1.5m respectively. A young subjects' average normal walking speed of 1.425 m/s [3] was used for the walking condition. For the running condition a speed of 4 m/s was used. All videos were compressed using XviD MPEG-4 Codec (single-pass encoding, target quantizer: 3.00). The animations were divided into three *walking animations* and two *running animations* in both directions, each lasting for five seconds.

For audio, as the animation related sound effect in both camera conditions (walking and running), the sound of footsteps, produced by capturing the sound of leather soled shoes against a firm tiled floor, was used. To synchronise the sound effects with the animation, the length of the silence between the ON signals was varied. The amount of echo in the effect was adjusted according to the nature of the scene and did not change during the animations. Sounds were delivered uncompressed, using two channels (stereo), sample rate 44100Hz and bit rate of 1411kbps. We did not play any background music in order to avoid any subjective side effects. If any of such additional factors are present, the perception and therefore the results could be affected [32,39].

For the audio-visual presentation we developed a framework with support for frame rate and audio control. All results from each trial were saved in separate text files.

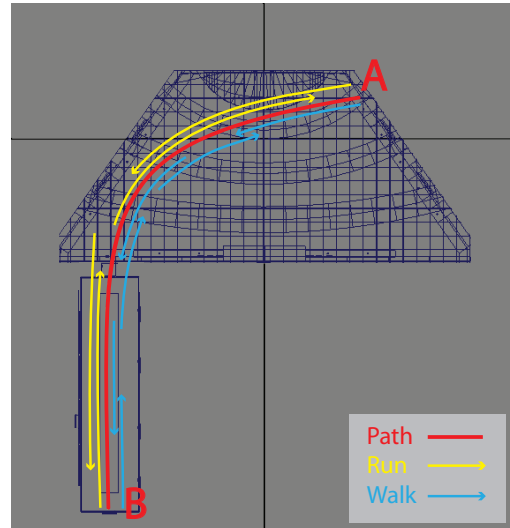


Fig. 4 Camera path used for the animations (red). Four running animation sequences (yellow) and six walking animation sequences (blue) were used.

3.5 Procedure

Prior to the experiment each participant was asked to read and sign a consent form, in which they agreed to voluntary and anonymously participate in the experiment. They were told that they could withdraw from it at any time. Participants were also given a questionnaire to fill in. Next, they were presented with instructions, followed by two sample animation pairs played at 10 and 60fps. They were told that these were the worst and the best cases respectively, but not what frame rates the animations were. In the instructions they were further explained what is frame rate using Figure 1, and that they were going to watch pairs of animations for which they will have to evaluate their smoothness. They were shown 22 and 37 pairs of animations in the first and second experiment respectively.

The randomly ordered animation pairs were presented sequentially, see Figure 3. Each animation was preceded by a grey box (RGB: 0.3, 0.3, 0.3) lasting for two seconds. The length of each animation was five seconds. After each pair the A and B boxes were shown

on the screen. Participants were instructed to choose the smoother animation by clicking on one of the two boxes, after which the next cycle started automatically. Each trial in the first experiment lasted for about six minutes and in the second experiment for about 10 minutes. The participants were debriefed on the nature of the study after the experiment.

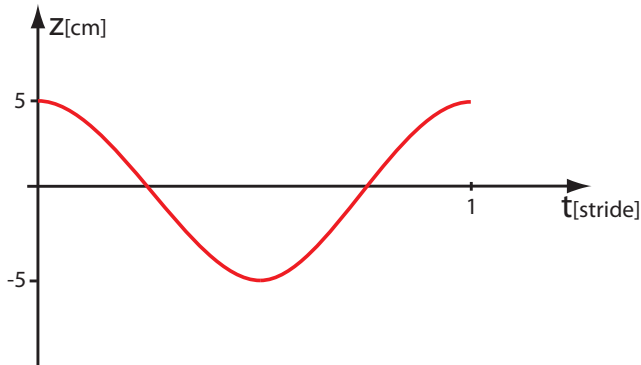


Fig. 5 Oscillating camera motion along the vertical (z) axis

4 Results

In order to test our hypotheses, the data was analysed using descriptive statistics and non-parametric tests. For Test 1 and Test 2 Chi-square test was used. Wilcoxon 2 Related Samples test was used for the Test 3.

4.1 Test 1: Camera movement speed influence on animation smoothness perception

The first research hypothesis was that the speed of the camera movement will affect the smoothness perception. The null hypothesis was that camera movement speed will have no effect on visual smoothness perception. The hypothesis was tested comparing the walking (Walk) and running (Run) animations, both played at 60fps, which was the gold standard. *NoAudio* (silent) and *Audio* condition were tested separately. The test had a single independent variable - camera movement speed.

Tabular data with observed and expected frequencies, χ^2 and p - values are given in Table 2. The Chi-square test found the relationship between Running and Walking animations for NoAudio condition as significant ($p=.020$). Therefore, the null hypothesis was rejected in favour of the research hypothesis that the speed of the camera movement affects the animation

NoAudio			
	Observed N	Expected N	Residual
Run	11	18.0	-7.0
Walk	25	18.0	7.0
Total	36		
$\chi^2(1) = 5.444, df = 1, p=.020$			

Audio			
	Observed N	Expected N	Residual
Run	14	18.0	-4.0
Walk	22	18.0	4.0
Total	36		
$\chi^2(1) = 1.778, df = 1, p = .182$			

Table 2 Test 1: Observed and expected frequencies for the Run - Walk animation smoothness perception comparison.

smoothness perception. The participants preferred the walking rather than the running animation.

The same test for Audio condition showed no significance ($p=.182$), and thus the null hypothesis cannot be rejected. The fact that the results were different for the Audio and NoAudio condition, indicates that audio might affect perception of animation smoothness. Therefore, we investigated further the influence of sound effect of footsteps on temporal visual perception in Section 4.2 and Section 4.3.

4.2 Test 2: Sound effect's influence on perceived smoothness threshold

In this test, we investigated the perceived smoothness threshold when watching the animations accompanied by the audio effects. The difference in the preferences between the discrepant frame rate pairs and the control group (60r-60w fps) were compared, see Table 1. Lower frame rates were used with the running animation in each test pair. The null hypothesis for each test pair was that animations in discrepant frame rate pairs played at 60 fps will not be perceived as smoother. Since we assume no bias, this means that each test pair will have the same preference compared to the control group.

For the analysis we used a one-tailed Chi-square test. Therefore, in order to test the validity of the one-tailed hypothesis, we compared corresponding means. The mean values for test pairs 10r-60w, 20r-60w and 30r-60w were 1.83, 1.52 and 1.63 respectively, where lower and upper bounds were 1 and 2, see Table 3. The mean value for our control group was 1.61. Since the mean value of 20w-60w condition was lower than the mean value of the control group, the null hypothesis

	Mean value	p -value
10r-60w	1.83	.032
20r-60w	1.52	N/A*
30r-60w	1.63	.5
60r-60w	1.61	N/A

Table 3 Test 2: Mean and p values for Audio condition. p -value is given for difference in preference between the test pairs and 60r-60w condition. Lower and upper bounds were 1 (first animation preferred) and 2 (second animation preferred) respectively. *Not inline with a 1-tailed test

for this pair cannot be rejected. For the 10r-60w and 30r-60w pairs, the difference was in line with our research hypothesis such that we can carry on with the test.

The results show that only for 10r-60w pair there was significant difference in preference ($p=.032$), and thus the null hypothesis can be rejected. For 30r-60w ($p=.5$) pairs there was no significant difference in preference comparing to the 60r-60w control group. Hence, the null hypothesis in this case cannot be rejected. These results appear to show that the perceived smoothness threshold when watching the animations with movement-related sound effects is somewhere between the 10 and 20 fps. This further indicates that walking animations, rendered at 60 frames per second, when accompanied by the movement-related sound effects, were not perceived significantly smoother than the same animation rendered at 20 and 30 fps.

4.3 Test 3: Sound effect’s influence on animation smoothness perception

The results from the first experiment showed that movement related sound effects could affect the perception of animation smoothness. Therefore, to test for that effect, the second experiment was conducted. Our research hypothesis was that movement related sound effects (i.e. footsteps) will increase the perception of smoothness. For the analysis, the data was first divided into two groups: run and walk. Then the Wilcoxon 2 related samples test was performed on each group separately, comparing the animation pairs given in Table 1. The animation played at higher frame rate in each pair was always played without sound (NoAudio condition), while the sound effect of footsteps was used for the animation played at lower frame rate. For example, for the pair 10-30 we compared users’ preference between the two test pairs: *10/Audio vs 30/NoAudio* and *10/NoAudio vs 30/NoAudio*.

The results showed that for the fast camera movement (running) there is no significant effect of audio

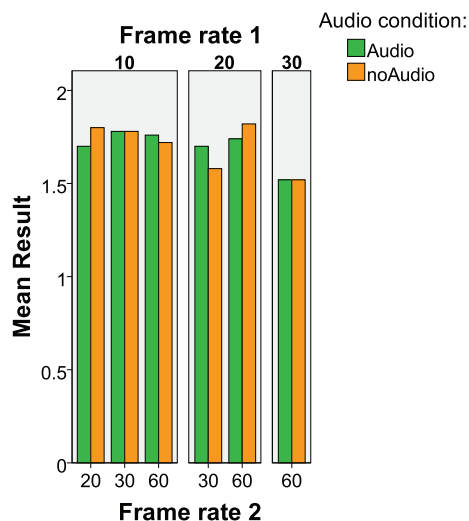


Fig. 6 Mean values of the compared running animation test pairs.

on the perceived visual quality, see Table 4. The mean values for this test are shown in Figure 6.

Statistical analysis revealed different results for the walking condition. Animations played at 10 fps with sound effects have been found as smoother than animations played at 30 ($p = .010$) or 60 fps ($p = .029$) with no audio, see Table 4. Additionally, the same test showed that animations presented at 20 fps with audio were rated as smoother than silent animations played at 30 ($p = .002$) or 60 fps ($p = .044$). For these animations pairs, the null hypothesis can be rejected, which means that movement related sound effects increase the temporal visual perception. Figure 7 depicts the mean values for the walking animations across Audio and NoAudio sound conditions.

5 Conclusion and Future Work

In this study, we investigated for the effect of three factors: the influence of camera movement speed on perceived smoothness quality of animation; the influence of movement-related audio effects on perceived smoothness threshold of video animations; and movement-related sound effect’s influence on animation smoothness perception. The three factors that could influence our results: camera movement speed, frame rate and auditory condition were considered.

The results from the analyses showed that both camera movement speed and movement-related audio effects influence animation smoothness perception. In Section 4.1, we showed that camera speed significantly affects the perception of animation smoothness when

RUN						
	10-20	10-30	10-60	20-30	20-60	30-60
Mean (Audio)*	1.70	1.78	1.76	1.70	1.74	1.52
Mean (NoAudio)*	1.80	1.78	1.72	1.58	1.82	1.52
<i>p</i> - value	.098	N/A**	N/A**	N/A**	.173	N/A**

WALK						
	10-20	10-30	10-60	20-30	20-60	30-60
Mean (Audio)*	1.92	1.78	1.78	1.50	1.52	1.32
Mean (NoAudio)*	1.90	1.92	1.88	1.78	1.68	1.40
<i>p</i> - value	N/A**	.010	.029	.002	.044	.197

*Lower bound = 1; Upper bound = 2

**Not inline with a 1-tailed test

Table 4 Test 3: Mean and p values for Audio condition. *p* - value is given for difference in preference between the test pairs and 60vs60 condition. Lower and upper bounds were 1 (first animation preferred) and 2 (second animation preferred) respectively.

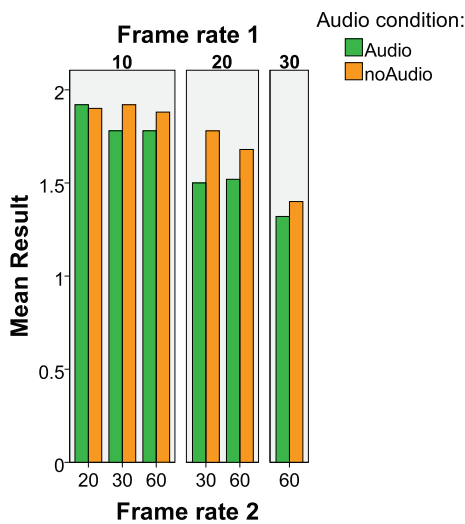


Fig. 7 Mean values of the compared walking animation test pairs.

presented with no audio. However, there was no significance of this effect in case of the Audio condition. In both audio conditions, the slower (walking) animation was preferred over the faster (running) animation. The different results for the Audio and NoAudio conditions indicate that audio effects could indeed influence smoothness perception. In Test 2 we showed that the perceived smoothness threshold of the pre-rendered animations, when accompanied by a movement-related sound effects, lies between 10 and 20 frames per second. Lastly, in Section 4.3 we looked at the significance of the auditory influence on perceived smoothness of the animations. The results showed that movement-related sound effects do increase the animation smoothness perception. The effect was, however, significant only for the slow animations. The reason for this might be in the

fact that in running animations, the vertical oscillating motion of the camera introduces a jitter effect, making the frame rate of the animation hard to distinguish.

These results represent another step towards understanding auditory-visual cross-modal interaction and its possible uses in computer graphics. Properly exploited in interactive scenarios, related sounds could be introduced or emphasised to maintain the same perceptual quality when the computational resources are insufficient.

In the future, we will compare the audio-visual content with lower frame rates with ones presented at higher frame rates directly. We would also like to investigate the same effect in interactive scenarios with and without user tasks. Such results also entail the possibility of building decision-theoretic systems [10] based on cross-modal effects that ensure a constant perceived frame rate rather than the commonly used fixed frame rate [12].

Acknowledgements The authors would like to thank all the participants from University of Warwick, UK and Universidade de Trás-os-Montes e Alto Douro, Portugal who volunteered for the experiments. We would also like to thank Maximino Esteves Correia Bessa and Martinho Gonçalves for their assistance in running the experiments and Piotr Dubla for helping us with the framework development.

References

- Alais, D., Morrone, C., Burr, D.: Separate attentional resources for vision and audition. *Proc Biol Sci* **273**(1592), 1339–1345 (2006). DOI 10.1098/rspb.2005.3420
- Allport, D.A., Antonis, B., Reynolds, P.: On the division of attention: a disproof of the single channel hypothesis. *Q J Exp Psychol* **24**(2), 225–235 (1972)
- Arif, M., Ohtaki, Y., Nagatomi, R., H., I.: Estimation of the effect of cadence on gait stability in young and elderly people using approximate entropy technique. *MEASUREMENT SCIENCE REVIEW* **4**, 29–40 (2004)

4. Bonneel, N., Suied, C., Viaud-Delmon, I., Drettakis, G.: Bimodal perception of audio-visual material properties for virtual environments. *ACM Transactions on Applied Perception* (2009). URL <http://www-sop.inria.fr/reves/Basilic/2009/BSVD09>
5. Bonneel, N., Suied, C., Viaud-Delmon, I., Drettakis, G.: Bimodal perception of audio-visual material properties for virtual environments. *ACM Trans. Appl. Percept.* **7**(1), 1–16 (2010). DOI <http://doi.acm.org/10.1145/1658349.1658350>
6. Bonnel, A.M., Hafter, E.R.: Divided attention between simultaneous auditory and visual signals. *Percept Psychophys* **60**(2), 179–190 (1998)
7. Burr, D., Alais, D.: Combining visual and auditory information. *Prog Brain Res* **155**, 243–258 (2006). DOI [10.1016/S0079-6123\(06\)55014-9](http://dx.doi.org/10.1016/S0079-6123(06)55014-9)
8. Cater, K., Chalmers, A., Ward, G.: Detail to attention: exploiting visual tasks for selective rendering. In: *EGRW '03: Proceedings of the 14th Eurographics Workshop on Rendering Techniques*, pp. 270–280. Eurographics Association, Leuven, Belgium (2003)
9. Choe, C.S., Welch, R.B., Gilford, R.M., Juola, J.F.: The "ventriloquist effect": Visual dominance or response bias? *Perception and Psychophysics* **18**(1), 55–60 (1975)
10. Dumont, R., Pellacini, F., Ferwerda, J.A.: Perceptually-driven decision theory for interactive realistic rendering. *ACM Trans. Graph.* **22**(2), 152–181 (2003). DOI <http://doi.acm.org/10.1145/636886.636888>
11. Duncan, J., Martens, S., Ward, R.: Restricted attentional capacity within but not between sensory modalities. *Nature* **387**(6635), 808–810 (1997)
12. Funkhouser, T.A., Séquin, C.H.: Adaptive display algorithm for interactive frame rates during visualization of complex virtual environments. In: *SIGGRAPH '93: Proceedings of the 20th annual conference on Computer graphics and interactive techniques*, pp. 247–254. ACM, New York, NY, USA (1993). DOI <http://doi.acm.org/10.1145/166117.166149>
13. Gebhard, J.W., Mowbray, G.H.: On discriminating the rate of visual flicker and auditory flutter. *Am J Psychol* **72**, 521–529 (1959)
14. Getzmann, S.: The effect of brief auditory stimuli on visual apparent motion. *Perception* **36**(7), 1089–1103 (2007)
15. Grelaud, D., Bonneel, N., Wimmer, M., Asselot, M., Drettakis, G.: Efficient and practical audio-visual rendering for games using crossmodal perception. In: *Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games* (2009). URL <http://www-sop.inria.fr/reves/Basilic/2009/GWAD09>
16. Howard, I.P., Templeton, W.B.: *Human spatial orientation* [by] I.P. Howard and W.B. Templeton. Wiley, London, New York, (1966)
17. Hulusic, V., Aranha, M., Chalmers, A.: The influence of cross-modal interaction on perceived rendering quality thresholds. In: Skala, V. (ed.) *WSCG 2008 Full Papers Proceedings*, pp. 41–48 (2008)
18. Hulusic, V., Czanner, G., Debattista, K., Sikudova, E., Dubla, P., Chalmers, A.: Investigation of the beat rate effect on frame rate for animated content. In: Hauser, H. (ed.) *Spring Conference on Computer Graphics 2009*, pp. 167–174. ACM SIGGRAPH Press (2009)
19. Hulusic, V., Debattista, K., Aggarwal, V., Chalmers, A.: Exploiting audio-visual cross-modal interaction to reduce computational requirements in interactive environments. In: *Proceedings of the IEEE conference on Games and Virtual Worlds for Serious Applications*. IEEE Computer Society (2010)
20. Humphreys, G.W., Bruce, V.: *Visual Cognition: Computational, Experimental and Neuropsychological Perspectives*. Lawrence Erlbaum Associates Ltd, East Sussex, BN3 2FA, UK (1989)
21. James, W.: *The principles of psychology*. Holt, New York (1890)
22. Kajiya, J.T.: The Rendering Equation. *SIGGRAPH Comput. Graph.* **20**(4), 143–150 (1986)
23. Kamitani, Y., Shimojo, S.: Sound-induced visual "rabbit". *J. Vis.* **1**(3), 478–478 (2001). URL <http://journalofvision.org/1/3/478/>
24. Kayser, C., Petkov, C.I., Lippert, M., Logothetis, N.K.: Mechanisms for allocating auditory attention: An auditory saliency map. *Current Biology* **15**(21), 1943–1947 (2005). DOI [10.1016/j.cub.2005.09.040](http://dx.doi.org/10.1016/j.cub.2005.09.040). URL <http://dx.doi.org/10.1016/j.cub.2005.09.040>
25. Keetels, M., Vroomen, J.: Tactile-visual temporal ventriloquism: no effect of spatial disparity. *Percept Psychophys* **70**(5), 765–771 (2008)
26. Kelly, M.C., Tew, A.I.: The continuity illusion in virtual auditory space. In: *in proc. of AES 112th Convention*. Munich, Germany (2002)
27. Larsen, A., McIlhagga, W., Baert, J., Bundesen, C.: Seeing or hearing? perceptual independence, modality confusions, and crossmodal congruity effects with focused and divided attention. *Percept Psychophys* **65**(4), 568–574 (2003)
28. Lecuyer, A., Burkhardt, J.M., Henaff, J.M., Donikian, S.: Camera motions improve the sensation of walking in virtual environments. In: *VR '06: Proceedings of the IEEE conference on Virtual Reality*, pp. 11–18. IEEE Computer Society, Washington, DC, USA (2006). DOI <http://dx.doi.org/10.1109/VR.2006.31>
29. Mack, A., Rock, I.: *Inattentional Blindness*. The MIT Press (1998)
30. Massaro, D.W., Warner, D.S.: Dividing attention between auditory and visual perception. *Perception & Psychophysics* **21**(6), 569–574 (1977)
31. Mastoropoulou, G.: *The effect of audio on the visual perception of high-fidelity animated 3d computer graphics*. PhD in Computer science, University of Bristol (2006)
32. Mastoropoulou, G., Chalmers, A.: The effect of music on the perception of display rate and duration of animated sequences: An experimental study. In: *TPCG '04: Proceedings of the Theory and Practice of Computer Graphics 2004 (TPCG'04)*, pp. 128–134. IEEE Computer Society, Washington, DC, USA (2004). DOI <http://dx.doi.org/10.1109/TPCG.2004.34>
33. Mastoropoulou, G., Debattista, K., Chalmers, A., Troscianko, T.: Auditory bias of visual attention for perceptually-guided selective rendering of animations. In: *GRAPHITE '05: Proceedings of the 3rd international conference on Computer graphics and interactive techniques in Australasia and South East Asia*, pp. 363–369. ACM Press, New York, NY, USA (2005). DOI <http://doi.acm.org/10.1145/1101389.1101462>
34. Mastoropoulou, G., Debattista, K., Chalmers, A., Troscianko, T.: The influence of sound effects on the perceived smoothness of rendered animations. In: *APGV '05: Proceedings of the 2nd symposium on Applied perception in graphics and visualization*, pp. 9–15. ACM Press, New York, NY, USA (2005). DOI <http://doi.acm.org/10.1145/1080402.1080404>
35. McGurk, H., Macdonald, J.: Hearing lips and seeing voices. *Nature* **264**(5588), 746–748 (1976). DOI [10.1038/264746a0](http://dx.doi.org/10.1038/264746a0)
36. Moeck, T., Bonneel, N., Tsingos, N., Drettakis, G., Viaud-Delmon, I., Alloza, D.: Progressive perceptual audio rendering of complex scenes. In: *I3D '07: Proceedings of the 2007 symposium on Interactive 3D graphics and games*, pp. 189–196. ACM, New York, NY, USA (2007). DOI <http://doi.acm.org/10.1145/1230100.1230133>

37. Moore, B.C.: An Introduction to the Psychology of Hearing. 2nd edn. Academic Press (1982)
38. Morein-Zamir, S., Soto-Faraco, S., Kingstone, A.: Auditory capture of vision: examining temporal ventriloquism. *Brain Res Cogn Brain Res* **17**(1), 154–163 (2003)
39. Parke, R., Chew, E., Kyriakakis, C.: Quantitative and visual analysis of the impact of music on perceived emotion of film. *Comput. Entertain.* **5**(3), 5 (2007). DOI <http://doi.acm.org/10.1145/1316511.1316516>
40. Ramic-Brkic, B., Chalmers, A., Boulanger, K., Pattanaik, S., Covington, J.: Cross-modal affects of smell on the real-time rendering of grass. In: Hauser, H. (ed.) *Spring Conference on Computer Graphics 2009*, pp. 175–179. ACM SIGGRAPH Press (2009)
41. Recanzone, G.H.: Auditory influences on visual temporal rate perception. *Journal of neurophysiology* **89**, 1078–1093 (2003)
42. Sekuler, R., Sekuler, A.B., Lau, R.: Sound alters visual motion perception. *Nature* **385**(6614), 308 (1997). DOI <http://dx.doi.org/10.1038/385308a0>. URL <http://dx.doi.org/10.1038/385308a0>
43. Shams, L., Kamitani, Y., Shimojo, S.: What you see is what you hear. *Nature* **408**, 788+ (2000)
44. Shams, L., Kamitani, Y., Shimojo, S.: Visual illusion induced by sound. *Brain Res Cogn Brain Res* **14**(1), 147–152 (2002)
45. Shams, L., Kamitani, Y., Shimojo, S.: Modulations of visual perception by sound. in the handbook of multisensory processes (eds. calvert, g.a., spence, c. and stein, b.e.) pp. 27–33 (2004)
46. Shipley, T.: Auditory flutter-driving of visual flicker. *Science* **145**, 1328–1330 (1964)
47. Simons, D., Chabris, C.: Gorillas in our midst: sustained inattentive blindness for dynamic events. *perception* **28**, 1059–1074 (1999)
48. Suied, C., Bonneel, N., Viaud-Delmon, I.: Integration of auditory and visual information in the recognition of realistic objects. *Experimental Brain Research* (2009)
49. Tsingos, N., Gallo, E., Drettakis, G.: Perceptual audio rendering of complex virtual environments. *ACM Trans. Graph.* **23**(3), 249–258 (2004). DOI <http://doi.acm.org/10.1145/1015706.1015710>
50. Vroomen, J., Bertelson, P., Gelder, B.d.: A visual influence in the discrimination of auditory location (1998)
51. Vroomen, J., de Gelder, B.: Perceptual effects of cross-modal stimulation: Ventriloquism and the freezing phenomenon. in the handbook of multisensory processes (eds. calvert, g.a., spence, c. and stein, b.e.) pp. 140–150 (2004)
52. Wada, Y., Kitagawa, N., Noguchi, K.: Audio-visual integration in temporal perception. *Int J Psychophysiol* **50**(1-2), 117–124 (2003)
53. Welch, R.B., Warren, D.H.: Immediate perceptual response to intersensory discrepancy. *Psychological bulletin* **88**(3), 638–667 (1980)
54. Yee, H., Pattanaik, S., Greenberg, D.P.: Spatiotemporal sensitivity and visual attention for efficient rendering of dynamic environments. *ACM Trans. Graph.* **20**(1), 39–65 (2001). DOI <http://doi.acm.org/10.1145/383745.383748>. URL <http://pdiff.sourceforge.net/ypg01.pdf>