Audio Engineering Society

Convention e-Brief 247

Presented at the 140th Convention
2016 June 4–7 Paris, France

Investigation into the perceptual effects of image source method order

Dale Johnson and Hyunkook Lee

Applied Psychoacoustics Laboratory, The University Of Huddersfield, Huddersfield, West Yorkshire, HD1 3DH
dale.johnson@hud.ac.uk, h.lee@hud.ac.uk

ABSTRACT

This engineering brief explores the perceived effects and characteristics of impulse responses (IRs) generated using a custom, hybrid, geometric reverb algorithm. The algorithm makes use of a well known Image Source Method (ISM) and Ray Tracing methods. ISM is used to render the early reflections to a specified order whilst ray tracing renders the remaining reflections. IRs rendered at varying ISM orders appear to exhibit differences in perceptual characteristics, particularly in the early portion. To understand these characteristics, an elicitation test based was devised in order to acquire terms for the different characteristics. These terms were grouped in order to provide attributes for future grading tests.

1. INTRODUCTION

Virtual acoustics software is often used for architectural design and more recently, virtual or augmented reality. Software packages often use geometric methods for modeling the propagation of sound waves, where commonly used are the Image Source Method (ISM) [1] and Ray Tracing [2]. ISM can only model specular reflections, whilst ray tracing can model both specular and diffuse reflections. It could been seen at first glance, that ray tracing would be a preferable method however, the major difference between the two algorithms is their detection method. ISM is deterministic and uses a point receiver, so all valid reflections will be detected, however ray tracing is stochastic and requires the use of a receiver sphere, which if not optimal, may lead to systematic errors such as missed rays or detection of invalid reflections [3].

However, ISM can become computationally expensive at increasing orders, and so it is desirable to combine it with another algorithm to improve efficiency, for example [4] combine ISM with a filter delay network to model a shoebox room, and [5] combine ISM with ray tracing to model arbitrarily shaped models.

The ISM order is usually dependent on the desired range of early reflections. Concert hall acoustic research [6] generally suggests that the early lateral reflections, or first 50-80ms of an impulse response (IR) are an important factor in the perception of spatial impression, whereas late reverberation contributes to the perception of listener envelopment (LEV). However, there are some research questions that have not yet been formally answered:

(i) as ISM order increases, what is the perceived effect of the increased amount of specular energy?
(ii) at what point should ISM stop rendering?

As the first step towards addressing the questions, an elicitation test was conducted in such a way that subjects freely described the perceived qualities of various IRs generated with varying ISM orders. These terms were then grouped into a common set of attributes through a group discussion.

2. VIRTUAL ACOUSTICS SOFTWARE

For the purposes of concert hall acoustics and psychoacoustics research, a 3D virtual acoustics rendering tool was developed in C++. This tool is able to render the acoustics of a virtual three-dimensional room model and produce a multichannel impulse response.

2.1. Rendering Algorithm

The tool implements a hybrid system consisting of two sub-algorithms: the Image Source Method (ISM), and Ray Tracing. The ISM is based on the work developed by [1][4][5][7][8] and the ray tracing sub-algorithm is based on work developed by [2][9][10]. To briefly explain the algorithms, firstly the ISM computes the impulse response by the means of recursive image expansion to a specified order. The impulse response is formed by computing the distance from each virtual image source to the receiver. ISM is limited to rendering specular reflections only. Ray tracing emits rays in all directions from the source. These are reflected indefinitely until their energy is negligible. The impulse response is formed by *detecting,* if any rays cross a receiver sphere. It is possible to render diffuse reflections by randomising the ray direction upon reflection. This is based upon the vector scattering method described by [11].

The rendering process involves calculating the early to intermediate reflections with ISM. All the possible images to a specified order are calculated, then using a process of back-tracing and visibility testing, each reflection path is validated. Once this has been completed, the remaining reflections are calculated using ray tracing. During ray tracing, rays that would be calculated using ISM are exempt from detection until the reflection order overtakes the ISM order. As each ray is 'detected' by a receiver, it is cached into a two-dimensional vector. Each row of the vector belongs to a receiver in the model. This enables any post-processing and analyses to be performed upon a raw impulse response without the need for re-rendering the entire impulse response from scratch.

2.2. Impulse Response Forming

A discrete multichannel IR is formed by extracting the delay times of each ray in reference to the speed of sound. The energy is response is formed by multiplying the amplitude by the corresponding receiver's polar pattern, and forming octave-band channels for each energy band. These are then filtered using linear phase, Butterworth band-pass filters and summed to form the final impulse response. This is done for each receiver.

2.3 ISM order & receiver size

A set of factors to consider is the approach ISM and ray tracing use in order to render room acoustics. ISM is deterministic, in which it is guaranteed to find all valid reflections. Ray tracing is stochastic, whereby it relies on probability that rays will cross a receiver sphere. Systematic errors may occur if the ray count and receiver radius are not optimal. The receiver size can be determined using equation 1, described by [12]:

$$r = \left(\frac{15V}{2\pi N}\right)^{\frac{1}{3}} \quad (1)$$

where r is the receiver radius, V is the room volume and N is the ray count.

This implies that it is likely that the density of reflections may increase as ISM order increases, up to the point where ray tracing will continue the rendering process. Therefore, this may have an effect on the perceptual characteristics of the IR.

3. EXPERIMENTAL DESIGN

In order to understand what the perceived effect of ISM order is, a free elicitation test based upon the QDA method [9] was devised. The experiment was split into two phases: an individual testing phase, followed by a group discussion phase. The listening tests were conducted in the ITU-R BS.1116-compliant listening room at the University of Huddersfield.

3.1 Room model

The room model constructed for the experiment is a typical concert hall sized room who's dimensions are 16.18m x 10.0m x 26.18m (W x H x D). All surfaces in the model have an absorption coefficient of 0.01 across the entire frequency spectrum. The reason for this is to minimise the effect absorption may have upon the experiment. A single sound source is positioned at (8.09, 2.0, 23.0). Two omni-directional receivers configured as a 0.5m spaced AB pair are positioned at 3m, 6m and 12m from the receiver, and are also raised to a height of 2m.

3.2. Stimuli

For the experiment, three anechoic recordings from the Bang & Olufsen Archimedes project CD [10] were used: trumpet, spoken dialogue, and conga excerpts. These recordings were chosen in order to examine the effects that different characteristics such as continuous notes, transients, and complex harmonic content may exhibit when processed with the artificial reverb.

A total of thirty six IRs were generated. These can be categorised into three groups depending on the source-receiver distance: 3m, 6m and 12m. Each IR in each group was generated using a different ISM order from 1 to 6, however with a constant number of 75,000 rays, with a calculated receiver radius of 0.6m. The IRs are interleaved into eighteen, stereo, 44.1kHz WAV files. Each sound source is then convolved with each IR to produce fifty four stimuli in total.

Figure1 shows an example of the IRs produced at ISM order 1 and ISM order 6 at 3m distance.

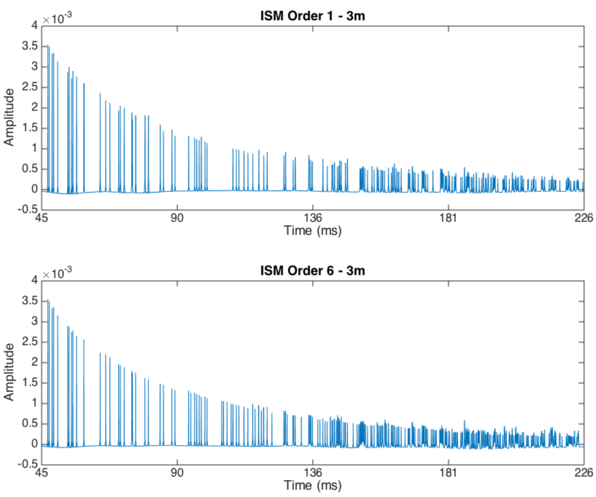

Figure 1. ISM Order 1 and 6 between 45 – 226 ms

The density from 100ms is higher in the IR rendered at ISM order 6 than ISM order 1. This difference may be caused systematic errors in an IR rendered at order 1, essentially solely with ray tracing.

3.2. Subjects

The experiment requires experienced critical listeners. Five members of the Applied Psychoacoustics Laboratory of the University of Huddersfield participated in the elicitation test and group discussion. All had much experience in spatial audio evaluation.

3.3. Elicitation test

Phase one was a listening test consisting of nine trials, and in each trial there are six audio samples. The test interface was designed in Cycling 74's Max 7. The order of the samples was randomised per trial.

In each trial, subjects are asked to play the six samples, and then describe and grade the apparent effects they were able to perceive by typing into a text box. The samples are looped and play in synchronisation so it is possible to switch between samples freely. After each test the data is analysed and any elicited terms are extracted and added to a spreadsheet.

4.4. Group discussion

The objective of the group discussion is to debate the meanings behind each elicited terms and under what spatial or timbral attribute to group them as. After a short period of time after the end of phase one, the subjects were asked to attend a group discussion in the same listening room. The terms gathered from phase one were displayed to the subjects. The objective was explained to the subjects, and over a period of two hours the terms from phase one were interpreted, discussed, and then grouped into categories that were deemed suitable the by entire group.

5. RESULTS AND DISCUSSION

The terms elicited from phase one were grouped into the categories extracted from phase two, as shown in table 1. The table shows how listeners are able to describe each attribute using many different terms. Although this was not a preference test, it was also interesting to note how terms such as 'rough' and 'muddy' were described as being a negative terms. On the other hand, terms such as

Table 1. Attributes and terms elicited from the experiment

Attribute	Description	Terms
Environmental image spread (51)	Environment related component. Horizontal / vertical spread.	Boxy (11), Diffuseness (1), Echoey (8), Frontal (1), Larger (5), 'Opening Up' (2), Reverberant (4), Room size (5), Short Reverb (1), Smaller (3), Spaciousness (8), Spatial Impression (2)
Roughness (38)	How rough / smooth after initial transient	Comb filtered (1), Natural (14), Phasey (4), Rattley (2), Roughness (11), Smooth (2), Unnatural (4)
Distance (24)	The impression of source-listener distance	Closer (9), Distance (9), Nearer (2), Presence (1), Proximity (1), Wetness (2)
Clarity (13)	Perception of intelligibility	Clarity (3), Clear (2), Intelligibility (1), Mid resonance (1), Muddy (6)
Thin / Full (6)	Amount of low frequency energy	Fullness (1) , Low frequency energy (2), Thin (2), Weight (1)
Dull / Bright (15)	Amount of high frequency energy	Bright (5), Dark (1), Dull (6), High frequency energy (2), High end (1)
Vertical Shift (4)	Perceived vertical shift in image	Localisation shift (1), Vertical image shift (3)
Loudness (15)	The apparent loudness of the source	Amplitude (2), Direct sound increase (3), Louder (3), Loudness (2), Lower mix of reverb (1), Presence (1), Stronger direct sound (1), Wetness (2)
Resonant (4)	Emphasis on certain frequencies	Nasal (2), Resonant (2)
Echo Perception (5)	The perception of additional echoes	Echo (2), Echoey (3)
Source Focus (3)	Apparent separation of source from reverb	Source focus (3)

'natural', 'smooth' and 'bright' were described as positive terms. This gives rise that the attributes are bipolar. The attributes with the largest occurrence are 'Environmental Image Spread', 'Roughness', 'Distance' and 'Loudness'.

The high occurrence of the 'Environmental Image Spread' attribute implies that the ISM order has an effect over the perceived environmental properties of the room. It could be for example that a low ISM order leads to a 'boxy' and 'small' characteristic whilst a higher order leads to an apparent 'opening up', increased 'Diffuseness' or an increased, 'Echoey' quality. The increasing density of specular energy at higher orders could potentially be playing an important role.

Another highly occurring attribute is 'Roughness', and the most elicited terms in this attribute are 'rough' and 'natural'. Again, this suggests that the ISM order has an effect upon the apparent timbral quality and plausibility of the IR. It may be the case that sub-optimal ISM orders will result in a 'rough', unnatural sounding reverb, whilst higher orders will lead to a more 'natural' or 'smooth' reverb.

The 'Distance' attribute is likely related to differing distances the IRs have been rendered at, however it may be working in conjunction with the ISM order in which higher orders could be adding a sense of depth to the perceived reverb. This attribute may also be working in conjunction with the 'Loudness' attribute in which the source-listener distance is merely affecting the apparent loudness of direct sound against the reverb.

'Clarity', 'Thin / Full' & 'Dull / Bright' appear are timbral attributes. The ISM order here may be contributing to the overall intelligibility in which an increase in order may lead to an increase in intelligibility, however absorption may also be a contributing factor.

The next stage in the investigation would be to firstly look into what extent the attributes obtained from the group discussion are perceived. This will be performed through the use of grading tests in which listeners will grade the strength in which they perceive the attribute. Secondly, the effects of the ISM order should be investigated using objective measures in order to compare analytical results with the results from the grading tests. These measures will include ICCC, C80, D50 and G.

It is worth noting that the terms elicited in this engineering brief can be seen as being applicable only to the room model described earlier. Therefore a future stage would investigate further what effect the dimensions and geometry of a room has on the attributes. If an objective measure can be formed, it would be useful in choosing an optimum ISM order for a perceptually plausible IR.

As the experiment was performed using only a stereophonic setup, other attributes such as ASW and LEV cannot be truly investigated, therefore future work will include performing ISM order experiments in multichannel 3D and binaural setups. This future work may lead to an attribute grading that may differ from stereophonic experiments.

6. SUMMARY

The aim of this engineering brief was to gain insight into the perceptual effects of artificial IRs rendered using differing ISM orders. To obtain perceptual attributes associated with the perceived effects, a free-elicitation method was used to obtain terms. The terms were then categorised into attributes that would describe the type of effect that was being perceived.

Analysis of the results show that the most elicited terms belonged to the attributes 'Environmental Image Spread', 'Roughness', 'Distance' and 'Loudness'. These results imply that ISM order may have connection to the degree of perception of these attributes. This can be investigated further through the use of grading tests and objective measures.

7. REFERENCES

[1] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics", *J. Acoust. Soc. Am*, vol. 65, pp 943-950 (1979).

[2] A. Krokstad, S. Strøm and S Sørsdal, "Calculating the acoustical room response by the use of a ray tracing technique", *J. Sound and Vib*, vol. 8, pp 118-125 (1968).

[3] S. Laine, S. Siltanen, T. Lokki and L. Savioja, "Accelerated beam tracing algorithm", *App. Acoust*, vol. 70, pp 172-181 (2009).

[4] T. Wendt, S. Van De Par and S. D. Ewert, "A Computationally-Efficient and Perceptually-Plausible Algorithm for Binaural Room Impulse Response Simulation." *J. Audio Eng. Soc*, vol. 62, pp 748-766 (2014), http://dx.doi.org/10.17743/jaes.2014.0042

[5] L.Aspöck, S. Pelzer, F. Wefers and M. Vorländer, "A Real-Time Auralization Plugin for Architectural Design and Education", *Proc. Of the EAA Joint Symposium on Auralization and Ambisonics*, (2014).

[6] J. S. Bradley and G. A. Soulodre, "Objective measures of listener envelopment", *J. Acoust. Soc. Am*, vol. 98, pp 2590-2597 (1995).

[7] 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, pp 172-178 (1989), http://dx.doi.org/10.1121/1.398336

[8] E. A. Lehmann and A. M. Johansson, "Prediction of Energy Decay in Room Impulse Responses Simulated with the Image-Source Model", *J. Acoust. Soc. Am*, vol. 124, pp 269-277 (2008), http://dx.doi.org/10.1121/1.2936367

[9] A. Kulowski, "Algorithmic Representation of the Ray Tracing Technique", *Applied Acoustics*, vol. 18, pp 449-469, (1985).

[10] D. O. Elorza, "Room acoustics modeling using the raytracing method: implementation and evaluation", *Licentiate Thesis Department of Physics. University of Turku* (2005).

[11] J. H. Rindel, "The Use of Computer Modeling in Room Acoustics", *J. Vibroengineering*, vol. 3, pp 41-72 (2000).

[12] Z. Jiang, C and X. Qiu, "Receiving radius determination in ray-tracing sound prediction of rectangular enclosure", *J. Sound and Vib.*, vol 301(1), pp 391-399 (2003).

[13] H. Stone and J. L. Sidel, "Sensory Evaluation Practices", 3rd ed, *Elsevier Academic Press*, pp 215-235 (2004).

[14] V. Hansen and G. Munch, "Making Recording for Simulation Tests in the Archimedes project", *J. Audio Eng. Soc,* vol. 39, pp 768-774 (1991).