

Investigation of the beat rate effect on frame rate for animated content

Vedad Hulusić^{1*}

Gabriela Czanner²

Kurt Debattista¹

Elena Sikudova¹

Piotr Dubla¹

Alan Chalmers¹

¹ The Digital Laboratory, WMG, University of Warwick, UK

² Cancer Epidemiology Unit, University of Oxford, UK

Abstract

Knowledge of the Human Visual System (HVS) may be exploited in computer graphics to significantly reduce rendering times without the viewer being aware of any resultant image quality difference. Furthermore, cross-modal effects, that is the influence of one sensory input on another, for example sound and visuals, have also recently been shown to have a substantial impact on viewer perception of image quality.

In this paper we investigate the relationship between audio beat rate and video frame rate in order to manipulate temporal visual perception. This represents an initial step towards establishing a comprehensive understanding for the audio-visual integration in multi-sensory environments.

Keywords: cross-modal, frame rate, beat rate, temporal perception

1 Introduction

The high-fidelity rendering of complex scenes in real time is still unachievable even on current high performance desktop machines. Researchers have thus started exploring the interaction across different senses to exploit any limitations of human perception that may then be applied to reduce computational requirements while still maintain a *perceptually* high-fidelity result. One particular cross-modal effect which has been successfully exploited in the past in the field of computer graphics is that of vision and audition [Mastropoulou 2006] [Hulusić et al. 2008]. This work was based on previous psychological research such as [Welch and Warren 1980] [Recanzone 2003] [Shams and Shimojo 2004].

In current interactive virtual environments, such as video games, more computation time is spent on computing compelling visuals than calculating audio. This is because much of the audio may be pre-recorded, unlike the physically-based illumination and character animation in the visuals that may have to be computed on a frame by frame basis. In such virtual environments audio, potentially music, may be used to create an emotional involvement with the situation. If the influence of audio can be exploited, it may be possible to reduce the computation required for the visuals without reducing the perceived visual quality. The emotional involvement caused by music is a very complex phenomenon. We propose to investigate the more straightforward relationship between specific

aspects of audio and video that function in the temporal domain, in particular beat rate and frame rate. If understood and then harnessed correctly, this relationship should make it possible to have a graphics engine that can change the beat rate and the frame rate on-demand to reduce its work load whenever required. This would effectively reduce the computational time of rendering, without the user noticing any perceptual loss in quality. In this paper we present an initial series of psychophysical experiments to explore this relationship. In the experiments with 49 participants the beat rate and frame rate were manipulated for different rendered animations and the perceptual difference in temporal quality was acquired and analysed using multiple statistical analysis methods.

2 Related Work

In the research area of perception much more attention has been given to the exploration of the individual senses rather than to the cross-modal interaction between them. Even the research that has been done has been more focused on sound influence on spatial image quality than on its temporal domain.

2.1 Audio-visual cross-modal interaction in psychology

The cross-modal interaction is apparent in both directions: as video can influence auditory perception, audition is also very influential while being presented along with the visual stimulus. Psychology has investigated many of these phenomena simulating them in the laboratory. Welch et al. [1980] introduced the modality appropriateness hypothesis which was the solid base for further investigation of cross-modal interaction. This hypothesis states that the modality which is more accurate and appropriate for a given task will dominate the perception in that particular task. The ventriloquism effect [Howard and Templeton 1966] for example, is the direct consequence of the greater spatial acuity of vision. It shows that humans, while watching TV or a puppet show, are associating sound source to a speaking person/puppet's mouth even though it originates from the speakers/ventriloquist positioned at a different location. Another example of the visual impact on auditory perception is the McGurk effect [McGurk and Macdonald 1976] where a sound of /ba/ is perceived as /da/ when accompanied with lip movement corresponding to the pronunciation of /ga/.

Other research has investigated the influence of audio on visual perception. This work showed that this effect is apparent and particularly strong in the temporal domain since the auditory system in that domain has greater acuity. One example of auditory dominance over vision is the auditory driving effect [Gebhard and Mowbray 1959] showing that auditory flutter influences the apparent rate of the visual flicker, but not the reverse. Another type of auditory influence on vision is presented in [Sekuler et al. 1997]. In this case two visual targets move towards each other and cross. When presented to participants without any sound, the majority report that

*e-mail: V.Hulusić@warwick.ac.uk

the targets pass through each other. After introducing a sound at the time of collision, the targets are reported to be bouncing off each other. Recanzone in [Recanzone 2003], when presenting auditory and visual stimuli at different frequencies, showed that audio can significantly influence perception of visual temporal rate, but visual stimulus has no effect on auditory perception. Morein-Zamir et al. [2003] showed 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. Other interesting phenomena are the illusory flash effect [Shams et al. 2000] [Shams et al. 2002] and the Rabbit illusion [Kamitani and Shimojo 2001]. These phenomena provide strong evidence of auditory influence on temporal visual perception. Here a single flash accompanied with multiple beeps is perceived as multiple flashes and three visual targets i.e. Rabbits, shown in sequence with the horizontal offsets are perceived as five flashes, and thus much smoother, when coupled with five beeps - three beeps presented synchronously with flashes and two inserted between them. Similar work was conducted by Getzmann [2007] in which he investigated visual apparent motion. For a complete overview see [Shams and Shimojo 2004].

2.2 Audio-visual cross-modal interaction in computer graphics

In the last decade there has been increased research in the field of cross-modal interaction. However, this work is still very much in its infancy. Mastoropoulou et al. [2005a] using selective rendering [Debattista 2006] and the inattention blindness phenomenon [Mack and Rock 1998] showed that when there is a sound emitting object in the scene it is enough to render that object in high quality and remainder of the scene in much lower quality without significant perceivable difference in quality. Further research by the same group [Mastoropoulou et al. 2005b] showed that using sound effects e.g. phone ringing or thunder as a distractor in animated content can reduce temporal visual perception. Hulusic et al. [2008] investigated the impact cross-modal interactions has on perceivable rendering thresholds. They showed that this threshold is reduced when using unrelated audio and increased in case of the related sound.

In his PhD thesis Storms [1998] showed that using high quality audio with video increases the perceived quality of the visual displays. More recent work using cross-modal interaction can be found in [Grelaud et al. 2009] [Bonneel et al. 2009] [Suied et al. 2009].

3 Experiments

The aim of our study is to investigate how audio and beat rate can influence frame rate perception. In this section we describe the set up for the experiment, design, participants, apparatus, stimuli, procedure, study questions and the analysis methods.

There are a number of key considerations when preparing any such experiments. The first is scene complexity. This involves many dimensions, for example geometry, materials and lighting complexity, as described in [Gero and Kazakov 2004] and [Ramanarayanan et al. 2008]. Additionally in computer graphics very often the camera is not static. Camera movement can be in the form of translation, rotation or panning, or combination of any of these movement types. Another parameter that is of a great importance in determining the complexity of the visual images presented to the subjects is the speed of the camera movement.

All this applies to both static and dynamic scenes, which doubles the number of independent variables involved in the experiment. Hence, we need to decide whether to use static scene with camera movement or dynamic scene with fixed camera or combination of both i.e. dynamic scene with moving camera.

Regarding sound, it is important to choose the sound which will not introduce any subjective side effects to the participants, such as emotions, anxiety, excitement, boredom, etc. Any of these factors could influence the perception and response of the participant during the psychophysical experiment [Mastoropoulou and Chalmers 2004].

Another considerable difference in experiments done in field of psychology (see Section 2.1), compared to what we propose, is that for visual stimulus they mostly use too low visual frequencies (frame rate) and too high audio frequencies (beat rate). Hence it is not possible to directly use the experimental setup that a psychologist may typically use for such audio-visual experiments. Synchrony-asynchrony discrimination is possible only for the temporal frequencies below 4Hz [Fujisaki and Nishida 2005].

3.1 Design

We studied the perceptual responses of subjects in a complete randomised design with three factors (independent variables): scene, frame rate and beat rate. The scene contains three major elements, scene content/complexity, rendering technique used for image generation and camera movement type. Further in the text we will refer to it as a scene factor. We used two different scene groups: four static scenes with moving camera and one dynamic scene with fixed camera and a moving object, see Figure 1. The latter is used for an initial study investigating the behaviour beat rate effect on dynamic scenes. We considered four different frame rates: 10, 15, 20 and 60fps (frames per second), and three different audio conditions: no sound (in analysis marked as 0), 2 and 6bps (beats per seconds). Further, we asked the participants how many hours weekly they spend playing video games, to study and control for the effect of familiarity. To measure the perceptual responses of subjects we used a single stimulus non-categorical method [International-Telecommunication-Union 2002] also known as Interactive Rating Scale method in sound quality studies [Giudice et al. 2006].

3.2 Participants

49 people volunteered for the experiment, 43 of whom were university students studying a variety of subjects, and the rest from university staff. Out of the 49 participants 36 were male and 13 female. The participants were aged between 18 to 41 with an average age of 23. All of them had normal or corrected to normal vision and no hearing impairments.

3.3 Apparatus

The experiments were conducted in a dark, quiet room. Visual stimuli were presented on a calibrated Dell E198FPB 19" monitor with 1280 × 1024 pixel resolution and a refresh rate of 60 Hz. The stimuli were positioned at eye level, 60-70 cm from the participants' eyes. A LTB Magnum 5.1 AC97 Headphone set was used for audio stimuli.



Figure 1: A simple frame from each of the animations used in the study. From left to right - static scenes: Kiti, Rabbit, Kalabsha, KitementalRay. Right - dynamic scene: Ball

3.4 Stimuli

Five different animations, see Figure 1, were used as the visual stimuli in order to avoid boredom and to analyse the behaviour of the participants according to scene variations. We used four different movements: translation for Kiti scene, panning for KitementalRay scene, rotation around own axis for Kalabsha scene, rotation around the object for Rabbit scene. For all translational types of camera movement we used the same speed, which corresponds to the young subjects' average normal walking speed of 1.425 m/s [Arif et al. 2004]. Additionally, we used a static camera with a dynamic scene in Ball scene. In this scene a room with the brown box is shown. There are two red balls bouncing through the room one after another, entering from opposite directions, i.e. left and right.

For auditory stimuli we used sounds which do not have emotional effect on participants, but do have a rhythmical significance. We also tried to minimise the synthetic nature of the sound. We used a looped rhythmical sample created in Propellerhead Reason software with two different conga kick sounds. All sounds were produced using two channels (stereo), sample rate 44100Hz and bit rate of 141 kbps. These audio samples varied just in the beat rate. Both audio and video files were uncompressed. The correlations between beat and frame rates are shown in Figure 2.

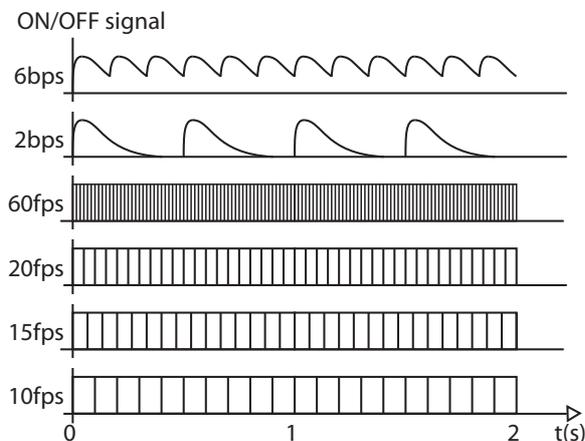


Figure 2: Correlation between beat rates (bps) and frame rates (fps). The diagram shows the number of frames that fit within a beat.

3.5 Procedure

Prior to the experiment the participants were asked to read and sign a consent form. They were then asked to read the instructions and were verbally explained the experimental procedure and how to rate the smoothness of the animations. They were shown a sample animation (Figure 3) at 10 and 60fps with no sound, and told that these



Figure 3: Two frames from the sample animation

are the worst and the best cases respectively. Participants were not told what frame rates were involved in either the experiment or the sample animations. After reading and listening to the instructions participants were asked to rate the smoothness of each animation using a slider bar, see Figure 4.

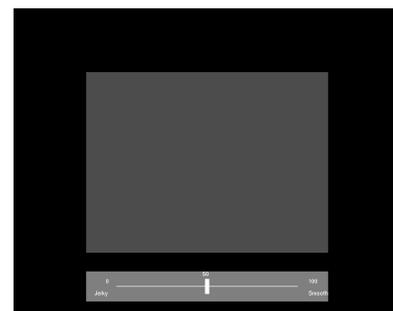


Figure 4: Preview of the slider bar used in the experiment

All 49 participants were shown randomly ordered 60 animations, each lasting for 10 seconds. Before each session five dummy animations were shown to "stabilise" the participants' opinion. The data from the dummy presentations were not taken into account when analysing the results of the experiments. After each animation a slider bar, ranging from 0 to 100, was shown at the bottom of the screen as shown in Figure 4. Initially the slider bar was set to the middle i.e. value 50. The time for the evaluation was restricted to 5 seconds. The next animation started automatically at the end of this 5 seconds. The total trial time was 16.3 minutes.

3.6 Study questions

We investigated four factors that could potentially influence the perception of smoothness of the animation: frame rate, familiarity, scene and beat rate. The main aim was to investigate the relationship between beat rate and frame rate and their effect on temporal quality perception. Our hypothesis was that the beat rate affects the

perception of the frame rate, and if the hypothesis was confirmed, we wanted to further on investigate the size of this effect.

Secondly, we examined whether there was significant difference between our chosen frame rates for all audio conditions together. This would show us if the frame rates we chose were appropriate for the experiment. The next question was *Does the scene influence the smoothness perception?* The fourth factor that we wanted to investigate was the familiarity with the task, taking into account the number of hours the participant on average spends weekly on playing video games.

3.7 Analysis methods

The most commonly used quality and perception evaluation method is pairwise forced choice comparisons. However, this has disadvantages such as the reported subjective preference may depend on the order of the presented videos, the experiment is time demanding (the number of comparisons increases rapidly with the inclusion of new levels of factors e.g. with 2, 3, 4, 5 scenes we need 2, 6, 12, and 20 comparisons) and these comparisons do not give the estimate of the size of effects of considered factors. Yet the forced choice pairwise comparison is often used (especially in static pictures with no camera rotation or movement) and the statistical methodology is well established for these comparisons.

In this study we used a slider bar to measure the perception of smoothness. The main challenge of slider bars is that subjects use it in a different way introducing another source of between subject variability into data. In our study about 5-10% of subjects tended to use only the middle 50-80 of the 0-100 scale, while others used the whole scale. Some authors recommend standardising the scores to reduce the variability, but there is a controversy about this. The first reported analysis of slider bar data is by [Allman-Ward et al. 2004] which gives a good example of how the results can be analysed via exploratory data analysis, using statistics such as the mean and standard deviation. However, this does not give an indication if the observed differences are significant i.e. the differences are due to the effect of a factor or are due to the random variability in the subjective feedback collected. Later sound perception studies suggest using Analysis of Variance (ANOVA) together with non-parametric methods [Giudice et al. 2006].

To study the effect of the sound condition on the perception of the smoothness we proceeded in two steps. First we performed a non-parametric Friedman test [Siegel and Castellan 1988] on three dependent subjective score samples defined by three studied sound conditions. In this test, for each combination of participant, frame rate and scene we calculated the order of preference of the sound conditions: e.g. if BR6 was rated with lowest score we give the BR6 the order 1, if BR0 was the second preferred sound condition it gets the order 2 and BR2 gets order 3. Thus this test is based on relative order between sound conditions. It does not assume any specific distribution of the subjective scores and it is robust against variability across subjects. However, as a non-parametric test it has lower power than parametric tests and hence it is harder to detect differences.

If the non-parametric Friedman test found a significant effect of a particular factor (sound condition, scene, frame rate and familiarity) we investigated further, the size of this effect by characterising the actual subjective scores via parametric methods. For each frame rate we aimed to assess the relationship between the subjective measures of smoothness perception and sound condition while controlling for the effect of the scene and familiarity. For this purpose we used Linear Mixed-Effects Models (MIXED) (also known as multilevel regression or mixed effects regression models) in SPSS

17.0 [SPSS]. This was estimated via restricted maximum likelihood. This approach takes into account the multilevel structure of the data: for each participant we had 12 observations (3 sound conditions per each of four scenes). We assumed random effect of subjects; and fixed effect for sound condition, scene and familiarity. If we found a significant effect of a factor we then performed mean comparisons for different levels of the factor with Sidak multiple comparison correction to ensure the 95% simultaneous confidence. MIXED models expand on the ANOVA models previously used in visual perception comparisons by for example [Mastoropoulou et al. 2005b] and audio quality in [Giudice et al. 2006]. The extension is in that the data from the same subject is permitted to be correlated and allowed to have random effect. In our study we allow for random effect of subjects to account for the variability across subjects caused by slider-bar measurements. The MIXED models have been recently used in studies of sound perception [Gomez and Danuser 2004].

4 Results

First, in preliminary investigations, we studied the main effects of sound condition, frame rate, scene and familiarity via non-parametric methods (Section 4.1). Then we studied further the effects in greater detail via parametric methods for static (Section 4.2) and dynamics scenes (Section 4.3) separately.

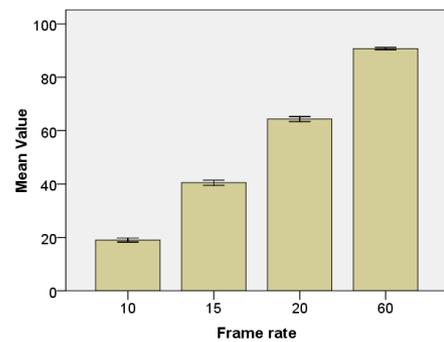


Figure 5: Mean values of subjective scores across frame rates with standard error. All scenes and beat rates are pooled.

4.1 Preliminary analysis

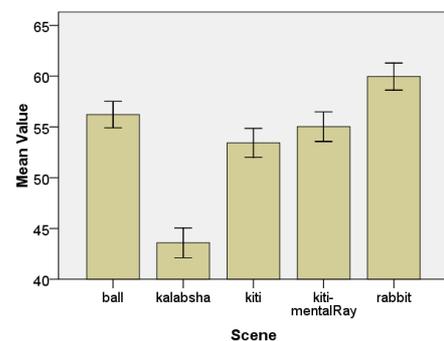


Figure 6: Mean values of subjective scores across scenes with standard error. All frame and beat rates are pooled.

First we investigated if the four frame rates were perceptually distinguishable in all audio conditions. Figure 5 shows that the mean subjective scores increase with frame rate. Results gained from the Friedman test show the significance in differences across frame rates ($p < .001$). Moreover, these differences were found to be significant via a linear mixed model test and multiple pairwise comparisons for all pairs (all p -values $p < .001$).

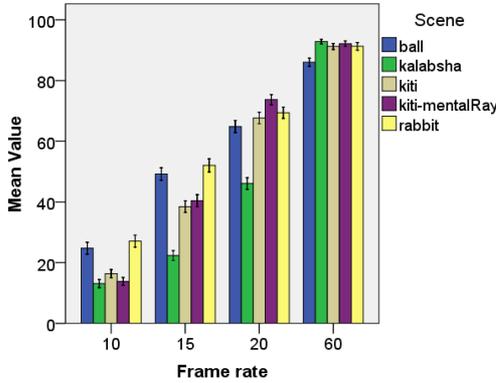


Figure 7: Mean values of subjective scores across scenes and frame rates with standard error

Next we studied how users’ responses varied across scenes. As we mentioned in Section 3 scene content and complexity are important factors in visual perception. Figure 6 indicates that there are differences in perceptual smoothness across scenes with the following order from smoothest to jerkiest: Rabbit, Ball, Kiti-mentalRay, Kiti and Kalabsha. The non-parametric Friedman test found the perceptual differences across scenes as significant ($p < 0.001$), with order Rabbit, Kiti-mentalRay, Ball and Kiti, Kalabsha (mean ranks 3.25, 3.10, 3.02, 3.02 and 2.61 respectively). Further, Figure 7 indicates that there may be interaction of scene and frame rate e.g. the moving ball scene is perceived much smoother than the other scenes at lower frame rates. This preliminary analysis confirms that the factor “scene” is important and needs to be included and further tested simultaneously with other factors in MIXED analysis (see Section 4.2).

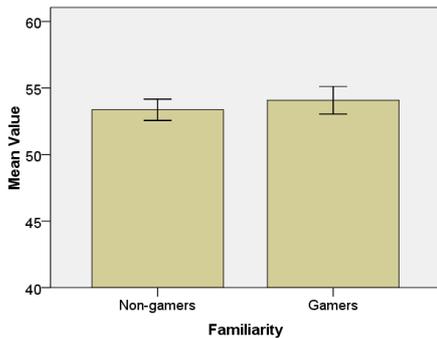


Figure 8: Mean values of subjective scores across familiarity with standard error. All scenes, beat and frame rates are pooled.

Since perception is affected by experience and familiarity [Bradley 2006], we decided to take this factor into account in our paper and study the differences in responses for the two groups: gamers and non-gamers. We divided the participants according to the average weekly number of hours they play video games. The threshold for this division was 6 hours per week. This data was acquired from

the questionnaire participants filled in prior to the experiment. As illustrated in Figure 8, the difference between the average scores of the two groups was rather subtle showing no indication of a familiarity factor. Using a 2-sample t test for independent samples we did not find any evidence ($p=0.519$) for the effect of familiarity on perceived smoothness. Also, the Wilcoxon non-parametric test did not find evidence for the effect of familiarity ($p=0.617$).

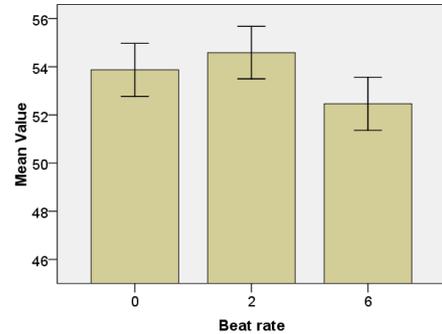


Figure 9: Mean values of subjective scores across beat rates with standard error. All scenes and frame rates are pooled.

The sound condition seems to affect the smoothness perception in our 49 subjects. The plot of mean subjective scores, Figure 9, across beat rates shows the highest ratings at BR2 and the lowest at BR6 i.e. the participants tend to prefer the slow to fast beat. The mean orders for BR0, BR2 and BR6 were 2.00, 2.06 and 1.94 respectively. The non-parametric tests found significant difference between the mean subjective ratings across these sound conditions ($p=.020$). We will further test the significance and size of these differences in the next two sections for static and dynamic scenes separately.

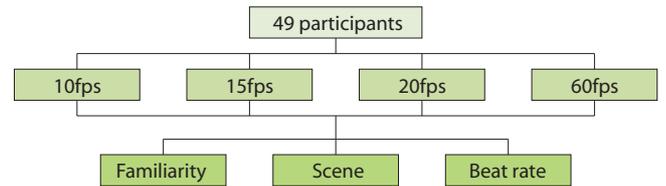


Figure 10: Diagram showing the data group split

This preliminary analysis suggested the important factors: scene, frame rate and beat rate. As a next step we study these factors via a more detailed Linear Mixed Analysis, which allows us to investigate the effect of these factors simultaneously. Since the number of factors is large, in order to facilitate the comparisons, we divided our data into four groups according to frame rate, Figure 10. We chose the frame rate as the dividing factor because our main goal is to study what affects the smoothness perception at the chosen frame rates.

4.2 Static scenes

In this section we present the results of the analysis with all the static scenes i.e. Kiti, Kalabsha, Kiti-mentalRay and Rabbit.

4.2.1 Frame rate 10 analysis

The parametric tests at beat rate 10 found strong evidence for significant effect of beat rate ($p=.051$) and scene ($p < 0.001$) but not

the familiarity ($p=.233$). With no sound, 2 and 6bps the subjective scores were on average 18.1, 17.8 and 15.1 respectively, i.e. at FR10 participants preferred silence to fast beat BR6 ($p=.08$). The perceived smoothness was the highest for Rabbit scene. All the other scenes were perceived significantly less smooth ($p < 0.001$) with no significant difference in the smoothness perception across them, Figure 11.

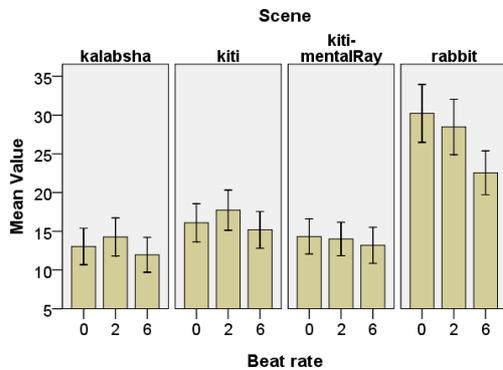


Figure 11: Mean values of subjective scores at FR10 across scenes and beat rates

4.2.2 Frame rate 15 analysis

Analysing the data tests showed that for frame rate 15 there is significant evidence for the beat rate ($p=.026$) and scene ($p < 0.001$) effect on smoothness perception, see Figure 12. At this frame rate, Rabbit scene was perceived as the smoothest ($p < 0.001$) and Kalabsha scene as the jerkiest ($p < 0.001$). The average ratings for each scene were: Kiti(38.9), Kiti-mentalRay(40.9), Rabbit(52.8) and Kalabsha(22.1).

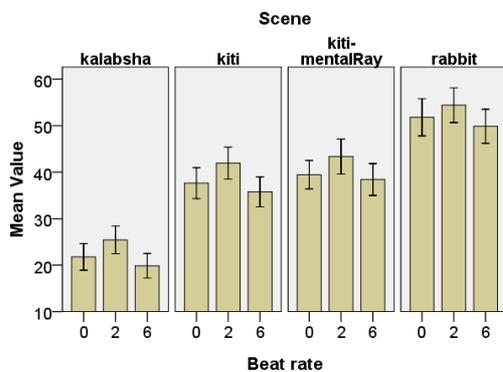


Figure 12: Mean values of subjective scores at FR15 across scenes and beat rates

Using the pairwise comparison based on estimated marginal means, we showed that frame rate 15 with no sound or accompanied with sound at BR2 were rated higher than when it was shown with BR6. The strong significance ($p=.007$) was found between BR2 and BR6.

4.2.3 Frame rate 20 analysis

According to the results we gained investigating the subjective ratings for frame rate 20, Figure 13, scene dependency was high with

the lowest preference for Kalabsha scene ($p < 0.001$). The difference in ratings between Kiti and Kiti-mentalRay was also found to be significant ($p=.043$). There was no significant difference between either the audio conditions nor familiarity with the task. The experienced gamers had a slightly higher average ratings (66.1) than unfamiliar participants (63.0).

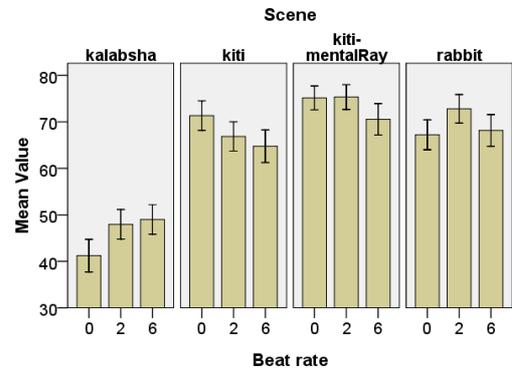


Figure 13: Mean values of subjective scores at FR20 across scenes and beat rates

4.2.4 Frame rate 60 analysis

We assumed that frame rate 60 would behave differently from the other frame rates. The reason is that it is considered as a high frame rate and usually taken as the gold standard in psychophysical experiments. That means that neither scene, familiarity nor beat rate should have any significant effect on the perceived smoothness of the animation.

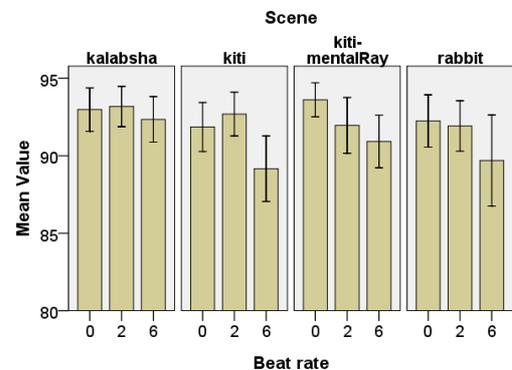


Figure 14: Mean values of subjective scores at FR60 across scenes and beat rates

In case of 60fps, parametric tests showed no significance for neither the effect of scene ($p=.365$) nor effect of gaming ($p=.735$). However, these tests showed a significant effect of audio on visual perception of smoothness ($p=.04$), as can be seen in Figure 14.

4.3 Dynamic scenes

The analysis here has been done according to the diagram shown in Figure 15.

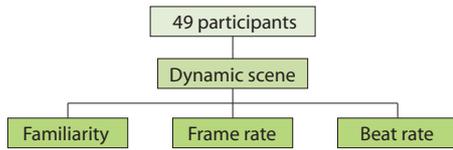


Figure 15: Diagram showing the data group split for dynamic scene

We first conducted a series of non-parametric analyses using the Friedman test. The difference in subjective ratings across frame rates was significant ($p < 0.001$). No effect of familiarity was found for any of the frame rates. The effect of audio was found as insignificant for all four frame rates (Figure 16) with following p-values: 10fps($p=0.927$); 15fps($p=0.254$); 20fps($p=0.802$); 60fps($p=0.920$).

Using a parametric approach and linear mixed models in particular we found the following results. The frame rate has a significant ($p < 0.001$) effect on smoothness perception. The mean scores for frame rates going from 10 to 60fps were 24.9, 49.3, 64.9 and 86.1 respectively. No effect of beat rate ($p=0.510$) and no effect of familiarity ($p=0.786$) was found in these analyses.

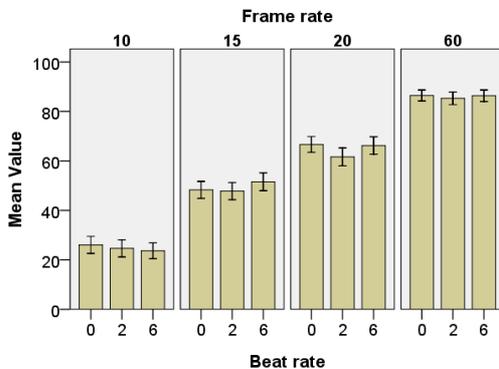


Figure 16: Mean values of subjective scores for dynamic scene across scenes and beat rates

5 Discussion

In this study we investigated the relationship between beat rate and the perception of frame rate. We also looked at the size of the effect between them. In order to generalise the results we chose 4 different factors (scene, frame rate, familiarity and beat rate) that could affect the smoothness perception. Since multiple factors and a correlation between them existed (each participant gives multiple scores), it was important to choose a statistical method which allows testing of one factor while controlling the effect of the others. We used the Linear Mixed Model Analysis together with non-parametric methods. Linear Mixed Models have been recently improved and added to most statistical software and they are an extension to ANOVA in two ways: they allow for correlated observations and allow to model the variability across subjects via random effects assumption. Hence, one of the contributions of our study is in recommending the statistical framework for video perception/quality comparisons via slider-bar numerical scores. This method is interactive and quick, but until now its validity had not been established [Giudice et al. 2006].

The results from our analysis have shown that there is an indication that lower beat rates affect the perception of the lower frame rates.

In our preliminary analysis, see Section 4.1, we inferred that there is a tendency of higher rating of the animations accompanied with low speed audio (2bps) and that high speed audio (6bps) decreases the smoothness perception. Using parametric analysis for static scenes, a significant effect of the beat rate on frame rate was found at frame rates of 10, 15 and 60 frames per second. At all frame rates BR6 decreased the smoothness perception, while for frame rates 10, 15 and 60 results at BR2 and BR6 differed significantly. At frame rate 20 no effect of sound was found. In the case of the dynamic scene no effect of beat rate was found to be significant. Overall this indicates that there is an effect of beat rate on the smoothness perception of an animation and that it might be possible, in future work, to find the direct relationship between them.

This paper represents an extension of the work by Mastoropoulou [2006]. However, we based our study on the investigation of a particular segment of the sound - beat rate, and its effect on perception of the frame rate. Further investigation could identify the direct relationship between beat rates and frame rates. This knowledge could then be used to influence the rendering system, saving computation while maintaining the same perceptual experience within a multi-modal environment. The relationship can be used in interactive systems when scene complexity increases rapidly such that the frame rate is affected. In such a system the beat rate could be manipulated interactively, somewhat akin to time-constrained level-of-detail manipulation [Funkhouser and Séquin 1993], such that the drop in frame rate is not perceivable by the viewer. An alternative scenario where the beat rate and frame rate relationship is useful is when computing high-quality animations which are highly computationally expensive and demand huge resources such as render farms. Under such conditions, if the animation score is known beforehand, the minimal required frame rate could be calculated for each audio segment, drastically reducing the total rendering time without compromising the visual experience.

In the paper we showed that familiarity did not effect the smoothness perception throughout the experiment. We tested this effect according to the number of hours participants play computer games per week. Scene complexity, rendering technique and camera movement type together played an important factor in smoothness quality perception for all frame rates except frame rate 60. Kalabsha scene was rated as the least smooth at all frame rates, while the animations with Rabbit scene at lower frame rates were rated as the smoothest.

6 Conclusion and future work

In this paper we presented some initial findings on how sound can be used to change the perception of the smoothness of an animation. Since this study is, to the best of our knowledge, the first to investigate this phenomenon, for a more complete understanding of the potential benefit, further investigation is required. In the future, higher granularity of the beat rates is suggested in order to find the direct relationship between the beat rate and the frame rate. Once found, this relationship would lead to creation of a metric which could be used for increasing the efficiency of high-fidelity rendering in multi-modal environments. We would also like to consider more dynamic scenes, since most of the video games and other visual contents are based on dynamic environments. Finally, it would be valuable to investigate the effect of emotions caused by music, so it can be brought together with the effect of the beat rate in order to increase the animation smoothness perception.

7 Acknowledgments

The Kalabsha scene was used with the permission of the University of Bristol. We would like to thank all the volunteers from Sarajevo School of Science and Technology and Faculty of Electrical Engineering - University of Sarajevo for participating in the experiments.

References

- ALLMAN-WARD, M., WILLIMAS, R., DUNNE, G., AND JENNINGS, P. 2004. The evaluation of vehicle sound quality using an nvh simulator.
- ARIF, M., OHTAKI, Y., NAGATOMI, R., AND INOOKA, H. 2004. Estimation of the effect of cadence on gait stability in young and elderly people using approximate entropy technique. *MEASUREMENT SCIENCE REVIEW* 4.
- BONNEEL, N., SUIED, C., VIAUD-DELMON, I., AND DRETTAKIS, G. 2009. Bimodal perception of audio-visual material properties for virtual environments. *ACM Transactions on Applied Perception*.
- BRADLEY, S. 2006. *Ibn Al-haytham: First Scientist*. Morgan Reynolds Publishing.
- DEBATTISTA, K. 2006. *Selective Rendering for High Fidelity Graphics*. PhD in Computer science, University of Bristol.
- FUJISAKI, W., AND NISHIDA, S. 2005. Temporal frequency characteristics of synchrony-asynchrony discrimination of audio-visual signals. *Exp Brain Res* 166, 3-4 (October), 455-464.
- FUNKHOUSER, T. A., AND SÉQUIN, C. H. 1993. 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*, ACM, New York, NY, USA, 247-254.
- GATZMANN, S. 2007. The effect of brief auditory stimuli on visual apparent motion. *Perception* 36, 1089-1103.
- GEHBARD, J., AND MOWBRAY, G. 1959. On discriminating the rate of visual flicker and auditory flutter. *The American journal of psychology* 72, 521-529.
- GERO, J. S., AND KAZAKOV, V. 2004. On measuring the visual complexity of 3d objects. *Journal of Design Sciences and Technology* 12, 1, 35-44.
- GIUDICE, S., JENNINGS, P., FRY, J., DUNNE, G., WILLIAMS, R., AND ALLMAN-WARD, M. 2006. The evaluation of vehicle sound quality using an nvh simulator.
- GOMEZ, P., AND DANUSER, B. 2004. Affective and physiological responses to environmental noises and music. *International Journal of Psychophysiology* 53, 2, 91 - 103.
- GRELAUD, D., BONNEEL, N., WIMMER, M., ASSELOT, M., AND DRETTAKIS, G. 2009. Efficient and practical audio-visual rendering for games using crossmodal perception. In *Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*.
- HOWARD, I. P., AND TEMPLETON, W. B. 1966. *Human spatial orientation [by] I.P. Howard and W.B. Templeton*. Wiley, London, New York,.
- HULUSIC, V., ARANHA, M., AND CHALMERS, A. 2008. The influence of cross-modal interaction on perceived rendering quality thresholds. In *WSCG2008 Full Papers Proceedings (WSCG'08)*, 41-48.
- INTERNATIONAL-TELECOMMUNICATION-UNION, 2002. Methodology for the subjective assessment of the quality of television pictures. Recommendation ITU-R BT.500-11.
- KAMITANI, Y., AND SHIMOJO, S. 2001. Sound-induced visual "rabbit". *J. Vis.* 1, 3 (12), 478-478.
- LUCK, S. J., AND VOGEL, E. K. 1997. The capacity of visual working memory for features and conjunctions. *Nature* 390, 6657 (November), 279-281.
- MACK, A., AND ROCK, I. 1998. *Inattentional Blindness*. The MIT Press.
- MASTOROPOULOU, G., AND CHALMERS, A. 2004. The effect of music on the perception of display rate and duration of animated sequences: an experimental study. In *In Theory and Practice of Computer Graphics 2004 (TPCG04)*, 128-134.
- MASTOROPOULOU, G., DEBATTISTA, K., CHALMERS, A., AND TROSCIANKO, T. 2005. 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*, ACM, New York, NY, USA, 363-369.
- MASTOROPOULOU, G., DEBATTISTA, K., CHALMERS, A., AND TROSCIANKO, T. 2005. 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*, ACM, New York, NY, USA, 9-15.
- MASTOROPOULOU, G. 2006. *The Effect of Audio on the Visual Perception of High-Fidelity Animated 3D Computer Graphics*. PhD in Computer science, University of Bristol.
- MCGURK, H., AND MACDONALD, J. 1976. Hearing lips and seeing voices. *Nature* 264, 5588 (December), 746-748.
- MOREIN-ZAMIR, S., SOTO-FRANCO, S., AND KINGSTONE, A. 2003. Auditory capture of vision: Examining temporal ventriloquism. *Brain research. Cognitive brain research* 17 (Jun), 154-163.
- PETERSON, L. R., AND PETERSON, M. J. 1959. Short-term memory retention of individual items. *Journal of Experimental Psychology*.
- RAMANARAYANAN, G., BALA, K., FERWERDA, J. A., AND WALTER, B. 2008. Dimensionality of visual complexity in computer graphics scenes. SPIE, B. E. Rogowitz and T. N. Pappas, Eds., vol. 6806, 68060E.
- RECANZONE, G. H. 2003. Auditory influences on visual temporal rate perception. *Journal of neurophysiology* 89 (Feb), 1078-1093.
- SEKULER, R., SEKULER, A. B., AND LAU, R. 1997. Sound alters visual motion perception. *Nature* 385, 6614 (January), 308.
- SHAMS, L., K. Y., AND SHIMOJO, S. 2004. Modulations of visual perception by sound. in the handbook of multisensory processes (eds. calvert, g.a., spence, c. and stein, b.e.). 27-33.
- SHAMS, L., KAMITANI, Y., AND SHIMOJO, S. 2000. Illusions. what you see is what you hear. *Nature* 408, 6814 (December).
- SHAMS, L., KAMITANI, Y., AND SHIMOJO, S. 2002. Visual illusion induced by sound. *Cognitive Brain Research* 14, 147-152.
- SIEGEL, S., AND CASTELLAN, N. 1988. *Nonparametric statistics for the behavioral sciences*, 2. ed ed. McGraw-Hill.
- SPSS. Spss statistics 17.0. <http://www.spss.com>.
- STORMS, R. 1998. *Auditory-Visual Cross-Modal Perception Phenomena*. PhD thesis, Naval Postgraduate School, Monterey, California.
- SUIED, C., BONNEEL, N., AND VIAUD-DELMON, I. 2009. Integration of auditory and visual information in the recognition of realistic objects. *Experimental Brain Research*. <http://www.springerlink.com/content/jj2w6g7366271237/fulltext.pdf>.
- WELCH, R. B., AND WARREN, D. H. 1980. Immediate perceptual response to intersensory discrepancy. *Psychological bulletin* 88, 3 (November), 638-667.