PERCEPTUALLY OPTIMISED VIRTUAL ROOM ACOUSTICS

Dale Johnson and Hyunkook Lee

Applied Psychoacoustics Laboratory
The University of Huddersfield
{Dale.Johnson, H.Lee}@hud.ac.uk

ABSTRACT

This paper presents the development of a method of perceptually optimising the acoustics and reverb of a virtual space. A spatial filtering technique was developed to group artificially rendered reflections by what spatial attribute they contribute to e.g. apparent source width, distance, loudness, colouration etc. The current system alters the level of different reflection groups depending on the desired type of optimisation. It is hoped that in the future this system could be coupled with machine learning techniques, such that it is able to determine the initial perceptual qualities of the artificial reverb, then optimise the acoustics depending on the user's needs. Such a system could ultimately be used to universally identify what spatial qualities are good and bad, then generically optimise the acoustics automatically.

INTRODUCTION

In recent years there has been a rise in automated mixing and optimisation techniques. Geometric acoustics algorithms give access to individual reflections during rendering of the impulse response, which is not otherwise normally possible with existing algorithms. This would then allow for control or optimisation techniques to be applied to the acoustics, such as enhancement of apparent source width (ASW) or listener envelopment (LEV), reduction in tonal colouration, or even improvement of locatedness of an auditory source. ASW refers to how wide the auditory source appears to the listener, whilst LEV refers to how enveloped by the resulting reverberance a listener may feel. It is generally agreed that high ASW and LEV in a concert hall is the sign of good acoustical quality [1].

However, to achieve any of those techniques, the underlying mechanisms of psychoacoustic phenomenon such as 'Spatial Impression' or the perception of 'timbre' and 'colouration' must be thoroughly understood. Further to this, it should be understood what attributes are perceived as being positive or negative. This will allow for analysis techniques to be developed that could identify groups of reflections based upon criteria such as their direction of arrival or delay time. For example, a group of early reflections that would arrive within horizontal or lateral region to the listener would have the greatest influence over ASW which, according to [1], is a practical measure of acoustic quality of a concert hall.

After identification, the reflections can then be manipulated in a variety of ways to affect their associated attribute whilst, ideally, not affecting others in order to maintain the perceptual plausibility of perceived reverb. Whilst a few reverberation algorithms with perceptual control exist, for example [2] and [3], they do not provide 'granular' control over individual reflections or classify them as contributors to particular phenomena such as ASW and LEV. The custom virtual acoustics program developed for this study allows for processing of arbitrary room models, as well as analysis and manipulation of independent reflections.

This system could ultimately be coupled with intelligent algorithms or machine learning techniques such as artificial neural networks (ANNs), so that it can automatically improve the subjective quality of the artificial reverb without modifying the geometry of the model. This paper will discuss and summarise a body of currently on-going research into the development of perceptual control and optimisation techniques for virtual room acoustics.

1. VIRTUAL ROOM ACOUSTICS

Virtual room acoustics, or alternatively artificial reverberation, deals with modelling the behaviour of sound in an enclosed spaced. This was initially attempted by [4] who modelled room reverberation using a bank of parallel comb filters cascaded with two all-pass filters. By utilising differing delay times for each filter, the algorithm was designed to increase the echo density of the audio signal, much like what would occur in real enclosed space. This method is known as an algorithmic reverberator and was later improved by [5] and [6]. The main limitation with algorithmic reverberators is they are not able to model any particular type of space, so serve as generalised models.

An alternative to this type is the geometric reverberator. This uses geometry to trace the exact path that sound takes as it propagates through a room and reflects of surfaces, often using a three-dimensional, computerised model of a real space. The two main types of geometric reverberation methods are the Image Source Method (ISM) initially developed by [7], and Ray Tracing, adapted for modelling acoustics by [8]. These two techniques are often used together to create a hybrid method [9], where the ISM accurately models the early reflections, whilst ray tracing is used to model the late reverb tail.

For this study, a custom, hybrid geometric acoustics program was developed in C++. It implements both ISM and Ray Tracing, is able to model any arbitrarily shaped space, and simulates octave band material absorption and diffusion. It is able to produce a multi-channel room impulse response (RIR) in the form of a WAV file, or as a custom raw format known as a 'Raw Impulse Vector' (RIV). This format stores metadata of each captured reflection, such as the direction of arrival, distance travelled, surface history and octave band energy. The benefit of this format is that it allows post processing and analysis that is not possible with WAV based RIRs, and thus serves as the primary motivation for the creation of the custom algorithm.

2. CREATING THE OPTIMISATION TECHNIQUE

The main use for the RIV format in this study was to allow for analysis and manipulation of early reflections arriving between 50-80ms in order to control, enhance or supress certain psychoacoustic attributes, primarily Apparent Source Width (ASW), and Tonal Colouration in order to optimise the acoustical quality of the room.

ASW describes the how wide and auditory source may appear to a listener, and it can be measured using the Inter-Aural Cross-correlation Coefficient (IACC) [1] or Lateral Fraction (LF) [9]. IACC measures the correlation between the two ear signals of a Binaural RIR, whilst LF measures the ratio of lateral reflection energy to all reflection energy, and thus is reflection angle dependent. LF assumes that the ASW continuously increases as reflection angle approaches 90°, where it would be at maximum. However, after an interpretation of results produced by [10], a test conducted by the authors [11] found that region of maximum, or saturated ASW exists in the horizontal plane between ~40°–135°. Any reflection arriving within 30ms inside this region, denoted hereafter as ASW_{max}, will produce the maximum amount of perceived ASW.

Tonal colouration on the other hand is described by [12] as a change in timbral quality of a sound source. It manifests itself as a 'boxy', 'metallic', 'phasey' or 'comb filter' artefact. Whilst this colouration naturally occurs due to the interaction between reflections and the direct sound, if the degree of colouration is large enough it can be perceived as being unacceptable.

A series of currently unpublished tests performed by the authors observed the effect of reflection angle and level on the acceptability of colouration. It was found that for a -6 dB reflection arriving between 2.5 to 30ms produced unacceptable colouration between ±45° in front of the listener, and ±135° at the rear. The tests also found that if the reflection level under -6 to -7 dB, the colouration is perceived as acceptable, whilst still being audible.

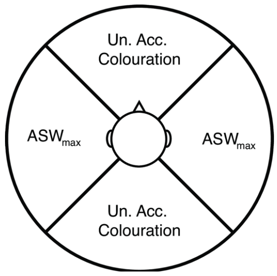

Figure 1 – The two proposed reflection grouping regions mapped around the listener

These regions can now be used as part of a potential optimisation technique. Using the region boundaries, the reflections extracted from a RIV produced by the custom geometric reverberator can be sorted into two groups depending on their arrival time and azimuthal direction. The ASW_{max} region focuses on reflections that will have the greatest influence on ASW, whilst the other region focuses on reflections that create perceptually unacceptable tonal colouration. To enhance the ASW, the energy of the lateral reflections arriving in the ASW_{max} region can be boosted by 6 dB. Alternatively, to reduce the ASW the energy can be reduced by 6 dB. According to [10], the level of lateral reflections directly affects the perceived ASW, thus increasing the lateral reflection energy would increase the amount of LF. Likewise, to improve the acceptability of any perceived colouration, the reflections that arrive in the two regions of unacceptable colouration can be lowered to an acceptable level by reducing the level to under -6 dB.

3. TESTING THE TECHNIQUES

After creating the two grouping regions, the following research questions were asked: To what degree does manipulating lateral reflection energy alter ASW? And to what degree does manipulating front-back reflection energy change the perceived colouration? To test the effectiveness of the above methods, a series of listening tests were devised. The first test was an elicitation test to deduce attribute associated with the change in colouration,

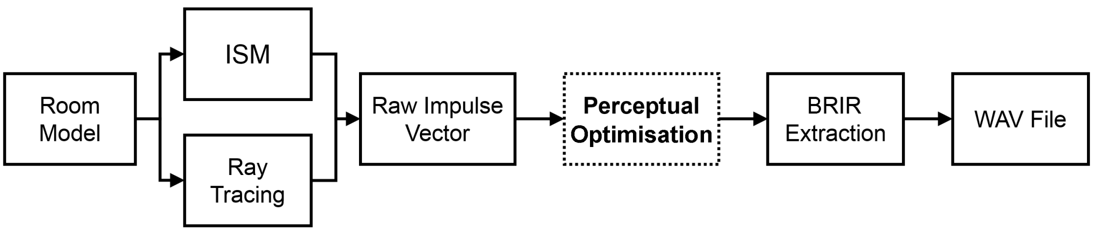

Figure 2 – Overview of the rendering pipeline including the geometric reverb and perceptual optimisation.

and other differences that may occur when the reflection levels are changed. The second will then test effectiveness of the control methods upon the artificially generated Binaural RIRs (BRIRs).

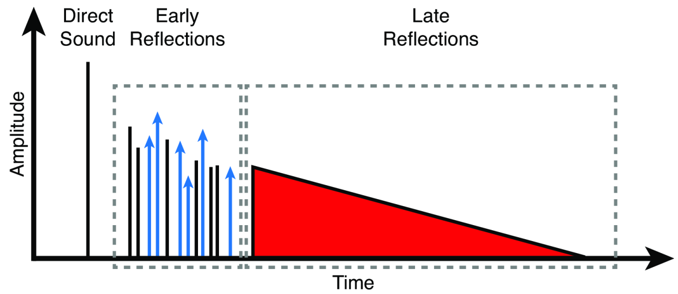

Figure 3 – Manipulation of the level of selected reflections (shown as blue arrows) in order to a control a perceptual attribute.

3.1. Stimuli

RIVs of two simple room models were created, one of a concert hall and the other a large, shoebox shaped room. Within each room, an omnidirectional source was placed towards the end of the rooms, whilst an omnidirectional receiver was placed at either 6m or 12m away from the source. This created four RIVs in total, two for each space. The RIVs were converted into BRIRs using a spatialisation process used by [13]. Using Head Related Transfer Functions (HRTFs) from the MIT KEMAR database [14], each reflection is convolved with a HRTF that corresponds with its direction of arrival, then summed together to form a BRIR. During the spatialisation process, the grouping regions are used to separate out either lateral or front-back reflections. The level of either of group is boosted or reduced by up to maximum of 6 dB in 3 dB steps, see Fig. 3. To create the actual stimuli, three anechoic sources, guitar, male speech and orchestra, were convolved with each BRIR. This would highlight how effective the control methods are with either continuous, transient or melodious sound sources.

3.2. Methodology

The second test was a multiple comparison test. For each trial, 5 stimuli with either lateral or front-back gain adjustment were presented, and subjects were asked to compare the stimuli against the reference that has no gain adjustment, then grade the stimuli by how much more or less ASW or Phasiness they perceived. The test was performed in HULTI-GEN [15] over Sennheiser HD650 headphones using a Merging Horus audio interface at 44.1 kHz, taking place in the ITU-R BS.1116 compliant critical listening room at The University of Huddersfield.

Twelve subjects took part in the experiment, four of which are critical listeners whilst the rest are music technology students from the same university. The students have moderate critical listening ability and an interest in psychoacoustics and concert hall acoustics.

3.3. Main findings

Statistical analysis on the test results found that for all rooms and source-receiver distances, only a 6 dB increase in lateral reflection energy caused a significant increase in ASW for orchestra. However, for guitar it was only effective at 6m in both rooms, and for speech in the large room at 12m. The experiment also found in certain cases, namely for guitar in the concert hall at 6m and the orchestra at 12m, an increase in front-back reflection energy also caused a significant increase in ASW. A reduction in energy did not cause a significant reduction in ASW. This therefore means that a fairly large, 6 dB boost in lateral reflection energy can enhance the perceived width, yet a reduction in energy did little to diminish it. It is possible that when the reflection energy is reduced, it could be dropping below a threshold for just noticeable difference in width.

As for phasiness, statistical analysis found that in certain cases gain adjustment had a significant effect. However, a Bonferroni adjusted Wilcoxon signed-rank test that in comparison to the 0dB reference, neither type of gain adjustment had any significant effect on the perceived phasiness.

4. COMPARISON BETWEEN METHODS

The tests performed in the previous section found that only a 6dB increase in lateral reflection energy had a significant effect on ASW. However, the same increase applied to front-back reflection energy also had an unforeseen effect on ASW, although it was unclear if the change in ASW due to front-back adjustment was as large as a change due to the adjustment of lateral reflection energy. Therefore, an identical multiple comparison test was performed that compared adjustment of the two reflection groups against each other and an unaltered reference in terms of difference in ASW and phasiness.

Furthermore, the elicitation test found that the manipulation of reflection energy of either set of regions also appeared to affect the perceived source ‘Loudness’ and ‘Distance’. Therefore, the multiple comparison test also investigated the effects of gain adjustment upon these attributes.

4.1. Methodology

The same pool of stimuli from the previous experiment were used although, however, both types of gain adjustment were combined per trial so that they could be directly compared, rather than being independently tested. The ±3 dB difference conditions were excluded as they were found to not cause a noticeable or significant effect. This kept the number of stimuli per test down to 5.

4.2. Findings

Statistical analysis of the test results found that increasing the energy of either lateral or front-back reflections both had a significant effect on ASW, and no significant effect on phasiness, which supports the findings from the previous experiment. However, it does show that the change in front-back reflection energy did not have the anticipated effect on phasiness, and instead had an effect on ASW. This was unexpected as lateral fraction [10] would predict a decrease in ASW if the front-back reflection energy was increased.

It was also found that an increase in energy of both reflection groups on the perceived source loudness and distance, where either the source was perceived to be significantly louder and closer when the energy was increased. This highlights a limitation with the current control methods. From a perceptual control viewpoint the methods should only affect their corresponding attribute, whilst this experiment found that they also affected loudness and distance.

5. CONCLUSION

This paper proposes a system that is designed to optimise the acoustics of a virtual room by enhancing either ASW or improve the acceptability of any perceived tonal colouration. This was achieved via manipulation of the energy of specific groups of reflections that are generated by a custom geometric, virtual acoustics algorithm. Whilst the experiments found that increasing lateral reflection energy enhanced ASW, it was also enhanced when front-back reflection energy were increased. Furthermore, it was found that reducing front-back reflection energy did not effectively reduce the perceived tonal colouration, if any. Secondly, experimentation found that manipulation of reflection energy in this fashion affected loudness and distance perception. This shows that further psychoacoustic research must be conducted to understand what adaptations to the current method must be performed in order to achieve the original goal.

Whilst the system is still in development, the current research presented here demonstrates that it has the potential to become an important tool for intelligently improving the acoustics of arbitrary virtual rooms. The algorithm could be coupled with cutting edge machine learning techniques that could better identify reflections and features in the RIR that cause ASW, colouration, and other perceptual parameters that are part of the spatial impression paradigm, and then manipulate those to improve and optimise the acoustics of the room. The system would be highly useful for virtual reality scenes that require a high quality and immersive experience. The acoustical quality of the scene could be automatically improved by an intelligent system in order to make it more perceptually pleasing.

6. REFERENCES

[1] T. Hidaka, L. L. Beranek and T. Okano, “Interaural cross-correlation, lateral fraction, and low-and high-frequency sound levels as measures of acoustical quality in concert halls,” *J. Acoust. Soc. Am,* vol. 98, no. 2, pp. 988-1007, 1995.

[2] Z. Rafii and B. Pardo, “Learning to control a reverberator using subjective perceptual descriptors,” in *10th International Society for Music Information Retrieval Conference (ISMIR)*, 2009.

[3] T. Carpentier, M. Noisternig and O. Warusfel, “Hybrid reverberation processor with perceptual control,” In *Proc of the 17th Int. Conference on Digital Audio Effects-DAFx-14,* Erlangen, Germany, 2014.

[4] M. R. Schroeder, “Natural sounding artificial reverberation,” *J. Audio. Eng. Soc,* vol. 10, no. 3, pp. 219-223, 1962.

[5] J. A. Moorer, “About this reverberation business”, *Computer Music Journal*, vol. 3, no. 2, pp. 13-28, 1979.

[6] J. M. Jot and A. Chaigne, “Digital delay networks for designing artificial reverberators,” In *Proc. of the 90th Audio Engineering Society Convention,* Paris, France, 1991.

[7] J. B. Allen and D. A. Berkley, “Image method for efficiently simulating small-room acoustics,” *J. Acoust. Soc. Am,* vol. 65, no. 4, pp. 943-950, 1979.

[8] A. Krokstad, S. Strøm and S. Sørsdal, “Calculating the acoustical room response by the use of a ray tracing technique,” *J. Sound Vib,* vol. 8, no. 1, pp. 118-125, 1968.

[9] M. Vorländer, “Simulation of the transient and steady-state sound propagation in rooms using a new combined ray-tracing/image-source algorithm,” *J. Acoust. Soc. Am,* vol. 86, no. 1, pp. 172–178, 1989.

[10] M. Barron and A. H. Marshall, “Spatial impression due to early lateral reflections in concert halls: the derivation of a physical measure”. *J. Sound Vib,* vol. 77 no. 2, pp. 211–232, 1981.

[11] D. Johnson and H. Lee, “Just noticeable difference in apparent source width depending on the direction of a single reflection,” In *140th Audio Engineering Society Convention*, Berlin, Germany, 2017.

[12] T. Halmrast, “Orchestral timbre: Comb-filter coloration from reflections,” *J. Sound Vib,* vol. 232, no. 1, pp. 53-69, 2000.

[13] L. Savioja, “Modeling techniques for virtual acoustics”, *PhD Thesis*, Helsinki University of Technology, 2000.

[14] W. G. Gardner and K. D. Martin “HRTF measurements of a KEMAR,” *J. Acoust. Soc. Am,* vol. 97, no. 6, pp. 3907-3908, 1995.

[15] C. Gribben and H. Lee. “Toward the development of a universal listening test interface generator in max,” In *138th Audio engineering society convention*, Warsaw, Poland, 2015.