Effect of Vertical Microphone Layer Spacing for a 3D Microphone Array

HYUNKOOK LEE, *AES Member,* **AND CHRISTOPHER GRIBBEN,** *AES Student Member*
(h.lee@hud.ac.uk)

University of Huddersfield, Applied Psychoacoustics Lab, Huddersfield, HD1 3DH, UK

Subjective listening tests were conducted to investigate how the spacing between main (lower) and height (upper) microphone layers in a 3D main microphone array affects perceived spatial impression and overall preference. Four different layer spacings of 0m, 0.5m, 1m, and 1.5m were compared for the sound sources of trumpet, acoustic guitar, percussion quartet, and string quartet using a nine-channel loudspeaker setup. It was generally found that there was no significant difference between any of the spaced layer configurations, whereas the 0m layer had slightly higher ratings than the more spaced layers in both spatial impression and preference. Acoustical properties of the original microphone channel signals as well as those of the reproduced signals, which were binaurally recorded, were analyzed in order to find possible physical causes for the perceived results. It is suggested that the perceived results were mainly associated with vertical interchannel crosstalk in the signals of each height layer and the magnitude and pattern of spectral change at the listener's ear caused by each layer.

0 INTRODUCTION

The recently proposed multichannel audio formats such as 22.2 [1] and Auro-3D [2] employ height channels to provide the auditory sensation of a "three-dimensional (3D)" space. For cinema sound or pop music production, the height channels could be used for creative panning of source image in the vertical domain as well as for providing extra ambience. On the other hand, for acoustic recordings made in a concert hall, the use of height channels is likely to be focused on extra ambience since source images would not need to be elevated in most cases (an exception of which could be choir singers on high stands).

In recent years a few main microphone techniques employing height channels have been introduced [3–5]. For example, Theile and Wittek [3] proposed a technique called "OCT-9" that employs four upward-facing cardioid microphones that are placed above the front left, front right, rear left, and rear right microphones of the main microphone array "OCT-5." The recommended spacing between the main and height microphones for this technique is 1m or wider. Williams [4] also designed a 3D microphone array with four height microphones that are vertically spaced from the main microphones. The proposed spacing between the lower and upper layers is 1m and the polar pattern of the height microphones is figure-of-eight. On the other hand, Geluso [5] proposed using a "coincident" microphone technique as a method to capture height information; a vertically oriented figure-of-eight "side" microphone is configured with a front-facing "mid" microphone without any spacing between the two.

To date, however, no formal experimental data has been provided on the effect of spacing between main and height channel microphones on perceived spatial impression. In the context of horizontal stereophony, it is widely known that a more spaced microphone pair would produce a greater spatial impression in reproduction [6–8]. This is due to the fact that a larger spacing between the microphones would lead to a lower degree of interchannel correlation between the signals [7]. However, research suggests that the principles of horizontal stereo might not be directly applicable to vertical stereo. In terms of localization, it is well known that vertical localization relies on spectral cues rather than interaural cues [9, 10]. The amplitude panning of phantom image in vertical stereophonic reproduction has been reported to be unstable [11]. It has also been found that the precedence effect does not fully operate between vertically arranged loudspeakers regardless of the time difference applied to them [12, 13], and that time panning in the vertical plane is ineffective [14]. With respect to spatial impression, the present authors investigated the effectiveness of interchannel decorrelation for controlling the perceived image spread of band-passed pink noise, using two loudspeakers arranged vertically in the median plane as well as those horizontally arranged [15]. It was found that the effectiveness of vertical decorrelation was not as strong as that of horizontal decorrelation, depending on frequency. However, the perceptual mechanism of vertical spatial impression has not been fully explored yet and therefore needs further investigation.

In the present study a series of listening tests has been carried out in order to investigate the effect that the microphone

spacing between main and height channel microphone layers has on the magnitude of perceived spatial impression and the subjective preference. Objective measurements of recorded and reproduced signals have also been carried out in order to examine possible physical causes for subjective results. The scope of the study was focused on acoustic recordings made in a concert hall, using a main microphone array. It is expected that the findings of this study will not only provide a useful basis on which to develop a 3D main microphone technique but also extend the knowledge of auditory perception in vertical stereophony.

This paper is organized as follows. Section 1 describes the process of experimental stimuli creation and the method of listening experiment. The results of the statistical analysis of the test data are presented in Section 2. Section 3 describes the objective measurements conducted and presents the results. Section 4 discusses the subjective results in relation to the objective measurements. Finally, Section 5 summarizes and concludes the paper.

1 EXPERIMENTS

1.1 Recording Setup

Two different types of recordings were made in a shoebox shaped concert hall called St. Paul's in Huddersfield, UK (V = approx. 5700m^3; RT = avg. 2.1s): one for obtaining multichannel room impulse responses (MRIRs), and the other for the recording of virtual ensembles. Fig. 1 shows the physical setup of the loudspeakers and microphones used in the recording. The impulse responses were obtained using the exponential sine sweep method [16] with a single Genelec 8040A loudspeaker placed in the stage center. The MRIRs were later convolved with anechoic trumpet and acoustic guitar signals to create single source stimuli. They were then used for the analyses of various acoustical characteristics in different time segments, which are described in Section 3. There were two types of virtual ensemble performances generated through the other four loudspeakers of the same type: percussion (conga/bongo) quartet and string quartet. These four sources were chosen to give a varied assessment between solo and ensemble instruments. They also allow for the investigation of the influences of the temporal and spectral characteristics of sound. The recordings were made in the PCM wave format at 44.1 kHz/16 bits. The loudspeakers used have reasonably flat on-axis frequency responses, from 48 Hz to 20 kHz within the range of $\pm$3 dB deviation. The off-axis response of a loudspeaker radiation typically has reduced high frequencies. Nevertheless, this was considered to be acceptable since most musical instruments also tend to be directional at high frequencies with reduced energy above about 4 kHz.

All of the microphones used were the same type (AKG C-414 B-XLS in cardioid polar pattern), and they were recorded with an identical amplification level. The microphones of the main (lower) layer were configured based on a multichannel microphone array called PCMA [17], as shown in the upper panel of Fig. 1. The distance and angle between microphones in the front triplet have been selected to provide a continuous and linear localization curve across the front three channels in reproduction, as well as to produce sufficient interchannel decorrelation for frontal spatial impression [17]. The stereophonic recording (coverage) angle calculated for this configuration at the 3m source-array distance was 110°. The 3m distance between the front triplet and the rear microphone pair, as well as between the two rear microphones, was determined based on [7, 8] to ensure sufficient interchannel decorrelation. The microphones were placed at 2m from the stage floor level. The frontal microphones were tilted 60° downwards while the loudspeakers were tilted 30° upwards. This was to ensure the original frequency spectra of both the exponential sine sweep and anechoic source signals, radiated by the loudspeakers, was captured as accurately as possible.

The positions and directions of height channel microphones were selected based on the approaches of the previously proposed techniques described in Section 0 [3–5]. Four height channel cardioids were positioned directly above the front left, front right, rear left, and rear right microphones of the main array as suggested by Theile and Wittek [3]. These were placed at four different heights of 0, 0.5, 1, and 1.5m from the main array in order to investigate the effect of different layer spacings. A spacing of 1m or higher is suggested in [3, 4] whereas a vertically coincident configuration (0m spacing) is proposed in [5]. Since the scope of the current study was a main microphone array design, which does not tend to have extreme microphone spacings, the 1.5m spacing was considered to be large enough for channel separation and yet close enough for microphone rigging in a practical recording situation. As with the approaches proposed in [3, 4], all of the height channel microphones were positioned directly upwards to capture reflections from the same direction. Theile and Wittek [3] suggest the use of the cardioid polar pattern for ceiling-facing height channel microphones. On the other hand, Williams's [4] technique employs vertically positioned figure-of-eight microphones. This polar pattern might be beneficial in terms of suppressing the direct sound if the microphone was angled so that its null point was toward the sound source. However, if the microphone was configured vertically as suggested in [4], the rear lobe of the microphone could potentially pick up strong direct sound and floor reflections, which might lead to inaccurate vertical localization and undesired tonal coloration. Therefore, for the current experiment using upward-facing height microphones, the cardioid was considered to be more suitable than the figure-of-eight.

1.2 Reproduction Setup

Listening tests were conducted in a dry listening room (8.3m (W) × 5.4m (L) × 3.4m (H); RT = 0.2s; NR 15) at the University of Huddersfield. Nine Genelec 8040A loudspeakers were arranged based on the nine-channel Auro-3D layout [2] as shown in Fig. 2. Five Genelec 8040A loudspeakers were situated in the conventional five-channel arrangement, and an upper layer of height channel loudspeakers of the same type was placed directly above the left, right,

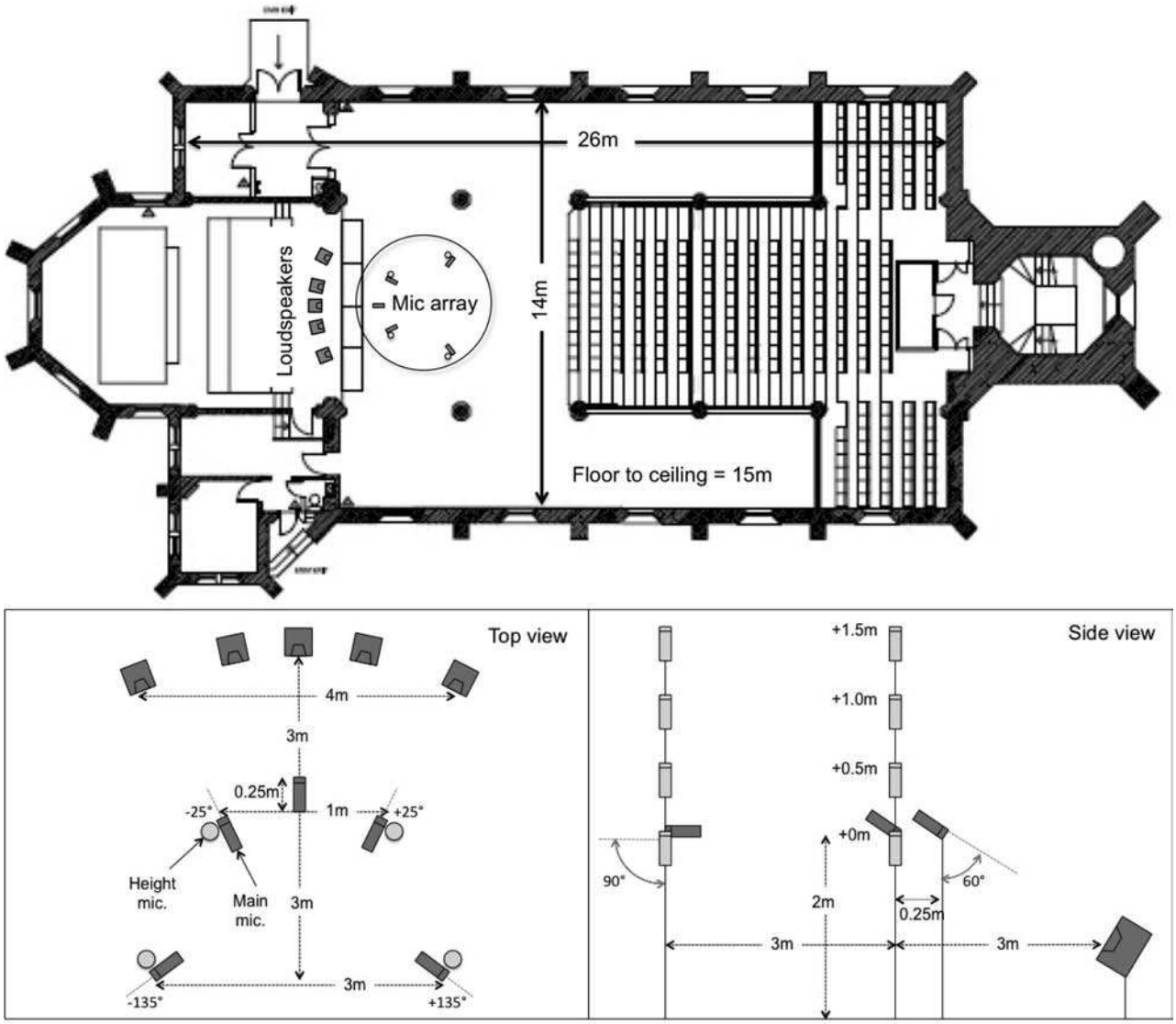

Fig. 1. Recording setup in St. Paul's concert hall

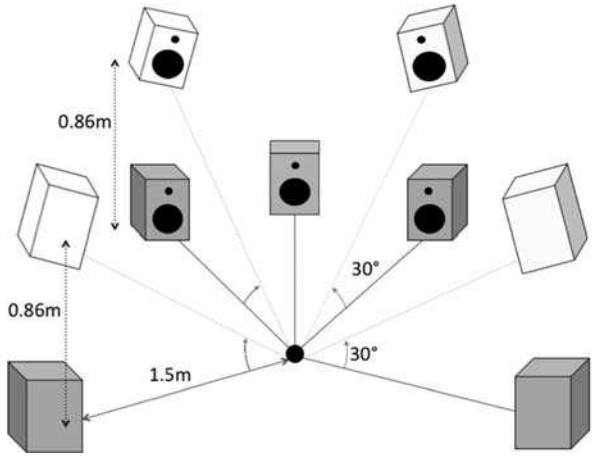

Fig. 2. Loudspeaker setup based on the 9-channel Auro-3D

surround left, and surround right loudspeakers at a vertical angle of 30° from the listening position. The main and height loudspeakers reproduced the five-channel main and four-channel height layer microphone signals, respectively. In order to match the arrival times and sound pressure levels (SPLs) of the main and height loudspeaker signals, necessary delay and level alignments were applied to the main channels with reference to the height channels; each of the main channel output signals was delayed by 0.68 ms and attenuated by 1.2 dB.

1.3 Test Stimuli

The multichannel impulse responses collected in the concert hall were convolved with anechoic recording excerpts of a solo acoustic guitar and a solo trumpet, which were taken from the Bang & Olufsen's Archimedes CD [18]. The other two samples used were the virtual percussion and string quartets, recorded directly in the concert hall. Therefore, a total of 16 stimuli were produced for the listening tests (four sources times four microphone spacings).

The playback level of the entire main microphone layer for each source was calibrated to the LAeq of 72 dB SPL, measured at the listening position over the entire length

of the source using a Casella CEL-450 real time analyzer. Height layer signals were reproduced at the same gain as the main layer ones, so that the original level relationships between the main and height microphones' signals were maintained. It should be noted that height microphones with different vertical spacings had different level relationships with the corresponding main microphone due to their cardioid polar pattern and fixed direction toward the ceiling. This resulted in a slight variation of SPL when the signals of each height microphone layer was reproduced together with the main layer signals; when both main and height channels were reproduced, the LAeq decreased from 75 to 73.7 dB SPL as the vertical microphone spacing changed from 0 to 1.5m. These differences were not compensated since they were produced inherently from the direction and polar pattern of the height microphones, which were experimental constants. This is also an ecologically valid scenario in practical recording situations where vertical microphone spacings are experimented using the same polar pattern and direction. The influence of this level difference relationship on perceived results will be discussed in detail in Section 4.

1.4 Test Method

There were two sets of tests conducted: "spatial impression" and "preference" tests. In the context of concert hall acoustics research, spatial impression is usually understood as an attribute that has two sub-dimensions of apparent source width (ASW) and listener envelopment (LEV) [19]. However, these sub-attributes only describe the two-dimensions of width and depth. Since the purpose of a microphone array with height channels is to add the height dimension in reproduced sound, the term spatial impression in the current experiment was defined as a global attribute that describes all possible spatial percepts from the three-dimensions (3D) of width, depth, and height for both source and environment-related sound components. It was not within the scope of this study to test possible sub-attributes of 3D spatial impression individually—future research is necessary to fully elicit and define different types of 3D spatial attributes.

A total of 12 subjects from the University of Huddersfield's music technology courses participated in the listening tests. They comprised staff members, researchers, and final year undergraduate students, all of whom had previous experiences in critical listening of various spatial audio attributes in listening test environments.

Multiple stimuli comparison tests were conducted using a graphical user interface (GUI) produced by the authors using Max-MSP software, shown in Fig. 3. The same GUI was used for both the spatial impression and preference tests. For each test, the subject was to complete a total of four trials, each of which contained the stimuli of the four microphone spacings for each sound source. All stimuli were played synchronously so that the subjects could switch between them continuously. The subject was asked to grade three stimuli against a reference stimulus on a bipolar continuous rating scale, where the reference was to be

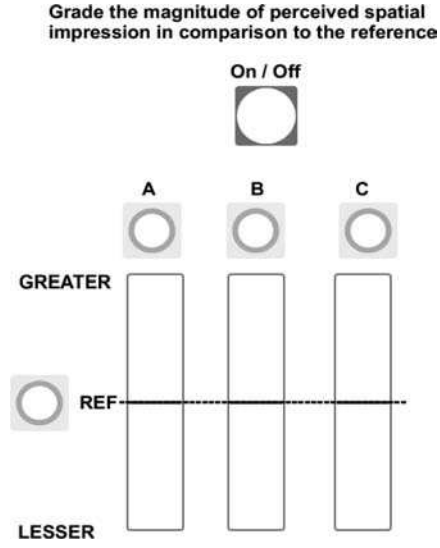

Fig. 3. Graphical user interface used for the listening tests

taken as 0, giving the subject a reference point in the scale when all stimuli were judged to be similar. One of the four stimuli was chosen to be the reference in each trial, and this was randomized in order to avoid potential psychological biases. There was an equal chance for each stimulus to be the reference for each source. The presentation orders of the stimuli, trials, and tests were also randomized.

For both spatial impression and preference tests the scale values ranged from –50 to +50 with a step size of 1, but these were internal quantities and not shown to the subjects. The labels "greater" and "lesser" were used to indicate directions for grading. The subjects were instructed that the end points of the bi-polar scale represented extreme differences against the reference (e.g., extremely greater or lesser magnitude of perceived spatial impression in comparison to the reference). When using this kind of continuous scale without semantic labels, there is a risk that subjects might use the scale inconsistently in each trial. To mitigate this, the subjects were given a familiarization trial including all stimuli before starting the actual tests.

2 LISTENING TEST RESULTS

Data collected from the listening test was first normalized with respect to mean and standard deviation as recommended in ITU-R BS.1116 [20]. This was to reduce potential inter-subject differences in the use of a scale range. Shapiro-Wilk and Levene's tests performed using SPSS suggested that the data for each microphone spacing had normal distribution and equal variance, respectively. Repeated Measure (RM) ANOVA tests were carried out to statistically analyze the main effects of vertical microphone spacing and sound source on the perceived spatial impression and preference, the results of which are presented in Table 1. Paired-samples t-tests were also performed to test the significance of differences between each layer. The Bonferroni correction was applied to the p values obtained from the t-tests in order to avoid potential type-I errors that could

Table 1. Results of RM ANOVA

	Spatial impression			Preference		
	F	p	Partial Eta Squared	F	p	Partial Eta Squared
Mic spacing	13.980	0.000	0.738	5.877	0.002	0.329
Source	1.573	0.213	0.381	1.040	0.387	0.080
Interaction	3.414	0.001	0.973	1.974	0.049	0.141

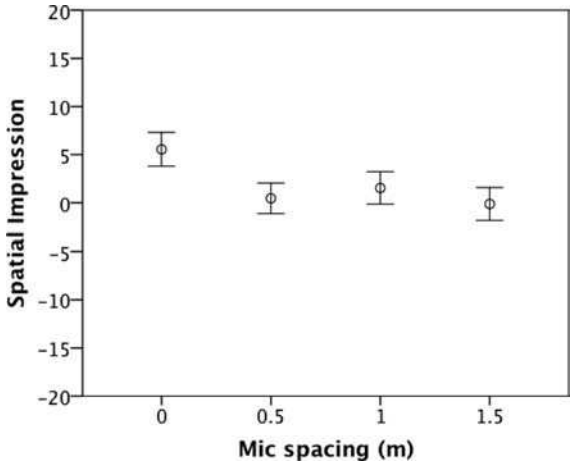

Fig. 4. Microphone spacing vs. spatial impression for all sources: Mean values and associated 95% confidence intervals

be caused in multiple comparison tests. The Bonferroni-corrected results are summarized in Table 2.

2.1 Spatial Impression

The RM ANOVA results in Table 1 indicate that the microphone spacing had a significant main effect on the magnitude of perceived spatial impression at the 1% significance level. Fig. 4 plots the mean values and associated 95% confidence intervals of the differences of 0.5, 1, and 1.5m to 0m for all sources. It can be seen that the 0.5m, 1m, and 1.5m spacings were all graded lower than the 0m one. The paired samples t-test results in Table 2 confirm that these differences are statistically significant ($p < 0.05$), although they appear to be only slight. Fig. 4 and Table 2 also suggest that the differences among the 0.5m, 1m, and 1.5m spacings in perceived spatial impression are insignificant.

Table 1 also indicates that interaction between microphone spacing and sound source had a significant effect ($p < 0.01$). This can be observed from Fig. 6. For the acoustic guitar and percussion quartet, which were the most transient stimuli among all, the magnitude of perceived spatial impression for the 0m microphone layer is slightly, but significantly, greater than those for the spaced layers as the t-test results confirm ($p < 0.05$). Among the spaced layers there is no significant difference observed with the exception of 1.5m graded significantly lower than the other spacings for the trumpet.

2.2 Preference

RM ANOVA (Table 1) suggests that the main effect of microphone spacing was significant ($p < 0.01$). Fig. 5 plots the mean values and associated 95% confidence intervals of the preference test data for all sources. It can be seen that the 0m spacing was graded slightly higher than all the other spacings overall. The t-test results in Table 2 suggest that the 0.5m and 1.5m results were significantly different from the 0m, whereas the 1m was not. The source-dependency of the microphone spacing effect can be observed in Fig. 7, which plots the data for each source separately. For the guitar and string quartet there is no significant difference between the 0m and any other spacings. For the trumpet and percussion quartet, on the other hand, there is a general trend that the 0m spacing is slightly preferred to the larger spacings. The differences among the spaced microphone layers are found to be insignificant, regardless of the source type.

Table 2. Results of paired samples t-tests for each sound source: Bonferroni corrected p values

	Mic pacing pair (m)	Trumpet	Acoustic guitar	Perc. quartet	String quartet	All sources
Spatial impression	0 – 0.5	1.000	0.030	0.018	1.000	0.000
	0 – 1	1.000	0.024	0.000	1.000	0.024
	0 – 1.5	0.000	0.306	0.030	1.000	0.000
	0.5 – 1	1.000	1.000	1.000	1.000	1.000
	0.5 – 1.5	0.006	1.000	1.000	1.000	1.000
	1 – 1.5	0.054	1.000	1.000	1.000	1.000
Preference	0 – 0.5	0.534	1.000	0.198	1.000	0.042
	0 – 1	0.090	1.000	0.900	1.000	0.192
	0 – 1.5	0.000	1.000	0.120	0.588	0.000
	0.5 – 1	1.000	1.000	1.000	0.276	1.000
	0.5 – 1.5	0.330	1.000	1.000	1.000	0.876
	1 – 1.5	1.000	1.000	0.426	0.210	0.186

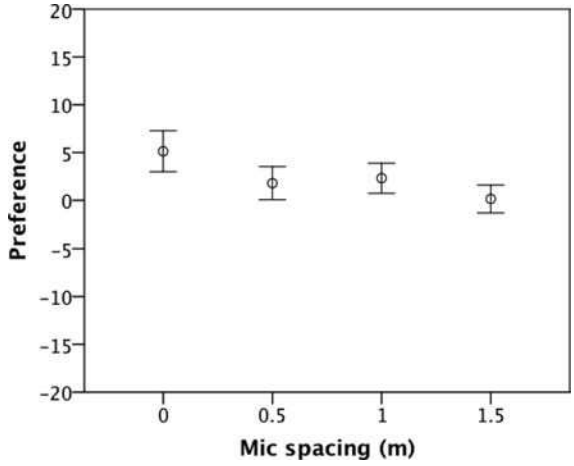

Fig. 5. Microphone spacing vs. preference for all sources: Mean values and associated 95% confidence intervals

3 POST-HOC SIGNAL ANALYSIS

In order to obtain insights into possible causes for the subjective results shown above, a series of post-hoc measurements has been carried out. Two types of signals were analyzed: original multichannel room impulse responses from the recording session (MRIRs) and binaural impulse responses of reproduced sounds (BIRRSs) captured in the listening room by a dummy head microphone. For the MRIRs, signal energies for different time segments, interchannel level differences (ICLDs), direct to reverberant (D/R) energy ratios, and interchannel cross-correlation coefficients (ICCCs) were computed. For the BIRRSs, signal energies for different time segments, spectral influence of height channels, and interaural cross-correlation coefficients (IACCs) were investigated. The methods used are described in the following sections, alongside the results, which will be discussed together with the listening test results in Section 4.

3.1 Channel Signal Analysis

3.1.1 Signal Energy

The signal energies were measured in decibel for two different time windows: 0 ms to 5 ms (direct sound) and 5 ms to 750 ms (ambient sound), with 0 ms being the arrival of the direct sound. This was to examine "interchannel level differences (ICLDs)" between different channels, as well as the "direct to reverberant (D/R) energy ratio" for each channel. Here, the ICLD is defined as the energy ratio between two impulse responses within 5 ms, and the D/R ratio is the energy ratio between sound arriving within 5 ms and that between 5 ms and 750 ms. The time windows were determined based on the research by Bronkhorst and Houtgast [21] and Hidaka et al. [19] and will be used to divide direct and ambient sound throughout this paper. ICLDs between front main and front height signals would be useful for understanding whether the direct sound included in

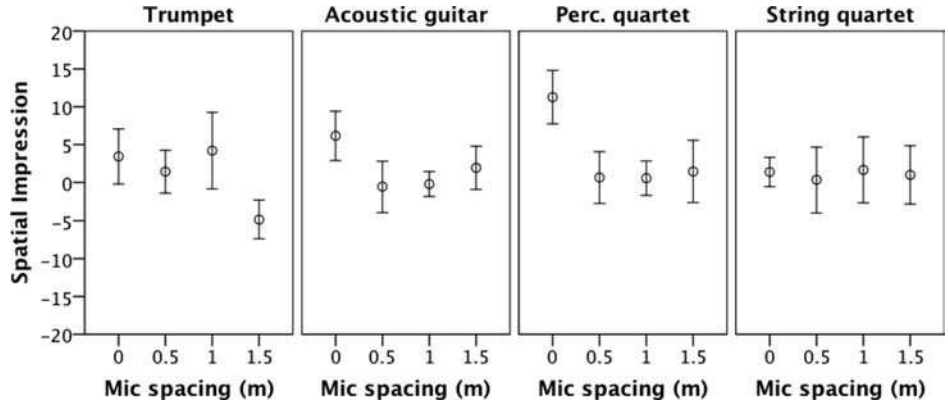

Fig. 6. Microphone spacing vs. spatial impression for each source: Mean values and associated 95% confidence intervals

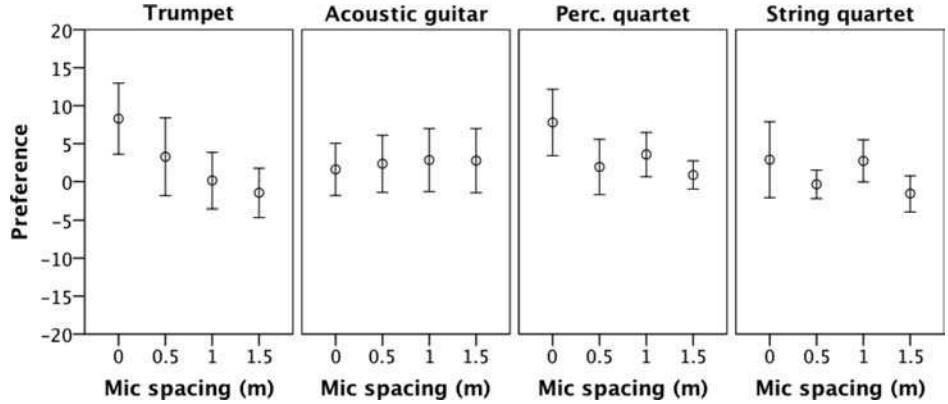

Fig. 7. Microphone spacing vs. preference for all sources: Mean values and associated 95% confidence intervals

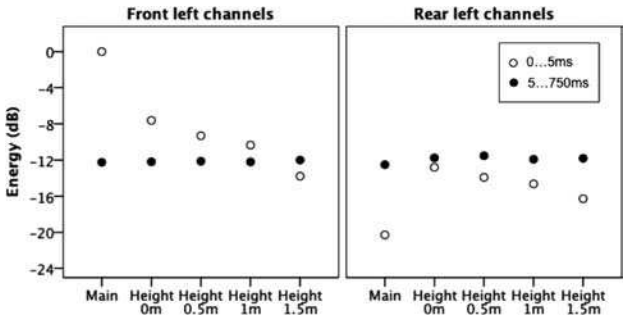

Fig. 8. Results of energy measurements for the direct and ambient sounds of multichannel impulse responses

Table 3. Interchannel Level Difference (ICLD) of each front left height microphone (FLh) to the front left main microphone (FL) and that of each rear left height microphone (RLh) to the rear left main microphone (RL)

Spacing	FLh to FL	RLh to RL
0m	−7.6 dB	−12.8 dB
0.5m	−9.3 dB	−13.9 dB
1m	−10.3 dB	−14.6 dB
1.5m	−13.8 dB	−16.3 dB

Table 4. Direct to Reverberant (D/R) Energy Ratio for front and rear left channel signals

Channel	Front left	Rear left
Main	12.2 dB	−7.8 dB
Height 0m	4.6 dB	−1.0 dB
Height 0.5m	2.8 dB	−2.4 dB
Height 1m	1.9 dB	−2.7 dB
Height 1.5m	−1.8 dB	−4.5 dB

the height channels has an effect on the perceived spatial impression (i.e., vertical interchannel crosstalk). The D/R ratios give an indication of the relative influence that direct and ambient sound energies have on the perceived spatial impression for the different microphone heights.

Fig. 8 plots energies measured for the impulse responses of the main and height microphones for the front left and rear left channels. The energy of the front main left signal was set to 0 dB, to which all other values were normalized as reference. As can be seen, the energies of ambient sounds are almost constant around –12 dB for both front and rear channels. On the other hand, the direct sound energy of the height channel signal decreases gradually as the microphone height increases. The direct energy of the rear main channel is substantially lower than those of the rear height channel, while the front channels show an opposite pattern—this is because the rear main microphone was facing backwards, thus having a greater rejection of direct sound. From the individual energy values, the ICLD between the main and each height signal and the D/R energy ratio for each signal were calculated; results of which are presented in Tables 3 and 4.

3.1.2 Interchannel Cross-Correlation Coefficient (ICCC)

The ICCC is an indicator of the degree of similarity between two channel signals. It is defined as the maximum absolute value of normalized cross-correlation function (NCF), which is defined below.

$$NCF_{t1,t2}(\tau) = \frac{\int_{t1}^{t2} x_1(t) \cdot x_2(t+\tau)dt}{\sqrt{\int_{t1}^{t2} x_1^2(t) \cdot \int_{t1}^{t2} x_2^2(t)dt}}, \quad (1)$$

where x_1 and x_2 are channel signals, t_1 and t_2 are the lower and upper boundaries of time segments, and τ is the time lag.

ICCCs for the MRIRs of main and height microphones with different spacings were measured in octave bands for four different pairs of vertical and diagonal channels: FL (front left) - FLh (front left height), FL and FRh (front right height), RL (rear left) and RLh (rear left height), and RL and RRh (rear right height). The measurements were taken for two time segments separately: $t_1 = 0$ms to $t_2 = 5$ms (direct sound) and $t_1 = 5$ms to $t_2 = 750$ms (ambient sound). The lag (τ) limit for the direct sound segment was ± 2.2 ms, which was the largest ICTD occurring between the direct sounds of the main and height channels, i.e., between FL and FLh with the 1.5m spacing. The lag limit for the ambient segment was taken to be ± 10 ms since it was the maximum ICTD that could occur for a reflected sound, i.e., between RL and RRh at 1.5m.

The results plotted in Fig. 9 are the average ICCCs of low (62.5 Hz, 125 Hz, and 250 Hz), middle (500 Hz, 1 kHz, and 2kHz) and high (4 kHz and 8 kHz) octave bands. Overall, it can be observed that the ICCCs for the ambient sounds are generally lower than those for the direct sounds. For the front channel pairs FL-FLh and FL-FRh, the ICCC results for the direct sounds vary in a relatively small range between around 0.7 and 0.9, regardless of frequency band. ICCCs for the ambient sounds tend to decrease in a slightly wider range as the microphone spacing increases, and this effect appears to be most obvious at the low frequency bands. It is also noticeable that the middle band ICCCs for the vertical pairs FL-FLh and RL-RLh show a steep decrease from 0m to 0.5m and then hardly vary as the spacing increases following that. The high frequency bands have lower ICCCs than the low and middle bands in general, but there is little difference caused by different microphone spacings. Results for the rear channel pairs RL-RLh and RL-RRh show similar patterns to those seen with the front ones, although the former has more irregular patterns for the direct sounds. Since musical sources typically have substantially reduced energies above 4 kHz, as pointed out by Hidaka et al. [19], the high band results seem to be least relevant when discussing the current results. On the other hand, the results for the low and middle band results are considered to be relevant depending on the spectral characteristics of the sound sources used. For example, the trumpet source has fundamental frequencies ranging between 500 Hz and 2 kHz only, and therefore the low band results seem to be irrelevant. In contrast, the guitar, percussion, and string

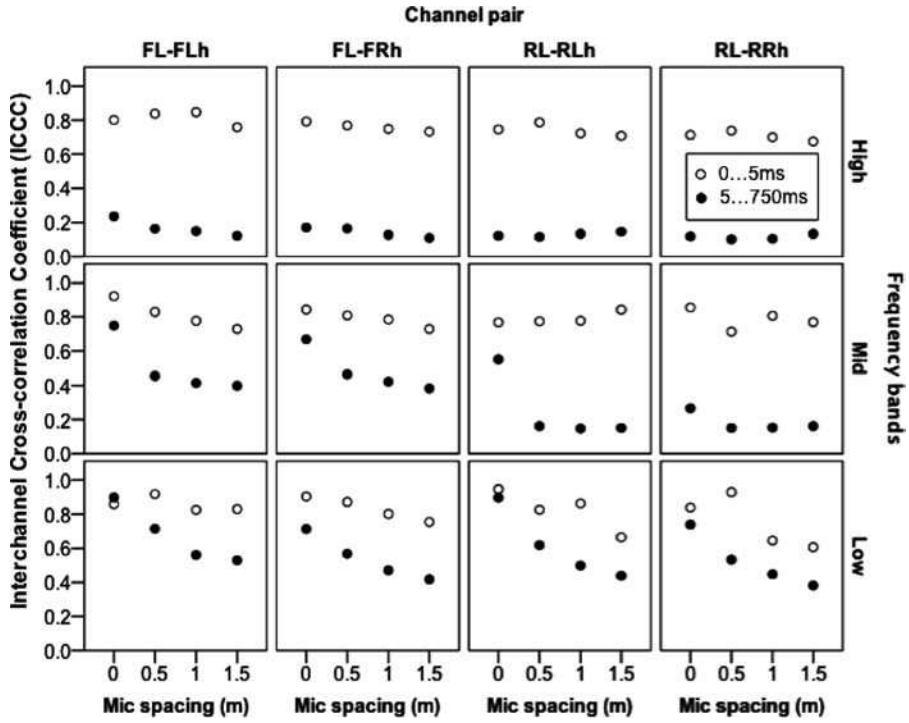

Fig. 9. Interchannel cross-correlation coefficients for vertical and diagonal pairs of microphone signals, measured for direct and ambient sounds separately; FL = front left, FLh (FRh) = front left (right) height, RL = rear left, RLh (RRh) = rear left (right) height

sources all have spectral peaks across the low and middle bands.

3.2 Binaural Signal Analysis

The original nine-channel MRIRs, as well as the convolved listening test stimuli, were reproduced by the corresponding loudspeakers in the room that was used for the listening tests. The loudspeaker configuration and the playback conditions were the same as the listening test. A reproduced sound field, created by a combination of the main and each height microphone layer, was recorded using a Neumann KU100 dummy head microphone placed at the listening position. This was also carried out for the main layer alone, in order to examine signal characteristic differences between 2D and 3D reproductions (i.e., main layer only vs. with height).

3.2.1 Energy of Ear Input Signal

The energies of the left ear input signals, resulting from the combinations of different microphone layers, were computed for the time segments of 0 ms to 5 ms (direct) and 5 ms to 750 ms (ambient) separately. Results of which are plotted in Fig. 10 using 0 dB, the energy for the main layer only, as reference. For the direct sound segment, the 0m height layer shows an energy increase of 2.5 dB. The energy appears to decrease linearly but only slightly from 0.6 dB to 0.2 dB, as the spacing increases from 0.5m to 1.5m. For the ambient parts of the signals, a more dramatic energy increase is observed between the 2D and 3D reproductions. Energy for the main layer is –6.8 dB and this increases by 7.4 dB with the 0m height layer. The ambient energies for the 0.5m, 1m, and 1.5m layers are 0 dB, 0.1 dB, and –0.1dB, respectively. D/R energy ratios for the spaced layers are considerably low and vary only slightly between 0.2 dB and 0.5 dB. The differences in energy increase observed for different microphone layer spacing seems to be associated with spectral changes caused by the different microphone spacings, which will be shown in Section 3.2.3.

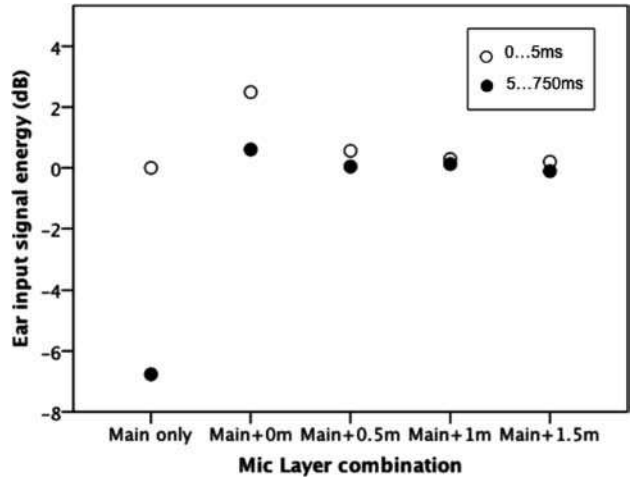

Fig. 10. Energies of left ear input signals with different vertical microphone spacings

3.2.2 Interaural Cross-Correlation Coefficient (IACC)

In order to examine whether the perceived results could have arisen as a result of horizontally perceived spatial impression, $IACC_{E3}$ and $IACC_{L3}$ have been computed for

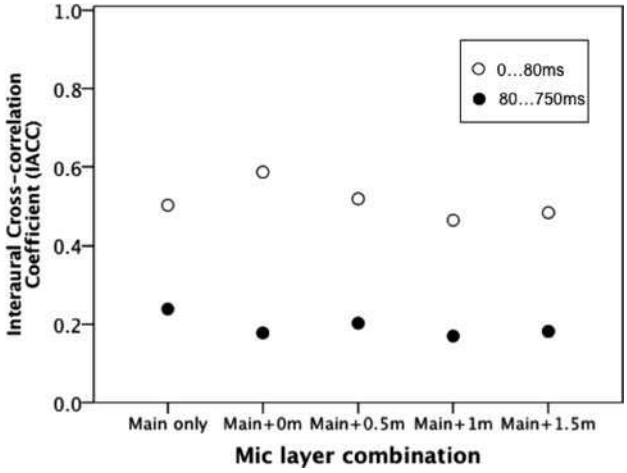

Fig. 11. Interaural cross-correlation coefficients for the binaural impulse responses of height channel reproduction with different vertical microphone spacings; $IACC_{E3}$ indicates integration from 0 ms to 80 ms and $IACC_{L3}$ from 80 ms to 750 ms.

the binaural signals produced by the main microphone layer combined with each height layer. This was also done for the main microphone layer alone, to see how added height channels would have affected horizontal spatial impression. These measures were proposed by Hidaka et al. [19] and are standard predictors for apparent source width (ASW) and listener envelopment (LEV), respectively. The IACC is the maximum absolute value of the normalized cross-correlation function Eq. (1) for binaural impulse responses calculated over the lag (τ) range of –1 ms and +1 ms [19]. $IACC_3$ is the average of the IACCs for three octave bands centered on 500 Hz, 1 kHz, and 2 kHz. $IACC_{E3}$ indicates the $IACC_3$ measured within the integration time window of 0 ms to 80 ms, and $IACC_{L3}$ from 80 ms to 750 ms.

IACC measurement results are plotted in Fig. 11. It is observed that the $IACC_{E3}$ values are greater than the $IACC_{L3}$ ones. The $IACC_{E3}$ for the main layer alone is 0.5 and the addition of the 0m height layer increases this by 0.09. $IACC_{E3}$ values for 0.5m, 1m, and 1.5m added to the main layer are 0.52, 0.46, and 0.48, respectively. For the $IACC_{L3}$ results, the height layers show slightly lower values than the main layer (0.24) in general, varying between 0.17 and 0.2.

3.2.3 Spectral Influence of Height Channel

Different spacings of main and height microphone layers introduce different time delays between the main and height loudspeaker signals arriving at the ear. This would cause differences in the frequency responses of the resulting ear input signals, depending on the phase relationship of the loudspeaker signals. In order to investigate the polarities and magnitudes of spectral changes in the ear signal, as caused by the addition of height microphones with different spacings, the frequency response of the left ear impulse response for the main layer has been subtracted from the main layer combined with each height layer. This was done for the direct (0 ms – 5 ms) and ambient (5 ms – 750 ms) sound components separately. The spectral influences of height layers were also measured for the binaural recordings of the listening test stimuli.

Fig. 12(a) shows the results obtained for the direct sound components. The bottom panel shows the frequency spectrum of the left ear input signal with only the main microphone layer reproduced. Each of the upper four panels shows the spectral magnitude differences of the left ear signal of the main and height layers to that of the main layer only. This represents the spectral changes caused to the main layer ear signal by the addition of each height microphone layer. It is observed that the 0m layer results show positive values at almost all frequencies, whereas the other layers have noticeable fluctuation in the polarity of magnitude difference. This means that the main and 0m height layer signals summed at the ear more constructively, while the other height layers introduced both addition and subtraction depending on frequency. This seems to be due to comb filtering effects caused by the interchannel time difference (ICTD) between the main and spaced height layer signals; the 0m layer did not suffer from this problem due to its vertically coincident nature. Above 5 kHz, the frequency spectrum of ear signal is largely influenced by the head-related transfer functions (HRTF) of the loudspeaker's elevation and azimuth angles. The high peaks around 8 kHz observed for the direct sound results were produced by the difference between the HRTFs of the main and height loudspeaker positions. For example, a HRTF for 0° elevation and 30° azimuth usually has a notch dip at around 8 kHz, whereas that for 30° elevation at the same azimuth has a notch peak at the same frequency [22]. From the results for the ambient sound components, plotted in Fig. 12(b), it is observed that the magnitude differences for every layer fluctuate less than those of the direct sound results and mostly have positive values up to about 10 kHz. The difference between each layer appears to be small, although the 0m layer tends to have slightly less fluctuations in magnitude than the spaced layers between 1 kHz and 5 kHz.

Fig. 13 shows the spectral magnitude differences measured for the left ear input signals of the listening test stimuli. As above, each panel represents the spectral influence of each height layer on the ear input signal for the main layer, of which the original spectrum is shown in the bottom panel. For the trumpet, the main difference among the height layers appears to be produced at frequencies between 400 Hz and 500 Hz, which are where the lowest spectral peaks lie, as can be seen in the bottom panel; the 0m layer produces magnitude gains of about 2 dB in that frequency region, whereas the 0.5m and 1.5m layers cause some reductions in magnitude. It is also apparent that the spaced layers reduce the magnitudes at around 2 kHz, whereas the coincident layer increases them. For the acoustic guitar, the magnitude gain of about 5 dB is produced at 130 Hz by the addition of 0m or 0.5m layer, whereas the 1 and 1.5m layers cause little change. The spectral peaks at 1.3 kHz, produced by the 0.5m and 1m layers, are 1.3 dB and 1 dB higher than that by the 0m layer. However, these spaced layers appear to cause magnitude losses at multiple frequency regions. In contrast, the percussion and string quartets do not tend to have considerable magnitude reductions with the spaced

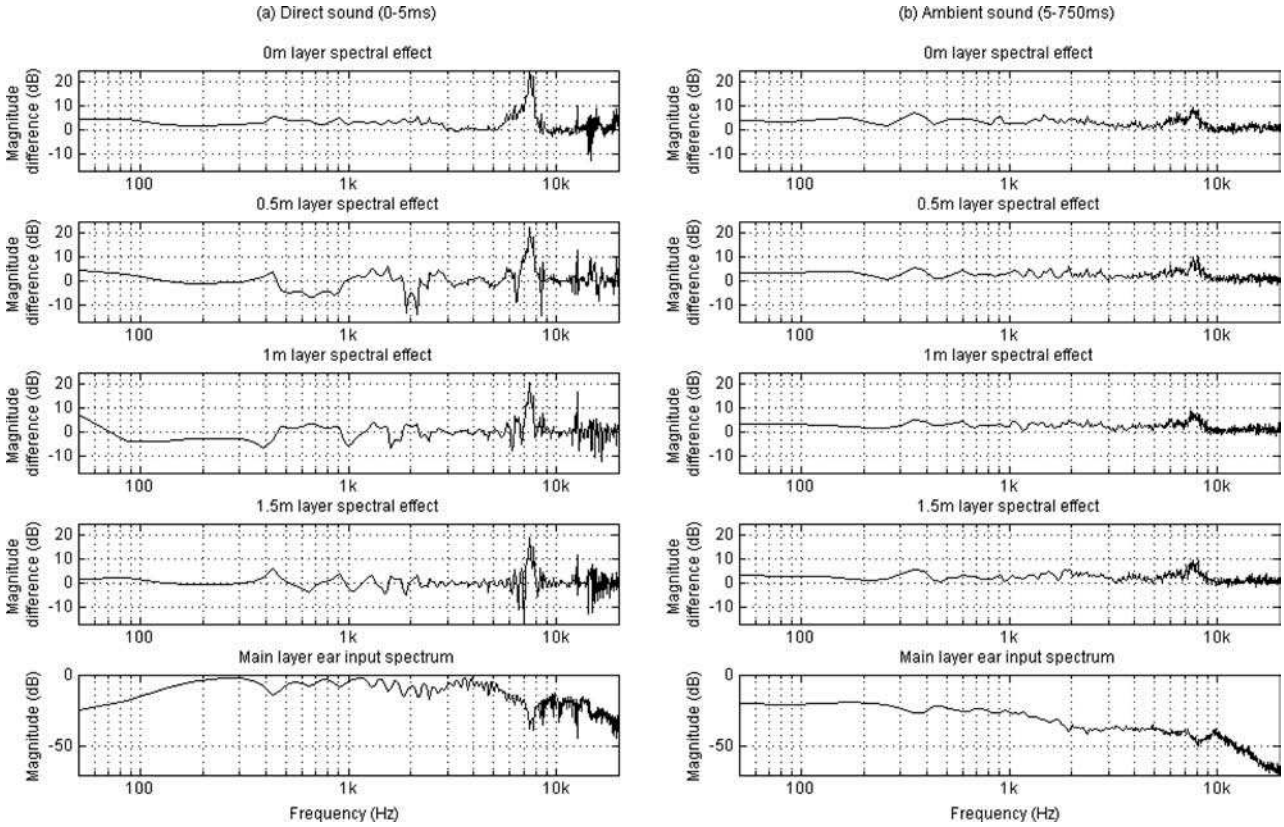

Fig. 12. Spectral magnitude differences of the left ear impulse response of the main and height layers to that of the main layer only; (a) direct sounds (0ms. . .5ms) (b) ambient sounds (5ms. . .750ms); 1024-point average FFT; the bottom panels show the magnitude frequency response of the ear input signal for the main layer.

layers. Nevertheless, the coincident layer still changes the spectrum most constructively for these sources, especially between 100 Hz and 300 Hz for the percussions and between 200 Hz and 800 Hz for the strings.

4 DISCUSSION

This section discusses the results of the listening tests based on those of the signal analysis presented above in terms of spatial impression, preference, and practical implications.

4.1 Spatial Impression

The results generally indicate that the effect of microphone spacing between main and height layers on 3D spatial impression was little or small, depending on the type of sound source. The 0m spacing produced a significantly, albeit slightly greater spatial impression than the larger spacings for the acoustic guitar solo and percussion ensemble, which have more transient characteristics than the trumpet solo and string quartet. It was also apparent that the 0.5m, 1m, and 1.5m spacings did not have significant differences in perceived spatial impression. Possible explanations for these results are provided as follows.

The perceived results should first be explained in relation to the level of direct sound picked up by the height microphone. As mentioned in Section 0, the primary purpose of the height microphones is to capture ambient sounds for height loudspeakers, whereas that of the main microphones is to localize the sound source image at the height of the main loudspeakers. In this regard, a direct sound component included in the height microphone signal can be regarded as a vertically introduced interchannel crosstalk. Table 3 showed that the interchannel level difference (ICLD) of the direct component of the front left height impulse response (FLh) to that of the front left main (FL) varied from –7.6 dB to –13.8 dB as the spacing increased from 0m to 1.5m, whereas the ambient sound level was almost constant across the main and all height signals (Fig. 8). As mentioned earlier, these crosstalk level differences were caused inherently due to the use of the constant polar pattern and direction for the microphones placed at different heights, and this is a practical recording situation. It could be argued that these level differences could have influenced the perceived results in such a way that a louder height microphone layer produced a greater source-related vertical image spread. However, it is considered that only the 0m spacing results might have been perceptually affected by the crosstalk based on the following explanation. Previous research [12] investigated the maximum level of delayed height channel signal, compared to the level of main channel signal, at which the perceived phantom image is localized fully at the position of the main-channel-only image (i.e., localization threshold), using two loudspeakers in the median plane with each elevated at 0° and 30° from the listener's eye level, respectively. The threshold at which the phantom image becomes completely inaudible (i.e.,

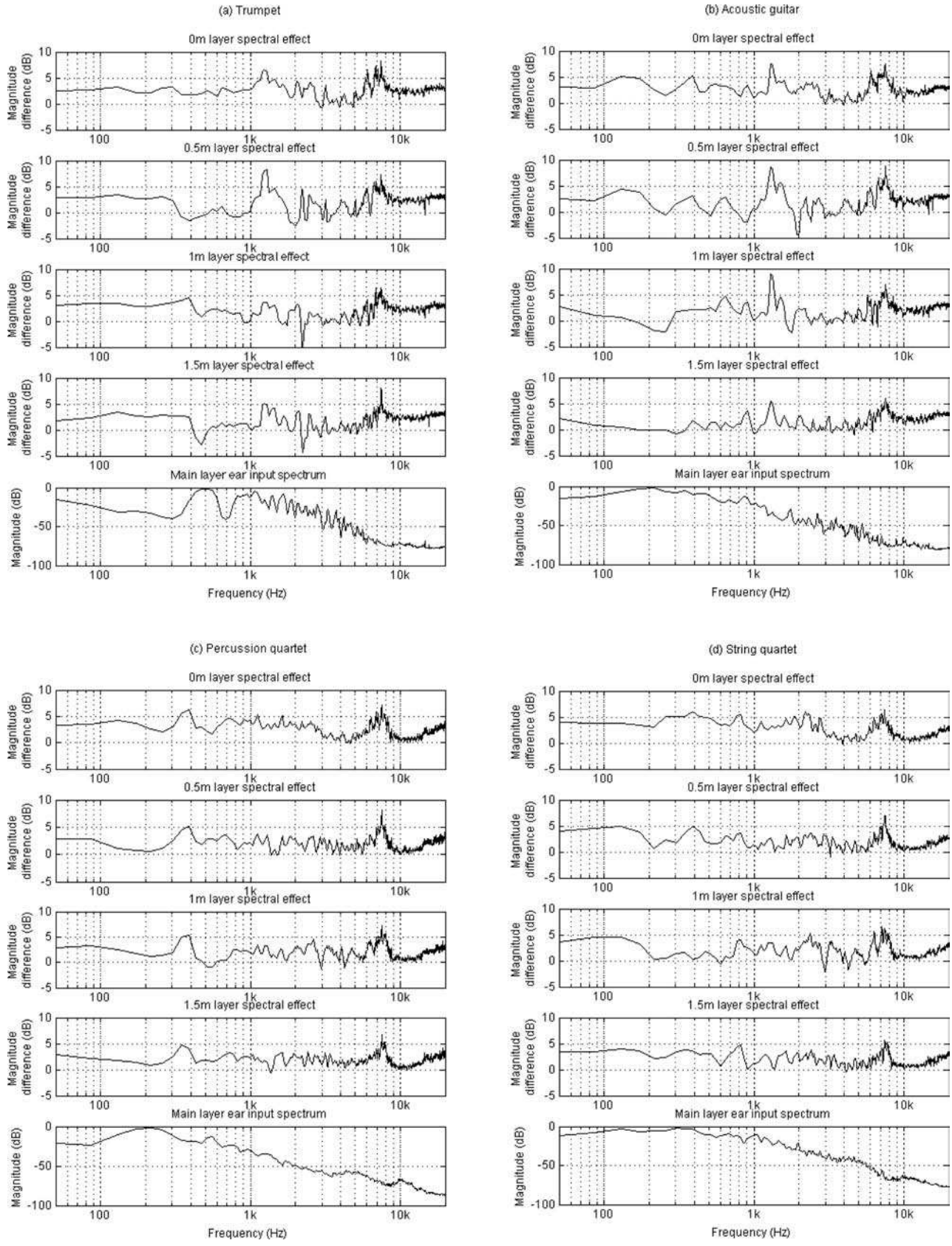

Fig. 13. Spectral magnitude differences for the left ear signal of each listening test stimulus of the main and height layers combined to that of the main layer only; (a) trumpet, (b) acoustic guitar, (c) percussion quartet, (d) string quartet; frequencies from 50 Hz to 20 kHz, 1024-point average FFT.

audibility threshold) was also investigated. The results showed that the localization threshold was between –6 dB and –7 dB for delay times up to 5 ms, whereas the audibility threshold was between –9 dB and –10 dB. Based on this, the ICLD of 7.6 dB for the 0m spacing in the current experiment would have been large enough for the source image to be localized at the base loudspeaker position but not enough to totally suppress potential audible effects caused by the vertical interchannel crosstalk. For the 0.5m, 1m, and 1.5m spacings, on the other hand, the ICLDs were greater than 9 dB (see Table 3) and would have therefore produced no or little perceptual differences. It is considered that if the front main layer microphones had been angled more downwards, thus making ICLDs between the main and height channels greater than, e.g., 9 dB, the 0m spacing might have not been significantly different from the other spacings in perceived spatial impression.

The result that the spaced microphone layers did not have significant differences seems to be associated with the ambient sound component rather than the direct component. It was shown in Fig. 10 that the ambient sound energies of the binaural impulse responses of reproduced sounds (BIRRSs) were almost constant for all spaced layers. Furthermore, the magnitudes and patterns of the spectral changes of ambient sounds that were caused by different layer spacings did not vary much, as shown in Fig. 12(a). However, the vertical interchannel decorrelation of ambient sounds does not appear to directly explain the perceived spatial impression results. It was shown in Fig. 9 that the microphone spacings of 0.5m, 1m, and 1.5m caused little variation in the middle frequency band ICCCs, between the main and height channel signals measured for the ambient sound components. This pattern initially seems to correspond to the perceived results. However, it was also apparent that the low band ICCCs decreased almost linearly as the microphone spacing increased, which does not explain the perceived results. The effectiveness of ICCC on the perceived vertical width change has not yet been fully investigated for musical sources. However, the present authors [15] found from subjective experiments using band-passed pink noise sources that perceived differences between different degrees of ICCC for vertical image spread were relatively small compared to that for horizontal effect. Based on this, it is suggested that the perceptual effect of vertical ICCC between the main and height microphones on the current results was not strong. This also gives rise to the question about whether ICCC would be an effective measure for predicting perceived vertical spatial impression in general, which requires a further investigation.

The results are also discussed in terms of horizontal spatial impression. Fig. 11 showed that the 0m height layer produced the highest $IACC_{E3}$; the difference between the 0m and 0.5m layers was 0.09, and those among the other spacings were in the region of 0.02 to 0.04. This initially seems to suggest that horizontal ASW perceived with the 0m layer would have been narrower than that with a more widely spaced layer. However, considering that the just noticeable difference (JND) of IACC is known to be around 0.075 [24], it is thought that the perceived differences in horizontal ASW caused by the IACC changes were minimal. For the $IACC_{L3}$ results, there was no obvious change observed for different layer spacings, thus suggesting no perceptible horizontal LEV change.

However, it should be noted that the above IACC measures only consider three middle octave frequency bands. Research by Morimoto and Maekawa [25] suggests that the levels of low frequency components of a source signal has an independent effect on the perception of ASW; increases of low frequency levels can cause greater increases of horizontal ASW than those at higher frequencies. This might be related to the dependency of the spatial impression results on the sound source, which can be seen in Fig. 6; the percussion and acoustic guitar sources had more obvious microphone spacing effects than the trumpet and strings. Fig. 13 showed that for the former, the ear input signal produced by the coincident layers had greater spectral magnitudes than that of the spaced layers at frequencies between 100 Hz and 300 Hz, whereas the latter did not show such differences. From this, it can be suggested that the spatial impression results were associated with horizontal ASW as well as vertical ASW, mainly due to the increase in low frequency level in the ear input signals. In addition, it is also suggested that such transient sources as the percussion and guitar also produced stronger LEV than the more continuous trumpet and string sources, since ambient sounds could be more clearly heard between the offset of one sound event and the onset of the following event.

4.2 Preference

Similarly to the spatial impression results, there was no significant difference observed between any of the spaced layers. The 0m height microphone layer was found to be slightly, but significantly preferred to the spaced layers, and this was most obvious between the 0m and 1.5m results for the trumpet and percussion sources. Formal elicitation of preference attributes was not conducted in the present study. However, the subjects were informally asked to comment on the main reasons for their preference judgment after the listening test. Most of them commented on extended height or vertically perceived image spread, but also a number of comments were given on tonal attributes such as clarity and fullness.

As in the discussion provided for the spatial impression results above, the preference results seem to be associated with the level of vertical interchannel crosstalk for each height microphone layer. It can be suggested that the spaced microphone layers had little preference differences since the levels of interchannel crosstalk for all of these layers were below the audibility threshold (see Section 4.1). On the other hand, the crosstalk for the coincident layer was more audible than those for the spaced layers, which might have raised the preference rating. However, it seems to be another important factor for the higher preference rating that the main and crosstalk signals of the coincident layer were summed constructively at the listener's ear without comb-filtering. As shown in Fig. 12, there was no spectral magnitude reduction caused by the addition of the

coincident layer to the main layer across the whole frequency. However, the spaced layers, which had a time delay between the main and crosstalk signals, caused somewhat destructive magnitude changes to the main layer at a number of frequency regions. Therefore, it is thought that if the crosstalk level of each spaced layer had been as high as that of the coincident layer, then there would have been more audible and negative coloration effects, potentially lowering the preference.

In addition, the spectrum with the coincident layer was shaped so that certain frequency regions were emphasized. For example, the coincident height layer for the percussion quartet produced magnitude gains at frequencies between 100 Hz and 300 Hz and also between 1 kHz and 2 kHz, which could have resulted in increases in fullness and clarity, respectively. Similarly, the trumpet source had magnitude gains mainly at frequencies between 400 Hz and 500 Hz and those around 2 kHz when the coincident layer signals were combined with the main layer signals at the ear.

4.3 Practical Implications for a 3D Microphone Array Design

From the discussions above it might be suggested that in practical recording situations where vertical interchannel crosstalk is inevitably present due to the desired angle and polar pattern of height microphone (e.g., upward-facing cardioids as in the current experiment), a vertically coincident 3D main microphone array could be beneficial compared to a vertically spaced array since coincident signals cause no comb-filtering at the ear. The coincident nature of main and height channel signals will also be useful for 3D to 2D downmix applications.

However, a fundamental solution to avoid a tone coloration would be to reject vertical crosstalk by choosing the polar pattern and angle of height microphone appropriately, although in this case the microphone could no longer face directly upwards as recommended in [3]. For example, in a coincident setup, a maximum ICLD between the main and height microphones could be achieved by using a so-called "back-to-back" cardioid configuration, with the microphones' subtended angle being 180°. Alternatively, a figure-of-eight height microphone could be configured in such a way that its null-point faces toward the source so that direct sound could be maximally suppressed. In this case, however, the rear lobe of the height microphone might pick up undesired floor reflections or audience noise.

Additionally, in cases where a vertically spaced array is utilized to achieve greater channel separation (e.g., lower ICCC), it is considered that the omni polar pattern would not be an ideal choice for height channel microphones since it would mainly produce a vertical ICTD rather than an ICLD. It is evident from [12, 13] that the precedence effect does not fully operate in a vertical stereophonic setup; a time delay applied to the height channel does not cause the phantom image to be fully localized at the position of main loudspeaker. The lack of vertical ICLD also means that the level of vertical interchannel crosstalk is not suppressed sufficiently. A delayed vertical crosstalk without any level suppression would cause strong comb-filtering when it is summed with the main channel signal at the listener's ear, which might be perceptually unpleasant.

4.4 Limitations and Further Works

Limitations of the current study and further works are discussed as follows. The direction of the height microphone was an experimental constant in this study; all of the height microphones were angled directly upwards to capture reflections from the same direction. This inherently gave rise to different degrees of interchannel crosstalk in the height channel signals due to the use of the cardioid polar pattern. The D/R energy ratios of the front height channel signals were also relatively high, which means that the influence of ambient sound on the perceived result might have been less dominant than that of the direct sound for those channels as discussed in the above sections. This suggests that the perceived results were mainly source-related. In order to investigate the effect of vertical microphone spacing for environment-related attributes without the influence of interchannel crosstalk, two types of experiments are proposed. First, the microphone setup of the current study will be modified in such a way that the null-points of the height microphones of each layer faces toward the sound source in order to maximally reject the direct sound form the source. Second, a 3D ambience microphone array will be designed and placed beyond the critical distance of a large recording venue, where the D/R ratio is below 1, in order to capture diffused sound mainly. For this, a conventional ambience microphone array called "Hamasaki Square" [7], employing four side-facing figure-of-eight microphones arranged in a square formation, will be augmented with four additional height channel microphones placed at different spacings from it.

Since the scope of the current study was a 3D main microphone array design, the height channel microphones were placed within a relatively small range of vertical spacing between 0m and 1.5m. In recording venues with high ceilings, however, some recording engineers might place height microphones at a large vertical distance from the main microphones for a maximum vertical channel separation. In order to test the effectiveness of this approach, a future experiment will include a wider range of height microphone spacings (e.g., microphones placed beyond the vertical critical distance).

The current experiment used only solo instruments and small ensembles as sound sources. Sound sources for future experiments will include large scale orchestra recordings since the use of such a horizontally wide ensemble might produce different results to the current results for the following reason. In the current experimental setup using upward-facing cardioid height microphones, as the distance between the array and sound source became larger, the differences in ICLD between the main and height microphone signals for different vertical layer spacings would become smaller. Consequently, there might be less perceived difference between different layer spacings. In addition, since it

is a vertically long instrument, the organ is considered to be a useful sound source for future 3D recording experiments. Height channel microphones should be configured so that the vertical spread of the original source image could be represented effectively.

5 SUMMARY AND CONCLUSIONS

The present study investigated the effect of spacing between main and height channel microphone layers on perceived spatial impression and preference in the context of a 3D main microphone array. Multichannel room impulse responses, as well as string and percussion quartets, were recorded in a concert hall using a nine-channel microphone array. A five-channel main array was vertically augmented, with four upward-facing cardioid microphones placed directly above the front left, front right, rear left, and rear right microphones. The spacings between the main and height microphones tested were 0m, 0.5m, 1m, and 1.5m. Impulse responses of each position were convolved with anechoic trumpet and acoustic guitar signals. Listening tests were conducted on perceived spatial impression and preference in a dry listening room using a nine-channel loudspeaker setup. The recorded impulse responses were analyzed for their signal energies, interchannel level differences (ICLDs), and interchannel cross-correlation coefficients (ICCCs). Binaural recordings of the test stimuli were also made at the listening position. The energies and interaural cross-correlation coefficients (IACCs) of the ear signals were measured. The magnitudes of spectral changes caused by the addition of each height microphone layer to the main layer were also investigated.

The listening test results were statistically analyzed and discussed together with the physical measurement results. It was shown that the layer spacings of 0.5m, 1m, and 1.5m did not produce significant differences in perceived spatial impression. The 0m layer was found to be slightly greater than or similar to the spaced layers depending on the type of source. These results were explained from a viewpoint of the perceptual effect of vertical interchannel crosstalk (direct sounds in the height channel signals); the ICLDs between the main and 0m height layer signals were not large enough to completely suppress the potential effects of crosstalk, whereas those for the spaced pairs were sufficiently large, thus no or little audible effects produced by crosstalk. The levels of ambient sounds analyzed for main and height microphone signals as well as those for ear input signals were found to be almost constant. IACCs measured for the ambient part of the ear input signals were also found to be similar for different layer spacings. These results suggested that vertical microphone layer spacing had little effect on the perception of environment-related spatial impression. Additionally, ICCCs measured for vertical and diagonal microphone pairs did not seem to explain the perceived results directly.

The preference results showed similar patterns to the spatial impression results overall; there was no significant difference between spaced layers, whereas the coincident layer was slightly preferred to the spaced layers depending on sound source. Informal comments collected from the subjects suggested that the main preference attributes were tonal quality as well as spatial quality. The results were discussed in relation to the delta spectrum measurements, which showed that the addition of the coincident microphone layer to the main layer had a positive spectral influence on the ear input signal, whereas that of a spaced layer caused substantial comb filtering effects in the resulting spectrum.

6 ACKNOWLEDGMENT

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC), UK, Grant Ref. EP/L019906/1. The authors thank the music technology students and staff at the University of Huddersfield who participated in the listening tests and Dr. Francis Rumsey for his advice on statistical analysis. They are also grateful to the editor and two anonymous reviewers of this paper for their insightful and constructive comments, which greatly helped improve the manuscript.

7 REFERENCES

[1] ITU-R, "Recommendations BS.2159-4: Multichannel Sound Technology in Home and Broadcasting Applications," *International Telecommunications Union* (2012).

[2] B. V. Daele and W. V. Baelen, Productions in Auro-3D, URL: http://www.auro-3d.com/professional/technical-docs/ 2012.

[3] G. Theile and H. Wittek, "Principles in Surround Recordings with Height," *presented at the 130th Convention of the Audio Engineering Society* (2011 May), convention paper 8403.

[4] M. Williams, "The Psychoacoustic Testing of the 3D Multiformat Microphone Array Design, and the Basic Isosceles Triangle Structure of the Array and the Loudspeaker Reproduction Configuration," *presented at the 134th Convention of the Audio Engineering Society* (2013 May), convention paper 8839.

[5] P. Geluso. "Capturing Height: The Addition of Z Microphones to Stereo and Surround Microphone Arrays," *presented at the 132nd Convention of the Audio Engineering Society* (2012 Apr.), convention paper 8595.

[6] S. P. Lipshitz, "Stereo Microphone Techniques: Are the Purists Wrong?" *J. Audio Eng. Soc.*, vol. 34, pp. 717–743 (1986 Sep.).

[7] K. Hamasaki, "Multichannel Recording Techniques for Reproducing Adequate Spatial Impression," *Proc. of the Audio Engineering Society 24th International Conference: Multichannel Audio—The New Reality* (2003 June).

[8] F. Rumsey and W. Lewis, "Effect of Rear Microphone Spacing on Spatial Impression for Omnidirectional Surround Sound Microphone Arrays," *presented at the 112th Convention of the Audio Engineering Society* (2002 Apr.), convention paper 5563.

[9] S. K. Roffler and R. A. Buttler, "Factors that Influence the Localization of Sound in the Vertical Plane," *J. Acoust. Soc. Am.*, vol. 43 , no. 6, pp. 1255–1259 (1968).

[10] J. Blauert, *Spatial Hearing*, rev. ed. (MIT Press, Cambridge, MA, 1997).

[11] J. L. Barbour, "Elevation Perception: Phantom Images in the Vertical Hemisphere," *Proc. of the Audio Engineering Society 24th International Conference: Multichannel Audio—The New Reality* (2003 June).

[12] H. Lee, "The Relationship between Interchannel Time and Level Differences in Vertical Sound Localization and Masking," *presented at the 131st Convention of the Audio Engineering Society* (2011 Oct.), convention paper 8556.

[13] R. Wallis and H. Lee, "Investigation into Vertical Stereophonic Localization in the Presence of Interchannel Crosstalk," *presented at the 136th Convention of the Audio Engineering Society* (2014 Apr.), convention paper 9026.

[14] F. Wendt, M. Florian, and F. Zotter, "Amplitude Panning with Height on 2, 3, and 4 Loudspeakers," *Proc. of the 2nd International Conference on Spatial Audio* (2014).

[15] C. Gribben and H. Lee, "The Perceptual Effects of Horizontal and Vertical Interchannel Decorrelation Using the Lauridsen Decorrelator," *presented at the 136th Convention of the Audio Engineering Society* (2014 Apr.), convention paper 9027.

[16] A. Farina, "Advancements in Impulse Response Measurements by Sine Sweeps," *presented at the 122nd Convention of the Audio Engineering Society* (2007 May), convention paper 7121.

[17] H. Lee, "A New Microphone Technique for Effective Perspective Control," *presented at the 130th Convention of the Audio Engineering Society* (2011 May), convention paper 8337.

[18] V. Hansen, and G. Munch, "Making Recordings for Simulation Tests in the Archimedes Project," *J. Audio Eng. Soc.*, vol. 39, pp. 768–774 (1991 Oct.).

[19] T. Hidaka, L. Beranek, and T. Okano "Interaural Cross-Correlation Lateral Fraction, and Low- and High-Frequency Sound Levels as Measures of Acoustical Quality in Concert Halls," *J. Acoust. Soc. Am.*, vol. 98, pp. 988–1007 (1995).

[20] ITU-R, "Recommendations ITU-R BS.1116: Methods for the Subjective Assessment of Small Impairments in Audio Systems including Multichannel Sound Systems," International Telecommunications Union (1994).

[21] A. W. Bronkhorst and T. Houtgast, "Auditory Distance Perception in Rooms," *Nature*, vol. 397, pp. 517–520 (1999).

[22] B. Gardner and K. Martin, URL: http://sound.media.mit.edu/resources/KEMAR.html. 2000.

[23] H. Lee and F. Rumsey, "Investigation into the Effect of Interchannel Crosstalk in Multichannel Microphone Technique," *presented at the 118th Convention of the Audio Engineering Society* (2005 May), convention paper 6374.

[24] T. J. Cox, W. J. Davies, and Y. W. Lam, "The Sensitivity of Listeners to Early Sound Field Changes in Auditoria," *Acoustica*, vol. 79, pp. 27–41 (1993).

[25] M. Morimoto and Z. Maekawa, "Effects of Low Frequency Components on Auditory Spaciousness," *Acustica*, vol. 66, pp. 190–196 (1988).

THE AUTHORS

Hyunkook Lee

Hyunkook Lee received a B.Mus. degree in music and sound recording (Tonmeister) from the University of Surrey, Guildford, UK, in 2002, and his Ph.D. degree in audio engineering and psychoacoustics from the Institute of Sound Recording (IoSR) at the same University in 2006. From 2006 to 2010, Dr. Lee was Senior Research Engineer in audio R&D at LG Electronics, South Korea. Currently, he is Senior Lecturer in music technology and the leader of the Applied Psychoacoustics Lab (APL) at the University of Huddersfield, UK. Dr. Lee has also been a freelance recording engineer since 2002. His current research includes spatial audio psychoacoustics, 3D sound recording and image rendering techniques, and interactive virtual acoustics. He is an active member of the Audio Engineering Society and a fellow of the Higher Education Academy, UK.

Christopher Gribben

●

Christopher Gribben is a Ph.D. student with the Applied Psychoacoustics Lab (APL) at the University of Huddersfield, UK. He graduated from the University's music technology and audio systems course with a First Class degree in 2013. During his undergraduate studies, Gribben undertook an industrial placement year working at Cass Allen Associates, UK, where his focus was on environmental acoustics and noise modelling. At present, his research concerns the psychoacoustics of 3D multichannel audio, in particular, methods to assess and predict the quality of 3D audio. He is a student member of the Audio Engineering Society.