Perceived cathedral ceiling height in a multichannel virtual acoustic rendering for Gregorian Chant

Peter Hüttenmeister and William L. Martens

Faculty of Architecture, Design and Planning, The University of Sydney, Sydney, Australia

ABSTRACT

Reverberation is considered a key factor in listener envelopment. Virtual acoustic renderings of architectural spaces presented via arrays of multiple loudspeakers distributed on the horizontal plane can provide for improved listener envelopment and immersion relative to conventional stereophonic reproduction. However, with a new 22.2-channel loudspeaker format that has been submitted by NHK for an ITU recommendation, incorporating speakers at various elevations (termed height channels), there are vast new opportunities for the separation and presentation of real world acoustic events in virtual acoustic renderings. This paper presents the results of an experiment testing the hypothesis that simulated indirect sound delivered from above the listener's head can create a distinct impression of a modelled virtual acoustic space with a tall ceiling. The experimental stimuli included spatialised indirect sound based upon an image model of the Grace Cathedral that allowed the direction and time of arrival of simulated ceiling reflections to be manipulated. A five-part, anechoically recorded Gregorian chant performance was processed to allow for spatially segregated presentation of indirect sound components (sets of discrete early reflections and global reverberation) via ear-level and height channels. Results show that listeners could distinguish well between six virtual acoustic renderings in terms of both 'ceiling prominence' and 'spatial width.'

1. INTRODUCTION

Reverberation plays a vital role in spatial impression (SI), with listener envelopment (LEV) and apparent source width (ASW) as the perceptual attributes defined by the acoustical characteristics of the space. The sound that arrives at a listener's ears is heavily influenced by a number of spatial factors, but of particular importance in the current work is the spatiotemporal distribution of reflections. Historically, in the field of concert hall acoustics, the distribution of lateral reflections has been shown to have a strong influence on SI (Barron & Marshall 1980 and Cremer 1989). Of course, in using a multichannel loudspeaker array to reproduce simulated reflections arriving from a limited set of spatial angle, the natural continuous spatial distribution is truncated in a manner that may lose important acoustical information. However, in the current work, a number of loudspeaker channels were chosen from the great number of potential loudspeaker angles represented in the 196-channel array located in the Spatial Audio Laboratory at The University of Sydney. This selection was based upon an informal evaluation performed in advance of the experimental trials in the formal investigation to be described in this paper.

Previous work into multichannel by the likes of Hiyama et al (2002) and Hamasaki et al (2001) has determined appropriate positioning of loudspeakers on the horizontal plane for sufficient LEV and localisation of sound in the horizontal plane. However, with the recent adoption of elevated loudspeakers in multichannel systems that have been designed for commercial use (Hamasaki et al. 2006 and Solvang & Svensson 2006), genuine interest has been garnered around the use of these channels for reproducing immersive content considering the additional three- dimensional spatial auditory cues that they provide. Of course, listeners in a natural acoustic environment typically is provided with adequate spatial auditory cues to allow them to perceive the spatial characteristics of the space without the limited spatial resolution of a reproduction system.

A room impulse response can be decomposed into three primary elements: direct sound, early reflections and late reverberation. Each of these three elements act both separately and conjointly on SI. Taking this into consideration, it might be reasonably suspected that SI is more complex than a singularly definable spatial attribute. Griesinger (1997) discusses, particularly in the context of musical performances, the relationships that exist between the musical stimulus, performance characteristics and indirect sound that influence in varying degrees overall SI. He deconstructs SI into three individual components, namely, early spatial impression (ESI), background spatial impression (BSI) and continuous spatial impression (CSI).

Taking the same approach as Griesinger, this study is predominantly focused on the perception of the listener rather than the physical acoustic characteristics of the space itself. Therefore, the importance of the signal received

at the listener's ear is of most importance. The approach toward this specific construction is discussed further below.

2. METHOD

2.1 Experimental conditions

Participants were asked to rate two perceptual attributes, being ceiling prominence – perceived ceiling due to the indirect sound arriving from overhead in the virtual acoustic environment – and spatial width – perceived width of the virtual acoustic environment – in two separate tasks for 12 stimuli presented in a randomised order. Six loudspeaker configurations were selected where a total of nine channels were used in each case. The sound field was created using the 196-channel hemispherical loudspeaker system in the Spatial Audio Laboratory at the University of Sydney's Faculty of Architecture, Design and Planning (Martens et al. 2015) where arbitrary spatial configurations of loudspeakers could be employed, enabling the azimuthal angles of subsets of loudspeakers to be held constant while their elevation angles were varied. The seven ear-level channels conformed to ITU-R BS.775-3 surround format. A subset of loudspeakers termed height channels were varied in their azimuth and elevation angle with three cases presented at ear-level with various azimuthal angles and three cases at elevation and various azimuthal angles. Table 1 presents the specific azimuth and elevation angles and figure 1 presents a visual representation of rising and lateral angle for the six spatial configurations of the height channels.

Table 1: Terminology and azimuth and elevation angles for the six subsets of height channels.

Abbreviation	Subset of channels	Azimuth *(°)*	Elevation *(°)*	Colour code
NE	Narrow, ear-level	±11	0	Orange
WE	Wide, ear-level	± 48	0	Yellow
SE	Side, ear-level	± 81	0	Green
NH	Narrow, with height	± 16	+40	Cyan
WH	Wide, with height	± 52	+40	Blue
VoG	Voice of God	± 59	+70	Magenta

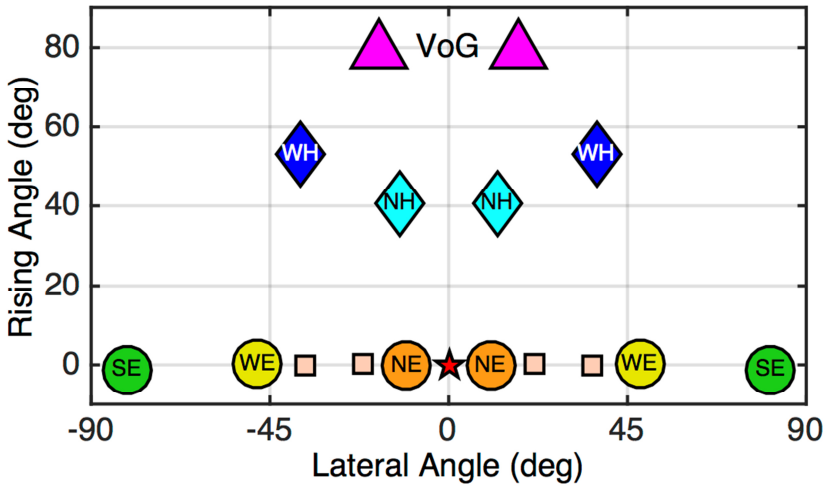

Figure 1: Frontal-hemisphere loudspeaker configuration with colour coded height channels.

2.2 Stimuli

Each stimulus consisted of three discrete core elements: the source signal of anechoic vocal performances, computer modelled discrete early reflections and diffuse reverberation.

2.2.1 Source signal

The program material consisted of two short phrases of the Magnificat, namely, 'Magnificat' and, 'Anima mea dominum' termed program A and program B respectively. However, for the purposes of this task, the 'dominum' of program B was omitted in order to emphasise spectral differences between programs through consonant articulation: program 1 having a higher number of vocal occlusions in the form of high frequency plosives and fricatives than program 2, which contained predominantly nasal articulation. Figures 2 and 3 provide a visual representation of the frequency over time of the vocal articulation.

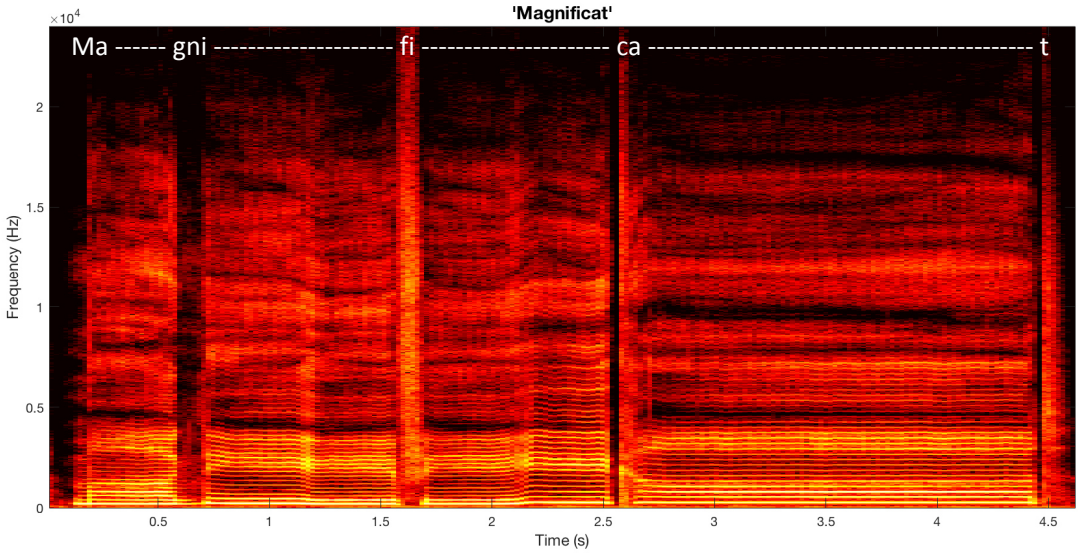

Figure 2: Spectrogram of program A 'Magnificat' displaying vocal articulation.

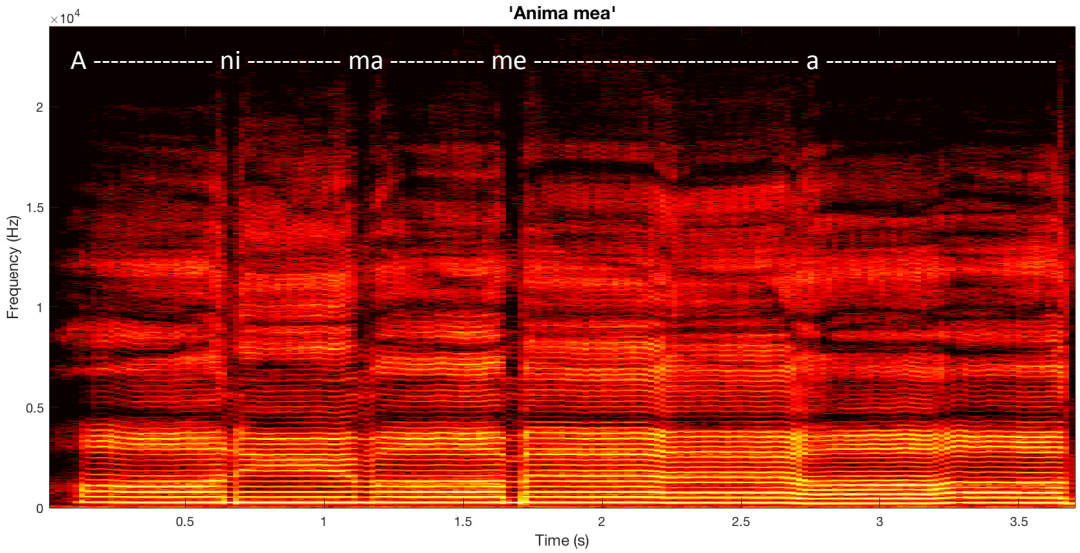

Figure 3: Spectrogram of program B 'Anima mea' displaying vocal articulation.

The vocal performances were recorded in an anechoic chamber in order to obtain source material devoid of reflections. Of course, the inherent nature of this process for performers, and more specifically vocalists who rely heavily on acoustic feedback, presents an unnatural experience that introduces difficulty in achieving a natural performance. In order to overcome this limitation, the performer was presented with a monitor mix containing contextually appropriate convolution reverberation. This provided acoustic feedback that the performer might expect in a Gothic cathedral which in turn informed a more comfortable and appropriate recording process.

Both program A and B contain five separate performances of the same excerpt of Gregorian chant. This was done in order to create an ensemble of five singers where slight variations in each performance of the vocalist gave rise to source separation through voice scatter (Ternström 1993) and so that each of the five performances might be presented through their own discrete channel.

2.2.2 Discrete early reflections

Early reflections were generated through the use of an image-source model in MATLAB. The spatial dimensions for the input of the model shown in table 2 were taken from Grace Cathedral, San Francisco (Hong et al. 2015). The source and receiver coordinates are relative to the bottom left hand corner of the rectangular room model. A slight spatial offset of both the source and receiver locations were used in order to increase variation in the angle of incidence of reflections arriving at the receiver position to reduce any symmetry of reflective patterns.

Table 2: Dimensions of the rectangular room input of the image model with corresponding source and receiver location coordinates.

	x *(m)*	y *(m)*	z *(m)*
Room	31.35	102.30	29.70
Source	16.50	79.20	2.31
Receiver	15.51	39.60	1.32

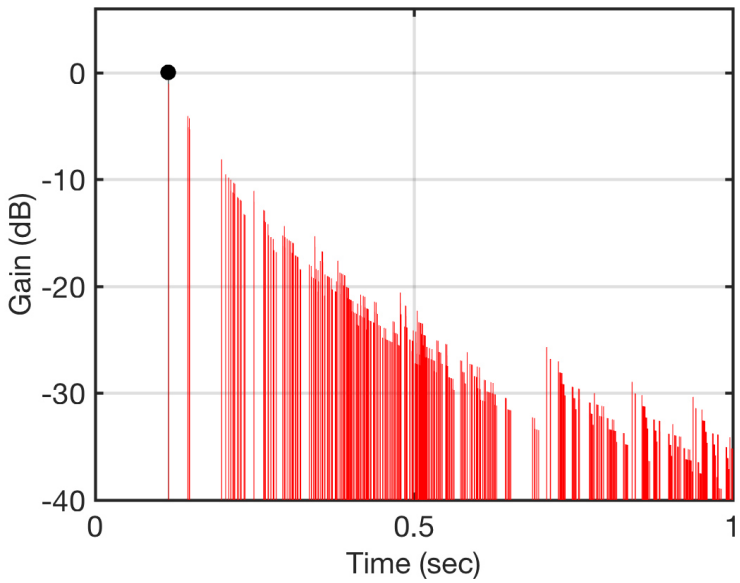

Figure 4: Reflectogram showing gain applied to discrete reflections generated by the image model.

Figure 4 provides a visual representation of the reflections generated by the image model. The reductions in gain occur primarily as a function of air absorption during propagation with a further attenuation applied for the corresponding order of each reflection, or the number of times a reflection intersects a virtual boundary before reaching the receiver position. For the sake of simplicity in the model the air absorption is considered broadband; however, the reflection order reduction is frequency dependent. A high-shelf filter with a cut-off frequency of 2 kHz and -2 dB attenuation is implemented for each reflection order. For example, a seventh order reflection is processed through the filter seven times. In addition to high-shelving, each time the filter is implemented the signal shifts slightly in time creating a temporal dispersion effect further limiting the cartoon-like effect associated with virtual acoustic renderings through image-source models.

Figure 4 also displays multiple decay slopes. These particular characteristics of reverberation within large enclosed spaces is discussed at length by Hong et al (2015). These multiple reverberant decays may be observed

where coupled spaces receive, store and then release sound energy back into the primary space at high concentrations from particular spatial locations. These concentrations can be observed in figures 5-8.

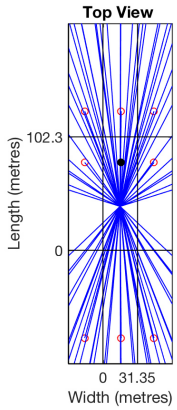

Figure 5: Top view of image model for input values in table 2 showing source position in black and virtual sources in red.

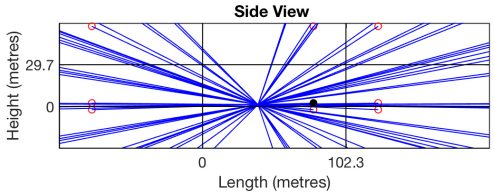

Figure 6: Side view of image model for input values in table 2 showing source position in black and virtual sources in red.

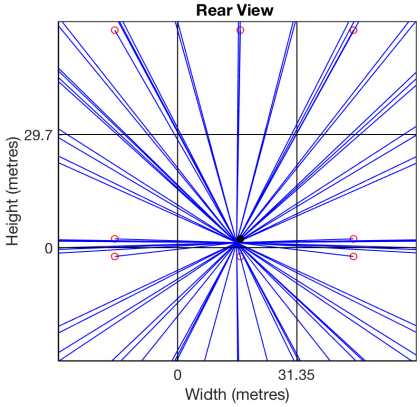

Figure 7: Rear view of image model for input values in table 2 showing source position in black and virtual sources in red.

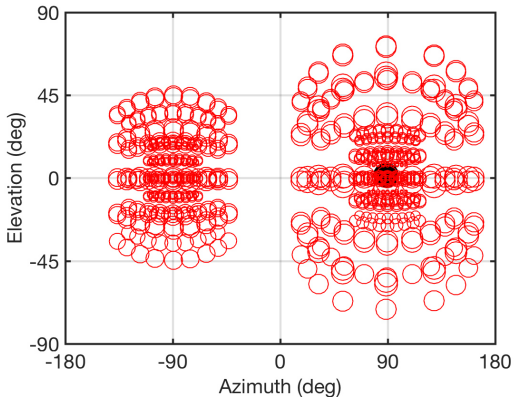

Figure 8: Spherical view of three-dimensional reflection distribution for input values in table 2 showing source position in black and virtual sources in red.

A spatial sorting algorithm based on azimuth and elevation angles was implemented in order to designate the reflections to their corresponding loudspeaker locations. Table 3 sets out the spatial catchment boundaries for each loudspeaker channel with the limits of the spatial selection defined by azimuth and elevation values that are spatially appropriate for their corresponding loudspeaker location.

Table 3: Boundary angular values for groups of discrete early reflections for each channel of the playback system.

Channel	Azimuth *(°)*	Elevation *(°)*
Left front	-90 < az < -55	-30 < el < 20
Left centre	-55 < az < -20	-30 < el < 20
Centre	-20 < az < 20	-30 < el < 20
Right centre	20 < az < 55	-30 < el < 20
Right front	55 < az < 90	-30 < el < 20
Left surround	-180 < az < -90	-30 < el < 20
Right surround	90 < az < 180	-30 < el < 20

Left front (height channel)	-90 < az < 0	20 < el < 90
Right front (height channel)	0 < az < 90	20 < el < 90

2.2.3 Diffuse reverberation

The diffuse reverberation was generated through the use of a commercial algorithmic reverberator, in this case, a Strymon Big Sky effects processor. This was chosen for its parameter manipulation flexibility and overall sound quality. A hall algorithm was selected with a reverberation time of four seconds. The ensemble of anechoic Gregorian chant was stereo processed through the unit and digitally captured. Each program contained three pairs of stereo reverberation for the fixed front and rear and variable height channels. This modulation generates a more natural and spatially diffuse stereophonic reverberation.

2.3 Procedure

A total of four listeners from the Audio and Acoustics program at the University of Sydney voluntarily participated in the listening task. Each participant was considered to be an expert listener and reported no hearing deficiencies. The test itself took approximately 10 minutes per subject with ceiling prominence and spatial width conducted as separate tasks. Subjects were provided with the SE and VoG cases for both programs A and B once each prior to completing the task in order to define the extremities of the virtual space and for brief listener training. Each iteration of the 12 stimuli were presented with a new randomised order to reduce the influence of program order on responses.

Participants then marked a dash intersecting either a vertical line for ceiling prominence or a horizontal line for spatial width that corresponded with their perception of these two attributes ranging from not prominent to strongly prominent for ceiling prominence and narrow to wide for spatial width (see figure 9 for example). These lines were five centimetres long providing 50 millimetre resolution. This was considered adequate for producing a sufficiently fine scale. The point along the vertical or horizontal line where the subject intersected with a dash was measured using a millimetre ruler. This data was then compiled and is presented in summary form below.

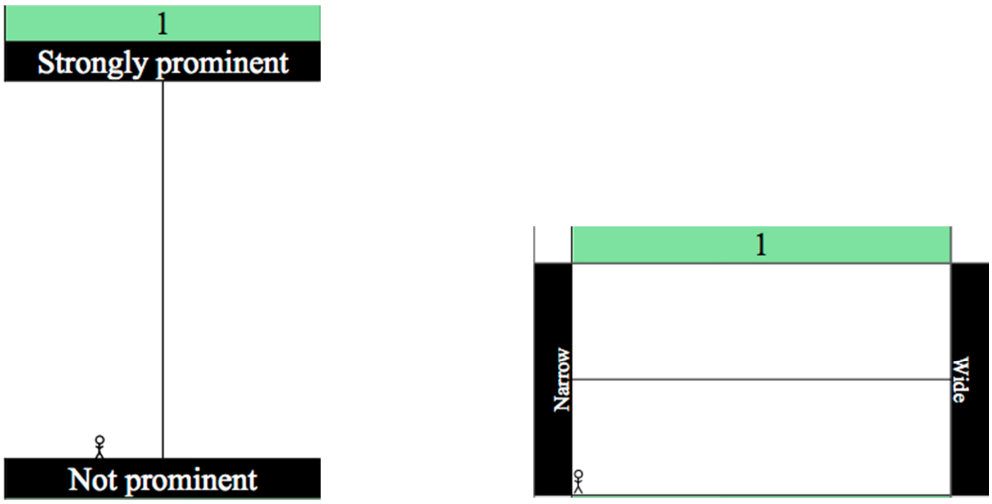

Figure 9: Ceiling prominence and spatial width response sheet examples. Note the embedded stick figure that was included to provide a reference for use of each scale.

3. RESULTS

Significant differences between program A and B in both ceiling prominence and spatial width ratings were not observed, therefore, the data for both programs have been combined for this analysis. Differences were expected due to the additional high frequency content of program A when compared to program B (see figures 2 and 3), however, it seems that the speech articulation did not play a large role in ceiling prominence. Also, as individual listeners used the graphic scales differently, all responses were standardised for each listener before construction of the summary graphics included in this paper. This standardisation practice is commonly employed in

such cases, where each individual listener's mean mark on the response scale is subtracted from all of their marks, so that 'centred' responses can be compared (each having a mean of zero for each scale). Furthermore, in order to ensure that all responses from all listeners cover the same range of the response scales, all 'centred' responses were divided by the standard deviation of each listener's responses so that the resulting set of responses would have the same value of 1 for standard deviation. Such standardised rating data are plotted for ceiling prominence and for spatial width in figures 10 and 11 respectively.

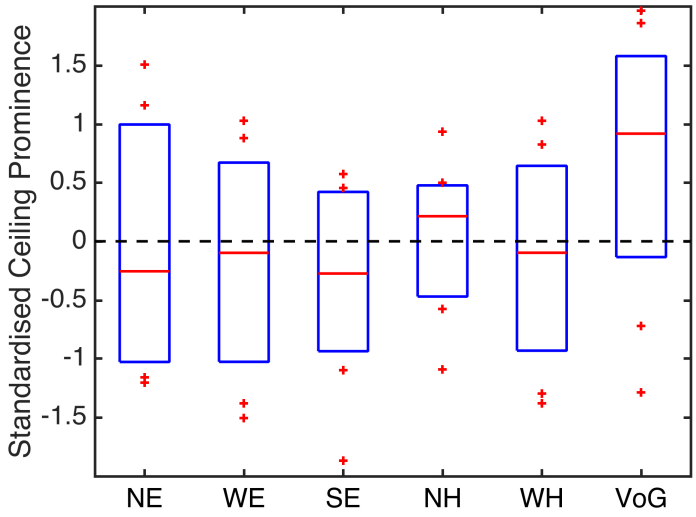

Figure 10: Box plots showing the distribution of the standardised ceiling prominence ratings for the six different loudspeaker configurations.

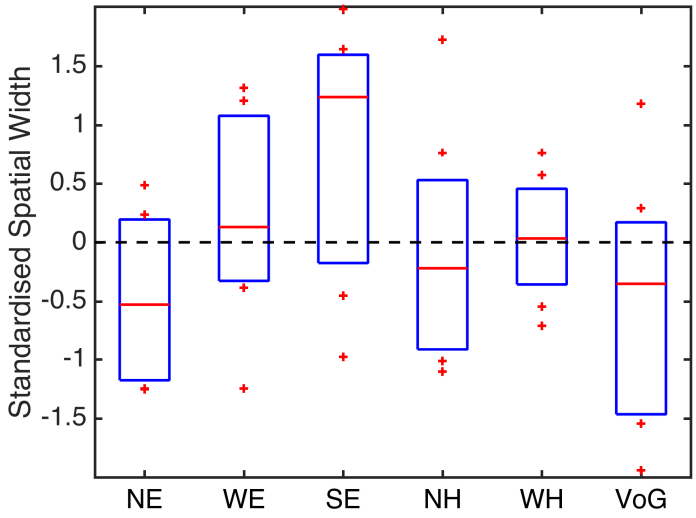

Figure 11: Box plots showing the distribution of the standardised spatial width ratings for the six different loudspeaker configurations.

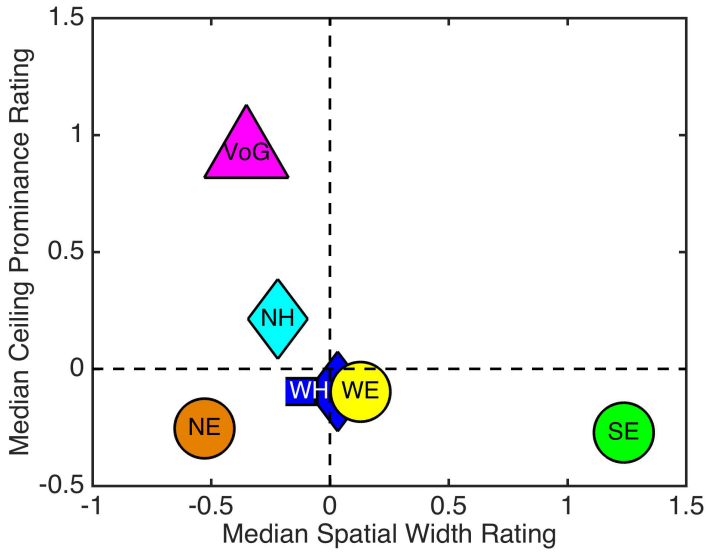

Figure 12: Median responses for both ceiling prominence and spatial width.

While the box plots shown in figures 10 and 11 illustrate a fairly wide distribution of standardised ceiling prominence ratings for the six different loudspeaker configurations the pattern of median values here, particularly that observed for the VoG loudspeaker configuration, correspond generally well with expected results. The outliers, however, seem to suggest that confidence between subjects was fairly low. That being said, both the narrow elevated configurations of NH and VoG yielded higher ratings of ceiling prominence; however, the WH configuration ratings were quite close to those of WE.

4. DISCUSSION

The auditory imagery experienced is affected by the spatial configuration of the loudspeakers. The pattern of variation in ratings is consistent with the hypothesis that elevated height channels conveying indirect sound provide the necessary cues for the listener to perceive a tall cathedral ceiling. The pattern also shows with height and width co-varying it is hard on the basis of ratings to separate these two factors out. This effect is clearly illustrated in the figure 12.

From what can be observed from the cursory data, the strongest conclusion that might be drawn is that strongly localisable elevated sources of indirect sound increase ceiling prominence albeit at a cost to LEV where spatial width ratings were impaired in the case of the VoG configuration. Where the height channels were presented at maximum width at ear level (SE) it was perceived as low and wide and where they were presented at minimum width and maximum elevation (VoG) tall and narrow. The difficulty, however, is that the intermediate configurations, and specifically that of WE, do not elicit responses of the stimulus being perceived as both wide and high. Ultimately, a loudspeaker configuration is yet to be determined where subjects may have a strong perception of a ceiling as well as a highly enveloping sound field. It is difficult to say at this stage whether the use of elevated loudspeakers has a positive affect on SI and further system configurations need to be explored.

There are a number of possible contributing factors to this outcome that are considered. Firstly, the playback system is limited somewhat in terms of obtaining sufficient lateral angle from the listening position at maximal elevation due to its fixed hemispherical structure. Ideally, nine potential configurations of the height channels, that is, one additional configuration at the intermediate height and two at the maximal elevation at lateral angles corresponding to those of the three ear-level conditions may elicit ratings that the virtual acoustic space was perceived to have both a prominent ceiling and wide spatial characteristics. Secondly, the masking of reflections arriving from elevated angles is a determinative factor in whether presenting the indirect sound from height channels increases SI. If the level of the indirect sound arriving above the horizontal plane is either too low or too similar compared to that of the lateral reflections, SI is less likely to be influenced by the elevated loudspeakers with the dominant cues being obtained from ear-level.

In informal discussions with subjects immediately following participation it was apparent that great difficulty

was encountered in the ceiling prominence task. This is reflected in the data where it appears fairly noisy with confidence in responses for ceiling prominence low shown by large distributions. Unfortunately, the term ceiling prominence was applied in a somewhat haphazard manner where error in responses may have been influenced by a lack of specificity and understanding in what was being asked.

5. CONCLUSION

A pilot study into the presentation of indirect sound utilising height channels in various multichannel loudspeaker configurations was executed. A correlation between ratings of 'ceiling prominence' and 'spatial width' was observed, which confirmed that highly uncorrelated signals at ear level give rise to the perception of lower ceiling prominence and greater spatial width, while highly correlated signals give rise to the perception of higher ceiling prominence and reduced spatial width. In order to establish the relative perceptual salience of these two factors, a study employing a global dissimilarity rating task is currently under way. Further studies are also planned in which it is hoped to determine the potential role of backwards masking from subsequent reverberation on the audibility of changes in the spatial distribution of simulated early reflections.

REFERENCES

Baron, M & Marshall, AH 1980, 'Spatial impression due to early lateral reflections in concert halls: The derivation of a physical measure', Journal of Sound & Vibration, vol. 77, no. 2, pp. 211-32.

Cremer, L 1989, 'Early lateral reflections in concert halls', Journal of the Acoustical Society of America, vol. 85, no. 3.

Griesinger, D 1997, 'The psychoacoustics of apparent source width, spaciousness and envelopment in performance spaces', Acta Acustica, vol. 83, pp. 721-31.

Hamasaki, K, Nishiguchi, T, Hiyama, K & Okumura, R 2006, 'Effectiveness of height information for reproducing presence and reality in multichannel audio system' Proceedings of the 120th Convention of the Audio Engineering Society, Audio Engineering Society, Paris, France.

Hamasaki, K, Shinmura, T, Akita, S, & Hiyama, K 2001, 'Approach and mixing technique for natural sound recording of multichannel audio', Proceedings of the Audio Engineering Society Conference: 19th International Conference: Surround Sound-Techniques, Technology, and Perception, Audio Engineering Society, Bavaria, Germany.

Hiyama, K, Komiyama, S & Hamasaki, K 2002, 'The minimum number of loudspeakers and its arrangement for reproducing the spatial impression of diffuse sound field', Proceedings of the 113th Convention of the Audio Engineering Society, Audio Engineering Society, Los Angeles, CA, USA.

Hong, JWJ, Woszczyk, W, Begault, DR & Benson, D 2015, 'Synthesis of moving reverberation using active acoustics – preliminary report', Proceedings of the 138th Convention of the Audio Engineering Society, Audio Engineering Society, Warsaw, Poland.

Martens, WL, Cabrera, D, Miranda, L & Jiminez D 2015, 'Potential and limits of a high-density hemispherical array of loudspeakers for spatial hearing and auralization research', Journal of Applied Mathematics and Physics, vol. 3, pp. 240-6.

Solvang, A & Svensson, UP 2006, 'Perceptual importance of the number of loudspeakers for reproducing the late part of a room impulse response in a listening room', Proceedings of the 121st Convention of the Audio Engineering Society, Audio Engineering Society, San Francisco, CA, USA.

Ternström, S 1993, 'Perceptual evaluations of voice scatter in unison choir sounds' Journal of Voice, vol. 7, no. 2, pp. 129- 35.