OPEN ACCESS Freely available online

C. Gribben and H. Lee, "The Frequency and Loudspeaker-Azimuth Dependencies of Vertical Interchannel Decorrelation on the Vertical Spread of an Auditory Image," *J. Audio Eng. Soc.*, vol. 66, no. 7/8, pp. 537–555, (2018 July/August.).

DOI: https://doi.org/10.17743/jaes.2018.0040

The Frequency and Loudspeaker-Azimuth Dependencies of Vertical Interchannel Decorrelation on the Vertical Spread of an Auditory Image

CHRISTOPHER GRIBBEN, *AES Member,* **AND HYUNKOOK LEE,** *AES Member*

(christopher.gribben@meridian.co.uk) (h.lee@hud.ac.uk)

Applied Psychoacoustics Laboratory (APL), University of Huddersfield, Huddersfield, HD1 3DH, UK

In horizontal stereophony, it is known that interchannel correlation relates to the horizontal spread of a phantom auditory image. However, little is known about the perceptual effect of interchannel correlation on vertical image spread (VIS) between two vertically-arranged loudspeakers. The present study investigates this through two subjective experiments: (i) a multiple comparison of relative VIS for stimuli with varying degrees of correlation; and (ii) the absolute measurement of upper and lower VIS boundaries for extreme stimuli conditions. Octave-band (center frequencies: 63 Hz to 16 kHz) and broadband pink noise signals have been decorrelated using two techniques: all-pass filtering and complementary comb-filtering. These stimuli were presented from vertically-spaced loudspeaker pairs at three azimuth angles (0°, ±30°, and ±110°), with each angle assessed discretely. Both the relative and absolute test results show no significant effect of vertical correlation on VIS for the 63 Hz, 125 Hz, and 250 Hz bands. For the 500 Hz band and above, there is a general tendency for VIS to increase as correlation decreases, which is observed for both decorrelation methods. This association is strongest at 0° azimuth for the 500 Hz and 1 kHz bands; at ±30° for 8 kHz and Broadband; and at ±110° for 2 kHz, 4 kHz, and 16 kHz. The 8 kHz band at ±30° has the strongest association of all conditions—post-hoc objective analysis indicates a potential relationship between HRTF localization cues (pinna filtering) and VIS perception within this frequency region. Furthermore, the absolute test results suggest that changes of VIS from interchannel decorrelation are fairly slight, with only the Broadband and 16 kHz bands showing a significant increase. The deviations of boundary scores also suggest a difficulty grading absolute VIS and/or potential disagreements among listeners.

0 INTRODUCTION

Over recent years, three-dimensional (3D) loudspeaker reproduction has been of much interest in the audio industry. Commercial 3D audio formats often see the inclusion of height-channel loudspeakers positioned above a main-layer of loudspeakers—examples of such systems include Auro-3D [1], Dolby Atmos [2], and DTS:X [3]. Given the upward extension of the sound field from established two-dimensional (2D) formats (e.g., 5.1 and 7.1 surround), it is necessary to conduct fundamental investigations into the perception of interchannel signal relationships in the vertical domain. From this, the present study observes the perceptual effect of signal correlation between vertically-spaced pairs of loudspeakers.

"Decorrelation" has been employed as a means of controlling the interchannel cross-correlation (ICC) (similarity) between two loudspeaker signals. The process of decorrelation is typically applied to a monophonic input signal, with the output consisting of two uncorrelated signals that sound sonically similar to the input, yet have different phase and/or amplitude relationships between the two signals.

When two correlated signals are presented between a stereophonic pair of left-right loudspeakers, a decrease of ICC increases the horizontal image spread (HIS) of the phantom auditory image [4, 5]. That is, a "pseudo-stereophonic" or image widening effect can be achieved when applying interchannel decorrelation along the horizontal plane [6]. This perceptual effect is due to a direct relationship between ICC and the interaural cross-correlation (IAC), where IAC is known to contribute to the perception of an auditory event's apparent source width (ASW), as dictated by lateral reflections within an enclosed space [7].

Decorrelation has also been applied in practical applications such as upmixing (e.g., one-to-two channel or two-to-five channel upmixing) [8–13]. Upmixing algorithms typically feature an ambience extraction stage, followed by interchannel decorrelation to synthesize additional ambient signals; however, it is not yet clear whether decorrelation could still be an effective method for vertical upmixing from 2D to 3D. It is therefore important to gain a better understanding of how decorrelation is perceived in the vertical domain.

A previous study conducted by the present authors suggested that a relationship between vertical ICC and vertical image spread (VIS) may exist [4]. Band-limited pink noise stimuli with varying degrees of ICC were presented between a vertically-spaced pair of loudspeakers in the median plane (with loudspeaker elevations of 0° and +30° to the listening position). Three frequency bands were assessed: "Low" (63–250 Hz octave-bands), "Middle" (0.5–2 kHz octave-bands), and "High" (4–16 kHz octave-bands). Results for each band showed a trend of VIS increasing as ICC decreased, which was most notable for the "Middle" band. Despite this, it is difficult to reveal or conclude the perceptual cues that gave rise to these results, since the frequency bandwidths were reasonably broad (tri-octave) and only one decorrelation approach was assessed (the amplitude-based complementary comb-filtering [6, 14, 15], as described in Sec. 1.2 below). Furthermore, given that the test was performed in the median plane, there was no consideration of vertical decorrelation off-axis, looking at how interaural differences might affect the perception of VIS.

From the above background, the current study examines the effect of vertical ICC on VIS further, with a particular focus on the frequency and loudspeaker-azimuth dependencies of VIS. Decorrelated octave-band and broadband pink noise stimuli have been presented from vertically-spaced loudspeaker pairs at three discrete azimuth angles around the listener (0°, ±30°, and ±110°)—as based on the Auro-3D 9.1 loudspeaker format, with the addition of a vertical pair in the median plane [1]. It is widely known that spatial hearing along the median plane is largely governed by spectral cues, rather than localization cues from interaural differences [16, 17]. Considering this, it is hypothesized that if vertical decorrelation gives rise to the perception of VIS in the median position (0° azimuth), then the associated perceptual mechanism would be frequency-dependent. On the other hand, since commercial 3D audio systems feature vertically-spaced loudspeaker sources off-center, it is also of interest to investigate a potential interaural influence on VIS perception, in addition to spectral cues.

In the first experiment of this study a multiple-comparison test was performed to grade the relative VIS among vertically decorrelated, correlated, and monophonic stimuli. Three degrees of decorrelation were assessed for two decorrelation methods, so that any relationship between ICC and VIS could be clearly observed. Perceptual evaluations of different audio systems or processing techniques in laboratory conditions are often conducted in a multiple comparison manner. This allows one to examine relative differences among different stimuli and their magnitudes more easily. However, in practical situations the listener would have no comparison to make and judge the perceived quality in an absolute sense. In order to gain an insight into how VIS would be perceived in an absolute sense, the second experiment measured the upper and lower boundaries of the perceived phantom image for extreme stimuli conditions from the first experiment. This approach reveals not only the overall magnitude of VIS but also the actual position of the perceived image in space. The combination of these two tests allows us to observe whether changes to VIS by vertical decorrelation are perceivable and significant at octave-band and broadband level. Further to these subjective experiments, the stimuli signals have been binauralized so that objective analysis of the ear input signals can be conducted.

The rest of this paper is organized as follows. Sec. 1 reviews the two decorrelation methods used in the present study: phase randomization and complementary comb-filtering. The first and second experiments are described in Secs. 2 and 3, respectively. Following this, objective analysis of the binauralized stimuli and binaural room impulse responses (BRIRs) is conducted in Sec. 4. Practical implications of the results from the experiments are then discussed in Sec. 5, with the conclusions of the study detailed in Sec. 6.

1 DECORRELATION METHODS USED IN THE STUDY

Conventional decorrelation methods can broadly be split into phase-based and amplitude-based techniques, with a summary of each provided in Secs. 1.1 and 1.2, respectively. The current study utilizes one phased-based technique proposed by Kendall [18] and one amplitude-based technique discovered by Lauridsen [14]. Through comparing both of these approaches side-by-side using the same controlled levels of ICC, the results should provide some indication whether a general relationship between vertical ICC and VIS perception exists. Furthermore, to assess the effect of vertical ICC on VIS, it is required that the decorrelation techniques used must achieve suitably low ICC coefficient (ICCC) values across all octave-bands, and that the degree of correlation is easily controlled to achieve the desired level of ICC.

The ICCC between two loudspeaker signals (s_1 and s_2) is calculated as the maximum absolute value of the interchannel cross-correlation function (ICCF) (Eqs. (1) and (2)) [5]— a lag time (τ) can be set with the ICCF to compensate for any delay between the signals. In the present study zero (0) lag is used as all signals are time-aligned, and ICCC is calculated as the average of running ICCCs ($ICCC_{avg}$) for 50 ms windowed signals.

$$ICCF\ (\tau) = \frac{\int_{-\infty}^{\infty} s_1(t)\, s_2(t+\tau)\, dt}{\sqrt{\left[\int_{-\infty}^{\infty} s_1^2(t)\, dt\right]\left[\int_{-\infty}^{\infty} s_2^2(t)\, dt\right]}} \tag{1}$$

$$ICCC = max\,|ICCF(\tau)| \tag{2}$$

1.1 Phase-Based Decorrelation

Considering phase-based decorrelation first, Kendall [18] proposes the use of all-pass filters to randomize the phase of frequencies within a signal, while maintaining the frequency magnitudes between the input and output. For this technique an original monophonic signal can simply be convolved with two impulses (short white noise bursts) of random phase and unit magnitude (Eq. (3)).

$$s_1(n) = x(n)^* h_1(n) \quad s_2(n) = x(n)^* h_2(n) \tag{3}$$

where s_1 and s_2 are the two output signals, x is the monophonic input signal, and h_1 and h_2 are the two FIR filter impulses (white noise bursts).

To create the FIR all-pass filters, two random number sequences are generated as the filter phase coefficients in the frequency domain (featuring random values between $-\pi$ and π), giving an inherent decorrelation between the two filters—the degree of this correlation can then be controlled by a mixing matrix, as seen in Eq. (4) below. It is thought that the correlation between the two impulses directly relates to the correlation between the two outputs; however, given the random generation of the number sequences, the actual degree of maximum ICC can vary drastically between each creation of filters, often requiring repetition until the desired ICC level is achieved.

$$h_1[n] = a[n] \quad h_2[n] = \frac{1}{1+k}\left(a[n] + (b[n] \cdot k)\right) \tag{4}$$

where h_1 and h_2 are the two random filter phase coefficients after mixing, a is the first random number sequence, b is the second random number sequence, and k is the mixing factor between 0 and 1, where 1 is maximum decorrelation.

It is assumed that the use of all-pass filter decorrelation should have little effect on the frequency response of the signals, however in practice, the length of the filter (white noise burst) can cause smearing of transient information [19]. In general, the longer the filter length, the greater or easier the decorrelation, yet at the expense of increased signal coloration. The waveform of the output can also be distorted by significant opposing phase-shifts of neighboring frequencies [20]. It is this random interaction of phase-shifts that leads one implementation of the all-pass filter to sound noticeably different from another.

Although it is relatively simple to achieve a low level of ICC with all-pass filters, the random element requires a trial-and-error approach in terms of both coloration and ICC, which may not be suitable for practical applications. Various suggestions have been made to improve phase-based approaches on coloration and transient handling—these include the use of exponentially decaying white noise bursts [21, 22], the random time-shifting of whole critical frequency bands [20, 23], and extraction of transient information [19]—however, the added complexity can make achieving low levels of ICC more difficult. Since the present study has a focus on controlling a broad range of $ICCC_{avg}$ values with continuous pink noise stimuli (i.e., limited transients and little consideration towards coloration), it was deemed that the original approach proposed by Kendall would be best suited for this purpose. Some existing upmixing models also indicate the generic use of all-pass filters to decorrelate signals by phase randomization [9, 10].

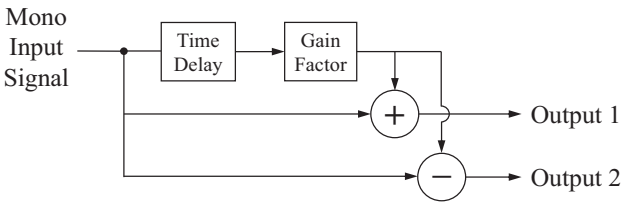

Fig. 1. Structure of the Complementary Comb-Filter decorrelator.

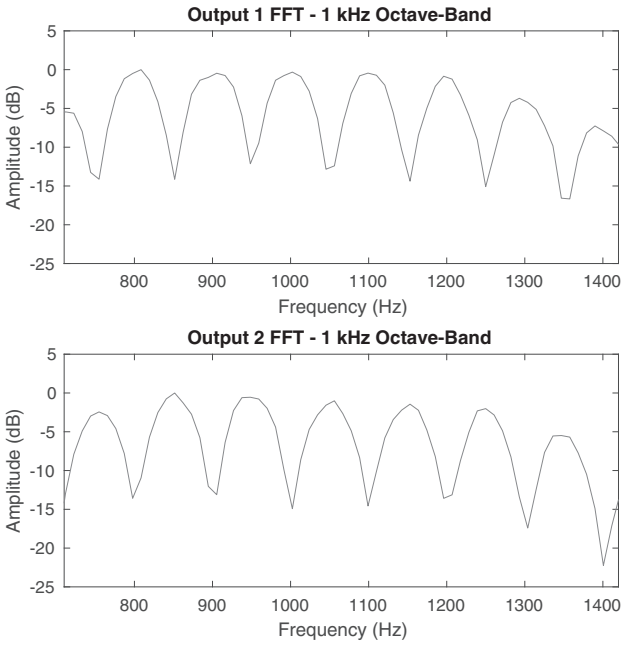

Fig. 2. FFT plots of the two output signals from the Complementary Comb-Filter decorrelator (4096 FFT-points) (1 kHz octave-band with a 10 ms time-delay and gain factor of 1.0).

1.2 Spectral-Amplitude-Based Decorrelation

Looking to amplitude-based methods, the simplest of these is in the form of frequency panning, where groups of frequencies are alternately panned between the output channels at regular intervals across the spectrum. Lauridsen [14] first discovered an effect whereby summing and subtracting a signal with a delayed version of itself created two comb-filtered signals that had opposing spectral amplitude differences. It was found that this amplitude decorrelation generated a "pseudo-stereophonic" impression of width when reproduced simultaneously from spaced positions. The structure of the Lauridsen decorrelation process can be seen in Fig. 1 below, where the monophonic input signal is time-delayed (T) and multiplied by a gain factor (G), before the summation and subtraction with the original signal.

The method was investigated further by Schroeder [6] and has since been termed "complementary comb-filtering" (CF) [15]. An example of the resulting complementary amplitude differences can be seen in Fig. 2 below, where

groups of frequencies are regularly panned between the two outputs for a 1 kHz octave-band. It is very simple and cost-effective to implement, and the degree of ICC is easily controlled by the gain factor applied to the delayed signal (where a gain factor of 1.0 is maximum decorrelation). Many other amplitude-based methods also work on a similar principle of frequency distribution between two channels [5, 24, 25]; however, suitably low levels of ICC can easily be realized with complementary comb-filtering, making it an appropriate technique to assess in this study. Furthermore, the method is featured in some proposed upmixing algorithms [8, 11], and was also the approach used in the authors' previous experiments [4].

2 EXPERIMENT ONE: RELATIVE GRADING OF VERTICAL IMAGE SPREAD (VIS)

The first subjective experiment looks to establish the relative perception of vertical image spread (VIS) between vertically decorrelated and correlated pink noise stimuli. Two decorrelation methods (as described in Sec. 1 above) were employed to control the degree of signal correlation (interchannel cross-correlation (ICC)) between pairs of vertically-spaced loudspeakers, testing average running ICC coefficients ($ICCC_{avg}$) ranging from 1.0 to 0.1 (calculated using Eqs. (1) and (2), with zero lag and 50 ms windowing). Stimuli were created by decorrelating octave-band and broadband pink noise, in order to observe the frequency-dependency of ICC on VIS—these stimuli were then presented at three different azimuth angles to the listening position (0°, ±30°, and ±110°). The results from testing also informed the second subjective experiment of this study, where absolute VIS was measured to examine changes of the upper and lower image boundaries.

2.1 Experimental Design

2.1.1 Physical Testing Setup

Ten Genelec 8040A loudspeakers (frequency response: 48 Hz – 20 kHz (±2 dB)) were used during testing, divided into two separate layers (main and height) with five loudspeakers in each. The five main-layer loudspeakers were spaced around the listener at azimuth angles of 0°, ±30°, and ±110°, in accordance with ITU-R BS.775-3 [26]. Each main-layer loudspeaker was positioned 2 m from the listening position, with the acoustic center at a height of 1.27 m and in line with the ear position. The height-layer loudspeakers were positioned directly above the main-layer azimuth positions at an elevation angle of +30° to the listener, as per Auro-3D 9.1 (with an additional center height) [1]. This resulted in five vertically-spaced loudspeaker pairs around the listener with a spacing of 1.15 m between the two layers (see Fig. 3).

Listening tests were conducted at the University of Huddersfield in a critical listening room that fulfils the specification of ITU-R BS.1116-3 [27] (6.2m × 5.6m × 3.8m; RT = 0.25 s; NR 12). Time and level alignment were applied between the two loudspeaker layers in order to compensate for interlayer difference of signal arrival at the listening position. An acoustically transparent curtain was also used to obscure the loudspeakers from view, so as to avoid visual bias during testing.

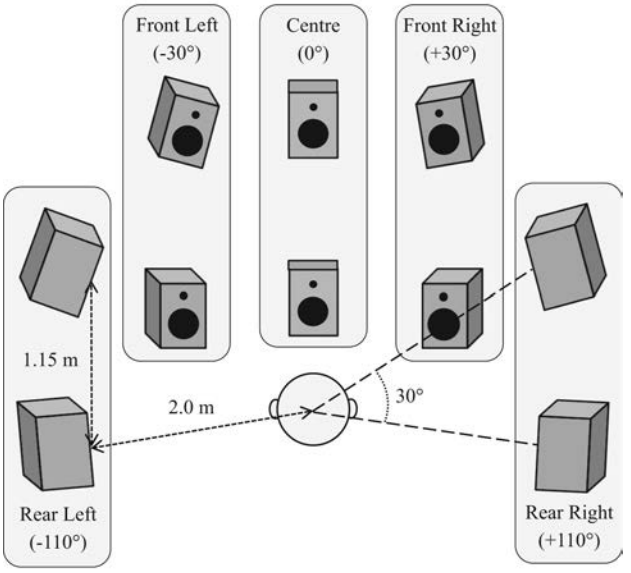

Fig. 3. Physical loudspeaker setup used during testing (based on Auro-3D 9.1 [1] with an additional Center height-channel). Five main-layer loudspeakers positioned 2 m from the listener at ear height with azimuth angles of 0°, ±30°, and ±110°. Five upper height-layer loudspeakers elevated directly above its main-layer pair by +30° to the listener.

2.1.2 Stimuli Creation

For stimuli creation, a continuous broadband pink noise sample was filtered into nine octave-bands, with center frequencies of 63 Hz to 16 kHz, using 16th-order linear phase Butterworth filters (96 dB/octave roll-off). To generate the stimuli, each pink noise octave-band and the original broadband sample were processed using the two decorrelation techniques reviewed in Sec. 1: Complementary Comb-Filtering (CF) and Phase Randomization by use of all-pass filters (PR). Although more sophisticated methods of decorrelation have been proposed in recent years to improve tonal quality (as mentioned in Sec. 1), the two approaches selected for this study allow for a simple controlled assessment of the ICC effect due to fewer parametric variables. Furthermore, both of the methods are able to achieve an $ICCC_{avg}$ of at least 0.1 for each of the octave-bands under testing (63 Hz – 16 kHz) (calculated using Eqs. (1) and (2), with zero lag and 50 ms windowing).

2.1.2.1 Phase-Randomization Stimuli (All-Pass Filtering) The two random number sequences used for creating the PR stimuli were 30 ms in length (1323 numbers at a sampling rate of 44.1 kHz). In theory, an all-pass filter method should not affect a signal's spectral response, since the frequency magnitudes remain constant throughout. However, it is found that a longer filter length can cause undesirable smearing of transients [19]—this is not of particular concern in the present study, though, as the assessment has been conducted on continuous noise sources. In the present

study, a length of 30 ms for the random number sequences was found to easily decorrelate the 63 Hz octave-band to $ICCC_{avg}$ 0.1, where the lowest frequency (44 Hz) has a cycle length of around 23 ms.

2.1.2.2 Complementary Comb-Filtering Stimuli As detailed in Sec. 1.2, the complementary comb-filtering technique features two parameters: time-delay and gain factor. The gain factor controls the degree of correlation, while the time-delay affects the comb-filter notch depth and the bandwidth between notches. The present study employs a time-delay of 10 ms, which was shown to generate an $ICCC_{avg}$ of 0.1 for the 63 Hz octave-band—it provides a maximum notch depth (interchannel level difference) of ∼12 dB for a 1.0 gain factor, with a relatively narrow notch bandwidth (100 Hz). This time-delay has also been suggested by other researchers as a compromise that gives the desired perception of widening, while avoiding any confusion that may be experienced with longer time-delays [8]. Furthermore, the 10 ms condition demonstrated a significant effect on VIS perception for the Middle and High frequency bands in the authors' previous vertical decorrelation research (described in Sec. 0) [4].

2.1.2.3 Stimuli Conditions For the techniques detailed above, three $ICCC_{avg}$ levels of decorrelation were generated with each (as calculated using Eqs. (1) and (2)): "0.1," "0.4," and "0.7." Additionally, $ICCC_{avg}$ "1.0" (fully correlated) and a monophonic signal routed to the main-layer loudspeaker only were also included as stimuli conditions. The monophonic stimuli were included with an interest to see whether decorrelation would actually increase VIS compared to the original condition in a vertical upmixing scenario. This resulted in eight stimuli for each of the ten frequency conditions (nine octave-band and one broadband). Both methods were implemented using MATLAB looping scripts that repeated the decorrelation process until the desired $ICCC_{avg}$ was achieved; in the case of Complementary Comb-Filtering the gain factor was incremented with each loop, and for Phase Randomization new random number sequences were generated each time (with the mixing matrix set to a reasonable level). Each of the two decorrelated output signals were RMS level-matched with the input signal to maintain an equal balance of energy between the channels. Of the two-channel decorrelated outputs, Output 1 was routed to the main-layer loudspeaker and Output 2 to the height-layer loudspeaker of a given vertical pair.

The level of each frequency band stimulus was determined by octave-band filtering a broadband pink noise signal at 75 dB LAeq, with each of the eight stimuli matched to the respective sound pressure level (LAeq) of the resultant bands. This ranged from 49 dBA for the 63 Hz octave-band to 68 dBA for 1 kHz and 2 kHz octave-band, with the Broadband stimuli level-matched to 75 dBA. Rather than matching all octave-bands to the same LAeq for loudness compensation, it was considered that maintaining the original inter-band energy relationship (equal energy per octave-band) would be more appropriate. The motivation here was to examine the effectiveness of decorrelation for each octave-band, while maintaining its inherent loudness within a broadband signal. This would also be more representative of a potential practical application, where some octave-bands are selectively decorrelated for vertical upmixing, without changing the spectral energy weighting of the original source signal.

2.1.3 Subjects

Twelve subjects took part in the first experiment of this study, comprising staff, postgraduate students, and final year students from the University of Huddersfield's Music Technology courses. All participants reported to have normal hearing and were familiar with critical listening exercises in spatial audio.

2.1.4 Testing Procedure

During the test, subjects were presented with a total of 30 multiple comparison trials in a graphical user interface (GUI), developed by the authors using Cycling '74's Max 7 (named HULTI-GEN [28]). Trials were based on an adapted MUSHRA format [29] with a continuous bipolar scale (ranging from –30 to 30), featuring visual markers every 10 points to help maintain consistency between trials. Each trial consisted of eight buttons and sliders to trigger and grade the eight stimuli, with another button next to the center of the scale (0) to trigger a reference stimulus. Six of the buttons/sliders corresponded to the decorrelated stimuli, with the other two controlling the correlated and monophonic stimuli. The reference signal was the correlated sample ($ICCC_{avg}$ 1.0), which was also included as a hidden reference among the stimuli. $ICCC_{avg}$ 1.0 was chosen as reference over the monophonic stimulus since the primary aim of the study was to observe the perceptual effect of interchannel decorrelation between vertically-arranged loudspeakers.

Each stimulus was to be graded for vertical image spread (VIS) against each other and the reference—where above the reference (at 0 on the scale) was labelled as greater and below the reference as lesser. VIS was described as the overall vertical extent of the auditory phantom image, with subjects instructed to focus on the VIS and ignore any vertical shifts of the image. The 30 trials were made up of the 10 frequency bands for 3 loudspeaker azimuth angle positions (0°, ±30°, and ±110°, see Fig. 3 above). Only one of the two 30° and 110° vertical loudspeaker pairs were tested for each band to limit the testing load—this was randomized between left and right for each 30°/110° trial. With every subject, the order of the 30 trials was also randomized and divided into 3 separate sessions, each of which took no more than 30 minutes for a listener to complete. Subjects were instructed to face forward and keep their head still throughout testing, which was ensured by the aid of a small headrest for the listener.

2.2 Results and Analysis

Results for the first experiment are presented in Fig. 4 below—all data has been normalized in accordance with ITU-R BS.1116-3 [27] and analyzed in SPSS. The graphs display the median scores of relative VIS with bars to

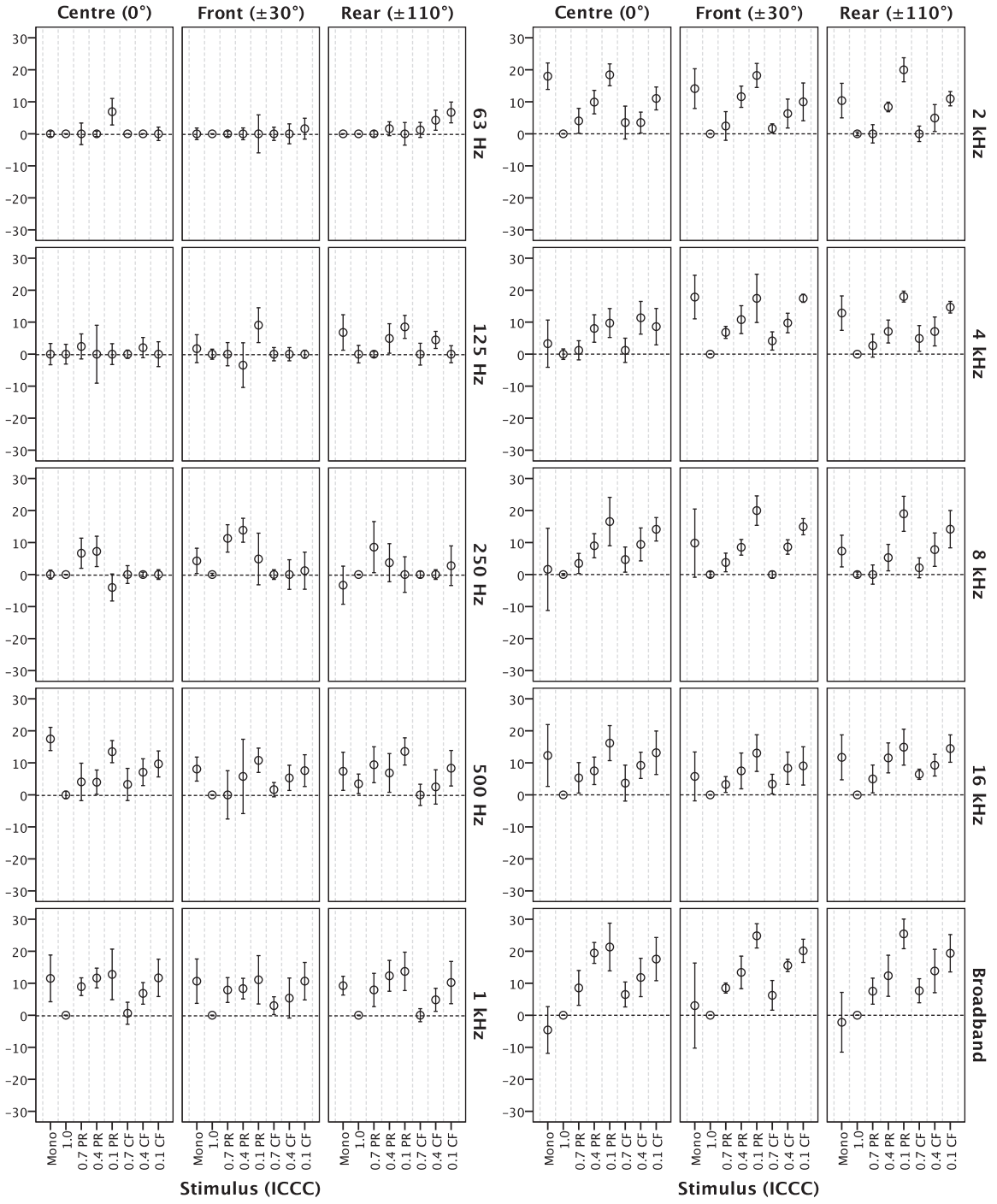

Fig. 4. Median Vertical Image Spread (VIS) normalized scores with 95% confidence notch edge bars. Each panel features the results of all average running interchannel cross-correlation coefficient ($ICCC_{avg}$) and monophonic conditions for a given multiple comparison trial (loudspeaker position × frequency band).

signify notch edges, representing non-parametric 95% confidence intervals [30]. Shapiro-Wilk tests for normality indicated that the data of each condition was not always normally distributed; therefore, non-parametric statistical tests were performed across all conditions for consistency and comparison.

Data was first analyzed for significant difference between the two decorrelation methods, using Bonferroni-corrected Wilcoxon signed-rank tests to compare the two methods at each of the three interchannel cross-correlation coefficient ($ICCC_{avg}$) levels. Friedman repeated measure tests were then conducted to assess the effect of ICC on VIS for either the methods combined, if no difference was established, or each method independently. The monophonic stimulus was not included in the Friedman testing, as it does not feature an $ICCC_{avg}$ value. Instead, Wilcoxon signed-rank tests with Bonferroni correction were carried out between the monophonic stimulus and all other stimuli to observe

significant differences, with the median and notch data of the monophonic stimuli also presented in the graphs for comparative purposes.

2.2.1 Comparing Decorrelation Methods

The Wilcoxon signed-rank test results indicated that there was little significant difference between the two decorrelation methods. In every case where a significant difference was present, PR was always greater than CF in VIS. For the 125 Hz, 2 kHz, and Broadband frequency bands at ±110°, "PR 0.1" had a significantly greater VIS than "CF 0.l" ($p < 0.05$); and for the 1 kHz band at ±110°, "PR 0.4" was significantly greater than "CF 0.4" ($p < 0.05$). Similarly, with the 250 Hz band, PR had a significantly greater VIS than CF for "0.7" and "0.4" at ±30°, and for "0.7" at ±110° ($p < 0.05$). All other ICCC / frequency band conditions demonstrated no significant difference between the two decorrelation methods ($p > 0.05$).

2.2.2 Interchannel Cross-Correlation Effect

Considering the effect of ICC on VIS, a significant change of VIS is observed for the majority of frequency band / azimuth angle conditions. With the 63 Hz octave-band, the Friedman test results indicate a significant ICC effect for 0° and ±110° azimuth ($p < 0.05$), but not at ±30° ($p > 0.05$). For the 125 Hz band, only the PR method at ±110° azimuth showed a significant ICC effect ($p < 0.05$), while the 250 Hz band has a significant ICC effect with PR at ±30° and both methods at ±110° ($p < 0.05$). Looking at Fig. 4, it is seen that the ICC effect on VIS tends to become more apparent and linear as frequency increases. In agreement with this, the Friedman results for the 500 Hz, 1 kHz, 2 kHz, 4 kHz, 8 kHz, 16 kHz, and Broadband frequency bands all demonstrate a significant ICC effect at each of the azimuth angles under testing (0°, ±30°, and ±110°) ($p < 0.05$).

2.2.3 Statistical Correlation between ICC and VIS

To further evaluate the apparent effect of ICC on VIS, the data sets of each decorrelation method have been combined, with the statistical correlation between ICC and VIS calculated for each condition (Table 1). The correlation coefficients were determined using both Spearman's rank-order and Pearson's product-moment measurement techniques. Spearman's rank-order test is non-parametric and can be applied to ordinal data, with the results indicating the strength of a monotonic relationship between variables—that is, as one variable increases, so does the other—whereas Pearson's approach assesses the linearity of correlation between two variables. The closer the coefficient is to 1, the stronger the correlation, with the authors assuming a coefficient greater than 0.7 to indicate a relatively strong correlation.

Observing the Spearman rank-order coefficients in Table 1, it can be seen that the strongest correlation between ICC and VIS is for the 8 kHz band at ±30° azimuth ($r_s = 0.83$), which is further reflected in the Pearson results ($r = 0.80$). The Spearman results also show strong correlation for the Broadband stimuli at ±30°, and 2 kHz, 4 kHz, 8 kHz, 16 kHz, and Broadband at ±110°. There appears to be general agreement between the two statistical correlation tests for each condition, demonstrating that the strong correlation observed is mostly linear. Furthermore, the correlation between ICC and VIS is significant at all azimuth angles for the 500 Hz octave-band and above, as well as for 63 Hz at 0°/±110° and 125 Hz at ±110° ($p < 0.01$), broadly agreeing with the significant ICC effects reported in Sec. 2.2.2.

2.2.4 Monophonic Results

The only decorrelated stimuli to have a significantly greater VIS than the monophonic conditions were 63 Hz, 2 kHz, and 8 kHz "0.1" at ±110°, and Broadband "0.1" at both 0° and ±110° ($p < 0.05$). In some cases, the monophonic condition was perceived as having a significantly greater VIS than decorrelated stimuli. For the 500 Hz band, Wilcoxon tests revealed that the monophonic sample was significantly greater than "1.0" at both 0° and ±30°, as well as greater than "0.7" and "0.4" for the PR method at 0° ($p < 0.05$). With the 1 kHz band, the monophonic sample was significantly greater than "1.0" at ±110° ($p < 0.05$). The 2 kHz monophonic stimulus was significantly greater than "1.0" and "0.7" (PR method) at 0° ($p < 0.05$), as well as "1.0" and "0.4/0.7" (CF method) at ±30° ($p < 0.05$). With the 4 kHz band, the monophonic sample was significantly greater than "1.0" and "0.7" at ±30° ($p < 0.05$), and greater than "1.0" at ±110° ($p < 0.05$). It can be seen from these results that an $ICCC_{avg}$ of "0.1" was always perceived as having a similar or greater VIS than the monophonic condition for every frequency band and azimuth angle.

2.3 Discussion of Relative Testing Results

Generally speaking, there is little linear interchannel cross-correlation (ICC) effect for octave-bands below the 500 Hz band; that is, where an increase of vertical spread is observed as correlation decreases. Looking at the median and notch edge plots in Fig. 4 above, it can be seen that a direct relationship between ICC and vertical image spread (VIS) begins to develop around the 500 Hz octave-band point—this is confirmed by the significant statistical correlation results shown in Table 1. In every case for the 500 Hz octave-band and above, either the ICC conditions of "0.1" or "0.4" (decorrelated) were significantly greater than "1.0" (the correlated condition) across all loudspeaker angles. This perceptual effect is likely due to the introduction of vertical localization cues generated by the pinna [16, 17], torso [31], and room reflections, as well as that of interaural cues by a head-shadowing effect when the source is presented off-center. Consequently, the presence of these features may allow the brain to interpret two largely uncorrelated signals from different directions simultaneously.

There also appears to be some direction-dependency on the relationship between ICC and VIS. For both the 500 Hz and 1 kHz bands, the statistical correlation between ICC and VIS is greatest at 0° azimuth, then appears to decrease as the azimuth angle increases. This suggests that the perception of VIS by decorrelation for these frequencies may

Table 1. Statistical correlation between the Interchannel Cross-Correlation Coefficient (ICCC) and the relative Vertical Image Spread (VIS) scores (** $p < 0.01$; * $p < 0.05$).

	Spearman's Rank-Order (r_s)			Pearson's Product-Moment (r)		
	Center (0°)	Front (±30°)	Rear (±110°)	Center (0°)	Front (±30°)	Rear (±110°)
63 Hz	0.29**	0.11	0.35**	0.24*	0.17	0.30**
125 Hz	0.01	0.05	0.30**	0.03	0.07	0.29**
250 Hz	–0.09	0.01	0.14	–0.10	0.02	0.14
500 Hz	0.49**	0.41**	0.33**	0.47**	0.36**	0.33**
1 kHz	0.52**	0.42**	0.42**	0.54**	0.42**	0.41**
2 kHz	0.63**	0.68**	0.74**	0.63**	0.66**	0.68**
4 kHz	0.51**	0.67**	0.75**	0.50**	0.68**	0.74**
8 kHz	0.65**	0.83**	0.73**	0.58**	0.80**	0.70**
16 kHz	0.52**	0.56**	0.73**	0.50**	0.55**	0.70**
Broadband	0.62**	0.77**	0.71**	0.54**	0.73**	0.69**

be most detectable when changes occur in both ears equally (i.e., the median plane). In contrast, the 2 kHz, 4 kHz, and 16 kHz bands show strongest correlation at ±110° where interaural differences are greatest from head-shadowing; while the 8 kHz and Broadband frequency bands indicate strongest correlation at ±30°, where there is a combination of both interaural differences and spectral filtering by pinna reflections. These points have been considered further in the objective analysis of Sec. 4 below.

Comparing the results for the two decorrelation methods, phase randomization (PR) and complementary comb-filtering (CF), there is generally little difference between them— 92% of all method comparisons exhibited no significant difference. In the cases where there was significant difference, the PR method was always perceived as having a greater VIS than CF. Given the general similarity of VIS between methods for each ICC level, and also the linear relationships seen with the 500 Hz octave-band and above, it can be suggested that the ICC relationship with VIS at middle to high frequency bands might be useful for practical applications such as vertical upmixing. Through further investigation and development, it may be found that vertical $ICCC_{avg}$ measurements at these frequencies can also contribute to the prediction of VIS in 3D audio content.

It is worth noting that, for the majority of monophonic results, the monophonic stimuli were not perceived as significantly different from the vertically decorrelated conditions (all of which were presented at the same SPL level). This similarity suggests that there may not be any benefit to vertically decorrelating signals when upmixing from 2D to 3D. One cause of an increase to the VIS of the monophonic condition could be the impact of vertical decorrelation on perceived loudness, where the interaction of two partially correlated signals at the ear may result in the cancelling of frequencies (e.g., by comb-filtering), leading to a perceived "weakening" of the signal. Early room reflections could have also influenced the perception of the monophonic stimuli, particularly for the middle frequencies around 500 Hz to 1 kHz—it has previously been shown that a single reflection can impact VIS [32]—this has been objectively investigated and discussed in Sec. 4.4. Furthermore, the directional bands phenomenon may be responsible for some confusion when perceiving the monophonic octave-band stimuli [33]; in particular, it has been shown that monophonic 1 kHz and 8 kHz octave-bands tend to be perceived behind and above respectively. Such confusion for some listeners may have caused the large error bars seen for monophonic stimuli in Fig. 4. It is possible that if the monophonic condition had been used as reference then the monophonic results may have been more consistent. However, the relative differences between conditions would have been broadly similar to the present results, i.e., a general monotonic relationship between ICC and VIS, and an increased VIS for some monophonic stimuli.

3 EXPERIMENT TWO: ABSOLUTE GRADING OF VERTICAL IMAGE SPREAD

From the first subjective experiment, it was established that the relative effect of interchannel cross-correlation (ICC) on vertical image spread (VIS) is both perceivable and significant, most notably for frequency bands of around 500 Hz and above. The aim of this second subjective experiment is to determine the actual extent of change between vertically decorrelated, correlated, and monophonic stimuli, through absolute measurement of the VIS upper and lower auditory boundaries. Results of the experiment will establish the accuracy and agreement of defining VIS boundaries, as well as in which direction any changes of VIS occur along the vertical plane.

3.1 Experimental Design

The second experiment used the same physical loudspeaker setup as the first; however, only the Center 0° and Front ±30° vertically-spaced loudspeaker pairs were tested (see Fig. 3). This was due to the practical difficulty of localizing and defining VIS in absolute terms from behind the listener. Listening tests for the second experiment were conducted in the same critical listening room as the first and under the exact same testing conditions, with time/level alignment and an acoustically transparent curtain. A vertical light-emitting diode (LED) strip was fixed beside the 0° loudspeakers, to aid the capture of the upper and lower auditory boundaries, while avoiding potential bias from visual markers on a physical scale. The LED strip was linked to a physical rotary controller through Cycling '74's Max 7,

which the user could rotate to control the position of each boundary separately (depressing the controller knob to switch between boundaries). This response method was originally proposed and evaluated by Lee et al. [34]—in their experiment on vertical phantom image localization, it was found that the LED method improved the subject's response consistency and shortened the test duration, when compared against the conventional visual marker method.

In order to assess the absolute difference of VIS between the narrowest and broadest perceived samples, three stimuli from the first experiment were tested for each frequency band. These three stimuli were the correlated stimulus ($ICCC_{avg}$ of 1.0), the monophonic stimulus (main-layer only) and the vertically decorrelated phase randomization (PR) stimulus ($ICCC_{avg}$ of 0.1). PR was chosen over the complementary comb-filtering (CF) method since the first experiment showed that, for every frequency band, PR with an $ICCC_{avg}$ of 0.1 consistently had the same or greater VIS than all other stimuli.

Nine subjects took part in the second experiment, all of who also participated in the first. In a single trial, subjects were asked to define the absolute upper and lower boundary of the auditory VIS, one stimulus at a time, using the LED strip and rotary controller connected to Max. There were 60 trials in total: 3 stimuli (decorrelated, correlated, and monophonic) for 10 frequency bands (63 Hz – 16 kHz octave-band and Broadband pink noise), tested at 2 loudspeaker azimuth positions (0° and ±30° (randomized between left and right)). The test was repeated twice for each subject and all trials were randomized over 2 listening sessions of around 15–20 minutes each.

3.2 Results and Analysis

Results from the absolute vertical image spread (VIS) testing can be seen in Fig. 5 below. Each stimulus condition features two box plots that represent the data for its lower and upper VIS boundaries—where the grey box plot on the left is the lower boundary and the white box plot on the right is the upper boundary. The line within the boxes signifies the median of subjects' responses, and the extent of the boxes indicates the 1st and 3rd quartiles (the interquartile range) of the data. To visualize the general spatial impact of the boundary positions between stimuli and frequency bands, the median overall VIS (the difference between the two boundaries) is displayed in Fig. 6—these values were obtained by calculating the overall VIS of each individual response, then plotting the median value from these calculations for each condition.

Bonferroni-corrected Wilcoxon signed-rank tests have been conducted between stimuli within each octave-band, to assess the perceived differences of the upper boundary position, the lower boundary position, and the overall VIS (the difference between the two boundaries). In the following results "Mono" refers to a signal from the lower main-layer loudspeaker only, "1.0" is the correlated stimulus, and "0.1" is the vertically decorrelated stimulus.

For octave-bands of 63 Hz to 500 Hz, the Wilcoxon tests revealed no significant change to the VIS boundary positions and the overall VIS between stimuli within each octave-band and loudspeaker position ($p > 0.05$). These results broadly agree with the relative test results in Sec. 2.2, where no significant ICC effect was observed for octave-bands below 500 Hz. Fig. 6 also shows that the overall VIS for each condition of the 63 Hz octave-band is consistently large for both loudspeaker angles (between 100—120 cm), when compared against the overall VIS of the other frequency bands.

Wilcoxon results for the 1 kHz band show that, for both the 0° and ±30° loudspeaker positions, the "0.1" upper boundary was significantly higher than the "1.0" upper boundary ($p < 0.05$), indicating an upward increase of VIS by decorrelation. On the other hand, looking at the 2 kHz octave-band from ±30°, the "1.0" condition had a significantly higher lower boundary than both the "0.1" condition ($p < 0.05$), demonstrating a downward increase of VIS following decorrelation. Furthermore, there is some suggestion of image shifting with the 1 kHz and 2 kHz octave-bands; in particular, the 1 kHz "0.1" condition at ±30° has significantly higher upper and lower boundaries than the "Mono" stimulus ($p < 0.05$).

For the 8 kHz band, both the 0° and ±30° results show a significant upward shift of the "1.0" image from the "Mono" condition ($p < 0.05$)—similarly, the "0.1" upper image boundary was significantly higher than the "Mono" condition for both positions ($p < 0.01$). This suggests that presenting an 8 kHz octave-band in vertical stereophony elevates the auditory image towards the height-channel loudspeaker. Figs. 5 and 6 also demonstrate that vertical decorrelation of the 8 kHz band increases VIS downwards, as with 2 kHz; however, this is not significant due to a large deviation of the lower boundary scores (the median absolute deviation (MAD) of these lower boundary results has been calculated as 36–50 cm). In contrast, decorrelation of the 16 kHz octave-band results in a significant upward increase of VIS from the "1.0" and "Mono" conditions at 0° azimuth ($p < 0.05$) (as shown in Figs. 5 and 6).

With the Broadband stimuli, there was some significant difference between boundaries for both loudspeaker positions. At 0° azimuth, the upper and lower boundaries of "0.1" and "1.0" were both significantly higher than the "Mono" boundaries ($p < 0.01$), indicating an upward image shift when introducing a height-channel. Whereas from ±30°, only the upper boundaries of "0.1" and "1.0" were significantly higher than the "Mono" sample ($p < 0.05$), with the lower boundaries remaining consistently below the lower loudspeaker position for all conditions. In terms of overall VIS for the Broadband stimuli, only "0.1" was significantly greater than the "Mono" stimulus at ±30° ($p < 0.05$) (supporting the results seen in the first experiment), although Fig. 6 also shows "0.1" as having the greatest VIS at 0° azimuth as well.

3.3 Discussion of Absolute Testing Results

The absolute results show no significant change to vertical image spread (VIS) boundaries or overall VIS for octave-bands of 500 Hz and below. This is broadly in

Fig. 5. Box plots displaying the location for the upper and lower boundaries of the auditory Vertical Image Spread (VIS) (cm), where each box features the median (the 2nd quartile) and interquartile range (1st to 3rd quartile) of a boundary position. The grey shaded boxes on the left show the lower boundary of a condition and the white boxes on the right show the upper boundary. The two dashed lines indicate the acoustic centers of the lower and upper loudspeakers (i.e., a 30° difference).

line with the results from the first experiment, where it is generally seen that the effect of interchannel cross-correlation (ICC) on VIS occurs significantly around the 500 Hz octave-band and above. The only cases of decorrelation influencing a significant increase to the overall VIS (the difference between the upper and lower boundaries) are for the 16 kHz octave-band at 0° and the Broadband condition at ±30°. However, the median values in Fig. 6 demonstrate that the "0.1" condition (the decorrelated stimulus) consistently had a greater overall VIS than "1.0" (the correlated stimulus) for octave-bands of 500 Hz and above (as well as for the Broadband condition), supporting the relative grading results in the first experiment.

In general, the changes of VIS appear to be slight, with large deviations about the median point for many boundary positions (see the interquartile ranges in Fig. 5)—this

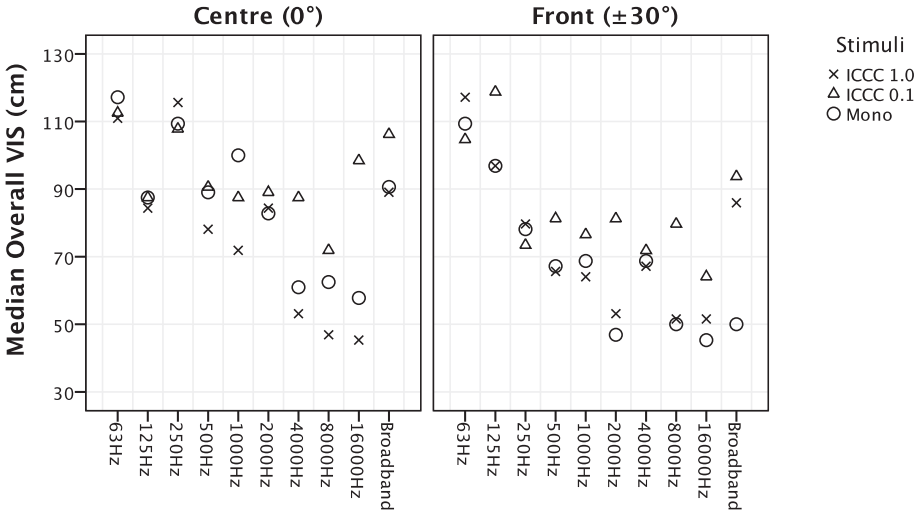

Fig. 6. The absolute median overall Vertical Image Spread (VIS) for each condition (cm), where the raw overall VIS scores were calculated as the difference between the upper and lower boundaries for each individual subject response before averaging.

is despite significant difference of VIS between the same conditions in the first experiment. These results suggest that perceiving changes to boundaries and vertical extent is likely to be easier when comparing stimuli relatively (as in the first experiment), whereas absolutely defining the image boundaries independently is a rather difficult task. In particular, the 8 kHz octave-band at $\pm 30^\circ$ had the strongest relationship between ICC and VIS in the first experiment (which is reflected by the absolute increase of overall VIS in Fig. 6); however, a noticeably high deviation of lower boundary responses for both "1.0" and "0.1" delivered a statistically insignificant change of absolute VIS (the MAD of these lower boundaries has been calculated as 36–39 cm).

Similarly, the Broadband results from the first experiment demonstrated a significant relative VIS increase by decorrelation, although the absolute results only show an insignificant increase of overall VIS from "1.0" and "0.1" in Fig. 6. It is possible that the absolute measures of the Broadband stimuli were judged based on the inherent large spread of low frequencies, as seen with the 63–250 Hz octave-bands in Fig. 6 (i.e., the entire spectral image was considered). On the other hand, the relative judgments of the first experiment could have been dictated by noticeable changes to VIS in the higher frequencies (e.g., within the 8 kHz octave-band). Significant movements of upper and lower boundaries (not necessarily increasing VIS) may have also led to a relative perception of VIS change in the first experiment. The 1 kHz result is one case where the perception of VIS could have been dictated by changes to a single boundary or shift of image, rather than an increase to the absolute VIS.

For the 2 kHz and 8 kHz octave-bands, the "1.0" condition (the correlated stimulus) at $\pm 30^\circ$ appears to be elevated towards the height loudspeaker (in comparison to the Mono condition)—the auditory image is then extended downwards from this elevated position towards the lower loudspeaker following decorrelation. In the case of the 8 kHz band, the bias towards the height-layer loudspeaker is very strong at both 0° and $\pm 30^\circ$ when a height-channel is present (i.e., vertical stereophony). The results presented here suggest that the vertical localization cues from an elevated 8 kHz octave-band source have dominance over a lower positioned correlated source, when both are presented simultaneously. A large spectral notch around 8 kHz (caused by the pinna from the main-layer loudspeaker direction) is likely to be the cause of this effect, resulting in more energy for the 8 kHz band from the height-channel direction [35]. Furthermore, the downward extension following decorrelation suggests that both loudspeaker signals might be perceived independently at the same time, with the decorrelation process reducing the height dominance to "un-mask" the main-layer loudspeaker. In contrast to this, all of the 16 kHz stimuli have a bias towards the main-layer loudspeaker, with a significant upward extension for the decorrelated stimuli. To investigate these findings further, analysis of the spectra for these key bands is presented in Sec. 4.2 below.

4 OBJECTIVE ANALYSIS AND DISCUSSIONS

4.1 Binaural Synthesis of Stimuli

To objectively analyze the stimuli signals, all 80 stimuli from the first subjective experiment (described in Sec. 2.1.2) have been convolved with binaural room impulse responses (BRIRs) of the listening room. These BRIRs were acquired using the exponential sine sweep (ESS) method [36] implemented in the HAART impulse response capture toolbox [37]. Sine sweeps were presented from each loudspeaker independently and recorded using a Neumann KU 100 dummy head in the listening position (with the ear height in line with the acoustic center of the main-layer loudspeakers, i.e., 0° elevation at a height of 1.27 m). This resulted in a total of 10 BRIRs, one for each of the 10 loudspeakers used during testing —5 main-layer and 5 height-layer loudspeakers, based on Auro-3D 9.1

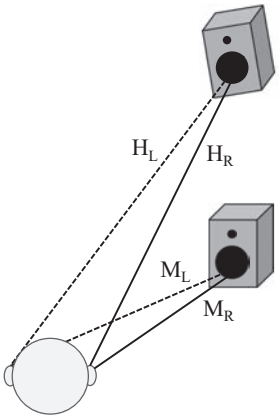

Fig. 7. Diagram of the convolution process for a single loudspeaker pair. M_L and M_R represent the left and right ear input signals of the main-layer (lower) loudspeaker, and H_L and H_R represent the left and right ear input signals of the height-layer (upper) loudspeaker. The two layers are then summed together to replicate the test stimuli.

with an additional center height-channel [1] (Fig. 3 above). During the convolution process, three conditions were created for each combination of stimulus and azimuth angle (0°, ±30°, and ±110°): the main-layer only (Eqs. (5) and (6)), the height-layer only (Eqs. (7) and (8)) and the layers combined (where the time/level-aligned main- and height-layer impulses were summed) (Eqs. (9) and (10)) (Fig. 7). Spectral analysis of the BRIR-convolved stimuli can be seen in Sec. 4.2, with calculation of the interaural cross-correlation coefficients (IACCs) presented in Sec. 4.3.

$$L_M = M_L * S_1 \quad R_M = M_R * S_1 \tag{5, 6}$$

$$L_H = H_L * S_2 \quad R_H = H_R * S_2 \tag{7, 8}$$

$$L_C = L_M + L_H \quad R_C = R_M + R_H \tag{9, 10}$$

where S_1 and S_2 are the two-channel stimuli signals, M is "Main," H is "Height," and L_x and R_x are the left and right ear signals of the convolved stimuli.

4.2 Spectral Analysis of Binauralized Stimuli

In the subjective relative testing results (Sec. 2.2), there was a significant interchannel cross-correlation (ICC) effect on vertical image spread (VIS) from around the 500 Hz octave-band and above, where VIS increased as the ICC coefficient (ICCC) decreased. At high frequencies, it is hypothesized that this change to VIS may be influenced by the head-related transfer function (HRTF) filtering of signals, particularly in the median plane [17].

To investigate this, delta spectra of the frequency amplitude difference between the BRIR-convolved stimuli is presented below for each azimuth angle (0°, ±30°, and ±110° in Figs. 8, 9, and 10, respectively). Spectra were calculated using 4096 FFT-points with a frame length of 4096 samples—a Hann window was used on the frames with 50% overlap, and 1/6-octave Gaussian smoothing was applied to the FFT output.

In the delta plots of Figs. 8, 9, and 10, the broadband FFT spectra of the decorrelated signals ($ICCC_{avg}$ of 0.1, 0.4, and 0.7) have been subtracted from that of the correlated condition ($ICCC_{avg}$ of 1.0), in order to observe the differences of spectrum as correlation decreases. A positive amplitude difference indicates a boost of frequency for that particular ICC condition in comparison to $ICCC_{avg}$ 1.0. Both decorrelation methods are presented (Phase Randomization (PR) in the upper panel and Complementary Comb-Filtering (CF) in the lower), with the left and right panels displaying plots for the left and right ears, respectively.

Observing the plots in Figs. 8, 9, and 10, general agreement can be seen between the upper and lower panels. This indicates that both the phase-based and amplitude-based decorrelation methods appear to produce similar spectral cues. Furthermore, it is demonstrated that the majority of spectral changes occur within the 8 kHz and 16 kHz octave-bands, presumably due to HRTF filtering from the pinna (the outer ear).

At 0° azimuth (Fig. 8), vertical decorrelation primarily results in a boost around 8–9 kHz for both methods. For the +30° azimuth angle (Fig. 9), boosts from decorrelation can be seen around 7–8 kHz and 16 kHz in the contralateral left ear and 8 and 12 kHz in the ipsilateral right ear. Whereas at +110° azimuth (Fig. 10), decorrelation seems to generally boost high frequencies in the contralateral left ear, while the ipsilateral right ear features two boosts around 8–9 kHz and 14–15 kHz. It is seen that all of these spectral boosts tend to increase systematically as ICCC decreases, where it appears the degree of vertical decorrelation relates directly to the degree of spectral change. This effect seems to be a reduction of phase cancellation at the ear, where the decorrelated signals cause the spectral notches from pinna filtering to be "filled in." It is well-known that spectral notches are important to the localization of sources in the vertical plane [16, 17]; therefore, a reduction of notch depth may decrease localization accuracy, resulting in a perceived increase of VIS. Further investigation is required to understand the significance of these spectral cues and also to determine whether simple amplitude manipulation of pinna notches can be used to influence the perception of VIS.

4.3 Interaural Cross-Correlation (IAC) of Binauralized Stimuli

Observing the spectral analysis in Sec. 4.2 above, there are clear interaural differences of spectral filtering at higher frequencies. It is already established that the interaural cross-correlation coefficient (IACC) has a strong relationship with the perceived horizontal extent of a sound source [7]; however, it is not currently known whether interaural differences also have a notable impact on the perception of vertical spatial extent. Furthermore, previous studies have indicated that horizontal decorrelation and IAC change can have an effect on the perceived elevation of an auditory source [38, 39]. Considering this, analysis of the IACC for the current stimuli may demonstrate an association between interaural differences and the perception of VIS. In

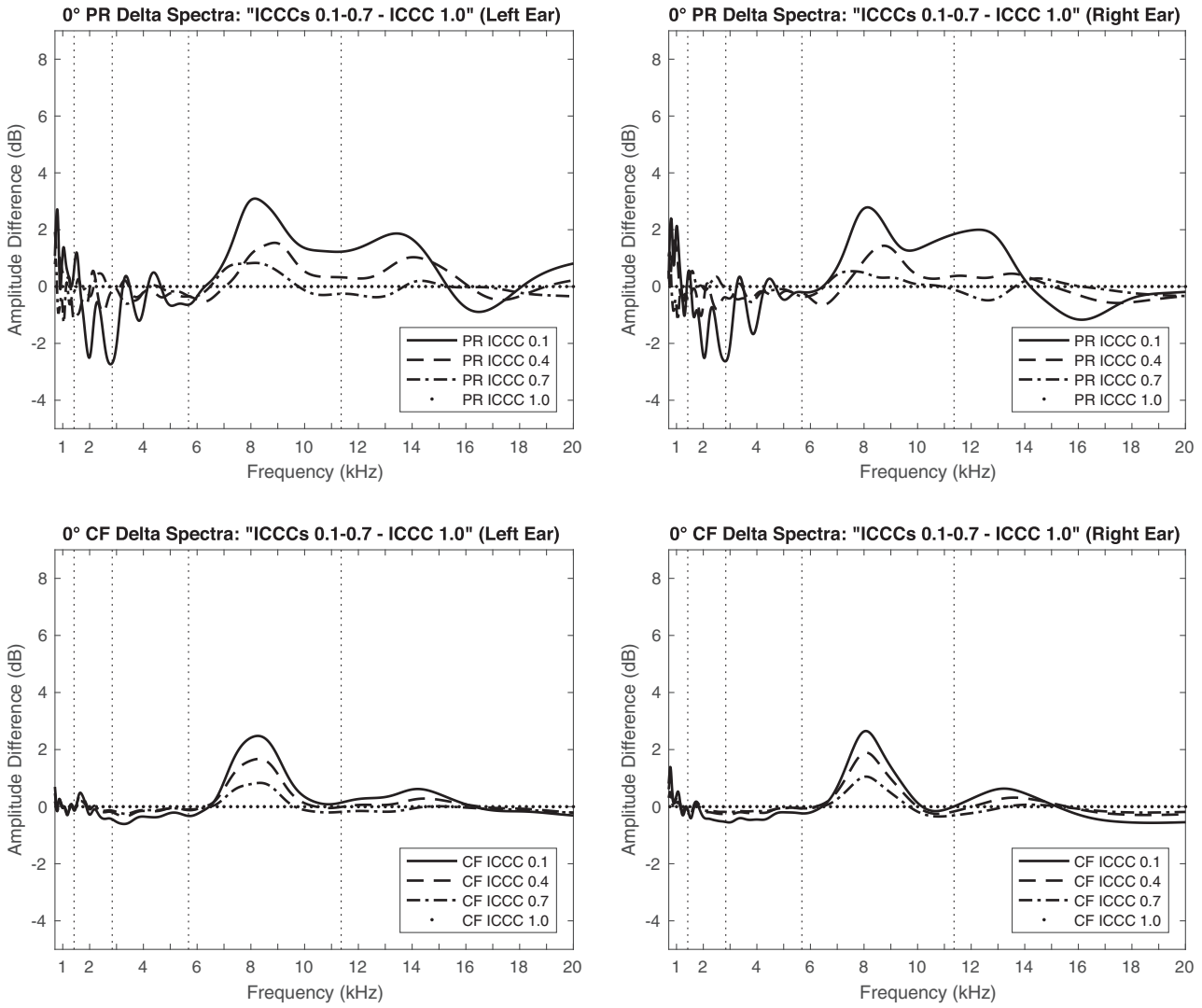

Fig. 8. 0° azimuth delta spectra of the FFT frequency amplitude difference between the BRIR-convolved correlated stimulus (ICCC 1.0) and the decorrelated stimuli (ICCCs 0.1-0.7). The vertical dotted lines signify the band limits of each octave-band. (Upper – PR Method; Lower – CF Method)

particular, the 500 Hz or 1 kHz octave-bands showed no spectral variation in line with ICC change (Sec. 4.2), despite these octave-bands having a significant ICC effect in the subjective testing (Sec. 2.2).

Using the convolved stimuli described in Sec. 4.1, $IACC_{avg}$ values have been determined for each of the convolved conditions using Eqs. (11) and (12) below, taking the average of time-varying IACCs calculated over 50 ms-long windows, with 1 ms lag to account for interaural time delay (ITD) [7]. It has previously been suggested that a 50 ms window length be used for IACC calculation based on the temporal resolution of the auditory system [40].

$$IACF\ (\tau) = \frac{\int_{-\infty}^{\infty} x_{left}\,(t)\,x_{right}\,(t+\tau)\,dt}{\sqrt{\left[\int_{-\infty}^{\infty} x_{left}^{2}\,(t)\,dt\right]\left[\int_{-\infty}^{\infty} x_{right}^{2}\,(t)\,dt\right]}} \qquad (11)$$

$$IACC\ =\ max\,|IACF\,(\tau)|\ \ where\ \ |\tau| \le 1ms \qquad (12)$$

A summary of the $IACC_{avg}$ results can be seen in Table 2 above. The table features $IACC_{avg}$ values for the extreme interchannel cross-correlation (ICC) conditions of "1.0" and "0.1" (for both decorrelation methods) and the monophonic stimuli used in the subjective experiment (Sec. 2.2). The $IACC_{avg}$ values are presented for each frequency band at the three azimuth angles (0°, +30°, and +110°, assuming symmetry between the left and right directions). Looking at the 0° results in Table 2, it is evident that the $IACC_{avg}$ generally begins to decrease around the 500 Hz octave-band and above ($IACC_{avg} < \sim 0.9$). This decrease is presumably due to the influence of room reflections summing at the ear input; whereas, little impact is seen at lower frequencies due to considerably longer wavelengths than the width of the head.

Observing the effect of vertical decorrelation on $IACC_{avg}$, a comparatively noticeable decrease of $IACC_{avg}$ ($> \sim 0.07$) between the ICCC "1.0" and "0.1" conditions is seen for the 0.5-1 kHz octave-bands at 0°, 1–8 kHz bands at +30°, and for the 0.5–4 kHz bands at +110°. That is, as correlation between the vertical pair of loudspeakers decreases, so does correlation between the two ear-input

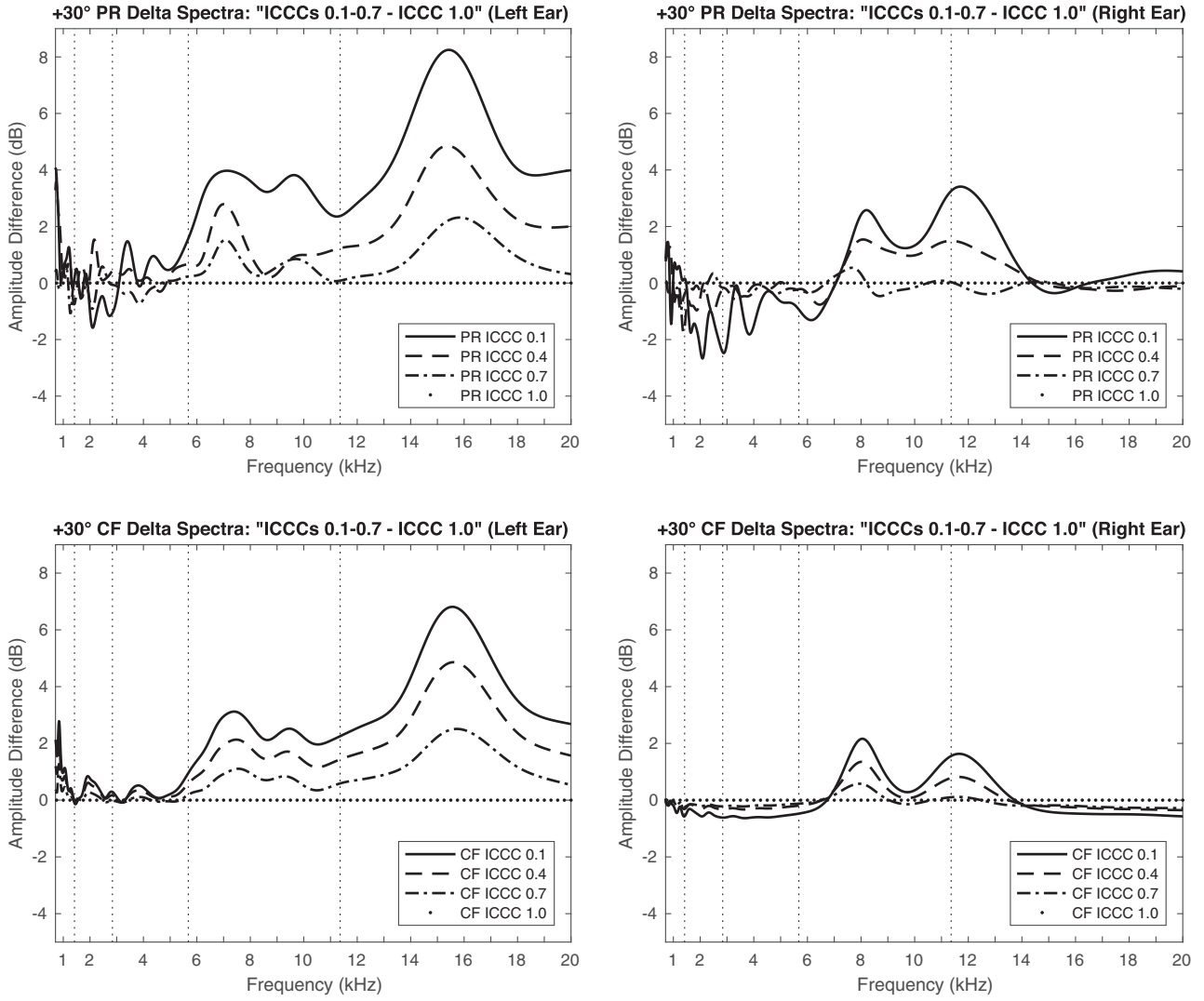

Fig. 9. +30° azimuth delta spectra of the FFT frequency amplitude difference between the BRIR-convolved correlated stimulus (ICCC 1.0) and the decorrelated stimuli (ICCCs 0.1-0.7). The vertical dotted lines signify the band limits of each octave-band. (Upper – PR Method; Lower – CF Method)

signals. These observations are broadly in line with the subjective results in Sec. 2.2, suggesting that IACC may contribute to the perception of VIS for both PR and CF vertical decorrelation. However, further investigation is required to determine whether a relationship between vertical ICCC and IACC does exist.

At 0° in the median plane, it is hypothesized that the decrease of $IACC_{avg}$ is dictated by decorrelated room reflections, given that direct sound is equal in both ears (as explored further in Sec. 4.4 below); whereas at ±30° and ±110°, it is possible that the changes in $IACC_{avg}$ may have also been influenced by interaural level and time differences. For instance, an effect whereby the height-channel is less shadowed by the head than the main-channel could affect IACC at wider azimuths (i.e., ±110°). In other words, the contralateral ear input is largely controlled by the height-channel signal, while the ipsilateral ear input is the sum of the main- and height-channel signals. Consequently, if both loudspeaker signals are correlated, the resulting IACC is relatively high; however, if the loudspeaker ICCC decreases, so does the IACC, due to the influence of the height-channel signal in the contralateral ear. Furthermore, an experiment comparing horizontal decorrelation at different heights found that presenting the same decorrelated stimuli from a greater elevation angle tended to increase the IACC, which may have also been caused by a decrease of head-shadowing [39]. The potential contribution of IAC to the perception of sources within the vertical domain is in need of further investigation.

4.4 The Influence of Early Reflection Energy

Taking into account the $IACC_{avg}$ results in Sec. 4.3, a relationship has been suggested between room reflections and the perception of VIS in the median plane (0° azimuth). This is particularly the case for the 500 Hz and 1 kHz octave-bands, where both the monophonic condition and vertically decorrelated conditions have a greater VIS and lower $IACC_{avg}$ than the correlated condition (as presented in Fig. 4 and Table 2, respectively). For the decorrelated condition, it is hypothesized that the decorrelation of

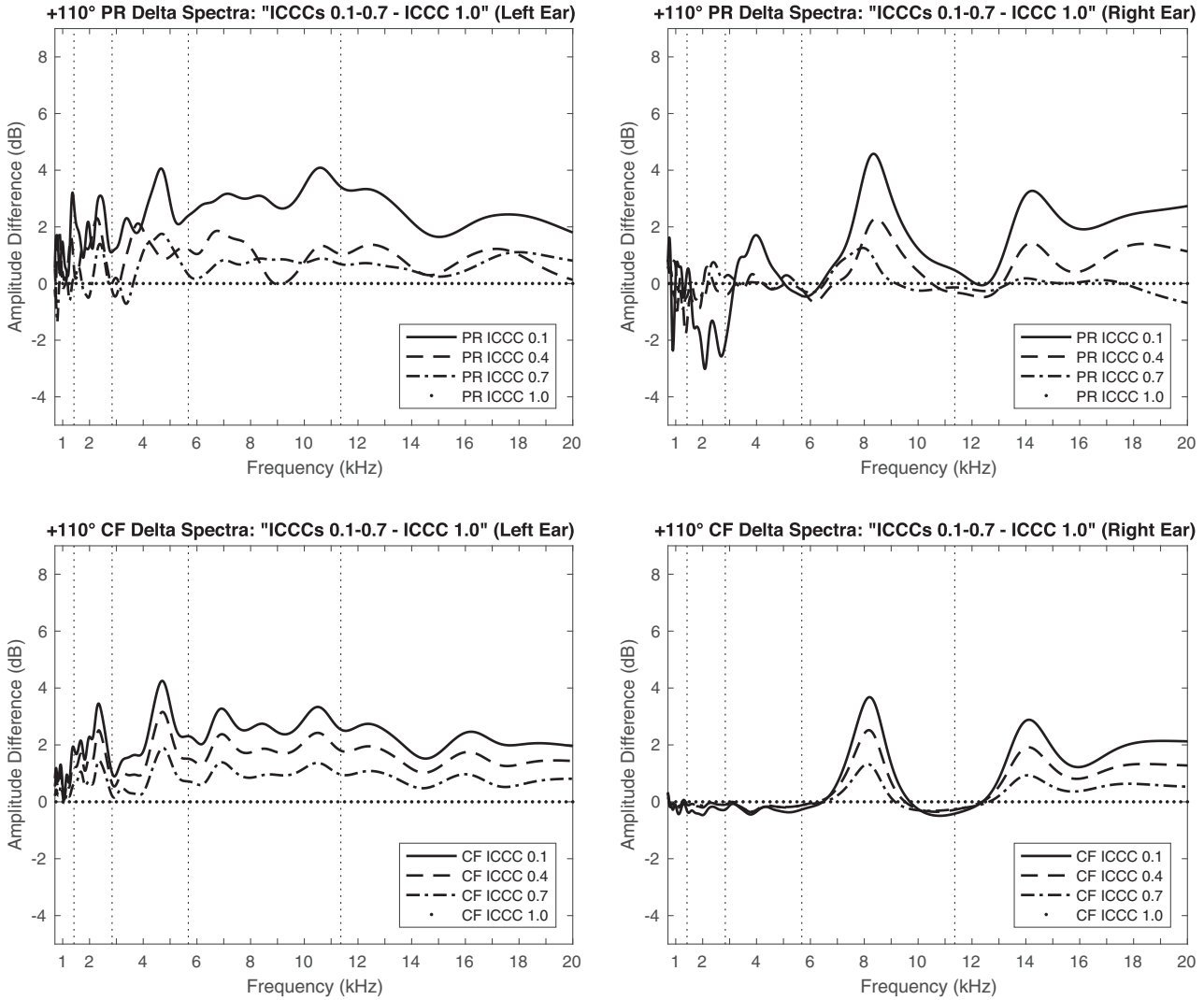

Fig. 10. +110° azimuth delta spectra of the FFT frequency amplitude difference between the BRIR-convolved correlated stimulus (ICCC 1.0) and the decorrelated stimuli (ICCCs 0.1-0.7). The vertical dotted lines signify the band limits of each octave-band. (Upper – PR Method; Lower – CF Method)

loudspeaker signals causes a greater decorrelation of reflections being summed at either ear input; however, the reason for the decrease of $IACC_{avg}$ with the monophonic condition is not clear.

To investigate the monophonic case further, Fig. 11 presents binaural room impulse responses (BRIRs) of the 500 Hz and 1 kHz octave-bands at 0° azimuth, captured as described in Sec. 4.1 above. A main-layer only BRIR (monophonic; left) is compared against the summed result of the time-aligned main- and height-layer BRIRs (correlated vertical stereophony; right) for each band. These two conditions have been RMS level-matched to replicate the SPL LAeq level-matching of stimuli in the subjective testing.

Looking at the plots in Fig. 11, an increase to the level of reflections is observed for the monophonic condition in comparison to the summed condition for both octave-bands. Calculating the ratio of early reflection energy (2.5 – 80 ms) to direct sound energy (< 2.5 ms) (ER/D ratio) confirms this, indicating that the monophonic condition has an increase of around 3–4 dB early reflection energy for both bands. It is possible that, when comparing the monophonic main-layer stimuli directly against the correlated stereophonic stimuli (ICCC 1.0) under controlled conditions, this increase in early reflection energy may be perceivable and contribute to an increased sense of VIS. In other words, the resulting decrease in $IACC_{avg}$ for the monophonic and decorrelated stimuli (Table 2) could have potentially influenced an ambiguous perception of greater extent both horizontally and vertically, given the weakness and inaccuracy of vertical localization at these frequencies [41].

Furthermore, responses in previous research suggest that a single ceiling reflection can increase the perceived VIS of an auditory event [32], indicating the effect a room's acoustic can have on vertical spatial perception. From the absolute VIS results (Sec. 3.2), it can be seen that there is downward extension of VIS for the 500 Hz and 1 kHz monophonic conditions at 0° azimuth. Considering this, the increase of VIS for the monophonic condition could be related to an increase of floor reflection energy from the

Table 2. Interaural Cross-Correlation Coefficient averages from 50 ms windows ($IACC_{avg}$). BRIR-convolved stimuli, captured using a Neumann KU 100 dummy head in the listening room.

	Center (0°)				Front (±30°)				Rear (±110°)			
	Mono	1.0	0.1 (PR)	0.1 (CF)	Mono	1.0	0.1 (PR)	0.1 (CF)	Mono	1.0	0.1 (PR)	0.1 (CF)
63 Hz	1.00	1.00	1.00	1.00	0.98	0.98	0.99	0.95	0.98	0.98	0.98	0.98
125 Hz	0.99	1.00	1.00	0.99	0.94	0.96	0.94	0.93	0.91	0.92	0.85	0.92
250 Hz	0.97	0.97	0.98	0.97	0.92	0.90	0.92	0.90	0.86	0.88	0.92	0.83
500 Hz	0.81	0.89	0.77	0.83	0.57	0.64	0.70	0.61	0.76	0.78	0.70	0.73
1 kHz	0.83	0.89	0.83	0.82	0.68	0.74	0.72	0.64	0.58	0.62	0.51	0.53
2 kHz	0.84	0.89	0.85	0.86	0.73	0.81	0.71	0.68	0.28	0.36	0.27	0.28
4 kHz	0.88	0.89	0.87	0.87	0.73	0.80	0.75	0.73	0.40	0.34	0.24	0.28
8 kHz	0.66	0.74	0.68	0.70	0.37	0.47	0.34	0.39	0.21	0.22	0.22	0.20
16 kHz	0.86	0.85	0.81	0.80	0.54	0.28	0.30	0.28	0.25	0.19	0.15	0.20
BB	0.95	0.97	0.96	0.94	0.79	0.81	0.81	0.77	0.70	0.73	0.73	0.71

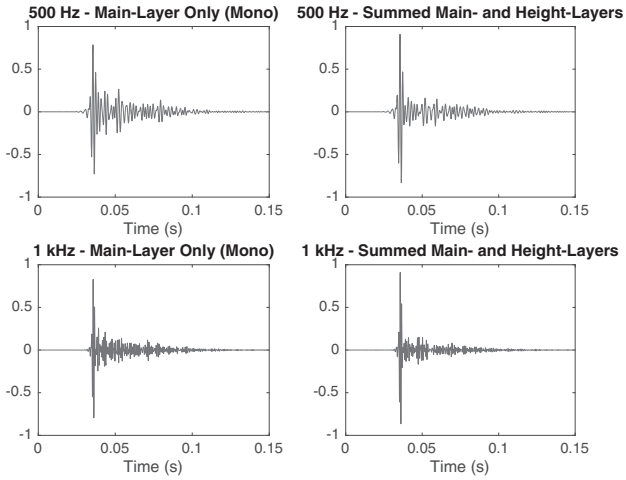

Fig. 11. "The Main-Layer Effect"—500 Hz and 1 kHz octave-band filtered binaural room impulse responses, with RMS level-matching between the mono (main-layer only) and summed conditions. Left = main-layer only impulse response (mono-phonic). Right = summed main- and height-layer impulse responses (vertical stereophony).

"main-layer effect" discussed above. Further investigation is required to observe the effect of room reflections on VIS perception; however, the results here suggest that VIS change at lower frequencies (i.e., 500 Hz and 1 kHz) may be influenced by environmental factors of the listening room.

5 PRACTICAL IMPLICATIONS

The results from the current study suggest that the effect of vertical interchannel decorrelation on vertical image spread (VIS) tends to be slight. Despite the significant interchannel cross-correlation (ICC) effect on VIS seen in the relative subjective test, the results from the absolute test appear to be more inconsistent, indicating a difficulty perceiving and defining VIS discretely. This suggests that vertical decorrelation of signals in a practical scenario may be largely ineffective, i.e., where a direct comparison between conditions is unavailable. In terms of 2D-to-3D upmixing, the inclusion of additional vertically-spaced pairs of loudspeakers may increase this difficulty further, due to an increase of decorrelated signals interacting at the ear. Nevertheless, there may still be some benefit of using vertical decorrelation to reduce phase cancellation from multiple similar signals at the ear, based on the spectral observations made in Sec. 4.2. Further investigation with complex sources is required to fully determine the effectiveness of 2D-to-3D upmixing and VIS rendering by vertical interchannel decorrelation.

Considering the frequency-dependency of vertical decorrelation, the subjective results appear to suggest that the 8 kHz octave-band is most effective. This is in general agreement with the literature that states frequencies around 8 kHz are important for vertical localization in the front median plane [16, 17]. A previous study [42] has shown that boosting a band around 8 kHz alone can increase the vertical elevation of a broadband signal, based on Blauert's boosted band hypothesis [33], so potentially it could also control the vertical extent of an image as well. However, there seems to be an inherent dominance of the height-channel signal when presenting the 8 kHz octave-band signals in vertical stereophony. As mentioned above, a reason for this could be due to an increase of energy for the 8 kHz band in the HRTF when presented from the height-channel direction (at +30° elevation) compared to the main-channel (0° elevation) [35]. A similar elevation effect for the 8 kHz octave-band (when presented in vertical stereophony) was seen by Wallis and Lee [43], who investigated the height-channel gain reduction required to localize stereophonic octave-band conditions at the height of a monophonic reference (the lower loudspeaker only) [44]—in the case of the 8 kHz octave-band, this was found to be around a –9 dB reduction.

If vertical decorrelation is found to be useful in a practical scenario, the results from both experiments presented here suggest that vertical decorrelation of frequencies below the 500 Hz octave-band may not be necessary. In general, the lower frequency bands (63–250 Hz) have an inherently broad VIS and are localized in similar positions between each condition (generally towards the lower loudspeaker)—this is regardless of whether they are presented monophonically or decorrelated vertically. Both of these points suggest that an increase of VIS upwards might still be achieved without these frequencies, e.g., for upmixing from 2D to

3D loudspeaker formats. It is possible that decorrelating a high-pass filtered broadband signal vertically, while routing the low-pass component to the main-layer loudspeaker only, would have a similar effect to decorrelating the entire broadband signal. This approach could have an impact on the tonal quality or clarity of the sound reproduction by reducing the number of low frequencies interacting at the ears.

6 CONCLUSIONS

This paper described a two-part investigation into the perception of vertical image spread (VIS) by vertical interchannel decorrelation. In both experiments, octave-band (center frequencies: 63 Hz to 16 kHz) and broadband pink noise stimuli were tested to assess the frequency-dependency of VIS at different loudspeaker azimuth angles to the listener.

For the first experiment, the relative VIS between stimuli was graded on a bipolar scale in multiple-comparison trials. Stimuli were presented through pairs of vertically-spaced loudspeakers, positioned at three azimuth angles to the listener (0°, ±30°, and ±110°). Two decorrelation methods were compared, each with three levels of average running interchannel cross-correlation coefficients ($ICCC_{avg}$) (0.1, 0.4, and 0.7). The results indicate that ICC begins to have a significantly linear effect on VIS for frequencies around the 500 Hz octave-band and above; that is, as correlation between the vertically-spaced loudspeaker pairs decreased, the VIS increased for all azimuth angles. The strength of association between ICC and VIS appears to be divided into three directivity groups: 500 Hz and 1 kHz are strongest in the median plane (0° azimuth, where energy is equal in both ears); 2 kHz, 4 kHz, and 16 kHz at ±110° azimuth (where head-shadowing is greatest); and 8 kHz and Broadband at ±30° azimuth (where both interaural differences and frontal HRTF filtering are present).

In the second experiment, absolute VIS was measured for the monophonic, correlated, and decorrelated stimuli from the first experiment. The same frequency bands were also tested but only for azimuth angles of 0° and ±30° due to practical reasons. Subjects defined the upper and lower boundaries of the VIS for each stimulus independently using a light-emitting diode (LED) strip. The absolute results demonstrate that significant changes to the boundaries occurred for stimuli above the 500 Hz octave-band, in line with the first experiment. Great deviations of boundary responses indicate an inherent difficulty with absolutely defining the image of an auditory event. The absolute results seem to suggest that the effect of vertical decorrelation is slight and may not be perceivable in a practical scenario.

Objective analysis of the binauralized stimuli signals generally indicates two groups of perceptual cues that separate the middle- and high-frequency octave-bands. For the 500 Hz and 1 kHz bands, VIS perception appears to be influenced by room reflections, whereas at higher frequencies, filtering from the head-related transfer function (HRTF) seems to have the most influence (either through head-shadowing or pinna-related spectral filtering). In general, vertical decorrelation causes spectral boosts that increase as correlation decreases—that is, a phase cancelling effect occurs when the signals are correlated (causing spectral notches), with decorrelation reducing the degree of phase cancellation (i.e., "filling in" the notches). A potential relationship between VIS and the average-running interaural cross-correlation coefficient ($IACC_{avg}$) has also been suggested, particularly with the 500 Hz and 1 kHz octave-bands. Lastly, the "main-layer effect" is proposed and discussed in the paper, whereby the early reflection energy of a main-layer only condition is greater than that of vertical stereophony, when both conditions are level-matched.

The experiments presented in the current paper focus on the perception of vertical interchannel decorrelation at discrete angles to the listener. If vertical decorrelation were employed in a practical 2D-to-3D upmixing scenario, it is likely that vertically decorrelated signals would be presented from multiple angles simultaneously—this situation is the subject of future investigations to provide a real-life context. Further work is also required to assess the vertical decorrelation of complex stimuli, in order to verify and generalize the observations made in the present study. Lastly, both subjective experiments suggest that the vertical decorrelation of low frequencies is unnecessary; considering this, it would be useful to investigate whether simply decorrelating high frequencies can control the VIS of a broadband signal.

7 REFERENCES

[1] Auro-3D, "Auro-3D Home Theater Setup: Installation Guidelines" (2015). URL: https://www.auro-3d.com/wp-content/uploads/documents/Auro-3D-Home-Theater-Setup-Guidelines_lores.pdf

[2] Dolby, "Dolby Atmos Home Theater Installation Guidelines" (2017). URL: https://www.dolby.com/us/en/technologies/dolby-atmos/dolby-atmos-home-theater-installation-guidelines.pdf

[3] DTS, "The First Step Toward 3D Audio: DTS Neo:X" (2012). URL: https://www.stormaudio.com/media/dtsneoxwhite_paper__019028000_1625_04032013.pdf

[4] C. Gribben and H. Lee, "A Comparison between Horizontal and Vertical Interchannel Decorrelation," *J. Applied Sciences*, vol. 7, no. 11, p. 1202 (2017). https://doi.org/10.3390/app7111202

[5] F. Zotter and M. Frank, "Efficient Phantom Source Widening," *Archives of Acoustics*, vol. 38, pp. 27–37 (2013). https://doi.org/10.2478/aoa-2013-0004

[6] M. R. Schroeder, "An Artificial Stereophonic Effect Obtained from a Single Audio Source," *J. Audio Eng. Soc.*, vol. 6, pp. 74–79 (1958 Apr.).

[7] T. Hidaka, L. L. Beranek, and T. Okano, "Interaural Cross-Correlation, Lateral Fraction and Low- and High-Frequency Sound Levels as Measures of Acoustical Quality in Concert Halls," *J. Acoust. Soc. Am.*, vol. 98, pp. 988–1007 (1995). https://doi.org/10.1121/1.412847

[8] R. Irwan and R. M. Aarts, "Two-to-Five Channel Sound Processing," *J. Audio Eng. Soc.*, vol. 50, pp. 914–926 (2002 Oct.).

[9] C. Avendano and J-M. Jot, "A Frequency-Domain Approach to Multichannel Upmix," *J. Audio Eng. Soc.*, vol. 52, pp. 740–749 (2004 Jul./Aug.).

[10] Y. Li and P. F. Driessen, "An Unsupervised Adaptive Filtering Approach of 2-to-5 Channel Upmix," presented at the *119th Convention of the Audio Engineering Society* (2005 Oct.), convention paper 6611.

[11] A. Adami, L. Brand, and J. Herre, "Investigations Towards Plausible Blind Upmixing of Applause Signals," presented at the *142nd Convention of the Audio Engineering Society* (2017 May), convention paper 9750.

[12] J. Herre, K. Kjörling, J. Breebaart, C. Faller, S. Disch, H. Purnhagen, J. Koppens, J. Hilpert, J. Rödén, W. Oomen, K. Linzmeier, and K. Seng Chong, "MPEG Surround—The ISO/MPEG Standard for Efficient and Compatible Multichannel Audio Coding," *J. Audio Eng. Soc.*, vol. 56, pp. 932–955 (2008 Nov.).

[13] A. Murtaza, J. Herre, J. Paulus, and L. Terentiv, "ISO/MPEG-H 3D Audio: SAOC-3D decoding and rendering," presented at the *139th Convention of the Audio Engineering Society* (2015 Oct.), convention paper 9434.

[14] H. Lauridsen, "Nogle forsøg reed forskellige former rum akustik gengivelse," *Ingeniøren*, vol. 47, p. 906 (1954).

[15] J. Breebart and C. Faller, *Spatial Audio Processing: MPEG Surround and Other Applications* (John Wiley, Chichester, 2007).

[16] S. K. Roffler and R. A. Butler, "Factors that Influence the Localization of Sound in the Vertical Plane," *J. Acoust. Soc. Am.*, vol. 43, pp. 1255–1259 (1968). https://doi.org/10.1121/1.1910976

[17] J. Hebrank and D. Wright, "Spectral Cues Used in the Localization of Sound Sources on the Median Plane," *J. Acoust. Soc. Am.*, vol. 56, pp. 1829–1834 (1974). https://doi.org/10.1121/1.1903520

[18] G. S. Kendall, "The Decorrelation of Audio Signals and its Impact on Spatial Imagery," *Computer Music J.*, vol. 19, pp. 71–87 (1995). https://doi.org/10.2307/3680992

[19] M-V. Laitinen, F. Kuech, S. Disch, and V. Pulkki, "Reproducing Applause-Type Signals with Directional Audio Coding," *J. Audio Eng. Soc.*, vol. 59, pp. 29–43 (2011 Jan./Feb.).

[20] M. Boueri and C. Kyirakakis, "Audio Signal Decorrelation Based on a Critical Band Approach," presented at the *117th Convention of the Audio Engineering Society* (2004 Oct.), convention paper 6291.

[21] C. Faller, "Parametric Multichannel Audio Coding: Synthesis of Coherence Cues," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, pp. 299–310 (2006). https://doi.org/10.1109/TSA.2005.854105

[22] V. Pulkki, "Spatial Sound Reproduction with Directional Audio Coding," *J. Audio Eng. Soc.*, vol. 55, pp. 503–516 (2007 Jun.).

[23] T. Pihlajamäki, O. Santala, and V. Pulkki, "Synthesis of Spatially Extended Virtual Sources with Time-Frequency Decomposition of Mono Signals," *J. Audio Eng. Soc.*, vol. 62, pp. 467–484 (2014 Jul./Aug.). https://doi.org/10.17743/jaes.2014.0031

[24] M. Fink, S. Kraft, and U. Zölzer, "Downmix-Compatible Conversion from Mono to Stereo in Time- and Frequency-Domain," presented at the *18th International Conference on Digital Audio Effects (DAFx-15)* (2015).

[25] C. Faller and F. Baumgarte, "Binaural Cue Coding—Part II: Schemes and Applications," *IEEE Transactions on Speech and Audio Processing*, vol. 11, pp. 520–531 (2003). https://doi.org/10.1109/TSA.2003.818108

[26] ITU-R, "Recommendations ITU-R BS.775-3: Multichannel stereophonic sound system with and without accompanying picture," International Telecommunications Union (2012).

[27] ITU-R, "Recommendations ITU-R BS.1116-3: Methods for the Subjective Assessment of Small Impairments in Audio Systems including Multichannel Sound Systems," International Telecommunications Union (2015).

[28] C. Gribben and H. Lee, "Towards the Development of a Universal Listening Test Interface Generator in Max," presented at the *138th Convention of the Audio Engineering Society* (2015 May), e-Brief 187.

[29] ITU-R, "Recommendations ITU-R BS.1534-3: Method for the Subjective Assessment of Intermediate Quality Level of Audio Systems," International Telecommunications Union (2015).

[30] R. McGill, J. W. Tukey, and W. A. Larsen, "Variations of Box Plots," *The American Statistician*, vol. 32, pp. 12–16 (1978). https://doi.org/10.1080/00031305.1978.10479236

[31] V. R. Algazi, C. Avendano, and R. O. Duda, "Elevation Localization and Head-Related Transfer Function Analysis at Low Frequencies," *J. Acoust. Soc. Am.*, vol. 109, pp. 1110–1122 (2001). https://doi.org/10.1121/1.1349185

[32] T. Robotham, M. Stephenson, and H. Lee, "The Effect of a Vertical Reflection on the Relationship between Preference and Perceived Change in Timbral and Spatial Attributes," presented at the *140th Convention of the Audio Engineering Society* (2016 May), convention paper 9547.

[33] J. Blauert, "Sound Localization in the Median Plane," *Acustica*, vol. 22, pp. 205–213 (1969/70).

[34] H. Lee, D. Johnson, and M. Mironovs, "A New Response Method for Auditory Localization and Spread Tests," presented at the *140th Convention of the Audio Engineering Society* (2016 May), e-Brief 240.

[35] H. Lee, "Perceptual Band Allocation (PBA) for the Rendering of Vertical Image Spread with a Vertical 2D Loudspeaker Array," *J. Audio Eng. Soc.*, vol. 64, pp. 1003–1013 (2016 Dec.). https://doi.org/10.17743/jaes.2016.0052

[36] A. Farina, "Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique," presented at the *110th Convention of the Audio Engineering Society* (2000 May), convention paper 5093.

[37] D. Johnson, A. Harker, and H. Lee, "HAART: A New Impulse Response Toolbox for Spatial Audio Research," presented at the *138th Convention of the Audio Engineering Society* (2015 May), e-Brief 190.

[38] W. L. Martens and D. A. Cabrera, "Perceived Elevation of Simultaneously Presented Sound Sources Depends upon the Correlation between the Source Signals," *J. Acoust. Soc. Am.*, vol. 341, no. 4, p. 3216 (2012).

[39] D. Stepanavicius and W. L. Martens, "Interaural Cross-Correlation Affects Perceived Elevation of Auditory Images Presented via Height Channels in Multichannel Audio Reproduction Systems," presented at *ACOUSTICS 2016, Second Australasian Acoustical Societies' Conference* (2016).

[40] R. Mason, T. Brookes, and F. Rumsey, "Creation and Verification of a Controlled Experimental Stimulus for Investigating Selected Perceived Spatial Attributes," presented at the *114th Convention of the Audio Engineering Society* (2003 Mar.), convention paper 5771.

[41] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*, Rev. Ed., (MIT Press, Cambridge (MA), London, 1997).

[42] C. J. Chun, H. K. Kim, S. H. Choi, S-J. Jang, and S-P. Lee, "Sound Source Elevation Using Spectral Notch Filtering and Directional Band Boosting in Stereo Loudspeaker Reproduction," *IEEE Transactions on Consumer Electronics*, vol. 57, pp. 1915–1920 (2011). https://doi.org/10.1109/TCE.2011.6131171

[43] R. Wallis and H. Lee, "The Effect of Interchannel Time Difference on Localization in Vertical Stereophony," *J. Audio Eng. Soc.*, vol. 63, pp. 767–776 (2015 Oct.). https://doi.org/10.17743/jaes.2015.0069

[44] R. Wallis and H. Lee, "Vertical Stereophonic Localization in the Presence of Interchannel Crosstalk: The Analysis of Frequency-Dependent Localization Thresholds," *J. Audio Eng. Soc.*, vol. 64, pp. 762–770 (2016 Oct.). https://doi.org/10.17743/jaes.2016.0039

THE AUTHORS

Christopher Gribben

Hyunkook Lee

Christopher Gribben recently completed his Ph.D. with the Applied Psychoacoustics Laboratory (APL) at the University of Huddersfield, UK. His thesis was on the perception of vertical interchannel decorrelation in 3D surround sound reproduction. Prior to this, he graduated with a First Class degree in music technology and audio systems from the same university. During his studies he undertook an industrial placement at Cass Allen Associates, UK, where his focus was on environmental acoustics and noise modelling. Since 2018 he has been working as an R&D engineer at Meridian Audio, UK. He is a member of the Audio Engineering Society.

•

Hyunkook Lee is Senior Lecturer (i.e., Associate Professor) in music technology and the leader of the Applied Psychoacoustics Lab (APL) at the University of Huddersfield, UK. From 2006 to 2010, Dr. Lee was Senior Research Engineer in audio R&D at LG Electronics, South Korea. He 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. His current research focuses on the perception, capturing, and processing of 3D and VR audio and the development of perceptually motivated virtual acoustics algorithm. Hyunkook is an active member of the Audio Engineering Society since 2001 and a fellow of the Higher Education Academy, UK.