The Perception of Band-Limited Decorrelation Between Vertically Oriented Loudspeakers

Christopher Gribben and Hyunkook Lee

***Abstract*—Two experiments have been conducted to investigate the perceptual effect of band-limited interchannel decorrelation between vertically oriented loudspeakers. The perceived vertical image spread (VIS) and tonal quality (TQ) of phantom auditory images have been subjectively assessed in multiple comparison trials. The aim of the article was to find a lower decorrelation boundary that provides a significant increase of VIS, whilst maintaining TQ close to that of the original source. For test stimuli, decorrelation was applied to natural sound sources and pink noise in groups of octave-bands, where the lowest band was varied between 63 Hz and 8 kHz and the upper band was fixed at 16 kHz, resulting in eight decorrelated conditions for each source. Unprocessed octave-bands below the lower boundary were reproduced simultaneously through the lower main-layer loudspeaker only, and a monophonic main-layer only condition was also included in the comparison alongside the decorrelated stimuli. Results reveal that vertical decorrelation of the 500 Hz octave-band and above tends to significantly increase VIS, similar to that of broadband decorrelation, with little impact on TQ. In some cases, decorrelation of higher octave-bands and above can also produce similar increases of VIS with less impact on TQ, however, this is shown to be largely source-dependent. These results suggest that vertical decorrelation of lower frequencies has little perceptual benefit, and band-limiting vertical decorrelation to higher frequencies is likely to reduce low frequency phase cancellation. Applications of such an approach include 2D-to-3D upmixing and binaural audio rendering, with additional implications for 3D audio recording.**

***Index Terms*—Digital signal processing, decorrelation, psychoacoustics, acoustics, acoustic applications, audio systems, loudspeakers.**

I. Introduction

IMMERSIVE audio via loudspeakers has gained a lot of interest over recent years. Loudspeaker reproduction formats that feature height-channels have become increasingly accessible to everyday consumers, through height-channel integration within many commercial cinemas and modern home-theatre systems. Examples of such multichannel audio formats include Dolby Atmos [2], Auro-3D [3] and DTS:X [4]. One particular question surrounding the development of 3D surround sound systems is how best to utilise the additional height-channels when reproducing existing content. Each of the mentioned audio formats have their own unique 'upmixing' algorithm to achieve a 3D output from a two-channel or 5.1 Surround input. These algorithms use digital signal processing to create new signals for the additional surround- and height-channels, generating audio that is typically ambient and largely uncorrelated with the input signal(s).

Decreasing the interchannel cross-correlation (ICC) between two signals is a common requirement for upmixing and spatial coding algorithms. This process is known as 'decorrelation,' and can be achieved through subtle changes to the interchannel phase and/or spectral-amplitude relationship [5], [6]. It is widely accepted that decorrelation in the horizontal plane (i.e. between a spaced pair of left and right loudspeakers) perceptually increases the horizontal spread of the phantom auditory image [7]. Given the recent need to upmix signals vertically for 3D multichannel systems, it is of interest to observe whether a similar perceptual effect exists in the vertical domain. That is, investigating whether a decrease of interchannel correlation between a vertically-spaced pair of loudspeakers results in an increase of vertical image spread (VIS).

The authors have previously found that a significant increase of VIS by vertical decorrelation is perceived for octave-band pink noise with centre frequencies of 500 Hz and above [8]. Furthermore, the same study demonstrated that the effectiveness of vertical decorrelation is dependent on the azimuth angle position of the vertically-spaced loudspeaker pair (where identical stimuli were judged from three independent azimuth angles: 0°, 30° and 110°).

The current study features two subjective experiments that build on these initial findings. In both experiments, interchannel decorrelation has been band-limited to higher frequencies and assessed from the same three azimuth angles (0°, 30° and 110°). The first experiment looks at the relative change of VIS from the band-limited vertical interchannel decorrelation, while the second assesses the tonal quality (TQ) of the same decorrelated stimuli. The stimuli consist of natural ambient sound sources (in addition to pink noise), as this is thought to be representative of a real-life upmixing scenario.

Since the previous investigation suggested that changes to VIS are most apparent at higher frequencies (the 500 Hz octave-band and above) [8], the present study has a particular focus on

Manuscript received June 6, 2019; revised September 17, 2019 and December 31, 2019; accepted January 11, 2020. Date of publication January 29, 2020; date of current version March 6, 2020. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Augusto Sarti. This paper was presented in part (vertical image spread (VIS) results) at the AES International Conference on Immersive and Interactive Audio, University of York, York, U.K., March 2019. This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) under Grant EP/L019906/1. *(Corresponding author: Christopher Gribben.)*

C. Gribben is with Meridian Audio Ltd., Huntingdon, Cambridgeshire PE29 6YE, U.K. (e-mail: chris.gribben@meridian.co.uk).

H. Lee is with the Applied Psychoacoustics Laboratory (APL), University of Huddersfield, Huddersfield, West Yorkshire HD1 3DH, U.K. (e-mail: h.lee@hud.ac.uk).

Digital Object Identifier 10.1109/TASLP.2020.2969845

2329-9290 © 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

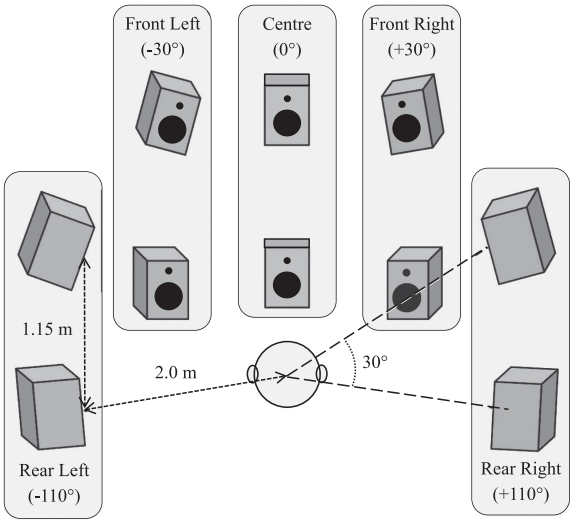

Fig. 1. Physical loudspeaker setup used during testing (based on Auro-3D 9.1 [3] with an additional centre height-channel). Five main-layer loudspeakers positioned 2 m from the listener at ear-height with azimuth angles of 0°, $\pm 30^\circ$ and $\pm 110^\circ$. Five upper height-layer loudspeakers elevated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

whether vertically decorrelating the upper frequencies of a signal can produce a similar effect to decorrelating the entire broadband signal, i.e., where mid-high octave-bands are decorrelated between a main- and height-layer loudspeaker pair, while the lower frequencies are routed monophonically to the main-layer loudspeaker only. The main aim is to find a lower frequency boundary for vertical decorrelation where a significant increase of VIS is achieved, whilst still maintaining TQ close to that of the original source.

Results from the previous study have also suggested that potential spectral cues from vertical decorrelation feature in the 8–16 kHz octave-bands at 0° azimuth, the 4–16 kHz octave-bands at 30° azimuth and the 2–16 kHz octave-bands at 110° [8]. If these spectral cues are particularly dominant for VIS perception, it could be found that vertical decorrelation of lower frequencies is not necessary for increasing VIS. Avoiding the decorrelation of low frequencies would also reduce the risk of phase-cancellation when two decorrelated signals are summed at the ear. As a result, a high-frequency band-limited approach to vertical decorrelation is likely to improve TQ over broadband vertical decorrelation.

The remainder of this paper is organised as follows. Section II describes the experimental design of the study, featuring the physical loudspeaker setup, the creation of stimuli and the testing procedure. Section III presents and analyses the results from the VIS part of the experiment, while Section IV presents and analyses the results from the TQ part. Both sets of results are then discussed in Section V, followed by a summary of the conclusions in Section VI.

II. Experimental Design

A. Physical Setup

The loudspeaker format used during testing is based on Auro-3D 9.1, with the addition of a centre height-channel loudspeaker (Fig. 1) [3]. This resulted in the discrete presentation of stimuli through vertically-spaced loudspeaker pairs at three azimuth angles to the listener: 0°, 30° and 110°. A total of ten Genelec 8040A loudspeakers (Frequency response: 48 Hz – 20 kHz (± 2 dB)) were used during testing. The five main-layer loudspeakers were positioned at a distance of 2 m from the listener at ear-height, with the height-layer loudspeakers positioned directly above the main-layer loudspeakers at an elevation angle of $+30^\circ$ to the listener (vertically-spaced by 1.15 m).

Testing was conducted at the University of Huddersfield in a critical listening room that fulfils the specification of ITU-R BS.1116-3 [9] (6.2 m $\times$ 5.6 m $\times$ 3.8 m; RT $=$ 0.25 s; NR 12). All main- and height-channel loudspeakers were time- and level-aligned at the listening position using impulse response and sound pressure level (SPL) measurements. An acoustically transparent curtain was also used to obscure the loudspeakers from view, so as to avoid visual bias during testing.

B. Stimuli Creation

The aim of the current investigation was to examine the spatial and tonal effects of vertically decorrelating solely the upper frequencies of a broadband signal, where the lower frequency boundary of the decorrelation band was as an independent variable of each trial. It was of interest to observe the effect in a practical context, therefore, ambient natural sources, as well as noise, were chosen as stimuli. In a typical upmixing scenario, it is the ambient part of an input signal that contributes most to the newly generated surround- and height-channel signals [10].

The original source signals for the current experiments comprised broadband Pink Noise and five anechoic recordings ($f_s =$ 44.1 kHz): Male Speech, Cello, Acoustic Guitar, a Drumkit and a String Quartet. Ambient signals were generated by applying the same artificial reverb (RT $=$ 2 s) to each of the natural sound sources (i.e. all but the Pink Noise). The reverb was applied using the 'ReaVerb' plug-in in the Reaper digital audio workstation, and the dry signal was removed from the output of the reverb to reproduce the ambient wet signal only. Algorithmic reverb was chosen over convolution with a real room impulse response as more control could be had over the frequency response of the output. It was important to maintain as much high frequency energy as possible, in order to fully assess the effect of decorrelation at high frequencies. Spectrograms of the resultant ambient stimuli after the reverb processing are presented in Fig. 2 (calculated using the short-time Fourier transform (STFT)).

The lower boundary of the vertical decorrelation was determined by octave-bands with centre frequencies ranging from 63 Hz to 8 kHz. This resulted in eight decorrelation conditions, where each octave-band condition featured decorrelation of that band and every band above it, up to the 16 kHz octave-band. For example, the 63 Hz condition signifies 'Broadband' decorrelation where every octave-band between 63 Hz and 16 kHz is decorrelated, while the 8 kHz condition only features decorrelation of the 8 kHz and 16 kHz octave-bands. A diagram demonstrating the splitting of octave-bands described here can be seen in Fig. 3, using the 1 kHz and above condition as an example.

For the decorrelation process, each of the six source stimuli were filtered into octave-bands (centre frequencies from 63 Hz

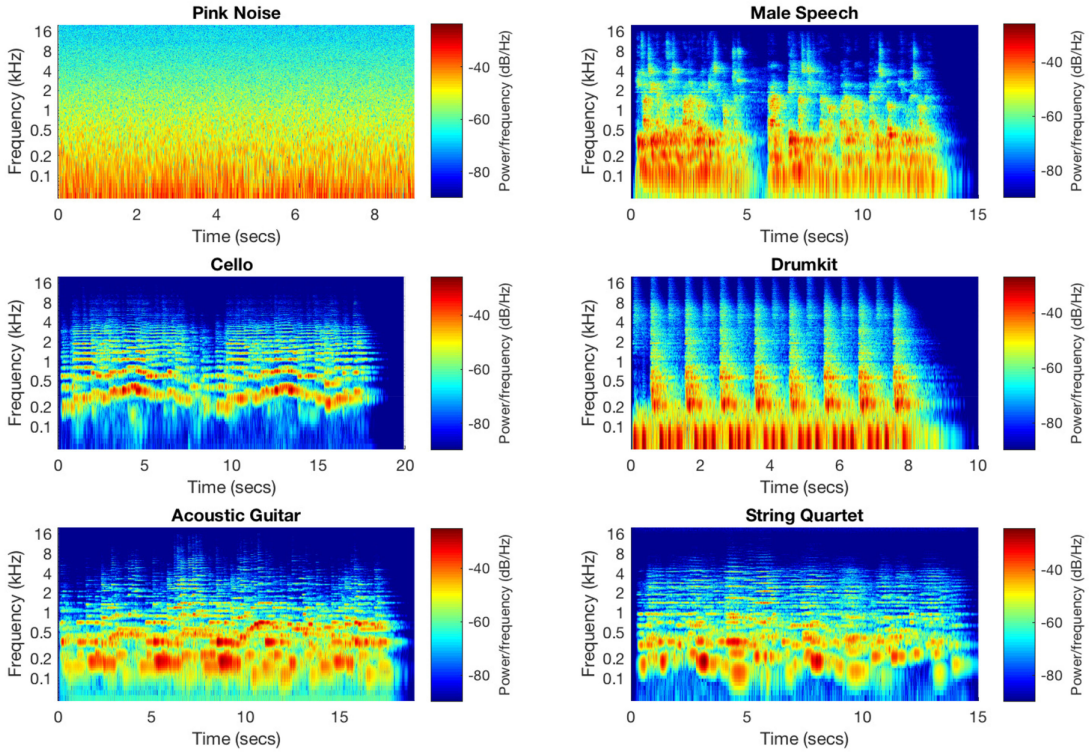

Fig. 2. Spectrogram showing the short-time Fourier transform (STFT) of the ambient source signals ($f_s = 44.1$ kHz); 4096 FFT-points calculated with a frame length of 1024 samples and 50% overlapping windows.

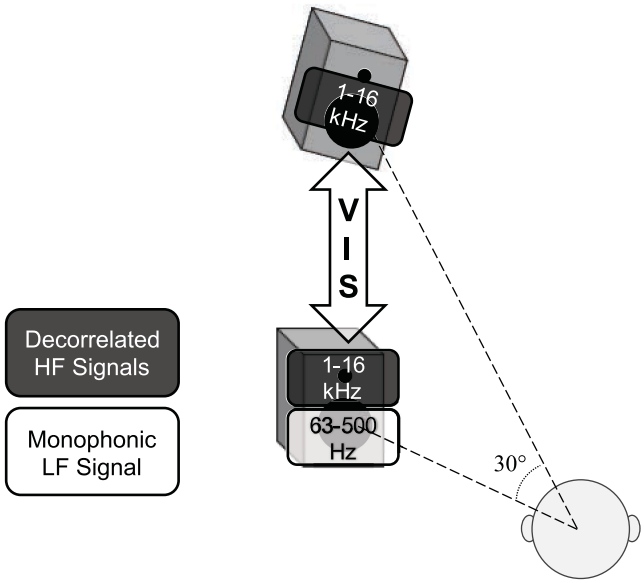

Fig. 3. Diagram demonstrating the splitting of octave-bands and signal routing, using the '1 kHz +' decorrelation condition as an example (i.e. where octave-bands with centre frequencies of 1–16 kHz are decorrelated between the main- and height-layer loudspeakers, while octave-bands with centre frequencies of 63–500 Hz are routed to the main-layer only).

to 16 kHz) using 16th-order linear phase Butterworth band-pass filters (96 dB/octave). The octave-band filtered signals were then decorrelated independently using an all-pass filter phase-randomisation method proposed by Kendall [5]. This approach involves convolving the source signal with two decorrelated impulses (short white noise bursts) of random phase and unit magnitude (Eqs. (1) and (2)). These impulses represent FIR all-pass filters, where the amplitude frequency response is identical between input and output.

$$s_1(n) = x(n) * h_1(n) \tag{1}$$

$$s_2(n) = x(n) * h_2(n) \tag{2}$$

To create the random phase coefficients of the FIR all-pass filters, random sequences of 1323 numbers between $-\pi$ and π were generated for each filter. The random numbers represent the phase component of each FFT bin in the frequency-domain, while the magnitude of each frequency bin is set to unity ($= 1$). The impulse response of a filter can then be obtained by performing an inverse FFT (IFFT) on these frequency components, resulting in a filter length of 30 ms (1323 taps at 44.1 kHz), as used in a previous study [8].

An interchannel cross-correlation coefficient running average (ICCC_{avg}) below 0.3 was achieved for each octave-band filtered signal (of every source stimulus)—the mean ICCC_{avg} for each octave-band across all sources was 0.19 (± 0.05 between sources). A script in MATLAB generated the two all-pass filters and convolved them with the source signal. This script was then looped until the two decorrelated output signals were calculated to be below the 0.3 ICCC_{avg} threshold.

The ICCC is defined as the maximum absolute value of the interchannel cross-correlation function (ICCF) (Eqs. (3) and (4)). A 50 ms window length was used since it has previously been

suggested as optimal for interaural cross-correlation coefficient (IACC) calculation, based on the temporal resolution of the auditory system [11]—it is thought that this is also applicable for ICCC calculation in the current context of auditory perception. The lag time (τ) was set to zero as all signals were time-aligned.

$$ICCF(\tau) = \frac{\int_{-\infty}^{\infty} s_1(t)s_2(t+\tau)\,dt}{\sqrt{[\int_{-\infty}^{\infty} s_1^2(t)\,dt][\int_{-\infty}^{\infty} s_2^2(t)\,dt]}} \tag{3}$$

$$ICCC = \max|ICCF(\tau)| \tag{4}$$

The two decorrelated output signals for each octave-band were RMS level-matched with the monophonic input signal and attenuated by -3 dB, in order to match the summed energy of the two-channel output with the energy of the monophonic source. That is, correcting for the $+3$ dB boost that occurs when the two decorrelated source signals are summed together in vertical stereophony at the ear. For each decorrelation condition, one decorrelated signal was routed to the main-layer and the other to the height-layer. The original monophonic octave-band signals that had not been decorrelated were then routed to the main-layer loudspeaker at ear-height only, as illustrated in Fig. 3. This octave-band decorrelation approach was chosen over simple high-pass filtering as a consistent ICCC_{avg} could be achieved for each band, while also maintaining the tonal characteristics of each band between the different decorrelation conditions.

All test conditions were level-matched in terms of L_{Aeq} within each source. The playback level of each source was determined through critical listening of the test stimuli by the authors, in order to match the perceived loudness between sources (ranging from 68 to 73 dB L_{Aeq}).

C. Testing Procedure

The same subjective testing procedure was performed on both VIS and TQ to collect comparable quantitative data. For each attribute, stimuli were presented in multiple comparison trials using an adapted version of the MUSHRA format [12]. Each trial featured nine stimuli, consisting of the eight decorrelated stimuli described in Section II-B above (with the varying octave-band cut-offs) and a 'Monophonic' condition. The 'Monophonic' condition was the original source stimulus reproduced from the lower main-layer loudspeaker only; this was included to represent the practical application of 2D-to-3D upmixing, where a monophonic main-layer signal is decorrelated vertically to generate a new height-channel signal.

A total of 18 trials were conducted for each attribute (VIS and TQ)—one for every source and loudspeaker azimuth angle combination (6×3). An adaptation of HULTI-GEN was used to create the graphical user interface (GUI) for testing, as shown in Fig. 4 [13]. Each of the nine stimuli under test were looped in synchrony and the user could freely switch between them throughout. Subjects were asked to grade the stimuli against each other and the reference stimulus (i.e. the 'Monophonic' condition). The sliders for grading the stimuli were on a continuous bipolar scale, ranging from -30 to $+30$, with the reference located at 0 on the scale. Each set of 18 trials were split over two sessions, totalling four sittings of around 30 minutes each to complete both parts.

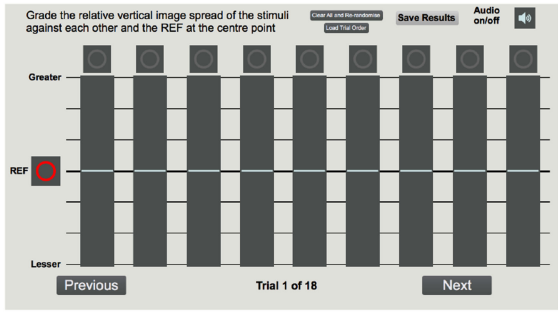

Fig. 4. Graphical user interface (GUI) used during the subjective testing, adapted from HULTI-GEN in Max [13].

For the grading of VIS, if the perceived extent of VIS was 'Greater' than the reference, subjects would grade that stimulus above 0 on the scale, and if it was perceived as 'Lesser,' it would be graded below. A familiarisation stage was also carried out before the VIS testing, which featured broadband pink noise sources of varying ICCC between a main- and height-channel loudspeaker pair. This familiarisation process introduced VIS to the listener and demonstrated the spatial changes that might be expected.

For the TQ part, TQ was considered as a global attribute that relates to the timbre and spectral balance of a stimulus, specifically in comparison to the unprocessed 'Monophonic' reference. To define TQ, subjects were presented with a list of opposing sub-attributes and corresponding definitions [14]. These sub-attributes were as follows: 'Clear / Muddy,' 'Natural / Unnatural,' 'Full / Thin,' 'Hard / Soft,' 'Bright / Dull' and 'Loud / Quiet,' along with definitions for 'Distortion' and 'Phasiness'. If a stimulus had perceptually 'Better' TQ than the reference, it would be graded above 0, and if it was perceived as having 'Worse' TQ, it would be graded below. To provide further insight into the subjective perception of TQ and support the quantitative results, subjects were also asked to write sub-attributes that best describe the highest- and lowest-rated samples for each TQ trial (not limited to those provided).

D. Subjects

In total, 17 subjects participated in the VIS part of the experiment, and 13 of the same subjects participated in the TQ part. Participants were comprised of staff, postgraduate and undergraduate researchers with the Applied Psychoacoustics Laboratory (APL) at the University of Huddersfield. All subjects reported normal hearing and were experienced in critical listening tests for spatial audio quality evaluation.

III. Experiment 1: Vertical Image Spread (VIS)

The relative vertical image spread (VIS) results can be seen in Fig. 5—all data was normalised in accordance with ITU-R BS.1116-3 [9] and analysed using the software package IBM SPSS Statistics V23.0 [16]. Shapiro-Wilk tests for normality

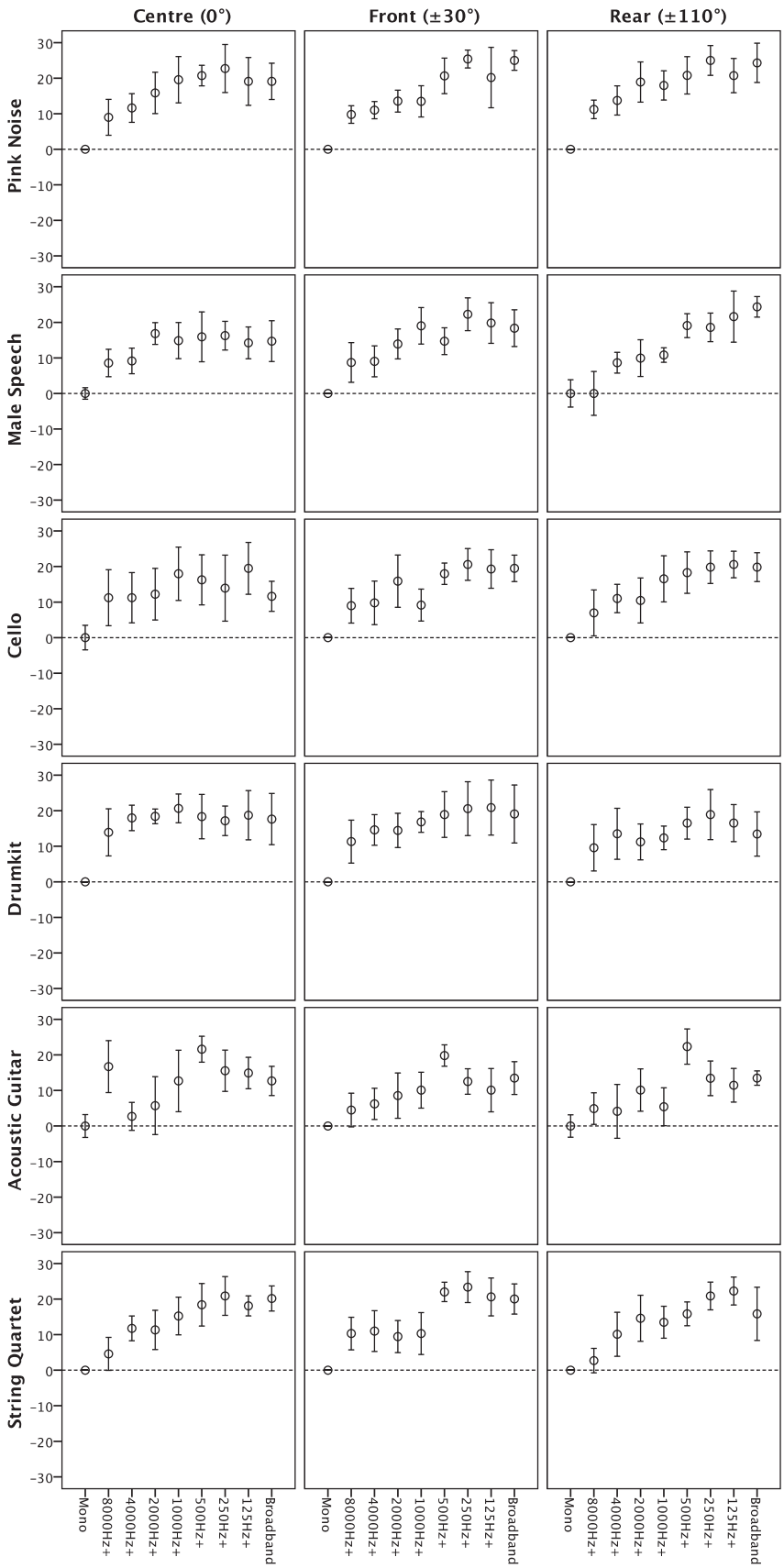

Fig. 5. Results of Experiment 1: Median scores of vertical image spread (VIS) ratings with non-parametric 95% confidence intervals [15].

indicated that the data of each condition was not always normally distributed—as a result, non-parametric statistical tests have been performed on all conditions. The graphs display the median VIS with notch edge bars (a non-parametric equivalent of 95% confidence intervals [15]). In the plots and following results, the format 'XXX Hz +' indicates decorrelation of the XXX Hz octave-band and above.

Friedman repeated measure ANOVA tests [17] assessed the main effect of the decorrelation octave-band cut-off on VIS perception. These tests included all eight decorrelated conditions (excluding the 'Monophonic' condition), and were performed for each source stimulus and loudspeaker angle combination. If a significant decorrelation cut-off effect was observed, post-hoc Wilcoxon pairwise tests with Bonferroni correction [14], [17] were performed to ascertain where the significant difference occurs. Further Wilcoxon tests with Bonferroni correction were conducted between the 'Monophonic' condition and each decorrelated condition, in order to reveal any significant increases of VIS for potential upmixing applications.

Analysis of the results for the Pink Noise source indicate a significant main effect of decorrelation cut-off on VIS for all three azimuth angles (0°, ±30° and ±110°) ($p < 0.01$). The pairwise tests reveal that there was no significant difference between any decorrelated conditions at 0° azimuth ($p > 0.05$), and no significant difference between 'Broadband,' '125 Hz +,' '250 Hz +' and '500 Hz +' at ±30° and ±110° ($p > 0.05$). Additional pairwise tests at 0°, ±30° and ±110° show that all decorrelated stimuli had significantly greater VIS than the 'Monophonic' condition ($p < 0.04$).

The Male Speech VIS data reveals a significant decorrelation cut-off main effect for the ±30° and ±110° azimuth angles ($p < 0.01$), but not 0° ($p > 0.05$). The pairwise tests show no significant difference between the 'Broadband,' '125 Hz +,' '250 Hz +' and '500 Hz +' conditions at 0°, ±30° and ±110° ($p > 0.05$), with each of these conditions perceived as significantly greater than the 'Monophonic' condition ($p < 0.05$).

Results for the Cello source indicate a significant decorrelation cut-off effect for the ±110° azimuth angle ($p < 0.01$), but not the 0° and ±30° azimuths ($p > 0.05$). Pairwise tests on the 0° data suggest that only '125 Hz +' was significantly greater than the 'Monophonic' condition ($p = 0.03$). Whereas, from ±30°, the 'Broadband,' '125 Hz +,' '250 Hz +,' '500 Hz +' and '4 kHz +' conditions all had significantly greater VIS than the 'Monophonic' condition ($p < 0.04$). For the ±110° data, there was no significant difference between any of the decorrelation conditions ($p > 0.05$), and all decorrelated conditions had significantly greater VIS than the 'Monophonic' condition ($p < 0.02$).

Analysis of the results for the Drumkit sample indicate a significant decorrelation cut-off effect for the ±30° and ±110° azimuth angles ($p < 0.05$), but not from 0° ($p > 0.05$). The pairwise test results show no significant difference between any decorrelated conditions for all azimuth angles ($p > 0.05$). However, when compared against the 'Monophonic' condition, all decorrelated conditions had significantly greater VIS from all angles ($p < 0.05$), except for 2 kHz at ±110° ($p > 0.05$).

The Acoustic Guitar analysis results demonstrate a significant effect for decorrelation cut-off at all azimuth angles ($p < 0.05$). From 0°, '500 Hz +' had significantly greater VIS than '4 kHz +' ($p < 0.03$), and the '500 Hz +' condition was the only decorrelated condition with significantly greater VIS than the 'Monophonic' condition ($p < 0.01$). With the ±30° results, the '500 Hz +' condition was significantly greater than '1 kHz +' and '8 kHz +' ($p < 0.03$); and for ±110°, '500 Hz +' was significantly greater than '1 kHz +,' '2 kHz +' and '8 kHz +' ($p < 0.03$). From both ±30° and ±110° azimuth, the 'Broadband,' '250 Hz +' and '500 Hz +' conditions all had significantly greater VIS than the 'Monophonic' condition ($p < 0.05$).

Results for the String Quartet sample show a significant decorrelation cut-off effect at all azimuth angles ($p < 0.01$). For all azimuth angles, the 'Broadband,' '125 Hz +,' '250 Hz +' and '500 Hz +' conditions were not significantly different from one another ($p > 0.05$), and each of these conditions had significantly greater VIS than the 'Monophonic' condition at 0° ($p < 0.05$).

IV. Experiment 2: Tonal Quality (TQ)

The results for the relative tonal quality (TQ) part of the investigation can be seen in Fig. 6. All data was normalised in accordance with ITU-R BS.1116-3 [9] and analysed using the software package IBM SPSS Statistics V23.0 [16]. Shapiro-Wilk tests indicated that not all conditions had normally distributed data, therefore, non-parametric statistical tests have been performed on all groups of data. In the graphs, the median TQ scores are plotted with notch edges (a non-parametric equivalent of 95% confidence intervals [15]). As with the VIS results, Friedman repeated measure ANOVA tests [17] were conducted on the data, in order to observe the main effect of decorrelation cut-off on TQ. If a significant cut-off effect was revealed, post-hoc Wilcoxon tests with Bonferroni correction [14], [17] were carried out between the decorrelated stimuli to determine any significant difference. Wilcoxon tests were also conducted between the decorrelated conditions and the 'Monophonic' condition, in order to assess for significant differences from the unprocessed reference.

Analysis of the Pink Noise data indicates a significant decorrelation cut-off main effect at all azimuth angles ($p < 0.05$). However, the Wilcoxon tests demonstrate that only the 'Broadband' condition had significantly worse TQ than the '4 kHz +' condition at 0° azimuth ($p < 0.03$). Additional Wilcoxon tests show that the 'Broadband,' '125 Hz +,' '250 Hz +' and '500 Hz +' conditions all had significantly worse TQ than the 'Monophonic' condition at 0° ($p < 0.05$); and from ±30°, the '125 Hz +' was also perceived as significantly worse than the 'Monophonic' condition ($p < 0.04$). Comments relating to all of the conditions with significantly worse TQ refer to the stimuli as 'Phasey' and 'Thin'—this potentially indicates a loss of frequencies when the decorrelated signals are summed at the ear. Although it is difficult to subjectively judge the TQ of pink noise (due to its unnatural features), it is clear that as the cut-off frequency is decreased (i.e. as more low frequencies are decorrelated), the perceived TQ appears to decrease almost

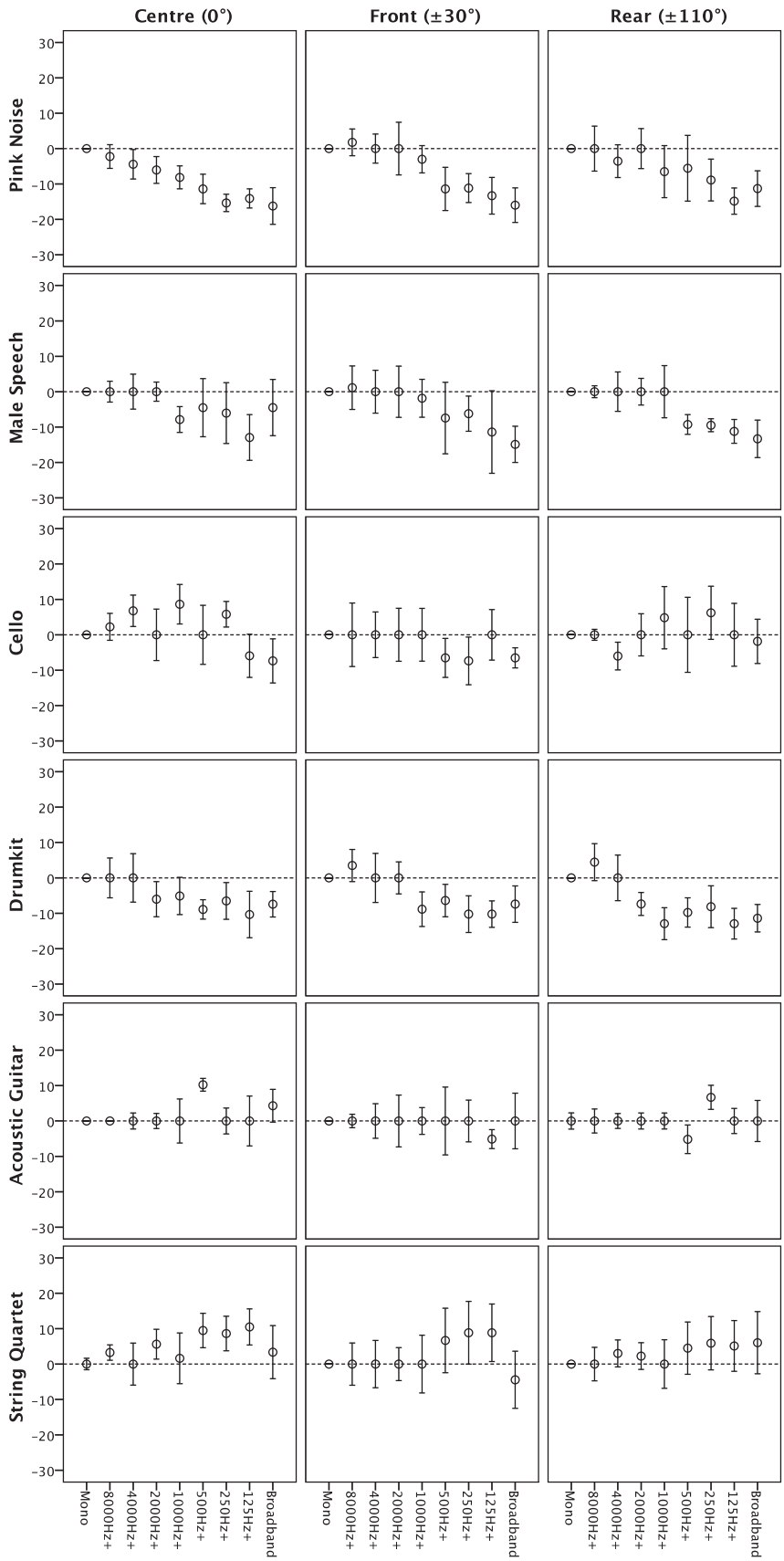

Fig. 6. Results of Experiment 2: Median scores of tonal quality (TQ) ratings with non-parametric 95% confidence intervals [15].

linearly, with a greater variation of responses as the azimuth angle increases.

The Male Speech results indicate that decorrelation cut-off has a significant main effect on TQ at $\pm 110^\circ$ azimuth ($p < 0.05$), but not at 0° and $\pm 30^\circ$ ($p > 0.05$). However, the Wilcoxon tests show no significant difference between the decorrelated conditions from all azimuth angles ($p > 0.05$). Compared against the 'Monophonic' condition, 'Broadband' at $\pm 110^\circ$ is the only decorrelated condition to be graded as significantly worse ($p < 0.03$). Comments for the 'Broadband' stimulus refer to terms such as 'Muddy,' 'Distorted,' 'Unnatural' and 'Dull,' suggesting that decorrelation of lower frequencies may have an impact on the intelligibility of speech sources. Observing the plots in Fig. 6, a general decrease of TQ is seen as the decorrelation cut-off frequency is decreased, similar to that of the Pink Noise sample—however, unlike the Pink Noise source, this decrease of TQ appears to be most prominent at $\pm 110^\circ$ azimuth.

With the Cello source, there is no significant decorrelation cut-off main effect at all azimuth angles ($p > 0.05$). There was also no significant difference between the 'Monophonic' condition and all of the decorrelated conditions ($p > 0.05$). Looking at the graphs in Fig. 6, a slight improvement of TQ is seen with some decorrelation conditions at both 0° and $\pm 110^\circ$ azimuth. The most common terms for the Cello sample with the best TQ were 'Full,' 'Clear' and 'Natural,' suggesting there may be some tonal benefit to vertical decorrelation.

The results for the Drumkit sample indicate a significant decorrelation cut-off main effect on TQ at all azimuth angles ($p < 0.05$). However, the Wilcoxon tests show no significant difference between any of the decorrelation conditions ($p > 0.05$). Additional Wilcoxon tests also demonstrate that the 'Broadband,' '125 Hz +' and '1 kHz +' conditions had significantly worse TQ than the 'Monophonic' condition at $\pm 110^\circ$ ($p < 0.05$). As with the Pink Noise and Male Speech sources, a general decrease of TQ as the cut-off frequency decreases can also be seen in Fig. 6 for the Drumkit. The comments for the Drumkit stimuli with significantly worse TQ were also similar to those for the Pink Noise and Male Speech: 'Phasey,' 'Thin,' 'Unnatural,' 'Distorted' and 'Muddy'.

Analysis of the Acoustic Guitar data reveals a significant decorrelation cut-off main effect for the 0° azimuth angle ($p < 0.05$), but not at $\pm 30^\circ$ and $\pm 110^\circ$ ($p > 0.05$). The post-hoc Wilcoxon tests suggest that the differences between the decorrelated conditions are not significant ($p > 0.05$). Observing the graphs in Fig. 6, it is seen that most conditions see no change of TQ from the 'Monophonic' reference. Some slight increase of TQ is observed for the '500 Hz +' condition at 0° and '250 Hz +' at $\pm 110^\circ$.

The results for the String Quartet sample show a significant main effect of decorrelation cut-off for the 0° azimuth angle ($p < 0.05$), but not $\pm 30^\circ$ and $\pm 110^\circ$ ($p > 0.05$). From 0°, Wilcoxon tests reveal that the '125 Hz +' condition had significantly better TQ than both the '8 kHz +' and 'Monophonic' conditions ($p < 0.04$). In contrast to the other samples, TQ generally appears to increase slightly as the decorrelation cut-off decreases (i.e. when more low frequencies are decorrelated). The subject comments for the significant '125 Hz +' condition suggest that decorrelation made the sample 'Warmer,' 'Softer,' 'Livelier,' 'Fuller' and 'Clearer'. This increase of TQ could be due to the musical nature of the source, where the frequency content from multiple parts varies over time, potentially leading to a richer sound.

V. Discussion of Results

From the VIS results presented in Fig. 5 and described in Section III, no significant VIS difference was seen between the 'Broadband,' '125 Hz +,' '250 Hz +' and '500 Hz +' conditions for all source signals and azimuth angles. Moreover, the '500 Hz +' condition also had significantly greater VIS than the 'Monophonic' condition for all sources and azimuth angles, except the Cello source at 0° azimuth. This suggests that vertical decorrelation of only the 500 Hz octave-band and above can significantly increase VIS in the vast majority of cases, similar to that of broadband decorrelation. Fig. 5 also indicates that a significant increase of VIS over the monophonic condition can be achieved with higher octave-band cut-offs, as discussed in Section V-A below.

Further to the VIS results, the TQ results presented in Fig. 6 and described in Section IV suggest that decorrelation of the 500 Hz octave-band and above has little negative impact on TQ. There is no significant difference between the 'Monophonic' unprocessed condition and the '2 kHz +,' '4 kHz +' and '8 kHz +' conditions for all sources, while the '500 Hz +' condition was only perceived as significantly worse than the 'Monophonic' reference for the Pink Noise source at 0°. Furthermore, the only complex sources to see a significant decrease of TQ from the reference at all were the Male Speech and Drumkit samples (both at $\pm 110^\circ$), which might be related to the character of speech and drum signals. For example, the familiarity of speech and the transient response of the drums may have revealed signal degradation more clearly.

In general, the TQ results appear to be particularly source-dependent, as discussed further in Section V-A below. It would seem that the optimal lower boundary for band-limited decorrelation is likely to be a compromise between maximum VIS and improved TQ. Observing the results above, it is thought that the lowest band required for maximum VIS across all sources is the '500 Hz +' condition. However, looking at the TQ results, the '1 kHz +' and '2 kHz +' conditions have noticeably better TQ than the '500 Hz +' in some cases. As previously mentioned, the TQ of the Male Speech sample is relatively poor for '500 Hz +,' yet improves considerably when using higher decorrelation cut-offs. Moreover, the VIS results for the Male Speech sample also show that '1 kHz +' and '2 kHz +' perform similarly to '500 Hz +'. So in this case, it is likely that a higher decorrelation cut-off would be preferred for increased vocal clarity. A similar trend can also be seen for the Drumkit sample, where '2 kHz +' and '4 kHz +' show improved TQ whilst producing similar levels of VIS to the '500 Hz +' and 'Broadband' conditions.

A. Sound Source Dependency

Observing the VIS results in Fig. 5, it is seen that the Cello sample generally has greater error bars at 0° compared to other positions, suggesting that grading the Cello at this angle may

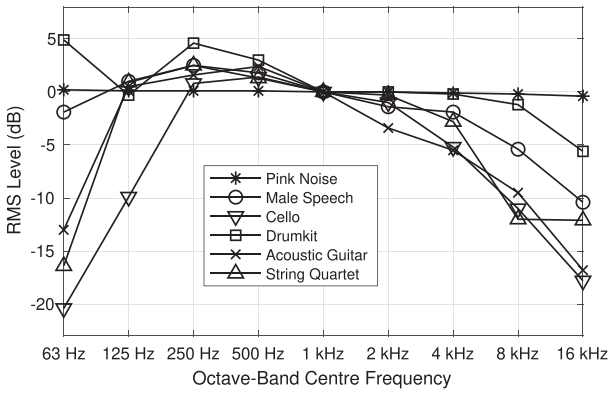

Fig. 7. Octave-band root mean square (RMS) energy values for the source signals (dB), normalised to 0.0 dB RMS at 1 kHz.

have been more difficult. A reason for this could be that the Cello is the only musical source with a monophonic melody, whereas the Acoustic Guitar and String Quartet samples have multiple parts playing at different frequencies (as visualised in Fig. 2). It is possible that the different musical lines are able to excite multiple VIS cues simultaneously, particularly when parts with greater energy at higher frequencies are present. For example, an inherent vertical spread may have occurred due to the different frequencies (melodic parts) being perceived from different heights, i.e., the pitch-height effect [18]. This hypothesis has previously been suggested by the authors in earlier studies, where a potential pitch-height effect was observed when vertically decorrelating band-limited pink noise at 0° azimuth [8], [19], however, additional investigations are required to assess the hypothesis further.

All samples except the Acoustic Guitar see some cases of significant VIS increase for decorrelation cut-offs higher than the 500 Hz octave-band. This demonstrates that decorrelation of even higher frequencies alone can have a significant impact on VIS perception, however, the effectiveness appears to be largely source-dependent. In order to compare the distribution of frequency energy between the different source signals, octave-band root mean square (RMS) energy values (normalised to 0.0 dB RMS for the 1 kHz octave-band) are presented in Fig. 7. It is seen that energy for the Acoustic Guitar sample is greatest in the 500 Hz octave-band, with relatively lower energy in the 2 kHz and 4 kHz bands compared to the other samples. This lack of high frequency energy may have dictated the insignificant change of VIS that was observed for higher decorrelation cut-offs with the Acoustic Guitar sample.

The results for the Drumkit and Pink Noise samples suggest that decorrelating solely the 8 kHz and 16 kHz octave-bands can significantly increase VIS over a monophonic signal. The octave-band RMS values in Fig. 7 show that the Drumkit and Pink Noise sources have relatively greater energy than the other stimuli within these bands. It is therefore assumed that effective vertical decorrelation with the '8 kHz +' condition is dependent on the source signal having sufficient high frequency energy. It is already known that pinna filtering within the 8 kHz octave-band provides strong cues for vertical localisation from the front [20], [21], with the results here suggesting that this frequency region may also contribute to the perception of VIS. Spectral analysis in a previous study by the authors provides support for this hypothesis, where it is shown that high frequency spectral notches are 'filled in' as two signals are decorrelated vertically [8].

Comparing loudspeaker positions, a similar relationship between the decorrelation cut-off and VIS is seen for all azimuth angles. That is, where VIS generally increases as the octave-band cut-off decreases (i.e. the bandwidth of decorrelation increases). A previous investigation by the authors on the absolute grading of VIS demonstrated that vertical decorrelation of octave-band pink noise is perceived differently depending on the frequency band and azimuth angle [8]. Since the results presented here are broadly similar for each angle, it suggests that the vertical decorrelation of multiple octave-bands may each contribute to the same broadband perception of VIS, regardless of presentation location. Furthermore, when the number of decorrelated octave-bands increases, the extent of perceived VIS generally increases too, indicating that VIS takes cues from multiple frequencies simultaneously. From this, it might be assumed that decorrelation of the 8 kHz and 16 kHz octave-bands still provides some contribution to the perception of broadband vertical decorrelation, even when sources lack high frequency energy.

B. *Influences of Frequency Spectrum*

Spectral analysis of the stimuli source signals has suggested that decorrelation by phase-randomisation may result in spectral distortion at the ear input. This is likely to occur when the randomisation process generates two phase components that are considerably out-of-phase with one another at a particular frequency, resulting in a cancellation effect when the two signals are summed. This effect is shown in Fig. 8, where the unprocessed monophonic source signals have been subtracted from the direct summation of the lower and upper source signals for the 'Broadband' decorrelation condition of each stimulus. These plots demonstrate how the frequency response of the summed outputs can vary throughout the spectrum, with the distortion appearing to be greater at lower frequencies in general. The subjective VIS results in Fig. 5 suggest that this low frequency decorrelation distortion could be avoided by decorrelating solely the mid-high frequencies of a broadband signal. This is supported by the results in Fig. 6, where TQ is broadly unaffected by decorrelation of the 500 Hz octave-band and above.

Fig. 7 shows that the Pink Noise, Male Speech and Drumkit sources have relatively greater energy in the 63 Hz band compared to the other samples. Given this increase of low frequency energy, it is possible that the low frequency distortions in Fig. 8 were more noticeable, resulting in the decrease of perceived TQ observed for these sources in Fig. 6. Relating this to the subjects' comments on TQ, a common term for the worst Pink Noise and Drumkit samples was 'Phasey,' suggesting that the degradation of TQ may have been related to the phase relationship when summing at the ear (particularly at lower frequencies). For the worst Male Speech stimuli, the most common terms were 'Muddy' and 'Phasey,' which indicates that decorrelation distortion may have also had an impact on intelligibility and

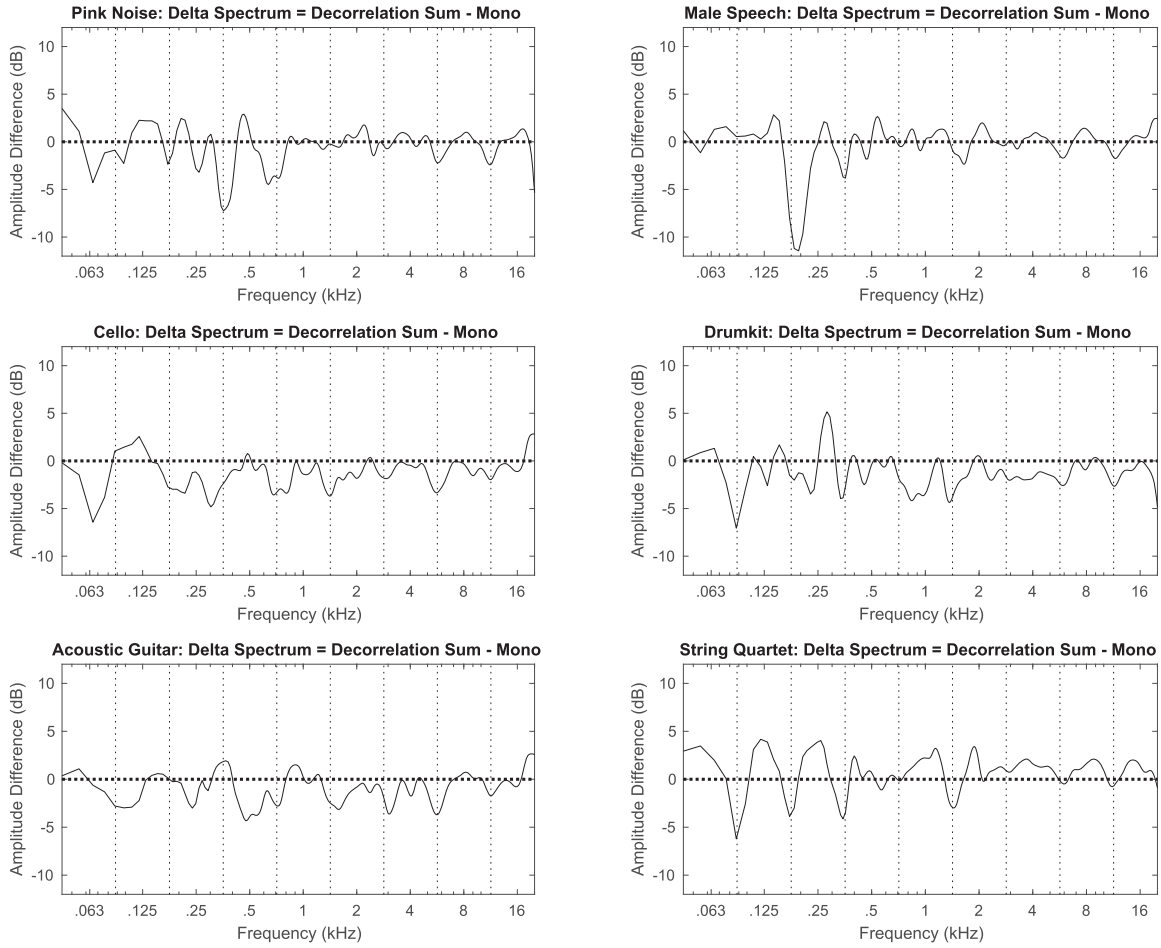

Fig. 8. Delta spectra: The unprocessed 'Monophonic' signal subtracted from the sum of the two 'Broadband' decorrelated main and height signals.

clarity. These observations suggest that further investigation into the perception of TQ when summing decorrelated signals is required.

With the Cello, Acoustic Guitar and String Quartet samples, Fig. 6 also shows some evidence of improved TQ following decorrelation. Observing the spectrograms in Fig. 2, a time-varying change in spectrum is clearly seen for these musical sources, and there are also strong harmonic tones above the fundamental frequency of the musical notes (particularly with the Cello sample). In contrast, the Male Speech and Drumkit samples have mostly consistent or repeating spectrums over time, particularly at lower frequencies, with these two samples also demonstrating the greatest tonal degradation across all sources. It is possible that the effect of vertical decorrelation on the frequency spectrum (e.g. the distortion in Fig. 8) is less perceivable, and in some instances preferable, for polyphonic musical sources with moving parts; whereas for relatively steady-state broadband sources with repetitive or familiar patterns (e.g. speech and drums), any spectral distortion may result in a noticeable degradation of TQ. Further investigation into the positive and negative effects of vertical decorrelation on TQ is required to test this hypothesis. However, considering a 2D-to-3D upmixing application, band-limited vertical decorrelation is likely to reduce some of the negative spectral effects observed in Fig. 8, and therefore improve the TQ of vertical decorrelation for a wider variety of sources.

C. Practical Applications

1) 2D-to-3D Upmixing: Since vertical decorrelation of solely high frequencies appears to be effective (even for '8 kHz +'), it may be useful to combine this approach with other upmixing techniques. One such technique is perceptual band allocation (PBA) [22], where octave-bands are routed discretely to either a main-layer or height-layer loudspeaker, based on their inherent vertical localisation (i.e. the so-called 'pitch-height' or Pratt effect [18]). For example, vertical decorrelation might be applied solely to the 8 kHz octave-band, which is known to feature a strong cue for auditory elevation perception [20], [21]. In this case, decorrelation could work to reduce the directionality of the band and give a more cohesive spread overall between the two loudspeaker sources. Such a reduction of vertical localisation cues by vertical decorrelation has already been demonstrated in a previous study by the authors [8].

2) Binaural Audio Rendering: In addition to upmixing, vertical decorrelation might also be applied when binaurally rendering audio for reproduction over headphones. This has potential application within virtual and augmented reality scenarios, giving greater control over the VIS of a virtual sound source. To

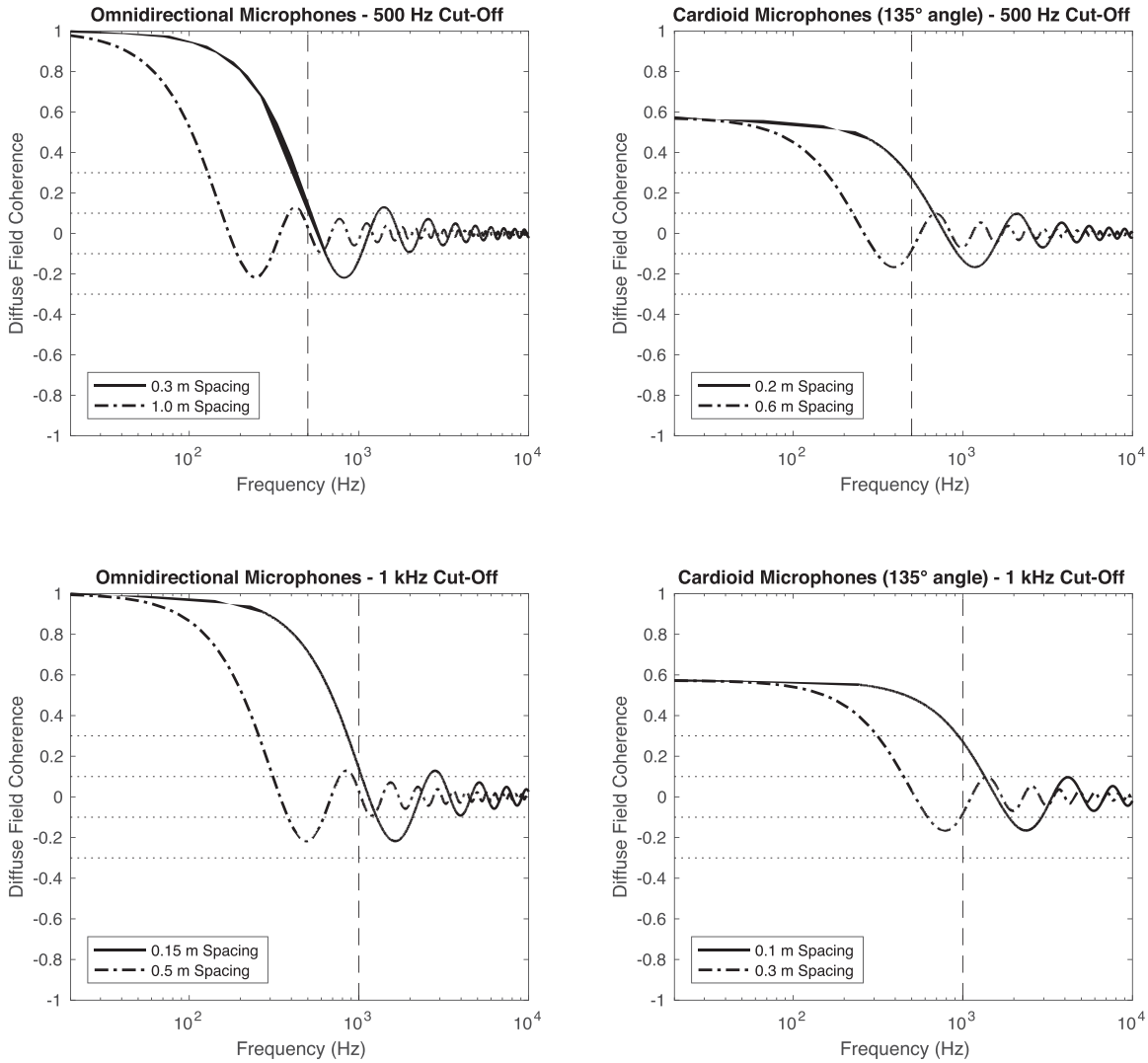

Fig. 9. Graphs presenting the diffuse field coherence between two microphones based on [23]. The Left plots show the results for a pair of omnidirectional microphones spaced at 0.15, 0.3, 0.5 and 1.0 m; and the Right plots show the results for a pair of cardioid microphones spaced at 0.1, 0.2, 0.3 and 0.6 m, where one is angled 135° from the other. The upper plots are marked with vertically dashed lines that highlight a 500 Hz cut-off, while the lower plots highlight a 1 kHz cut-off. The horizontally dashed lines on all plots mark $\pm$ 0.1 and $\pm$ 0.3 coherence.

achieve this, two decorrelated signals of the same source could be reproduced at two points within the virtual space, either between a pair of virtual vertically-spaced loudspeakers or between two HRTF points at vertically-spaced positions. Further investigation may also show that decorrelation at multiple points can control both the horizontal and vertical extent of a source simultaneously. Moreover, a similar decorrelation approach could be taken when processing sound objects within object-based multichannel formats e.g. Dolby Atmos [2]. It is possible that horizontal and vertical image spread parameters could be encoded for each audio object, in addition to the positional information, allowing for improved manipulation of a source's spatial image.

3) 3D Sound Recording: Understanding the perception of interchannel correlation between main- and height-layer loudspeakers can also inform the design of microphone arrays for 3D audio recording. In current microphone configurations, independent microphones are often used to record solely height-channel information. Such signals are typically the ambience or diffuse sound of a room, captured by positioning the microphones beyond the room's critical distance (i.e., when the reverberant sound is equal to or greater than the direct sound). Increasing the distance between two microphones in a diffuse field results in a decrease of the frequency coherence between the two recorded signals, where a greater distance increases decorrelation at lower frequencies.

The findings of the current study show that reproducing decorrelated low frequencies in height-channel loudspeakers has no perceptual benefit, and can often be detrimental due to phase cancellation at the listening position. It is therefore of interest to determine the optimal distance between two microphones in a diffuse field, where only a decrease of correlation is observed above the decorrelation cut-off identified in the current paper. Specifically, the subjective results above have demonstrated that decorrelation of the 500 Hz octave-band and above can

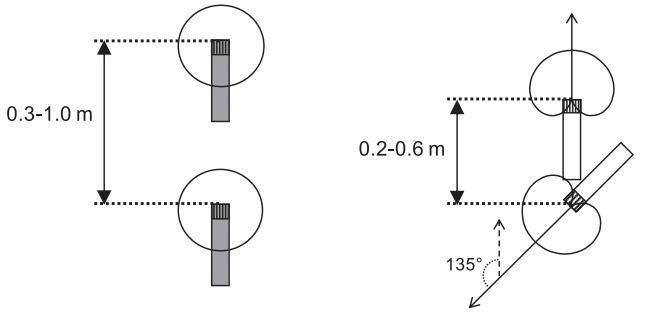

Fig. 10. Diagram illustrating the maximum vertical spacing of microphones required to achieve decorrelation at 500 Hz and above. The Left image shows a pair of omnidirectional microphones and the Right image shows a pair of cardioid microphones angled by 135°.

significantly increase the perception of VIS, similar to that of broadband decorrelation, for all sources under test.

The frequency coherence between two microphones in a diffuse field can be calculated as per the equations found in [23]. Results of these calculations are presented in Fig. 9, where the upper two panels show a vertical dashed line to mark a 500 Hz frequency cut-off, while the lower two panels mark a 1 kHz frequency cut-off. Moreover, the left two plots show the results for a pair of omnidirectional microphones in a diffuse field, and the right two plots show the results for a pair of cardioid microphones (where one is rotated 135° from the other, as shown in Fig. 10). All plots in Fig. 9 also feature horizontal dashed lines to mark $\pm$ 0.1 and $\pm$ 0.3 coherence, where $\pm$ 0.3 is the decorrelation level achieved for each octave-band in the current study.

Between a pair of omnidirectional microphones, the diffuse field coherence results in Fig. 9 demonstrate that only a 0.3 m vertical spacing is required to achieve coherence within $\pm$ 0.3 at 500 Hz and above. This distance can be reduced even further to 0.2 m when cardioid microphones are used with a 135° rotation angle, as illustrated in Fig. 10. The results also show that the maximum distance required to achieve coherence within $\pm$ 0.1 is 1.0 m for two omnidirectional microphones and 0.6 m for the cardioid pair—this result suggests that there is no perceptual benefit in using distances greater than these between main- and height-channel microphones in a 3D array. A previous study by the authors has also shown that there is little perceptual difference between interchannel cross-correlation coefficients (ICCCs) of 0.1 and 0.4 in terms of VIS [8], i.e., a distance of 0.3 m is likely to be equally as effective at decorrelating above 500 Hz as 1.0 m, when recording in a diffuse field with a pair of omnidirectional microphones.

Moreover, it was found in the current study that decorrelation cut-offs higher than the 500 Hz octave-band can significantly increase VIS in many cases, provided the source signal has sufficient high frequency energy. Given this, it is thought that smaller microphone spacings than those discussed above are also likely to be effective. A 1 kHz cut-off is shown in the lower panels of Fig. 9 to demonstrate this—the results show that as the target cut-off frequency is doubled from 500 Hz to 1 kHz, the distance required to achieve decorrelation at the target frequency is halved (i.e. from 0.3 m to 0.15 m). When recording in a space with a strong high frequency energy response, the target frequency could be increased even further, potentially allowing for a pair of cardioid microphones to be positioned in a near-coincident configuration. An earlier study by the authors has already shown that coincident main- and height-channel microphones gives similar or improved spatial impression, when compared against vertically-spaced microphone pairs [24].

VI. Conclusion

Two subjective listening tests have been conducted to investigate the effects of vertical interchannel decorrelation, assessing both the perceived vertical image spread (VIS) and tonal quality (TQ) of natural sound sources and pink noise. A 10-channel 3D loudspeaker array, based on Auro-3D 9.1 with an additional centre height-channel, was used for the experiments. The present study focused on finding the lower frequency band boundary of decorrelation within a broadband signal that can produce optimal results for both VIS and TQ. Eight decorrelation conditions were assessed where the lower band boundary was varied between eight octave-bands with centre frequencies from 63 Hz to 8 kHz. That is, 'XXX Hz +' signifies vertical decorrelation of the XXX Hz octave-band and all octave-bands above, up to the 16 kHz octave-band. Non-decorrelated octave-bands below the boundary were routed to the main-layer loudspeaker only, while the two-channel decorrelated bands were routed between the main- and height-layer loudspeakers. Six source signals were tested: Broadband Pink Noise, Male Speech, Cello, Drumkit, Acoustic Guitar and String Quartet. All stimuli were presented independently from vertically-spaced loudspeakers at three azimuth angles around the listener (0°, $\pm$30° and $\pm$110°), with an elevation angle of +30° between the main-layer and height-layer.

The subjective VIS results show that the '125 Hz +,' '250 Hz +' and '500 Hz +' octave-band decorrelation conditions have a similar perceived VIS to broadband decorrelation ('63 Hz +') for all sources and azimuth angles. In general, the VIS results were largely source-dependent in that those sources with greater high frequency energy required narrower decorrelation bands to significantly increase VIS. For example, the Pink Noise and Drumkit samples had greater energy at 8–16 kHz, resulting in the '8 kHz +' condition having significantly greater VIS than the monophonic sample—this suggests the importance of vertical localisation cues around 8 kHz towards VIS perception. It is also shown in the paper that summing two phase-decorrelated (all-pass filtered) signals can result in spectral distortion, most notably at lower frequencies. In some instances, the band-limited decorrelation conditions were perceived as having greater VIS than the broadband decorrelation condition, which may have been caused by the low frequency spectral distortion affecting spatial cues.

The subjective TQ results suggest that decorrelation solely of the 500 Hz octave-band and above tends to produce a TQ that is similar to the unprocessed monophonic condition. On the other hand, the lower cut-off and broadband decorrelation conditions generally show greater TQ degradation, presumably due to the low frequency spectral distortion mentioned above. In

many cases, TQ was improved by increasing the decorrelation cut-off to higher octave-bands, however, this was found to be dependent on the type of source. The degradation of TQ by vertical decorrelation was most apparent with the Pink Noise, Male Speech and Drumkit sources; whereas, in some cases, vertical decorrelation of the Cello and String Quartet samples actually improved TQ over the monophonic reference.

From the results of this study, it is evident that vertical decorrelation below the 500 Hz octave-band has no perceptual benefit. As such, it is thought that the optimal low frequency cut-off for vertical decorrelation is likely to be somewhere between '500 Hz +' (for maximum VIS) and a higher cut-off frequency (for potentially improved TQ), depending on the source. Furthermore, it is shown that if a source signal has sufficient high frequency energy in the 8–16 kHz octave-bands, a VIS similar to broadband decorrelation can be achieved with the '8 kHz +' condition, significantly reducing the impact of decorrelation on TQ. These findings can support the future development of 2D-to-3D upmixing algorithms, binaural audio rendering and 3D audio recording, as discussed in the paper.

REFERENCES

[1] C. Gribben and H. Lee, "Increasing the vertical image spread of natural sound sources using band-limited interchannel decorrelation," in *Proc. Audio Eng. Soc. Conf. AES Int. Conf. Immersive Interactive Audio*, Audio Eng. Soc., 2019, Art. no. 78.

[2] "Dolby Atmos home theater installation guidelines," [Online]. Available: https://www.dolby.com/us/en/technologies/dolby-atmos/dolby-atmos-home-theater-installation-guidelines.pdf, Accessed: Sep. 16, 2019.

[3] "Auro-3D home theater setup guidelines," [Online]. Available: https://www.auro-3d.com/wp-content/uploads/documents/Auro-3D-Home-Theater-Setup-Guidelines_lores.pdf, Accessed: Sep. 16, 2019.

[4] "The first step toward 3D audio: DTS Neo:X," [Online]. Available: https://www.stormaudio.com/media/dtsneoxwhite_paper__019028000_1625_04032013.pdf, Accessed: Sep. 16, 2019.

[5] G. S. Kendall, "The decorrelation of audio signals and its impact on spatial imagery," *Comput. Music J.*, vol. 19, no. 4, pp. 71–87, 1995.

[6] H. Lauridsen, "Nogle forsøg med forskellige former rumakustik gengivelske," *Ingeniøren*, vol. 47, 1954, Art. no. 906.

[7] F. Zotter and M. Frank, "Efficient phantom source widening," *Archives Acoust.*, vol. 38, no. 1, pp. 27–37, 2013.

[8] 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.

[9] ITU-R, "Recommendations ITU-R BS.1116-3: Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems," International Telecommunication Union, Geneva, Switzerland, Standard, 2015.

[10] H. Lee, "2D-to-3D ambience upmixing based on perceptual band allocation," *J. Audio Eng. Soc.*, vol. 63, no. 10, pp. 811–821, 2015.

[11] R. Mason, T. Brookes, and F. Rumsey, "Creation and verification of a controlled experimental stimulus for investigating selected perceived spatial attributes," in *Proc. 114th AES Convention*, 2003, pp. 22–25.

[12] ITU-R, "Recommendations ITU-R BS.1534-3: Method for the subjective assessment of intermediate quality level of audio systems," International Telecommunication Union, Geneva, Switzerland, Standard, 2015.

[13] C. Gribben and H. Lee, "Towards the development of a universal listening test interface generator in Max," presented at the 138th Convention Audio Eng. Soc., Warsaw, Poland, May 2015.

[14] S. Bech and N. Zacharov, *Perceptual Audio Evaluation—Theory, Method and Application*. Hoboken, NJ, USA: Wiley, 2007.

[15] R. McGill, J. W. Tukey, and W. A. Larsen, "Variations of box plots," *Amer. Statistician*, vol. 32, no. 1, pp. 12–16, 1978.

[16] IBM, "IBM SPSS Statistics V23.0 documentation," [Online]. Available: https://www.ibm.com/support/knowledgecenter/SSLVMB_23.0.0/spss/product_landing.html, Accessed: Sep. 16, 2019.

[17] A. Field, *Discovering Statistics Using IBM SPSS Statistics*. Newbury Park, CA, USA: Sage, 2013.

[18] D. Cabrera and S. Tilley, "Vertical localization and image size effects in loudspeaker reproduction," in *Proc. Audio Eng. Soc. Conf.: 24th Int. Conf.: Multichannel Audio, New Reality*, Audio Engineering Society, 2003, Art. no. 46.

[19] C. Gribben and H. Lee, "A comparison between horizontal and vertical interchannel decorrelation," *J. Appl. Sci.*, vol. 7, no. 11, 2017, Art. no. 1202.

[20] S. K. Roffler and R. A. Butler, "Factors that influence the localization of sound in the vertical plane," *J. Acoust. Soc. Amer.*, vol. 43, no. 6, pp. 1255–1259, 1968.

[21] J. Hebrank and D. Wright, "Spectral cues used in the localisation of sound sources on the median plane," *J. Acoust. Soc. Amer.*, vol. 56, no. 6, pp. 1829–1834, 1974.

[22] 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, no. 12, pp. 1003–1013, 2016.

[23] M. Kuster, "Spatial correlation and coherence in reverberant acoustic fields: Extension to microphones with arbitrary first-order directivity," *J. Acoust. Soc. Amer.*, vol. 123, no. 1, pp. 154–162, 2008.

[24] H. Lee and C. Gribben, "Effect of vertical microphone layer spacing for a 3D microphone array," *J. Audio Eng. Soc.*, vol. 62, no. 12, pp. 870–884, 2014.

Christopher Gribben received a degree in music technology and audio systems from the University of Huddersfield, Huddersfield, U.K., in 2013 and a PhD from the Applied Psychoacoustics Laboratory, University of Huddersfield, Huddersfield, U.K., in 2018. His thesis was on the perception of vertical interchannel decorrelation in 3-D surround sound reproduction. During his studies, he undertook an industrial placement with Cass Allen Associates, Ltd., an independent acoustic consultancy, where he focused on environmental acoustics and noise modeling. Since early 2018, he has been a Research Engineer with Meridian Audio Ltd., U.K. His current research aims to provide a better understanding of psychoacoustics, in order to optimise the development of new digital signal processing algorithms.

Hyunkook Lee received a degree in music and sound recording (Tonmeister) from the University of Surrey, Guildford, U.K., in 2002 and a PhD in spatial audio psychoacoustics from the Institute of Sound Recording (IoSR), University of Surrey, Guildford, U.K., in 2006. He is a Reader (i.e., Associate Professor) in Music Technology and the Director of the Applied Psychoacoustics Laboratory, University of Huddersfield, U.K. His recent research has advanced understanding about the perceptual mechanisms of vertical stereophonic localisation and image spread as well as the phantom image elevation effect. This helped develop new 3-D microphone array techniques, vertical mixing/upmixing techniques, and a virtual 3-D panning method named VHAP. His ongoing research topics include 3-D sound perception, capture and reproduction, virtual acoustics, and objective sound quality metrics. From 2006 to 2010, he was a Senior Research Engineer in audio R&D with LG Electronics, South Korea, where he participated in the standardisations of MPEG audio codecs and developed spatial audio algorithms for mobile devices. He is a Fellow of the Audio Engineering Society and a Fellow of the Higher Education Academy, U.K.