EG UK Theory and Practice of Computer Graphics (2010)
John Collomosse, Ian Grimstead (Editors)

The Effect of Discretised and Fully Converged Spatialised Sound on Directional Attention and Distraction

C. Harvey[†1], S. Walker[2], T. Bashford-Rogers[1], K. Debattista[1] and A. Chalmers[1]

[1]International Digital Laboratory, University of Warwick, UK
[2]Arup, UK

Abstract

A major challenge in Virtual Reality (VR) is to be able to provide interactive rates of realism. However this is very computationally demanding and only recently has high-fidelity rendering become close to interactive rates through a series of novel exploitations of visual perception; to render parts of the scene that are not currently being attended by the viewer at a much lower quality without the difference being perceived. This paper investigates the effect spatialised directional sounds, both discrete and converged have on the visual attention of the user with and without an auditory cue present in the scene. We verify the worth of investigating subliminal saccade shifts from directional audio impulses via a pilot study to eye track participant's free viewing a scene with an audio impulse and an acoustic identifier and also with an audio impulse and no acoustic identifier versus a control. By selecting look zones, we can identify how long users are spending attending a particular area of a scene in these scenarios. This work also investigates whether the effect prevailed, and if so to what extent, with discretised spatialised sound as opposed to a fully converged audio sample. We also present a novel technique for generating interactive discrete acoustic samples from arbitrary geometry. We show that even without an acoustic identifier in the scene, directional sound provides enough of an impulse to guide subliminal saccade shifts and affect perception in such a way that this can be used to guide selective rendering of the scenes.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation - Viewing Algorithms I.4.8 [Image Processing and Computer Vision]: Scene Analysis - Object Recognition I.4.8 [Image Processing and Computer Vision]: Scene Analysis - Tracking

1. Introduction

By rendering parts of the scene that are not currently being attended by the viewer in lower quality, we exploit visual perception such that the difference is not noticed and computation cost is saved. This is a major challenge in VR with the aim of generating perceptually accurate images at a higher throughput due to lower computation required.

The computational requirements of simulating physically accurate virtual environments in real-time are beyond the capabilities of even the latest hardware, and this is likely to be the case for a number of years to come. In order to have such 'realism in real-time' for use in virtual reality applications now, it is necessary to exploit knowledge of the Human Visual System (HVS) to significantly reduce computation without any loss in perceptual quality. The visually important features in a scene can be computed from Saliency Maps [IKN98, YPG01]. However, another key feature of human perception is cross-modality. Typical saliency models account for visual perception only. This paper considers spatialised sound and investigates how it affects eye saccade shifts and attention and whether this can be exploited to reduce computation without any perceptual loss in visual quality.

The paper is organised as follows. Section 2 describes previous work in the field of perceptually-based rendering, visual attention, cross-modal interactions and briefly covers prevalent work on spatialised sound. Section 3 outlines

† Carlo.Harvey@warwick.ac.uk

DOI: 10.2312/LocalChapterEvents/TPCG/TPCG10/191-198

the algorithm used to generate interactive, yet discrete, spatialised sound. Section 4 describes the experimental setup and techniques and concludes each experiment overlay with results. Finally, Section 5 draws conclusions and suggests future work in the area.

2. Related Work

Previous work [MDCT05b, Mas06] has shown that participants viewing an animation whilst an ambient sound was playing, were statistically significantly more likely to make incorrect choices as to the current frame rate when an acoustic stimulus was introduced together with a visual representation of that stimulus. Mastoropoulou et al. [MDCT05a] also showed that during an animation an ambient sound presented in the presence of a visual cue can be used to exploit perception to guide selective rendering.

Hulusic et al. [HAC08] explored this further by investigating the influence of related ambient sounds to a scene. They showed that participants' ability to notice the difference between lower Sample Per Pixel (SPP) images and images rendered in higher quality was significantly affected. In further work, Hulusic et al. [HCD*09] studied the effect of varying beat rates on users' perception of animations. They showed that in case of scenes where the camera moved, but there were no other moving objects, lower beat rates can have a significant effect on the perception of low frame rates.

2.1. Visual Attention

Coded into the retina [Dow87] is the sequential selection process the HVS uses to determine the hierarchy of visual cues used to deterministically select which objects in any given image are most important. This is necessary because there exists far too much information in any one image to remember it all with a single glance.

The first psychological study of human task-orientated saccades [Yar67] was undertaken by Yarbus who noted that under task-based scenarios the eyes jump in saccades to new points of interest in the scene. However, once the object of interest has been found and lies within the foveal region the eye tracks the object in a smooth manner. More recent psychological research [NBH*08] shows that concurrent audio stimuli increases the visual system's ability to distinguish brief visual events.

Cater et al. [CCL02,CCW03] look at how users perceived a selective quality animation versus a high quality animation under different task-based scenarios. Asked to count pencils during watching the animations, the participants were unable to notice the difference between high quality and selective quality where only the area surrounding the interest region was rendered in high quality.

2.1.1. Peripheral Vision

Spatial acuity is maximised around the fovea as shown in Figure 1. In [ML97,LM99,LMYM01] the authors had a gaze contingent multi resolutional display producing only high visual resolutions within the area attended by the fovea. Using pre-stored exhaustive combinations of possible images they showed an update was required before 5 miliseconds elapsed after a fixation of the fovea on a region following a saccade. This was necessary to maintain the integrity of the illusion without disturbing the cognitive process. Loshky et al. [LMYM01] ran a series of tests in order to adjudge that 4.1° representation for the foveal coverage of the multi resolutional display produced results which indicated the difference between it and a completely high resolution display was statistically indistinguishable.

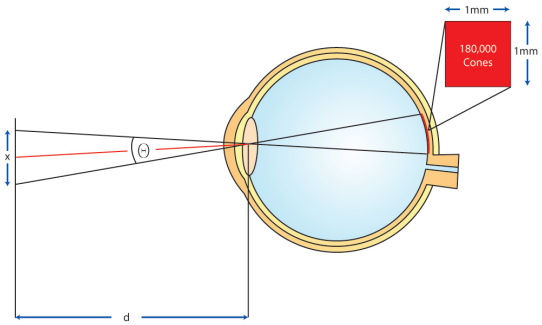

Figure 1: *The eyes' foveal angle. Due to the high number of cones (coloured light receptors) located in the fovea this area has the highest impact on visual perception of a colour image.*

2.2. Inattentional Blindness

Inattentional blindness is a failing of the human eye to see items not in the gaze [MR98]. Saccade shifts for the human eye are very much attenuated by this phenomena. Itti and Koch's [IK00] ideas of top-down (user driven) and bottom-up (scene stimulus driven) phenomena directing attention are prevalent to inattentional blindness. In that user driven task-orientated viewing can cause inattenional blindness to scene stimuli and vice versa.

2.3. Directional Sound

A Room Impulse Response (RIR) is the output of a dynamic environment to an input stimulus. This input stimulus attempts to emulate a Dirac Delta or a unit impulse function. Auralising a sound for a particular sound source, receiver, and environment can be achieved by convolving an RIR with an anechoic source signal to model the acoustical effects of sound propagation within that environment [Kut91]. This auralisation remains accurate only for the particular input position (sound source) and output position (listener) that

the RIR simulates. Convolution is the process of multiplying each and every sample in one audio file with the samples from another waveform. The effect is to use one waveform to model another. Mathematically this results in equation 1, where y is the output waveform, x_n are samples of the audio to be modelled and i_k are samples from the impulse response (the modeller).

$$y_n = \sum_{n}^{k} i_k \cdot x_{n-k} \tag{1}$$

Sound generated in this fashion can be used to manipulate audio to generate most of the spatial cues a listener requires to draw assumptions about size and scale of an environment due to wave pressure and directionality [Beg94]. RIR's can be synthetically generated from arbitrary geometry or recorded in real locations.

Lauterback et al. [LCM07] combine the efficiency of interactive ray tracing with the accuracy of tracing a volumetric representation to generate RIR's. The method uses a four-sided convex frustum and performs clipping and intersection tests using ray packet tracing.

There is a wide range of crossover sound synthesis has with Virtual Reality imaging techniques such as immersive video games, concert hall design and exploiting synergy to enhance immersitivity [MBT*07, RLC*07]. There is a limited collection of work into exploiting perceptual effects with the hybrid and synergy of graphics and spatial sound. More recently [BSVDD10, GBW*09] work has started to appear which exploit bimodal perception using level of detail selection based upon user studies to dynamically weight computational resources to generate audio or visuals based on the selection heuristics. This can fluctuate resources to generate sound if a lot of objects are colliding and reduce load to some visual components. In effect load balancing, this approach is directed towards game engines.

3. Framework for Generating Interactive Spatialised Sound

To investigate whether it was possible to further exploit any cross-modal effects we wanted to compare a fully converged RIR used to generate sound versus a discrete method of generating spatialised sound to extrapolate as to whether any effect, if any, persists with a lower quality of 3D audio reproduction.

3.1. Virtual Point Microphones

Extending the Instant Radiosity algorithm [Kel97] and with an approach similar to [LCM07], instead of propagating Point Lights along a path, discrete samples on a sound wave emanating from a source are propagated. Intersecting with the geometry in the scene and depositing a Virtual Point Microphone (VPM) within the scene should deposit conditions be met. Each VPM holds data on how far it has propagated and also how far the sound wave associated with it has been attenuated based on material absorption and distance travel decay.

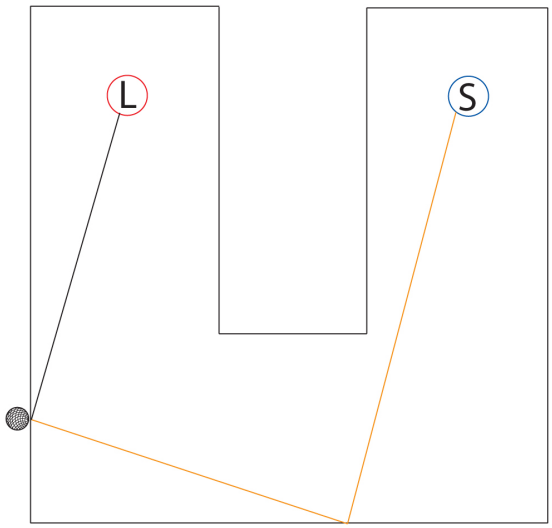

Figure 2: *Diagram highlighting one traversal step, a VPM being deposited and a shadow ray being shot to test if the VPM contributes. L is the listener and S the sound source within the scene.*

In Figure 2 we show the process of depositing one VPM via propagating rays through the enviroment until an absorption criteria is met and the VPM is deposited on a material surface. A shadow test to the listener determines if it is a contributing source. Each contributing VPM acts as a discrete sample on the wave form. A software sound channel with incoming vector and velocity information is assigned for each VPM. The anechoic source is then reattenuaed for that VPMs returned pressure and distance travelled. Each VPM played asynchronously yields a discrete approximation of the spatial sound for the geometry. As each VPM contribution is played asynchronously with each other to simulate spatial effects it is note worthy that the convolution step after RIR generation is optimised out due to the fact that each VPM acts as its own temporal $i_k.x_{n-k}$ as shown in Figure 3.

The waveforms shown in Figure 4 show how the discrete approach compares to the converged approach. For the low quality sound recorded for the experiment we used 1024 VPMs. It is possible to see where the granularity of the approach comes in over and under estimating sample contributions due to the sparse sample rate. It should be noted that this approach, due to levels of detail and discretisation would operate best under conditions of a gated RIR, that is to say an environment whereby late reflections are ignored past a certain time.

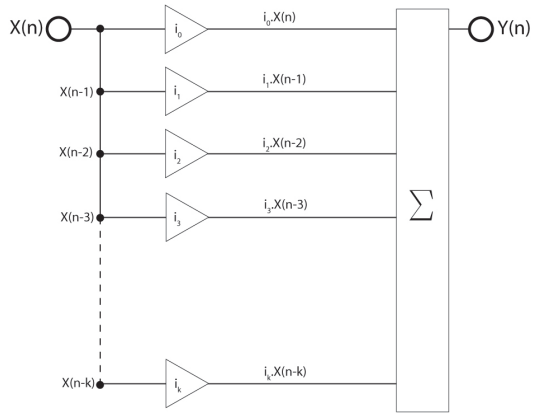

Figure 3: *The process of convolution, x represents audio samples and i impulse response samples.*

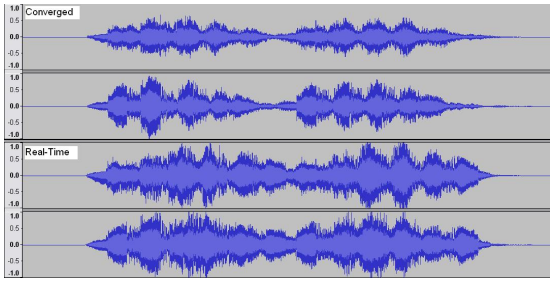

Figure 4: *Analysis of waveforms from both techniques. (Top): Stereo converged wave form from one phone ring. (Bottom): Stereo waveform for the real-time approach highlighting that it lacks the accuracy of a converged RIR due to the discretised approach used.*

4. Psychophysical Experimental Layout, Procedure and Results

The study consists of two separate experiments revolved around viewing 2 separate images. The first image is of a living room with a phone on the coffee table, in the second image the phone was removed. The first experiment, a pilot study, to ascertain the feasibility of a wider study was designed to investigate whether an effect existed between spatial sound impulses and attention to an image. The wider study was aimed at whether this effect could be exploited to guide selective rendering. The second experiment followed on from the feasibility study to ascertain whether spatial sound played a significant role in saccade and fixation control; for both scenes with a visual cue (phone) for the sound and scenes without a visual cue (no phone) for the sound. All images were rendered using the path tracing algorithm [Kaj86]. Exposure to a single image in both experiments was a total of 28 seconds accompanied by the audio. No audio was used for the control scenarios.

4.1. Feasibility Study

There were 3 scenarios for each image set. A Control test where the image was viewed with no sound playing, and also two with sound; one High Quality (HQ) rendered by convolving a converged RIR with the anechoic sound source and one Low Quality (LQ) which was generated via the means described in Section 3. Our hypothesis is that the introduction of spatial sound will direct attention to the pertaining areas of the scene. This set of experiments was conducted to see if this is true, and whether the effect persists under a lower quality of sound exposure. In addition it was researched if this effect persisted without the presence of a visual cue.

4.1.1. Participants

For the pilot study the sample set for the eye tracking consisted of 10 people, 7 males and 3 females with an average age of 25. Each participant recorded data for 3 of the 6 scenarios.

4.1.2. Equipment

Eye tracking was performed with participants. The eye tracker used was as uninvasive as possible. A passive measuring device with no extraneous materials connected to the participant making free viewing an image as natural as possible. faceLAB^{TM}provides a system which records in real time blink, saccade and fixation estimates. For these reasons this system was chosen for the eyetracking during the pilot study. Images were rendered using path tracing on a quad core Q6850@3Ghz and 2GB RAM. All images were displayed on the faceLAB^{TM}laptop of screen resolution 1400x1050 pixels. All participants wore 5.1 surround headphones.

4.1.3. Methodology

Each participant was shown either images from set Phone or set No Phone due to the fact sequential images with a salient object removed could cause visual attention distraction and skew results in favour one way or the other. This kept participant exposure down to one set of images.

To analyse the eye tracking data for the feasibility study Look Zones (LZs) provided a method to capture time spent attending within a specific area. The areas chosen for LZs can be seen in figure 5. We took an average time spent in each LZ across all participants for each group, this information can be seen in table 1. This data can be seen for one participant in figure 6. Different sized Look Zones in effect may introduce bias to the results and skew them, however the zones were created to collect data about view history, the important zones are only time spent within the Phone bounding box and time spent out of it, the LZs used were consistent throughout the control, HQ and LQ tests however.

Image Set	Phone			No Phone		
LookZone(LZ)	*Control*	*HQSound*	*LQSound*	*Control*	*HQSound*	*LQSound*
Phone	5.4	10.8	9.45	2.1	5.1	3.4
Desk and Chair	6.2	2.7	3.9	5.7	4.7	5.4
Book and Candle	2.1	2.1	4.1	3.2	2.9	4.2
Lights	1.3	0.6	0.9	1.4	1.1	1.5
Other	13.0	11.8	9.65	15.6	14.2	13.5

Table 1: *Table showing LZ gaze duration averages, in seconds, across pilot participants.*

4.1.4. Feasibility Results

It could be argued the set for the pilot study was not large enough to draw conclusion about LQ sound especially with the value of significance being so close to the p=0.05 mark as shown in table 2. However for the purposes of extrapolating an experiment to decipher whether this effect can be utilised for selective rendering techniques it was appropriate to solely focus on the HQ sound effects for the selective quality tests as this sound had the most statistically pronounced result.

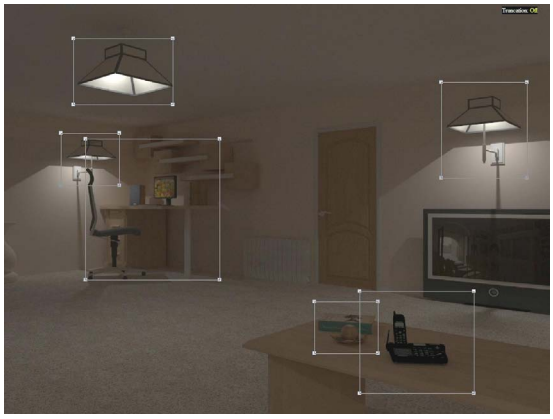

Figure 5: *Highlighting the Look zones used for data analysis. Sets are Lights, Desk and Chair, Book and Candle and Phone.*

Using the chi-squared metric upon this data (with Yates' correction due to degree of freedom) with expected values equal to our control results we see in table 2 that using HQ spatial sound was significant on the p = 0.05 level of significance,(p=0.0097, df=1) whilst there is a visual identifier for the sound present. Also we note that using LQ spatial sound was not deemed to be statistically significant in this sample set, however it is very much on the cusp of being statistically significant.

In table 3 it is shown that even without a visual identifier and with HQ sound present the excess time spent gazing in the direction of where the phone should have been was of statistical importance (p=0.0109, df=1). However

	Control	*HQSound*	*LQSound*
LZ Phone	5.4	10.8	9.45
LZ Not Phone	22.6	17.2	18.55
Chi-Squared p		**0.0097**	0.0524

Table 2: *Table showing the p values for the relevant gaze durations for Control, HQ and LQ sounds in the Phone category. Statistically significant results emboldened.*

	Control	*HQSound*	*LQSound*
LZ Phone	1.8	5.1	3.5
LZ Not Phone	26.2	22.9	24.5
Chi-Squared p		**0.0109**	0.1902

Table 3: *Table showing the p values for the relevant gaze durations for control, HQ and LQ sounds in the No Phone category. Statistically significant results emboldened.*

with LQ sound present this effect was mitigated to an extent as the result was not deemed to be important at the p = 0.05 level (p=0.19, df=1). Figure 7 shows one participants recorded gaze fixations and saccades whilst free viewing the No Phone set of images control vs. HQ sound.

4.2. Selective Quality Experiment

The second experiment used the same two images again. However this time they were rendered selectively outside of a foveal region at the following sample rates: (32SPP, 128SPP, 256SPP, 512SPP, 1024SPP and 2048SPP). The lowest sample rate chosen was at 32SPP because for the particular scene used noise rates were observed to be too high if a lower sample rate was used. The foveal region was rendered at the gold standard quality of 3072SPP as at this level the image appeared to be converged. The sample rates were chosen to give a wide test range of SPP levels to make it easier to determine at what point, if any, an effect came into play.

4.2.1. Participants

For the comparison of selective quality renderings we had 60 participants aged from between 13 to 66 with the average age 23.6. 39 males and 21 females constructed this set and the average knowledge level of computer graphics was

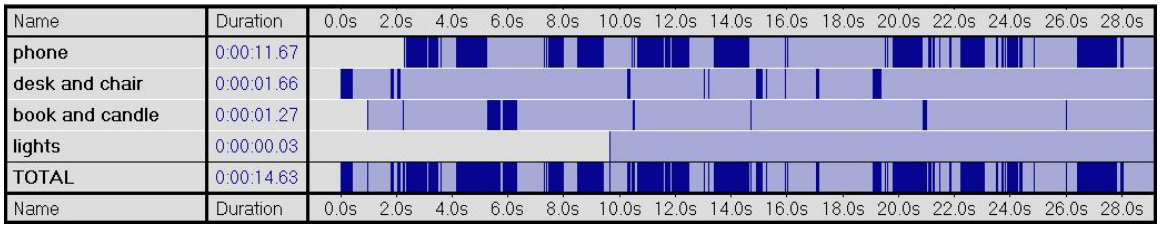

Figure 6: *Time spent attending within the specific Look Zones for one participant for the acoustic identifier and discrete spatial sound group.*

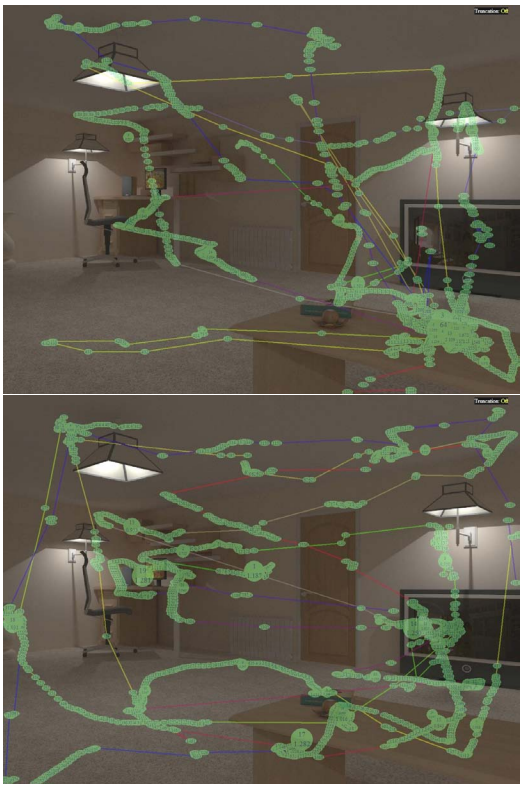

Figure 7: *Eye Track data from one participant of the pilot study. Lines are saccade shifts, small ovals represent gaze data and larger circles are fixations. (Top): No visual identifier and HQ spatial sound. (Bottom): Control for no visual identifier set.*

2.18 [range:1(low)..5(high)]. It was shown in [MDCT05b] that familiarity with computer graphics influenced participants decision accuracy. As is shown our sample set should not introduce too much bias due to this. Each participant took part in 6 of the 12 experiments meaning we retrieved 30 results per experiment.

4.2.2. Equipment

No eye tracking was necessary for this experiment, experiments took place on screens of similar screen dimension (30.6cm x 23cm) to the faceLAB^{TM}laptop and of the same screen resolution (1400x1050 pixels). All participants wore 5.1 surround headphones.

4.2.3. Methodology

Selective quality images were generated based upon a foveal angle surrounding where the phone (and thus directional sound) resided in screen space. This was the case for the image rendered omitting the phone also. Through trigonometry referencing Figure 1, it is possible to calculate the resolution of the eye at a specific distance away from the lens of the eye: $x = d\tan\theta$.

Participants underwent the experiment with the chair positioned such that their eyes were approximately 1 metre away from the screen and using the angle of 4.1° for the foveal region as shown in [LMYM01] we derive screen space of the foveal region which can be easily converted into pixel values for the screen (given we know the screen resolution and dimensions: 1400px, 30.6cm x-axis). This works out to be 7.16 cm of screen space, the equivalent of 328 pixels for the foveal diameter. The 4.1° foveal region was subtended by a further gradient of spp ramping 2° around the foveal region such that the rendering quality threshold change was much less spatially obvious. This can be seen at the 32 spp level in figure 8.

Subjects were asked to compare the quality of the images presented in each pair and choose the one they thought contained the higher rendering quality out of the two - two alternative forced choice (2AFC). They were forced to choose from one selective quality rendering and the reference image whilst exposed to high quality sound. The results were easier to intepret by consistingly choosing between SQ and HQ images. If the participants chose between a low SQ and a higher SQ image the results would be more ambiguous whilst maintaining the same level of data collection. Sets of images were displayed quasi-randomly such that each participant would make 10 choices in total, however 6 prevalent ones used for results, yet maintaining an even data distribution. 4 extra sets of images were thrown in as red herrings

with no difference between them (same SQ vs. same SQ or HQ vs. itself). This ensured the experiment was kept unpredictable. Images with accompanying spatial sound were presented sequentially separated by a black screen of 2 seconds duration.

4.2.4. Selective Quality Experiment Results

Based on previous research in this field we hypothesised that spatial sound would directly influence the HVS's top-down approach to the cognitive function of visualising the image. To verify our results we use chi-square statistical techniques due to the binary 2AFC responses. As was shown by the feasibility study, there is an effect worth considering when HQ sound is present for both scenes with a visual identifier for the sound and scenes without, however this effect did not prove to be significant using LQ sound for either case. This experiment proceeded just using HQ sound for the selective quality trials.

Table 4 shows the statistical significance against our null hypothesis for all cases. While a visual identifier, in our case a phone, is present there exists a statistical significance against the p=0.05 level for 32 and 128spp selective quality images (p=0.000012 and p=0.028, df=1 respectively). The null hypothesis that choosing images is distributionally random for spp levels of 256spp and higher for the Phone set cannot be rejected. For the No Phone set, the null hypothesis for selective quality levels of 32spp to 512 spp is rejected. Only for selective quality levels of 1024 and 2048spp can we not reject our null hypothesis.

ImageSet	*SPP*	X^2	*p*
Phone	32	**19.2**	**0.000012**
	128	**4.3**	**0.028459**
	256	2.13	0.144127
	512	2.13	0.144127
	102	0.53	0.465208
	204	0.53	0.465208
No Phone	32	**26.1**	**0.00000032**
	128	**19.2**	**0.000012**
	256	**22.5**	**0.000002**
	512	**8.53**	**0.003487**
	1024	2.13	0.144127
	2048	0.13	0.715000

Table 4: *Chi-squared results for quality comparison choices made between selecive quality images and a referance image. Statistically significant results emboldened.*

5. Conclusions and Future Work

This paper has presented a method which exploits the HVS's top-down approach. The fact that the HVS is guided by bi-modal impulses, allows us to selectively render the regions attended to in higher quality and the remainder of the image in a much lower quality without this quality difference being noticed. Our results show that even for test scenarios where a visual locator for the fovea to fixate on was missing from the image, there were still involuntary saccade shifts to the direction of the spatial sound. This was used to show that unattended areas of the scene could be rendered in lower quality than the location of the sound impulse capitulating on Inattentional Blindness. This result was similar to that of previous work, however it extended upon it by showing the results are applicable to spatial sound even when a visual cue is deliberately omitted.

Figure 8: *Showing the selectively rendered image which is 3072 spp around the predicted foveal region tailed off with a gradient displayed here to 32spp.*

There are many aspects of this work which require further investigation. Primarily, what was noticed during pilot testing was that saccade shifts occured much more frequently asynchronously with the introduction of audio and fixations seemed to last for the duration of the phone ring impulse and after the acoustic impulse faded eye movement often then carried on exploring the scene in free view mode. The next step is to investigate whether the effect described can be further exacerbated by exploting temporal selective quality shifts for the predicted foveal region based on the timing of the acoustic impulse: can we phase between high quality and selective quality renderings using a translation map based on acoustic impulse timings for the free view session. An investigation into whether this effect prevails when the acoustic identifier is animated with the sound attenuated accordingly should provide weight to this phenomena. This approach can accomodate any rendering algorithms which traces rays from the eye. We also intend to apply our work to a GPU based path tracer, and to final gathering from photon maps. Finally, the generation of a general saliency map applicable for bi-modality VR exposure would contribute greatly to the applications available to this research.

References

[Beg94] BEGAULT D.: *3-D Sound for Virtual Reality and Multimedia*. Academic Press Professional, 1994. 3

[BSVDD10] 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 (2010), 1–16. 3

[CCL02] CATER K., CHALMERS A., LEDDA P.: Selective quality rendering by exploiting human inattentional blindness: looking but not seeing. In *VRST '02: Proceedings of the ACM symposium on Virtual reality software and technology* (New York, NY, USA, 2002), ACM, pp. 17–24. 2

[CCW03] CATER K., CHALMERS A., WARD G.: Detail to attention: Exploiting visual tasks for selective rendering. In *Eurographics Symposium on Rendering 2003* (June 2003), ACM, pp. 270–280. 2

[Dow87] DOWLING J.: *The retina: An approachable part of the brain*. Cambridge: Belknap, 1987. 2

[GBW*09] GRELAUD D., BONNEEL N., WIMMER M., ASSELOT M., DRETTAKIS G.: Efficient and practical audio-visual rendering for games using crossmodal perception. In *I3D '09: Proceedings of the 2009 symposium on Interactive 3D graphics and games* (New York, NY, USA, 2009), ACM, pp. 177–182. 3

[HAC08] HULUSIC V., ARANHA M., CHALMERS A.: The influence of cross-modal interaction on perceived rendering quality thresholds. In WSCG 2008 Full Papers Proceedings, pp. 41–48. 2

[HCD*09] 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 *SCCG '09: Proceedings of the 25th Spring Conference on Computer Graphics* (2009). 2

[IK00] ITTI L., KOCH C.: A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research 40*, 10-12 (May 2000), 1489–1506. 2

[IKN98] ITTI L., KOCH C., NIEBUR E.: A model of saliency-based visual attention for rapid scene analysis, 1998. 1

[Kaj86] KAJIYA J. T.: The rendering equation. *SIGGRAPH Comput. Graph. 20*, 4 (1986), 143–150. 4

[Kel97] KELLER A.: Instant radiosity. In *SIGGRAPH '97: Proceedings of the 24th annual conference on Computer graphics and interactive techniques* (New York, NY, USA, 1997), ACM Press/Addison-Wesley Publishing Co., pp. 49–56. 3

[Kut91] KUTTRUFF H.: *Room Acoustics (3rd Edition)*. Elseiver Applied Science, 1991. 2

[LCM07] LAUTERBACH C., CHANDAK A., MANOCHA D.: Interactive sound propagation in dynamic scenes using frustum tracing. In *IEEE Trans. on Visualization and Computer Graphics* (2007), vol. 13, pp. 1672–1679. 3

[LM99] LOSCHKY L., MCCONKIE G.: Gaze contingent displays: Maximizing display bandwidth efficiency. In *ARL Federated Laboratory Advanced Displays and Interactive Displays Consortium* (1999), Advanced Displays and Interactive Displays Third Annual Symposium, pp. 79–83. 2

[LMYM01] LOSCHKY L., MCCONKIE G., YANG J., MILLER M.: Perceptual effects of a gaze-contingent multi-resolution display based on a model of visual sensitivity. In *ARL Federated Laboratory Advanced Displays and Interactive Displays Consortium* (2001), Advanced Displays and Interactive Displays Fifth Annual Symposium, pp. 53–58. 2, 6

[Mas06] MASTOROPOULOU G.: The effect of audio on the visual perception of high-fidelity animated 3d computer graphics. PhD Thesis, University of Bristol, 2006. 2

[MBT*07] 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* (New York, NY, USA, 2007), ACM, pp. 189–196. 3

[MDCT05a] MASTOROPOULOU G., DEBATTISTA K., CHALMERS A., TROSCIANCO 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* (2005), ACM Press, pp. 363–369. 2

[MDCT05b] MASTOROPOULOU G., DEBATTISTA K., CHALMERS A., TROSCIANCO T.: The influence of sound effects on the perceived smoothness of rendered animations. In *APGV 2005: Proceedings of the 2nd Symposium on Applied Perception in Graphics and Visualization* (August 2005), ACM SIGGRAPH, pp. 9–15. 2, 6

[ML97] MCCONKIE G., LOSCHKY L.: Human performance with a gaze-linked multi-resolutional display. In *ARL Federated Laboratory Advanced Displays and Interactive Displays Consortium* (1997), Advanced Displays and Interactive Displays First Annual Symposium, pp. 25–34. 2

[MR98] MACK A., ROCK I.: Inattentional blindness. *Massachusetts Institute of Technology Press* (1998). 2

[NBH*08] NOESSELT T., BERGMANN D., HAKE M., HEINZE H.-J., FENDRICH R.: Sound increases the saliency of visual events. *Brain Research 1220* (2008), 157 – 163. Active Listening. 2

[RLC*07] RAGHUVANSHI N., LAUTERBACH C., CHANDAK A., MANOCHA D., LIN M. C.: Real-time sound synthesis and propagation for games. *Commun. ACM 50*, 7 (2007), 66–73. 3

[Yar67] YARBUS A.: Eye movements during perception of complex objects. *In L. A. Riggs, Ed., Eye Movements and Vision 7* (1967), 171–196. 2

[YPG01] YEE H., PATTANAIK S., GREENBERG D. P.: Spatiotemporal sensitivity and visual attention for efficient rendering of dynamic environments. *ACM Trans. Graph. 20*, 1 (2001), 39–65. 1