Perceptual Simplification for Model-based Binaural Room Auralisation

Hüseyin Hacıhabiboğlu[†], Fionn Murtagh[‡]

† *Corresponding Author*, Centre for Communication Systems Research (CCSR), University of Surrey, Guildford, GU2 7XH, United Kingdom, Tel: +44 1483 683435, Fax: +44 1483 684701

‡ Department of Computer Science, Royal Holloway, University of London, Egham, Surrey TW20 0EX, United Kingdom, Tel:+44 1784 443429, Fax: +44 1784 439786

e-mail: h.hacihabiboglu@surrey.ac.uk, fmurtagh@acm.org

ABSTRACT

Design of computationally efficient yet perceptually realistic room auralisation algorithms require a careful selection of the early reflections to be auralised. A perception-based simplification algorithm is presented for the selection of the early reflections using a criterion which depends both on the arrival time and on the angle of incidence of the early reflection with respect to the listener. Results of two subjective tests for the evaluation of the proposed algorithm are presented. The proposed algorithm is shown to provide a significant reduction in the number of early reflections without significantly affecting the tested localisational or spatial qualities of auralisation.

Keywords: Room auralisation, binaural hearing, precedence effect, signal processing

PACS numbers: 43.55.Hy, 43.55.Ka, 43.66.Pn

Submitted: January 4, 2007, *Revised:* February 19, 2007

⋆ Part of this paper has been presented in the International Conference on Auditory Display (ICAD'06): *Hacihabiboglu, H., and Murtagh, F., (2006), "Perception-based simplification for binaural room auralisation", International Conference on Auditory Display (ICAD-06), pp. 268-271, London, UK.*

1 Introduction

The pressure fluctuation at an arbitrary location inside a room due to an impulsive omnidirectional sound source is called the room impulse response (RIR) at that location. A room impulse response consists of the direct sound, the early reflections, and the late reverberation. The direct sound refers to the pressure fluctuation due to the sound wave which arrives at the location before being reflected from the boundaries of the enclosure. The early reflections consist of the pressure fluctuation due to the sound waves that arrive in a temporal order after being reflected from at least one boundary of the enclosure. Late reverberation is characterized by the high-order reflections together with the diffuse reflections which form a chaotic sound field. Room auralisation refers to the process of making audible the binaural listening experience inside such a room by mathematical or physical means, over a suitable reproduction medium [1].

Auralisation systems have their place not only in room acoustics design tools, but also in computer games, telepresence/teleconference systems, and mission-critical virtual reality simulators [2, 3]. The advent of low-power mobile computing devices have made computational complexity, bandwidth efficiency and portability of a room auralisation system essential issues to be addressed. A bandwidth-efficient and portable auralisation method is binaural auralisation. Binaural auralisation uses only two channels to simulate the binaural listening experience over a pair of headphones. Computationally, direct sound is processed with digital filters modeling the head-related transfer functions (HRTFs) [4] and air attenuation [5], while early reflections need to be processed with wall absorption filters as well [6]. Late reverberation can be synthesised using any one of the rather simple, low-cost digital signal processing algorithms [7–9].

The number of early reflections in an actual RIR is proportional to the cube of its temporal length [10]. Therefore, it may be suggested that the main source of computational complexity in a binaural auralisation system is related to the number of processed early reflections. As the early reflections define most of the perceived qualities of a room [11], they need to be selected carefully to provide a good localisation, and a high level of realism. In other words, even if it is not possible to achieve a true-to-original rendering of the acoustics of an enclosure, the selection of the early reflections should allow for a realistic listening experience.

The effect of the direction of a reflection on its relative perceptual prominence was investigated

previously by the authors [12–14]. In particular, a mathematical model was presented within the context of sound source localisation under precedence effect conditions. This model allows the quantification of the effect of a reflection on the perceived auditory event. This paper presents an application of the mentioned psychoacoustical model to the perception-based selection of the early reflections in a binaural auralisation system.

This paper is organized as follows. Relevance of the present study to previous studies will be presented in Section 2. The Gaussian-mixtures model of sound source localisation under the precedence effect and the concepts derived therefrom will be briefly summarized in Section 3. The perception-based method for the selection of early reflections from a geometrical room acoustics model for auralisation will be presented in Section 4. Two subjective listening tests for the evaluation of the proposed method will be given in Section 5.

2 Background

The detection of the relative perceptual importance of reflections has been a topic of interest not only in the context of room auralisation but also in other areas of audio engineering such as acoustical design, optimization of virtual auditory displays, or the quest for intelligent loudspeakers. The previous work done on the subject emphasizes the threshold of detection and the just noticeable difference for a single reflection in a complex acoustic field. The effects of different reflections to the timbre and spatial perception in small rooms had been investigated thoroughly in a set of different studies.

Olive and Toole [15] used a simple setup consisting of four spatially separated loudspeakers to assess the *absolute threshold* and the *image-shift threshold* using speech, noise and castañet sounds. Absolute threshold is the relative level at which the reflection becomes completely inaudible. Image-shift threshold is the minimum relative level at which the reflection has the effect of causing a shift in the direction of the fused auditory event. For pink noise stimuli, the level of a reflection not coincident with the direction of the direct sound needs to be at least 20 dB lower than that of the simulated direct sound to be totally inaudible. If the direction of the reflection coincides with that of the direct sound, the absolute threshold is higher in comparison with the non-coincident case (around -10 dB with respect to the direct sound), signifying that

the precedence effect is stronger. The presence of reverberation and the type of the stimulus are also very effective in reducing the absolute thresholds. The image-shift thresholds are about 7 dB higher than absolute thresholds. This means that an early reflection may be perceived to be present but still not cause a pronounced shift in the direction of the auditory event nor have any significant effect on the perceived width of the auditory event. For band-limited stimuli like speech or music, both thresholds are higher.

Bech investigated the effect of early reflections on timbre [16, 17] and spatial perception [18] using a setup comprising 17 loudspeakers for simulating discrete early reflections and 6 loudspeakers for simulating late reverberation. The stimuli used in these experiments were speech and noise. Two questions were posed in these studies: 1) Which early reflections individually contribute to the perceived timbre and affect the spatial perception, and 2) what is the minimum change of the reflection level to trigger a change in the overall perception of timbre/spaciousness? Two thresholds related to those questions were investigated. The first is the threshold of detection which determines the level at which a reflection becomes audible. The second is the just noticeable level difference at which the early reflection can change the timbral/spatial perception individually. The simulated acoustical environment was a standard listening room with a single loudspeaker positioned asymmetrically with respect to the longitudinal axis. The early reflections (17 of which were auralised using spatially distributed loudspeakers) were calculated using the image source model (ISM) [19]. The effect of artificial reverberation was also investigated. The reported findings suggest that only the far lateral wall has the capacity to individually change the overall timbre perception. Reverberation decreases the threshold of detection (i.e. absolute threshold) for timbre. It was also shown that only the first order floor reflection can contribute individually to spatial perception of the auditory environment. Another finding was that the modeling of frequency dependent wall absorption did not affect the timbre threshold of detection for speech severely. Nevertheless, it had a pronounced effect on broadband noise.

Begault [20, 21] investigated the relative prominence of early reflections by using a virtual auditory display. In the first study, four conditions were tested with speech or music as stimuli. The absolute threshold for an isolated early reflection at a delay of 18 ms was found to be around -20 dB with respect to the direct sound source at an azimuth of 90° left. The absolute threshold (i.e. threshold of detection) is higher if the azimuth separation between the direct sound and the

early reflections are small. Further, when the number of early reflections is higher, the probed reflection is harder to perceive. In another investigation of early reflection thresholds, Begault et al. [22] used speech and tone burst signals as stimuli testing for isolated reflections (i.e. lead/lag pairs) with different azimuth and elevation separations. The results were in agreement with the findings of classical precedence effect experiments that the absolute thresholds decrease for increasing time delay between the direct sound and the early reflection. A rule-of-thumb is proposed such that early reflections will be inaudible when their level is less than 21 dB below the direct sound at a delay of 3ms or around 30dB below the direct sound at a delay of $15-30$ms.

Buchholz et al. [23] suggest a unified model of what they term the *room masking*. The model is physiologically motivated and is based on the concept of temporal masking. Although useful as a model, the proposed strategy depends on the assumption of a correspondence between temporal masking and precedence effect. This correspondence is not well-established [24].

In another study by Pellegrini [25, 26], the auralised reflections were chosen not from a set representing specular reflections exactly, but from a set which simulates distance and room size on a perceptual basis. In other words, the auralised reflections did not coincide with what the geometrical room model suggests, but were derived to give the perception of distance and room size approximately. This work also investigated the auralisation of small rooms (specifically a reference listening room).

As explained earlier, the selection of early reflections that need to be auralised constitutes an important difficulty in achieving a perceptually veridical rendering of room acoustics. However, given that a large number of reflections exist in a given sound field, it is not a straightforward task to make this selection. As an example, assume that it is needed to select 50 early reflections from among 100 that are available. The number of different selection options would be ${}_{100}C_{50} \approx 1 \times 10^{29}$. In other words, this is practically infinite. The essential question is, which one(s) of this infinite number of possibilities would provide a more realistic rendering than the others?

3 Gaussian mixtures view of the precedence effect

When we are listening to a sound source in a room, we can tell the location of it despite the high number of interfering reflections resulting from the room boundaries. This is possible by

a property of our auditory system which weighs the sound (which consists of a summation of the direct sound and its reflections) in favor of the first arriving wave front and suppresses the perception of redundant directional information conveyed in the reflections. Our auditory system gives precedence to the first arriving sound wave and discards most of the directional and spatial information of the reflections. This property is therefore called the "precedence effect" [27].

The precedence effect mainly depends on the time interval between the onsets of the leading and the lagging sound sources. If the time separation between the leading and the lagging sound sources is smaller than a lower temporal threshold (τ_{low}) a fused auditory event is perceived as incident from a direction between the actual directions of the leading and the lagging sound sources. This phenomenon is known as the *summing localisation*. When the time delay is between the lower temporal threshold and a higher temporal threshold (τ_{high}) the direction of the leading source dominates and the directional discrimination of the lagging source is suppressed. Location of this fused auditory event is largely determined by the leading source. When the time separation is greater than the higher temporal threshold, both the leading and the lagging sources are perceived seperately. Because of this reason, τ_{high} is also known as the temporal echo threshold. For broadband signals such as clicks or white noise bursts, $\tau_{\text{low}} \approx 1$ ms and $\tau_{\text{high}} \approx 5$ ms. In summary, any reflection arriving with a delay less than τ_{high} is perceived not as a separate auditory event, but contributes to the perception of the fused auditory event.

The perceptual prominence of a reflection is by and large determined by the strength of the precedence effect. There exist several aspects related to the classical precedence effect conditions with a lead/lag pair [28]: Fusion refers to the perception of a single auditory event instead of two. Localisation dominance refers to the dominance that the direction of the leading sound has on the mean direction of the fused auditory event. Depending on the directions of the leading and the lagging sound the perceived width of the auditory event increases. Further, the discrimination of the relative position of the lagging sound source with respect to the leading sound source is suppressed.

In our previous work, we have proposed a new analysis method utilizing Gaussian mixture models to assess subjective localisation performace for broadband lead/lag pairs under the precedence effect conditions [14]. Sound source localisation responses obtained in an observational

study for stimuli consisting of spectrally coherent lead/lag pairs presented at equal levels with a delay of 4 ms were modeled using a weighted sum of two Gaussian components. Each Gaussian component correlates to the internal (i.e. perceived) representation of the direction of the leading or the lagging sound source. Fusion, localisation dominance, lag discrimination suppression, widening aspects of the precedence effect were captured by the mathematical properties of the model. The utility of the model was tested with two subjective localisation experiments. The subjects localised sound sources by turning their heads to face the perceived auditory event. The responses were the displacement of head azimuth with respect to the subject's median plane. A thorough discussion of this model is outside the scope of this paper. The interested reader is directed toward our original research paper.

An important outcome of the mentioned study was the definition of a modality function which was used for modeling the lag discrimination suppression property of the precedence effect for the given stimuli conditions. It was argued that if the Gaussian mixture modeling the response distribution is unimodal, the discrimination of the lagging sound is suppressed. If it is bimodal, the lagging sound is likely to be discriminated. Therefore the modality of the distribution represented whether a simulated reflection could be discriminated. The results of our experiments revealed the following exponential form for the modality of the response distribution defined as a function of the azimuth separation of the leading and the lagging sound sources in degrees (Δ_θ):

$$F_{\mathrm{mod}}(\Delta_\theta) = 0.5\,|\Delta_\theta|\,e^{-k|\Delta_\theta|-l} \qquad \text{for} -\pi/2 \leq \Delta_\theta \leq \pi/2 \tag{1}$$

where $k > 0$ and $l > 0$ are constants. If $F_{\mathrm{mod},\theta} > 2\sigma$, where σ is the response standard deviation (in radians) for the lead-only condition, the response is bimodal. The response standard deviation (σ) is a measure of localisation acuity in that it represents the variability of the subject response in the single sound source case. As the discrimination of the lagging source is not suppressed, the effect of the lagging sound on the spatial aspects of the auditory event is significant. Average values of the constants k and l were found to be $k = 0.0343$ and $l = -0.5722$ for $\Delta_\theta < 0$, and $k = -0.0305$ and $l = -0.6892$ for $\Delta_\theta > 0$. It was suggested that larger k and l corresponds to a stronger precedence effect.

4 Perception-based Selection of Early Reflections

If we consider only the specular reflections, the impulse response of an enclosure consists of filtered and delayed impulses juxtaposed in time and space. Most of these reflections will only have a minor effect on the perceived acoustics of the room given that the precedence effect takes place. The precedence effect may function differently when there are subsequent reflections or when bandlimited signals are used instead of broadband signals. Such differences are not accounted for by the original Gaussian mixtures model of the precedence effect. Nevertheless, the assumption made here is that the thresholds and results of our prior experiments are applicable as a worst-case condition. This assumption would make the applied simplification algorithm valid for most cases involving band-limited real-world signals such as speech or music.

An assumption made here is that a single early-reflection may act as a suppressor for a group of temporally and directionally proximate reflections. The temporal proximity of reflections suggest similar mean free paths for each reflection. Therefore, each reflection in such a group will have a similar attenuation level at the listener position as long as values of wall absorption are within a reasonable range. This, coupled with localisation dominance aspect of the precedence effect suggests that, the earliest reflection in such a group of reflections would have perceptual dominance over the second earliest reflection in the group if isolated. However, this dominance is also direction dependent as previously shown [14] and as also summarised above. Therefore, a perception-based selection algorithm should also take this directional aspect into account. The method which is presented in this article is a generalisation of this reasoning.

The perceptual clustering method proposed in this paper uses the image source model (ISM) of the enclosure [19, 29, 30] as a starting point. Briefly, positions of the secondary sources which model the specular early reflections from the walls are calculated. These are represented as points in the 3D Cartesian space. The directions and the distances of the image sources to a fixed listener position are obtained. These image sources are clustered in three steps to obtain temporal and angular clusters. The image sources representing the salient early reflections are selected from these clusters by the application of the modality function explained in the previous section.

4.1 Temporal Clustering

The valid and visible image sources obtained using the ISM are first clustered according to their relative time of arrival with respect to the direct sound at the listener position. Here, it is assumed that a reflection can take over the role of the direct sound and suppress the localisation of subsequent early reflections arriving no later than the temporal threshold, τ_{high}, of the precedence effect.

Assume that the listener is positioned at $\mathbf{X}_{\mathrm{L}} = (x_{\mathrm{L}}, y_{\mathrm{L}}, z_{\mathrm{L}})$, the sound source at $\mathbf{X}_{\mathrm{S}} = (x_{\mathrm{S}}, y_{\mathrm{S}}, z_{\mathrm{S}})$, and an arbitrary image source at $\mathbf{X}_i = (x_{\mathrm{i}}, y_{\mathrm{i}}, z_{\mathrm{i}})$. The n^{th} temporal cluster γ_{n} can be represented as the set of image sources:

$$\gamma_{\mathrm{n}} = \left\{ \mathbf{X}_{\mathrm{i}} : (n-1) \cdot \tau_{\mathrm{high}} \cdot c \leq |\mathbf{X}_{\mathrm{i}} - \mathbf{X}_{\mathrm{L}}| - |\mathbf{X}_{\mathrm{S}} - \mathbf{X}_{\mathrm{L}}| < n \cdot \tau_{\mathrm{high}} \cdot c,\ n \in \mathbb{Z}^{+} \right\} \tag{2}$$

where τ_{high} is the higher temporal threshold (i.e. *echo threshold*) for the precedence effect, and c is the speed of sound. In other words, the early reflections arriving at the listener position ($\Delta t_{\mathrm{n\text{-}1}} = (n-1) \cdot \tau_{\mathrm{high}}$) later than the direct sound but not later than ($\Delta t_n = n \cdot \tau_{high}$) are grouped together (see Figure 1).

[FIGURE 1 HERE]

If the room is assumed to be rectangular, and a large number of image sources are calculated with the ISM, the number of image sources in the n^{th} cluster, N_{γ_n}, can be given as [10]:

$$N_{\gamma_{\mathrm{n}}} \approx \frac{4\pi c^3}{V} \int_{n \cdot \tau_{\mathrm{high}}}^{(n+1) \cdot \tau_{\mathrm{high}}} t^2 dt = \frac{4\pi (c \cdot \tau_{\mathrm{high}})^3}{3V} \cdot \left[(n+1)^3 - n^3 \right] \tag{3}$$

This shows that the number of image sources in a temporal cluster increases quadratically with the cluster's index, n. However, the image sources are not distributed evenly for complex room architectures. Further, some of the clusters may be empty because of the image sources that are invisible at the listener position. Therefore Equation 3 only approximately holds.

It is well-known that the suppression mechanism of the precedence effect is effective even if the level of the lagging sound signal is higher than the leading sound [27]. This suggests that temporal precedence is more important than the level difference. Therefore, the image source in the cluster which is the nearest to the listener position is regarded as the "primary suppressor" of the given temporal cluster. As mentioned in Section 3, lag discrimination suppression depends also on the angular separation between the leading and the lagging sound source. This property of the precedence effect is taken into account by

forming angular clusters and applying the modality function for obtaining the salient early reflections.

4.2 Selection of the Representative Image Sources

The modality function, F_{mod}, was defined for lead/lag pairs positioned on the listener's horizontal axis for discrimination suppression conditions (see Equation 1). If $F_{\mathrm{mod}} < 2\sigma_\theta$ for a given pair of image sources, the farther image source (i.e. lagging early reflection) is unlikely to contribute significantly to the overall spatial perception if it is not within 1 ms distance of the leading early reflection. Therefore, it is not included in the set of image sources to be auralised. If this condition is not met by the image source or if it is within 1 ms distance of the leading image-source, the lagging specular reflection is likely to contribute to the spatial perception of the auralised space and is included in the set of image sources to be auralised.

One practical problem with using the modality function is that it is only defined for $-\pi/2 < \Delta_\theta < \pi/2$. This necessitates further clustering based on azimuth angle of the image sources within the temporal clusters. This is done in the following way:

1. In a temporal cluster, γ_{n}, the image source that is the nearest to the listener position is the primary suppressor of that cluster. Assume that the image source at $\mathbf{X}_{\mathrm{n}}^{'} = (r_{\mathrm{n}}^{'}, \theta_{\mathrm{n}}^{'}, \phi_{\mathrm{n}}^{'})$ is the primary suppressor of γ_{n} where radius, r, represents the radial distance from the listener, θ and ϕ represent the azimuth and elevation of the image source with respect to the listener. Using the region where $F_{\mathrm{mod},\theta}$ is defined, the first azimuthal cluster, $\gamma_{\mathrm{n},\theta_1}$, is formed by grouping together the image sources, $X_{\mathrm{j}} = (r_{\mathrm{j}}, \theta_{\mathrm{j}}, \phi_{\mathrm{j}})$, for which $\theta'_{\mathrm{n}} - \pi/2 < \theta_{\mathrm{j}} < \theta'_{\mathrm{n}} + \pi/2$.

2. The remaining image sources form the reduced temporal cluster $\gamma_{\mathrm{n}^-} = \gamma_{\mathrm{n}} \setminus \gamma_{\mathrm{n},\theta_1}$ ($\setminus$ denotes set difference). The image source in this set which is nearest to the listener position is the *secondary suppressor* positioned at $\mathbf{X}_{\mathrm{n}}^{''} = (r_{\mathrm{n}}^{''}, \theta_{\mathrm{n}}^{''}, \phi_{\mathrm{n}}^{''})$. The second azimuth cluster, $\gamma_{\mathrm{n},\theta_2}$, is formed using the same strategy explained above using the reduced temporal cluster, γ_{n^-}.

3. The third azimuth cluster is simply the difference set between the reduced temporal cluster and the second azimuth cluster (i.e. $\gamma_{\mathrm{n},\theta_3} = \gamma_{\mathrm{n}^-} \setminus \gamma_{\mathrm{n},\theta_2}$). The image source within this cluster that is the nearest to the listener position is the *tertiary suppressor* positioned at $\mathbf{X}_{\mathrm{n}}^{'''} = (r_{\mathrm{n}}^{'''}, \theta_{\mathrm{n}}^{'''}, \phi_{\mathrm{n}}^{'''})$. The total number of non-empty azimuth clusters for a given temporal cluster can thus be at least 1 and at most 3 (see Figure 2).

[FIGURE 2 HERE!]

Therefore, the maximum number of non-empty clusters that can be obtained from a set of image sources defining the first T_{res} milliseconds of the room impulse response is $3 \times \frac{T_{\mathrm{res}}}{\tau_{\mathrm{high}}}$.

It may be suggested that if temporally precedent reflections are perceptually more significant in a complex sound field, a temporal cluster with a small index is perceptually more prominent than one with a large index. Similarly, an azimuth or elevation cluster within a temporal cluster is more prominent if its index is small.

Directional information conveyed in early reflections within 1 ms of direct sound are not suppressed. Such reflections are effective in summing localisation and cause a shift in the perceived auditory event [27]. Therefore, all of the early reflections within 1 ms delay of the primary suppressor are selected for auralisation. The modality function is used to select the early reflections whose discrimination is not likely to be suppressed by the suppressor of the cluster that they belong to. Consider the cluster $\gamma_{\mathrm{n},\theta_{\mathrm{i}}}$, with a suppressor at $\dot{\mathbf{X}} = (\dot{r}, \dot{\theta}, \dot{\phi})$. For any given image source $\mathbf{X}_{\mathrm{i}} = (r_{\mathrm{i}}, \theta_{\mathrm{i}}, \phi_{\mathrm{i}})$ in the cluster, which is not within 1 ms temporal distance from the suppressor (i.e. all the early reflections within 1 ms of the suppressor are included in the model), the early reflection corresponding to the image source is predicted to contribute significantly to the combined spatial impression of the cluster if the following condition is satisfied:

$$F_{\mathrm{mod}} = 0.5 \left| \theta_{\mathrm{i}} - \dot{\theta} \right| e^{-k\left|\theta_{\mathrm{i}} - \dot{\theta}\right| - l} > 2\sigma_{\theta} \tag{4}$$

where σ_{θ} is the standard deviation related to localisation blur in azimuth, k_{θ} and l_{θ} are exponential model parameters defining the precedence level difference for azimuth and elevation. It must be noted once again that the typical values of these parameters were found in our previous studies using data from localisation tests which tested the precedence effect exclusively for sound sources in the horizontal plane of the listener. However, interaural delay cues, despite being significantly less pronounced, are observed for different azimuth angles at different elevations as well. Therefore, we assume that a similar mechanism for the precedence effect exists for sources incident from different azimuth angles even if they are not on the horizontal plane.

5 Subjective Evaluation

Two separate subjective experiments were carried out to find out whether the proposed perceptual simplification strategy resulted in any degradation in certain qualities of the auralisation. The first evaluation assessed the localisation acuity using a virtual source-identification paradigm [31]. The second evaluation was a subjective rating experiment [32] which assessed the effects of the proposed method on presence, spaciousness, and envelopment.

5.1 Experiment 1: Localisation

The first experiment investigated the localisation performance of subjects with three auralisation models. The proposed strategy was compared with the original model obtained from the image-source method, and simplification as proposed by Begault [22, 33].

5.1.1 Method and stimuli

[FIGURE 3 HERE!]

The image-source model of a rectangular room (5 m×7 m×3 m), was calculated up to 4^{th}-order for seven virtual sources. The virtual sources were positioned in front of the listener on the azimuth plane with 5° separation, 15° to the left and 15° to the right forming a circular arc spanning 30°. The sources were equidistant from the listener at 3 m (see Figure 3). Three different binaural room impulse responses (BRIRs) were calculated for each source using the blocked-meatus KEMAR HRTFs from the CIPIC HRTF database [34]. These were obtained by the following strategies:

1. Original (*ori*): The BRIRs were obtained by using all the image-sources obtained from the model without any model reduction. This strategy, as it includes all the image-sources in the model resulted in the most detailed BRIRs.

2. Begault (*beg*): The BRIRs were obtained by selecting the image-sources according to the reduction strategy proposed by Begault [22, 33]. Namely, any early reflection with a level 21 dB below the direct sound at a delay of 3 ms and greater, and 30 dB below the direct sound at a delay of 15 ms or greater were eliminated.

3. Reduced (*per*): The BRIRs were obtained by selecting the ISs using the perceptual simplification

strategy proposed in this article. The value for the response standard deviation was selected to be, $\sigma_\theta = 1.5^\circ$ which is a representative value in subjective localisation studies under similar conditions.

The numbers of image sources used in the calculation of each BRIR were similar across all of the modeled source positions. The average number of image sources used in the calculation of the BRIRs were 438 (*ori*), 370 (*beg*), and 137 (*per*). It may be noted that for the given ISM order, the method *beg* cannot reduce the number of early reflections significantly at least for the modeled rectangular room.

The absorption coefficients of the room were selected to be the same for all surfaces ($\alpha = 0.3968$). This specific value of absorption coefficient was calculated from the Sabine's reverberation formula to approximate a room with a reverberation time of $T_{60} = 300$ms for a room with a volume equal to the one that is modeled (105 m^3). At that level of reverberation, the critical distance for reverberation (i.e. the distance at which the reverberant energy is equal to direct energy for an omnidirectional sound source) was 1.22 m. Such a reverberation time is typical of mid-sized listening rooms conforming to ITU-R BS-1116 recommendations for critical listening rooms. High-frequency absorption due to air humidity was not considered.

Frozen windowed white noise of 500 ms duration was convolved with the obtained left and right ear BRIRs for each different strategy and each virtual source direction as stated above. The reason for selecting a broadband signal such as windowed white noise is that it would allow the excitation of the virtual acoustical model at a wider band of frequencies than narrow band signals such as musical samples or speech. In addition, the cognitive aspects of localisation associated with stimuli such as speech and music have also been excluded from possible effects by employing windowed white noise which lacks any meaningful content. Artificial reverberation with $T_{60} = 300\,\text{ms}$ obtained using a feedback-delay network (FDN) was added to the obtained signal. The total reverberant energy was calculated from the direct-to-reverberant (D/R) energy ratio at the given listener position.

The localisation experiment was a double-blind virtual source-identification experiment with binaural stimuli presented over headphones. The stimuli had been grouped into blocks according to the model reduction strategy (i.e. *ori*, *beg*, and *per*) in order to facilitate conditions similar to real situations where adaptation to stimuli would take place. The presentation order of the blocks was randomised. The ordering of the stimuli in each block was also randomised. Over the course of each block, each stimulus was repeated 10 times. This resulted in a total of 210 presentations (7 sources $\times$ 3 methods $\times$ 10 repetitions) for each subject. The subject's task was to identify which source had been active given a

played back stimulus, by clicking the corresponding button on a user interface running under MATLAB. The subjects could only listen to a specific stimulus once and had to respond in order to listen to the next stimulus which was played back after 1 s of the subject's response. No feedback was given to subjects during the test as to whether the source they identified was the correct one.

5.1.2 Subjects and procedure

Six subjects (four males and two females; aged 26-32) with normal hearing, who were either members of staff or Ph.D. students in CCSR, University of Surrey, participated in the experiment. The first author of this article (HH) also took part as a subject in the experiment. Except two subjects, HH and BGH, who had extensive experience of subjective evaluation, the subjects were naïve.

The listening test was carried out in an acoustically treated studio space. The stimuli was played back over Beyer Dynamic DT-150 circumaural headphones using the MOTU 828 Mk-II audio interface which was also used as the headphone amplifier. The presentation level of each stimulus was 65($\pm$1) dBA (SPL) measured near the subject's eardrum using a DPA 4061 miniature omnidirectional microphone inserted into the ear canal. The aim of the experiment was explained to each subject prior to the experiment. A short training on how to use the user interface was given. A training run was then carried out to allow the subject to form an individual source localisation strategy. The actual experiment started after the subject was confident with both the experimental paradigm and the user interface and lasted around 30 minutes for each subject to complete.

5.1.3 Results

The obtained results were used in the calculation of two parameters for each subject, $\overline{s}$ and $\overline{C}$ that represent the response variability and the constant localisation bias respectively in a source identification paradigm. These values were calculated as suggested by Hartmann *et al.* [31] such that:

$$s^2(k) = A^2 \frac{1}{M_\mathrm{k}} \sum_{\mathrm{i}=1}^{M_\mathrm{k}} \left[R_\mathrm{i} - R(k)\right]^2, \tag{5}$$

$$\overline{s} = \sqrt{\frac{1}{N} \sum_{\mathrm{k}=1}^{N} s^2(k)} \tag{6}$$

where $s(k)$ is the localisation variability for the k^{th} source, A is the angular separation in degrees between each source, M_{k} is the total number of trials for the k^{th} source, R_{i} is the subject's response on the source-index scale in the i^{th} trial, and $R(k)$ is the average response for the k^{th} source. Similarly, the average constant localisation bias, $\overline{C}$, is calculated as follows:

$$C(k) = A\left[R(k) - k\right], \tag{7}$$

$$\overline{C} = \frac{1}{N}\sum_{\text{k}=1}^{N} C(k), \tag{8}$$

where $C(k)$ is the localisation bias associated with the k^{th} source.

Figure 4 shows the responses of each subject in the experiment. It may be observed that the localisation performance is not much different for the different strategies of model selection.Response variability averaged across all subjects for different selection methods are $\overline{s}_{beg} = 3.57°$, $\overline{s}_{ori} = 3.88°$, and $\overline{s}_{per} = 3.49°$. These values are in general higher than response variability observed in a real room as opposed to the virtual acoustics auralised in this study. The reason for this difference is due to the use of non-individualised HRTFs. The localisation bias averaged across all subjects for different selection methods are $\overline{C}_{beg} = -0.32°$, $\overline{C}_{ori} = -0.74°$, and $\overline{C}_{per} = -0.86°$.

[FIGURE 4 HERE!]

Figure 5 shows the response variation scores and the localisation bias scores for all the subjects. It may be observed that the localisation bias is negative in general for most subjects for all different methods. This suggests a left-right asymmetry biased towards left. This sort of asymmetry in spatial hearing, particularly with precedence effect experiments, has previously been reported in other studies [35, 36].

[FIGURE 5 HERE!]

Two one-way analysis-of-variance (ANOVA) models were fit to $\overline{s}$ and $\overline{C}$ values calculated for the factors *Subject* and *Method*. Post-hoc multiple comparisons were carried out using the Bonferroni correction. *Subject* was a statistically significant factor at the $\alpha = 0.05$-level ($\text{F} = 12.74,\ \text{df} = 5,\ \text{p} < 0.001$) in the ANOVA model for the response variability, $\overline{s}$. Multiple comparisons revealed that the subjects BGH, HH, and SD were better localisers, in that they had lower response variance scores in general. Almost all pairwise comparisons of these subjects with the others were significant at the $\alpha = 0.05$-level. *Method* was not a significant factor. Neither *Method* nor *Subject* were significant factors for the ANOVA model of the constant localisation bias scores, $\overline{C}$. This was also verified by the post-hoc multiple comparisons as no

statistically significant difference existed between different subjects and image-source selection methods.

The results lead to the conclusion that although subjective differences exist, the proposed perceptual simplification method does not have a significant degrading (or improving) effect on the localisation performance. The same conclusion also holds for the other selection method (i.e. *beg*). However, the number of selected image-sources is lower for the proposed method which makes it more desirable in this case.

Subjects were also informally asked whether they perceived a significant timbre difference between any block of sources. None of the subjects reported to have perceived such a difference.

5.2 Experiment 2: Presence, Spaciousness, and Envelopment

The second experiment investigated the effects that the previously mentioned simplification strategies (i.e. *per* and *beg*) have on several spatial properties of the auralisation. These included presence, spaciousness, and envelopment.

5.2.1 Method and Stimuli

The experiment was a subjective scaling experiment as suggested by Lokki *et al.* [32]. Image-sources for three small rooms were modeled up to 5^{th}-order for a single source position and three listener positions for each case. All the rooms had six surfaces and the absorption coefficients of the rooms were selected so that the total reverberation in each case was $T_{60} = 300$ ms regardless of room size. The modeled rooms were:

1. Rectangular (R1): The room had a rectangular shape with a volume of 105 m^3 (W = 5m, L = 7m, and H = 3m). The absorption coefficients for all surfaces were selected to be equal to $\alpha = 0.3968$ to obtain a reverberation time of $T_{60} = 300$ ms. The critical distance for reverberation was 1.22 m for the given reverberation time and room volume (see Figure 6(a)).

2. Trapezoidal (R2): The ceiling and floor of the second room were parallel isosceles trapezoids. The room had a volume of 148.5 m^3 ($W_{\text{long}} = 7$ m, $W_{\text{short}} = 4$ m, L = 9 m, and H = 3 m). The absorption coefficients for all surfaces were set to $\alpha = 0.4072$ for obtaining a reverberation time of $T_{60} = 300$ ms. The critical distance for reverberation was 1.4 m for the given reverberation time and room volume (see Figure 6(b)).

3. Corridor (R3): The third room was also rectangular, but with dimensions approximating a typically long corridor. The volume of the room was 82.5 m^3 (W = 3m, L = 11m, and H = 2.5m). The absorption coefficients for all surfaces were selected to be $\alpha = 0.3255$ for obtaining a reverberation time of $T_{60} = 300$ ms. The critical distance for reverberation was 1.2 m for the given reverberation time and room volume (see Figure 6(b)).

[FIGURE 6 HERE!]

For all the listener positions, the listener was modeled to be facing the virtual sound source inside the modeled enclosure. The selection strategies were the same as those used in the previous experiment. The average number of image-sources obtained using the tested simplification methods are summarised in Table 1.

[TABLE I HERE!]

The BRIRs were calculated from the image-sources using the blocked-meatus HRTF measurements of the KEMAR from the CIPIC HRTF database. Five different and realistic dry sound signals were used as stimuli. The stimuli consisted of *cello* (cel), *guitar* (gui), *trumpet* (trum) sounds, and *female* (fem) and *male* (mal) speech ($F_S = 44.1$ kHz) extracted from the Music for Archimedes CD [37]. The durations of the sound signals were 24, 30, 16, 11, and 17 seconds respectively. The sound signals were first convolved with the obtained BRIRs. Reverberant portions of the stimuli were obtained using a feedback delay network (FDN) type artificial reverberator and then added to the stimuli. Total number of factors which had been tested in this experiment is therefore 135 (3 Rooms $\times$ 5 Sounds $\times$ 3 Methods $\times$ 3 Positions).

The subjects were asked to rate the following three spatial properties of the auralised sounds:

1. Presence: The subjects rated the presence of the auralisation by rating how realistically they perceived that they were in the same environment with the sound source. The perceived presence of the stimuli was rated on a scale from 1 (*not realistic*) to 5 (*very realistic*).

2. Spaciousness: The spaciousness of a stimulus was explained to the subjects as how wide a sound source was perceived in a given auralisation scenario. The subjects rated the spaciousness on a scale from 1 (*narrow*) to 5 (*broad*).

3. Envelopment: The envelopment of a stimulus was explained to the subjects as how surrounded they felt for a given stimulus. The subjects rated the envelopment on a scale from 1 (*not enveloping*) to 5 (*very enveloping*).

The order of the stimuli was fully randomised. The subjects used a graphical user interface running under MATLAB in order to listen to the stimuli and record their responses. They were allowed to listen to the same stimulus as many times as they wish, and they were also encouraged to do so in order to obtain more robust results.

5.2.2 Subjects and Procedure

The same six subjects who participated in the first experiment also participated in this experiment. The experiment was carried out in the same acoustically treated studio space. The same equipment and setup used in the first experiment was used. The average level of presentation for each stimulus varied between 65 dBA (SPL) and 72 dBA(SPL) measured near the subject's eardrum. Different spatial qualities that the subjects were required to rate were initially explained. A short training run was then carried out in order to get subjects acquainted with the user interface and the listening test. At the end of the training run all the subjects stated that they were comfortable with the tested qualities and the user interface. Each stimulus was presented only once because of time constraints. The second experiment lasted between around 45 minutes and 1 hour. For one of the subjects (HO) the experiment took around 2 hours. A forced break of 5 minutes was given in the middle of the experiment to prevent fatigue.

5.2.3 Results

The subjects reported that the second experiment which required them to rate the spatial properties of stimuli was harder than the first experiment. The overall means of *Presence*, *Spaciousness*, and *Envelopment* were 3.04 (std = 1.11), 3.18 (std = 1.01), and 3.01(std = 1.17) respectively. The marginal means of *Presence* for the methods *beg*, *ori*, and *per* were 2.94, 3.11, and 3.06 respectively. The marginal means of *Spaciousness* for the methods *beg*, *ori*, and *per* were 3.01, 3.34, and 3.18 respectively. The marginal means of *Envelopment* for the methods *beg*, *ori*, and *per* were 2.94, 3.08, and 2.98 respectively. These values show that the simplification method, *beg* against which we have tested our method, was rated lower in average than both the non-reduced auralisation (i.e. *ori*), and the simplification method that is proposed in this paper (i.e. *per*).

Although the subjects were initially instructed to rate each quality independently of each other, the Pearson product-moment correlation coefficients, $r = 0.495$ (presence-spaciousness), $r = 0.641$ (presence-envelopment), and $r = 0.510$ (spaciousness-envelopment), were also statistically significant at the $\alpha = 0.01$

level.

Initial pairwise comparisons revealed that there were significant subjective differences in particular with the subject mean responses (see Figure 7). Therefore, the responses were transformed into z-scores by subtracting subject means and dividing by subject standard deviation for eliminating these inter-subject differences for sample mean and to the sample variance of each subject. The obtained results were then analyzed using a three-way multivariate analysis-of-variance (MANOVA) model with the factors *Method*, *Sound*, and *Room*.

[FIGURE 7 HERE!]

The results show that the factors, *Room* ($F = 14.24,\ df = 6,\ p < 0.001$), *Sound* ($F = 6.83,\ df = 12,\ p < 0.001$), and *Method* ($F = 3.13,\ df = 6,\ p = 0.005$) were statistically significant at the $\alpha = 0.05$ level as main effects in the multivatiate model. In addition, the interaction terms $Sound \times Room$ ($F = 2.43,\ df = 24,\ p < 0.001$) and $Sound \times Method$ ($F = 3.25,\ df = 24,\ p < 0.001$) were statistically significant at the $\alpha = 0.05$ level.

Univariate tests reveal that, the main effects were statistically significant at the $\alpha = 0.05$ level for the z-scores of ratings given to all different spatial qualities of the stimuli, except *Method*, which was a significant univariate effect only for *Presence* ($F = 3.08, df = 2, p = 0.047$) and *Spaciousness* ($F = 7.861, df = 2, p < 0.001$).

[FIGURE 8 HERE!]

Figure 8 shows the number of cases for which the z-scores for the *Presence*, *Spaciousness*, and *Envelopment* were greater than the overall mean of the respective z-scores. It may be observed that the auralisation using the method *beg* results in slightly higher z-scores for R1 for all of the tested spatial qualities. However, when the room is a corridor or when it has a trapezoidal shape, the number of z-scores above the mean decreases for *beg*.

The analysis points to the direction where we cannot reject the hypothesis that the reduction of the number of reflections using the proposed simplification method does not degrade any one of the tested spatial qualities of the original, non-simplified auralisation for the rooms, locations, and the sound signals used. In other words, the reproduction of the spatial qualities with the proposed perceptual simplification strategy were statistically not any worse than that of the original model including all of the calculated image sources. However, a certain amount of degradation was observed for the other simplification strategy.

6 Conclusions

This paper has introduced a data reduction strategy for binaural room auralisation based on the discriminability of specular reflections in a complex sound field. The modality function defined within the context of an observational precedence effect model explained in our previous work was used in the selection of early reflections that contribute significantly to the perceived sound field. Two subjective listening tests for the evaluation of the proposed algorithm were reported.

The first test evaluated the localisation performance in a virtual acoustics scenario for three different methods of selecting early reflections. The tested methods were, *ori*, in which all the image-sources in the model were used in the auralisation, *beg*, in which the image-sources were selected by their relative level, and *per*, in which the image-sources were selected on the basis of their predicted discriminability. The test was a virtual source-identification paradigm using windowed broadband noise bursts as stimuli. It was found that although the proposed simplification method allows for a significant reduction in the number of processed early reflections, it has no significant degrading effect on the localisation performance.

The second test evaluated the effect of the mentioned simplification methods on the perceived spatial qualities of the auralised sound. The test was a subjective rating experiment, in which the subjects rated the perceived *Presence*, *Spaciousness*, and *Envelopment* of the auralised stimuli. It was found as with the first experiment that the proposed simplification method had no significant degrading effect on the perceived spatial qualities.

In both of the subjective tests, all of the walls of modeled rooms had a fixed frequency-independent absorption coefficient. This is not fully representative of real-life conditions where different surfaces have different frequency dependent absoption. We suggest that the usage of frequency-dependent absorption would increase the perceived presence and make the auralisation sound more realistic. However, it should be noted that the modeled condition can be conceived as a worst-case scenario in which the early reflections are attenuated and delayed copies of the direct sound.

The proposed simplification method reduces the number of early reflections to approximately around only 30% of all early reflections in an image-source model. This is a significant reduction in the computational power required for auralisation making it suitable for interactive applications on mobile devices, advanced teleconferencing, virtual and augmented reality applications, and computer games. However, it must be noted that when it is not necessary to employ a model that corresponds to an actual enclosure,

there exist simpler methods that employ recursive simulation of global reverberation [38, 39].

The proposed method does not consider the level of a reflection as a selection criterion. Therefore, we suggest without further consideration that other selection methods based on the relative level of an early reflection may be used to complement the proposed method in order to achieve a greater reduction in the number of early reflections to be processed.

Acknowledgment

This work was supported by the EPSRC Research Grant GR/S72320/01.

References

[1] M. Kleiner, B.-I. Dalenbäck, and P. Svensson, "Auralization - an overview," *J. Audio Eng. Soc.*, vol. 41, no. 11, pp. 861–875, 1993.

[2] L. Savioja, J. Huopaniemi, T. Lokki, and R. Väänänen, "Creating interactive virtual acoustic environments," *J. Audio Eng. Soc.*, vol. 47, no. 9, pp. 675–705, 1999.

[3] D. R. Begault, *3-D Sound for Virtual Reality and Multimedia*. Cambridge, MA, USA: Academic Press, 1994.

[4] H. Hacıhabiboğlu, B. Günel, and F. Murtagh, "Wavelet-based spectral smoothing for head-related transfer function filter design," in *Proc. of the AES* 22^{nd} *Int. Conf. on Virtual Synthetic and Entertainment Audio (AES22), Espoo, Finland*, 2002, pp. 131–136.

[5] J. Huopaniemi, L. Savioja, and M. Karjalainen, "Modeling of reflections and air absorption in acoustical spaces—a digital filter design approach," in *Proc. of 1997 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'97), New Paltz, NY, USA*, 1997, pp. 19–22.

[6] J. Huopaniemi, "Virtual Acoustics and 3-D Sound in Multimedia Signal Processing," Ph.D. dissertation, Helsinki University of Technology, Espoo, Finland, 1999.

[7] M. R. Schroeder, "Natural-sounding artificial reverberation," *J. Audio Eng. Soc.*, vol. 10, no. 3, pp. 219–233, 1962.

[8] J.-M. Jot and A. Chaigne, "Digital delay networks for designing artificial reverberators," *Presented at the* 104^{th} *Convention of the Audio Engineering Society*, preprint 3030, Paris, France, 1991.

[9] W. G. Gardner, "Reverberation algorithms," in *Applications of Digital Signal Processing to Audio and Acoustics*, M. Kahrs and K. Brandenburg, Eds. Boston, MA, USA: Kluwer Academic, 1998, pp. 85–131.

[10] M. Vörlander, "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.

[11] H. Kuttruff, *Room Acoustics*. London, UK: Applied Science Publishers Ltd., 1979.

[12] H. Hacıhabiboğlu, "Modelling sound source localization under the precedence effect using multivariate gaussian mixtures," *Presented at the* 115^{th} *Convention of the Audio Engineering Society*, preprint 5976, New York, NY, USA, 2003.

[13] ——, "Perceptual room auralization for virtual auditory displays," Ph.D. dissertation, Queen's University Belfast, 2004.

[14] H. Hacıhabiboğlu and F. Murtagh, "An observational study of the precedence effect," *Acta Acustica united with Acustica*, vol. 92, no. 3, pp. 440–456, 2006.

[15] S. E. Olive and F. E. Toole, "The detection of reflections in typical rooms," *J. Audio Eng. Soc.*, vol. 37, no. 7/8, pp. 539–553, 1989.

[16] S. Bech, "Timbral aspects of reproduced sound in small rooms. I," *J. Acoust. Soc. Am.*, vol. 97, no. 3, pp. 1717–1726, 1995.

[17] ——, "Timbral aspects of reproduced sound in small rooms. II," *J. Acoust. Soc. Am.*, vol. 99, no. 6, pp. 3539–3549, 1996.

[18] ——, "Spatial aspects of reproduced sound in small rooms," *J. Acoust. Soc. Am.*, vol. 103, no. 1, pp. 434–445, 1998.

[19] 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.

[20] D. R. Begault, "Binaural auralization and perceptual veridicality," *Presented at the* 93^{rd} *Convention of the Audio Engineering Society*, preprint 3421, San Francisco, CA, USA, 1992.

[21] ——, "Audible and inaudible early reflections," *Presented at the* 100^{th} *Convention of the Audio Engineering Society*, preprint 4244, Copenhagen, Denmark, 1996.

[22] D. R. Begault, B. U. McClain, and M. R. Anderson, "Early reflection thresholds for virtual sound sources," in *Proc. of the 2001 Int. Workshop on Spatial Media, Aizu-Wakamatsu, Japan*, 2001.

[23] J. M. Buchholz, J. Mourjopoulos, and J. Blauert, "Room masking: Understanding and modelling the masking of room reflections," *Presented at the* 110^{th} *Convention of the Audio Engineering Society*, preprint 5312, Amsterdam, Netherlands, 2001.

[24] K. Hartung and C. Trahiotis, "Peripheral auditory processing and investigations of the "precedence effect" which utilize successive transient stimuli," *J. Acoust. Soc. Am.*, vol. 110, no. 3, pp. 1505–1513, 2001.

[25] R. S. Pellegrini, "A Virtual Reference Listening Room as an Application of Auditory Virtual Environments," Ph.D. dissertation, IKA, Ruhr-Universität Bochum, Germany, 2001.

[26] R. Pellegrini, "Perception-based design of virtual rooms for sound reproduction," in *Proc. of AES* 22^{nd} *Int. Conf. on Virtual, Synthetic and Entertainment Audio (AES22), Espoo, Finland.*, 2002, pp. 196–205.

[27] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization.* Cambridge, MA, USA: MIT Press, 1997.

[28] R. Y. Litovsky, H. S. Colburn, W. A. Yost, and S. J. Guzman, "The precedence effect," *J. Acoust. Soc. Am.*, vol. 106, no. 4, pp. 1633–1654, 1999.

[29] J. Kirszenstein, "An image source computer model for room acoustics analysis and electroacoustic simulation," *Appl. Acoust.*, vol. 17, no. 4, pp. 275–290, 1984.

[30] J. Borish, "Extension of the image model to arbitrary polyhedra," *J. Acoust. Soc. Am.*, vol. 75, no. 6, pp. 1827–1836, 1984.

[31] W. M. Hartmann, B. Rakerd, and J. B. Galaas, "On the source-identification method," *J. Acoust. Soc. Am.*, vol. 104, no. 6, pp. 3546–3557, 1998.

[32] T. Lokki and H. Järveläinen, "Subjective evaluation of auralization of physics-based room acoustics modeling," in *Proc. of Int. Conf. on Auditory Displays (ICAD 2001), Espoo, Finland*, 2001.

[33] D. R. Begault, B. U. McClain, and M. R. Anderson, "Early reflection thresholds for anechoic and reverberant stimuli within a 3-D sound display," in *Proc. 18^{t}h Int. Congress on Acoustics (ICA 2004)*, Kyoto, Japan, 2004, pp. (CD–ROM).

[34] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF database," in *Proc. of 2001 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, USA*, Oct 2001.

[35] D. W. Grantham, "Left–right asymmetry in the buildup of echo suppression in normal hearing adults," *J. Acoust. Soc. Am.*, vol. 99, no. 2, pp. 1118–1123, 1996.

[36] K. Saberi, J. V. Antonio, and A. Petrosyan, "A population study of the precedence effect," *Hearing Res.*, vol. 191, pp. 1–13, 2004.

[37] Bang and Olufsen, "Music for Archimedes," CD B&O 101, 1992.

[38] G. S. Kendall, W. L. Martens, D. J. Freed, M. D. Ludwig, and R. Karstens, "Image model reverberation from recirculating delays," *Presented at the 81st Convention of the Audio Engineering Society*, preprint 2408, Los Angeles, CA, USA, 1986.

[39] G. S. Kendall, W. L. Martens, and M. D. Wilde, "A spatial sound processor for loudspeaker and headphone reproduction," in *Proc. of the AES 8th International Conference: The sound of audio (AES8), Washington, D.C., USA*. New York, NY, USA: AES, 1990, pp. 8–027.

List of Tables

List of Figures

Table 1: Average number of image-sources used in the calculation of BRIRs for different rooms using the three image-source selection methods, beg*,* ori*, and* per*.*

	N_{beg}	N_{ori}	N_{per}
Rectangular (R1)	488	755	223
Trapezoidal (R2)	468	1821	558
Corridor (R3)	573	733	262

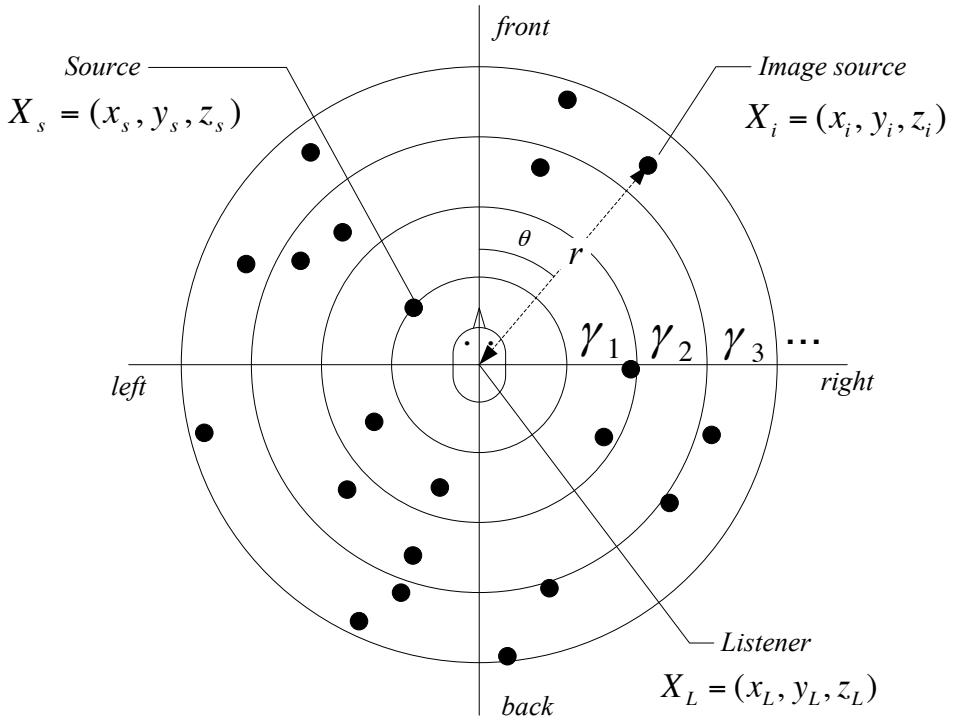


Figure 1: Temporal clustering of the image sources.

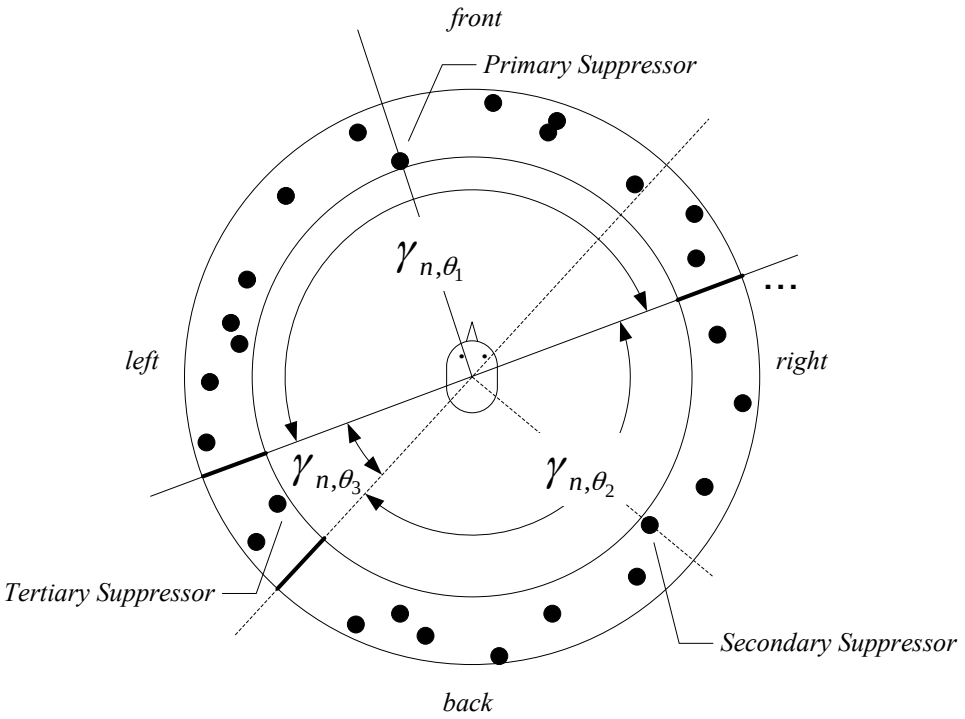


Figure 2: Azimuth clustering of the image sources.

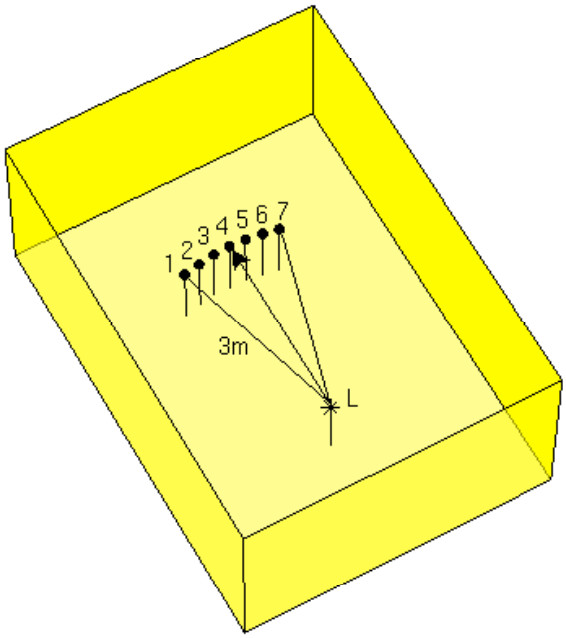


Figure 3: The setup of the modeled room for the source identification experiment. The virtual sources are marked with numbers, and the listener is marked with an L. The listener faces the front virtual source, as indicated by the arrow, at all stimulus conditions.

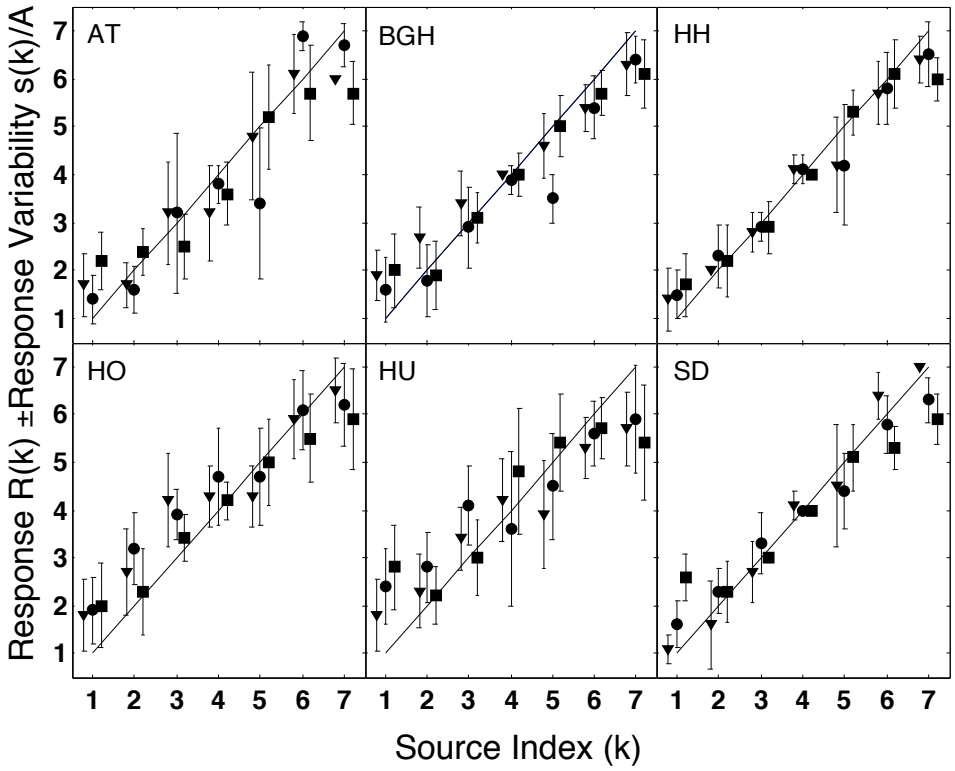
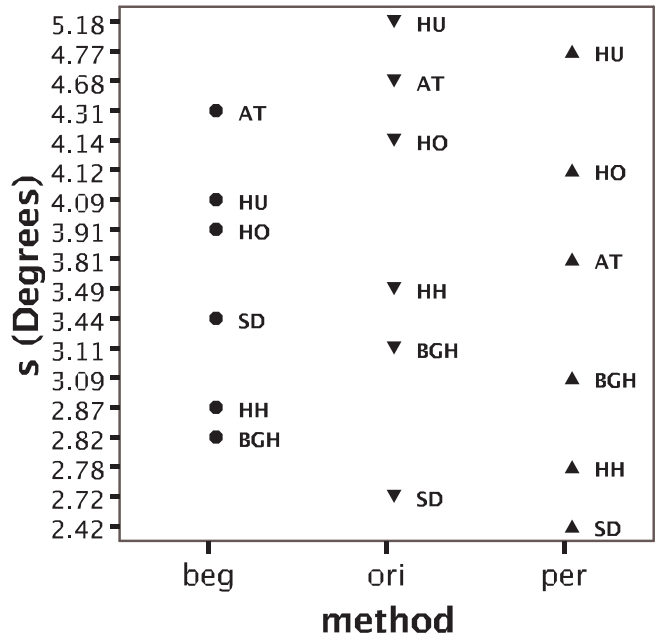
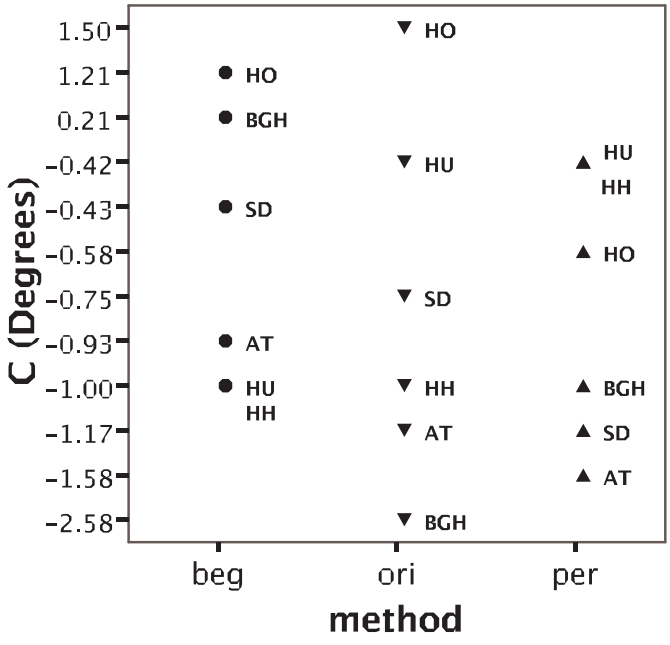


Figure 4: localisation responses by each subject. The markers represent in units of source index, k, the mean response, $R(k)$, given by each subject for each virtual source for the methods, beg *(▼),* ori *(•), and* per *(■). The lower and upper error bars represent the response variability, s(k)/A, in units of source index, k. The straight line shows the ideal response (i.e. $R(k) = k$)*



(a)



(b)

Figure 5: (a) The response variability, $\overline{s}$ and (b) the constant localisation bias, $\overline{C}$ for different subjects and methods.

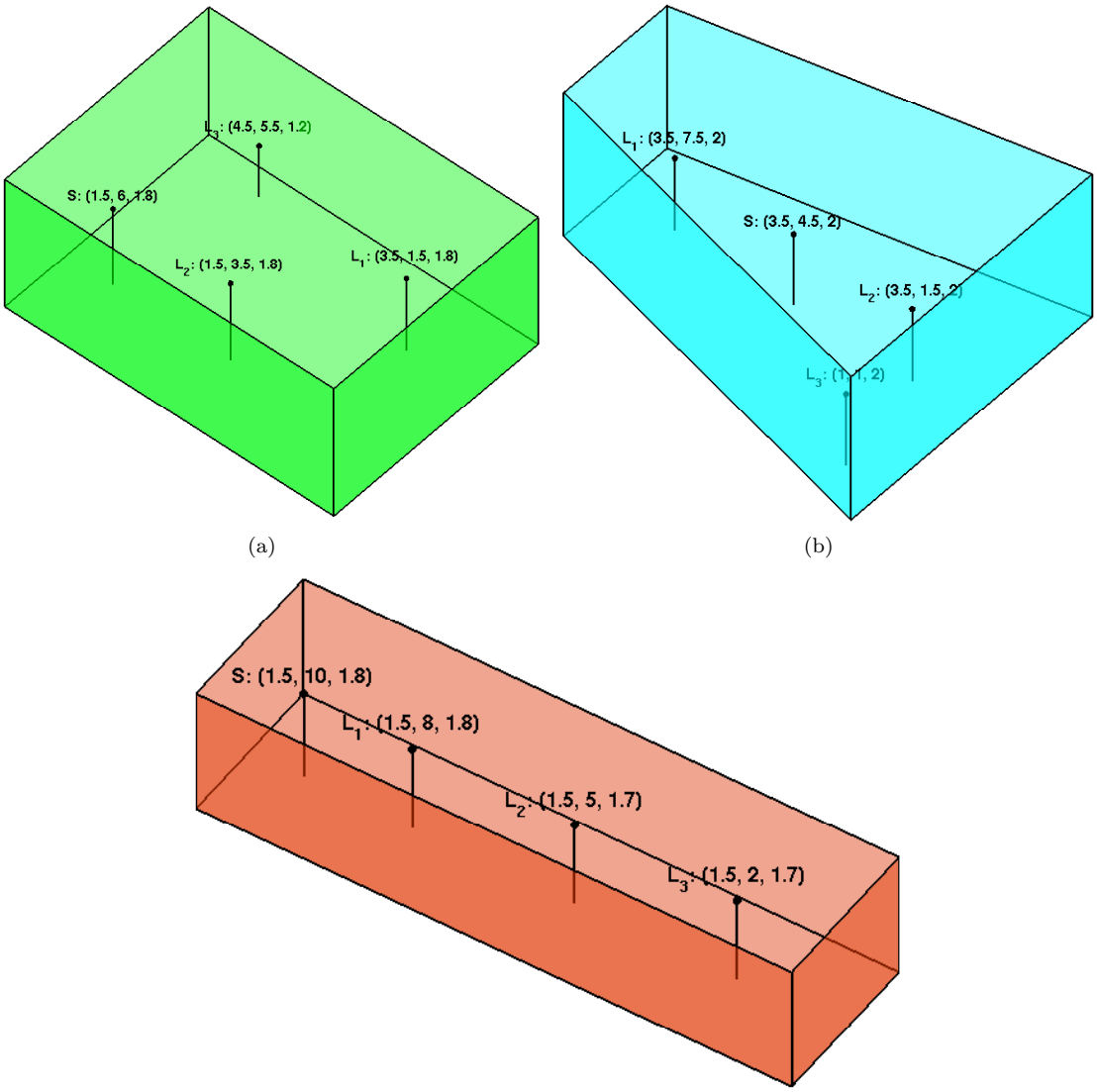
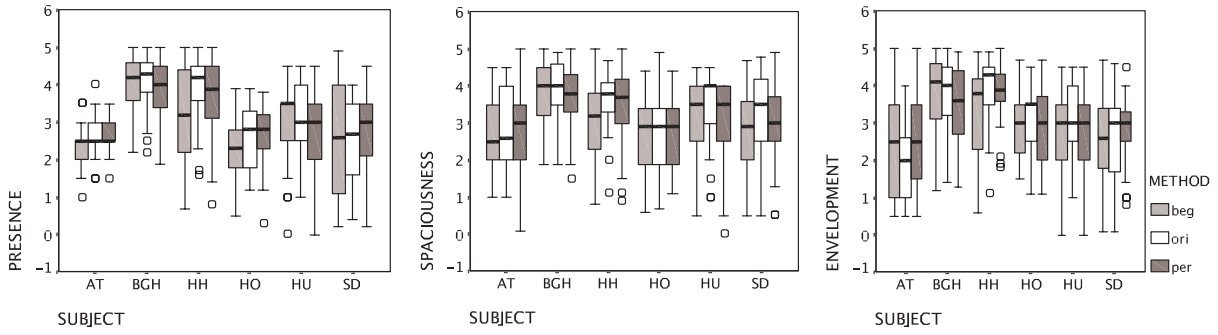
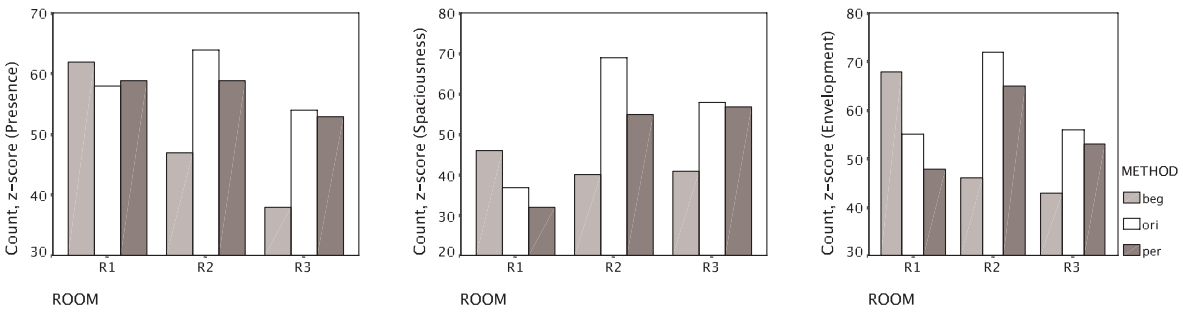


Figure 6: (a) The first (R1: Rectangular), (b) The second (R2: Trapezoidal), and (c) the third (R3: Corridor) rooms tested in the second experiment. The modeled source (S) and listener positions (L_1, L_2, and L_3,) are marked



(a) (b) (c)

Figure 7: The responses of the subjects for (a) Presence*, (b)* Spaciousness*, and (c)* Envelopment*. The boxplots depict the first and third quartiles (upper and lower edges of the box), median of the data (horizontal line in the box), the minimum and maximum values (whiskers), and the outliers (circles).*



(a) (b) (c)

Figure 8: The number of z-scores above the mean z-score for (a) Presence*, (b)* Spaciousness*, and (c)* Envelopment *for the tested rooms and methods.*