Journal of the Audio Engineering Society Vol. 64, No. 12, December 2016
DOI: https://doi.org/10.17743/jaes.2016.0052

Perceptual Band Allocation (PBA) for the Rendering of Vertical Image Spread with a Vertical 2D Loudspeaker Array

HYUNKOOK LEE, *AES Member*
(h.lee@hud.ac.uk)

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

Two subjective experiments were conducted to examine a new vertical image rendering method named "Perceptual Band Allocation (PBA)," using octave bands of pink noise presented from main and height loudspeaker pairs. The PBA attempts to control the perceived degree of vertical image spread (VIS) by a flexible mapping between frequency band and loudspeaker layer based on the desired positioning of the band in the vertical plane. The first experiment measured the perceived vertical location of the phantom image of each octave band stimulus for the main and height loudspeaker layers individually. Results showed significant differences among the frequency bands in perceived image location. Furthermore, the so-called "pitch-height" effect was found for two separate frequency regions, with most bands from the main loudspeaker layer perceived to be elevated from the physical height of the layer. Based on the localization data from the first experiment, six different PBA stimuli were created in such a way that each frequency band was mapped to either the main or height loudspeaker layer depending on the target degree of VIS. The second experiment conducted a listening test to grade the perceived magnitudes of VIS for the six stimuli. The results first indicated that PBA could significantly increase the perceived magnitude of VIS compared to that of a sound presented only from the main layer. It was also found that the different PBA schemes produced various degrees of perceived VIS with statistically significant differences. The paper discusses possible reasons for the obtained results in details based on the localization test results and the frequency-dependent energy weightings of ear-input signals. Implications of the proposed method for the vertical upmixing of horizontal surround content are also discussed.

0 INTRODUCTION

Various methods have been proposed for rendering auditory image spread in horizontal stereo, such as decorrelation techniques based on all-pass filtering [1–3] and comb-filtering [4], stereo shuffling [5], and frequency-dependent panning [6]. These methods are fundamentally based on the fact that our ears are spaced apart and horizontally arranged. An introduction of differences between horizontal channel signals leads to a change in the relationship between ear input signals. For example, a lower interchannel cross-correlation coefficient (ICCC) tends to produce a lower interaural cross-correlation coefficient (IACC), thus causing the perception of a wider auditory image [3].

However, since the perceptual mechanism of vertical stereophony does not rely on interaural cues, the conventional methods might not be suitable for rendering vertically perceived image spread. It was reported in [7] that interchannel decorrelation applied for pink noise with a vertically arranged loudspeaker pair was not as effective as that with a horizontal loudspeaker pair in controlling perceived image spread. Also in the context of 3D microphone array, it was found that the ICCC of vertically oriented ambient signals was not directly associated with perceived 3D listener envelopment (LEV) [8].

The literature generally suggests that vertical auditory localization relies on the frequency component of the ear-input signal. The so-called "pitch-height" or Pratt's effect, which suggests that a higher frequency tone tends to be perceived higher than a lower frequency tone, has been studied by many researchers [9–14]. Blauert [15] proposed a similar theory known as the "directional bands," which maps specific frequencies to front-overhead-back perceptions in the median plane, but his experiment did not exclusively consider the vertical heights of different frequencies. It has been shown in [12] that the pitch-height relationship between frequency and vertical localization was valid for band-passed noise signals also, but in this case the effect depended on the physical height of the loudspeaker that

presented the sound. That is, low frequency noise signals tended to be localized at a height around the listener's ear height, whereas high frequency ones or broadband signals containing high frequency components above 7 kHz were localized more accurately near the physical loudspeaker position. It was also shown in [13, 14] that the pitch-height effect operated for both octave-band noise and musical signals.

The current paper investigates a novel method for rendering vertical image spread (VIS) named "Perceptual Band Allocation (PBA)," which exploits the pitch-height effect mentioned above. The PBA aims to control the perceived spread of a phantom image created between vertically arranged loudspeakers by flexibly mapping frequency bands decomposed from an original signal to either the lower or upper loudspeakers depending on their perceived heights. With the PBA, the frequency spectrum of the original signal is reconstructed at the ear without comb-filtering since no identical frequency is presented from both loudspeakers. This could be an advantage over conventional image widening methods based on phase alteration, considering that comb-filtering introduced vertically tends to be perceptually unpleasant [16].

In the author's previous study [17], simple 2-band PBA scenarios have been subjectively tested with an Auro-3D [18] loudspeaker setup in the context of 2D (5-channel) to 3D (9-channel) ambience upmixing, using multichannel ambience signals recorded in a reverberant hall for various musical sources. The results demonstrated that the PBA-upmixed stimuli could produce a slightly greater or similar magnitude of 3D listener envelopment (LEV) compared to original 9-channel 3D recordings as well as original 5-channel recordings.

In the present study the ability of the PBA to render different degrees of VIS is investigated using octave-band pink noise stimuli. In contrast with previous localization studies using loudspeakers vertically arranged in front of the listener, the current study used a frontal two-dimensional (2D) stereophonic loudspeaker configuration, thus testing the vertical localization of "phantom" images rather than "real" images. This study required two experimental stages. First, the original signal was decomposed into octave-bands and the perceived height of each band was measured through a listening test (Experiment 1). Second, each octave-band was allocated to either the lower or upper loudspeaker layer to render different degrees of VIS, based on the results of the first experiment (Experiment 2). This paper will describe the experimental procedure and results for each experiment, followed by discussions and conclusions.

1 EXPERIMENT 1: VERTICAL LOCALIZATION OF PHANTOM IMAGES

The aim of the first experiment was to measure the perceived vertical location of each octave band signal filtered from broadband pink noise. The results were to be used for the rendering of various degrees of vertical image spread (VIS) in the second experiment.

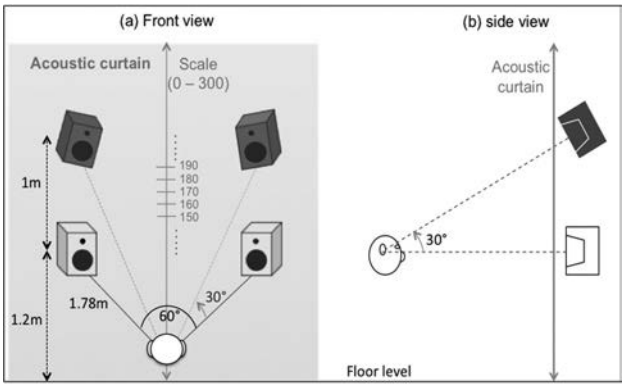

Fig. 1. Loudspeaker setup used for Experiment 1.

1.1 Experimental Design

1.1.1 Physical Setup

The listening tests were conducted in a dry listening room at the University of Huddersfield (8.3m × 5.4m × 3.4m; RT = 0.2s). Fig. 1 shows the loudspeaker setup used for the tests. Four Genelec 8040A loudspeakers (Frequency response: 48 Hz – 20 kHz (±2 dB), Crossover frequency: 3 kHz, Distance between the woofer and tweeter: 14 cm) were arranged in a frontal two-dimensional (2D) fashion. Two loudspeakers at the listener's ear height (main layer), with 1.78 m spacing between them, were configured with the standard 60° angle from the listening position. The middle position between the woofer and tweeter of each loudspeaker was 1.2 m high from the floor, which was also set as each subject's ear height in the listening test. Two height layer loudspeakers were placed directly above the main loudspeakers so that they were elevated by 30° as reference to the listener's ear position, which made the vertical distance from the floor to the middle position of the height loudspeaker 2.2 m. The height loudspeakers were tilted towards the listening position in order to ensure the on-axis frequency response. The main and height loudspeakers were aligned in terms of time delay and sound pressure level at the listening position. The frontal 2D stereophonic configuration was chosen in line with some of the current 3D reproduction formats utilizing a pair of front height channels, such as "Auro-3D" [18], "2+2+2" [19], and "Dolby Prologic IIz" [20]. Additionally, the 2D layout is considered to be also useful for loudspeaker arrangements for large sized televisions.

The loudspeakers were visually hidden to the listeners by using an acoustically transparent curtain. Vertically oriented number labels ranging from 0 to 300 with the interval of 10, representing the height from the floor in cm, were indicated on the curtain as reference points that the subject could use in measuring the height of a perceived image. A too wide gap between each visual label might potentially give rise to a coarse quantization bias in localization judgment. However, from the author's preliminary test, the 10 cm interval in the vertical scale, which was also used in Roffler and Butler [11, 12], was considered to be small enough to avoid such a bias. Due to a small spacing between the acoustic curtain and the loudspeakers, the position on

the scale that corresponded to the height loudspeaker position was 2.14 m rather than 2.2 m.

1.1.2 Stimuli

The sound source was broadband continuous pink noise. The broadband signal was filtered into nine octave-bands (–48 dB/octave) with the center frequencies ranging from 63 Hz to 16 kHz, using an 8th order linear-phase Butterworth filter. Although each band had the equal energy, the perceived loudness of each band was different due to the nonlinearity of human frequency perception. In previous studies on vertical localization using different tones or octave-bands of noise [11–14], a frequency weighting has been generally applied in order to compensate for the variable spectral sensitivity of the hearing system. However, since the purpose of the current localization experiment was to serve as the basis for the later PBA rendering that aims to render VIS while maintaining the original inter-band spectral relationship, it was aimed to measure the perceived vertical location of each band at its inherent loudness with equal energy per band, and therefore frequency weighting was not applied.

Test stimuli were created in two-channel horizontal stereo for each octave-band and the original broadband, which was fed into the left and right channels with the same level, thus producing a "phantom" mono image. The stimulus of each band was to be presented from each of the main and height loudspeaker pairs individually. The output level of the playback system was calibrated to 70 dB LAeq at the listening position using the broadband noise stimulus, and the same playback level was maintained for all octave-bands.

1.1.3 Subjects

Twelve critical listeners participated in the experiment. They were staff researchers, post-graduate researchers, and final year music technology students from the University of Huddersfield. They all had extensive experience in subjective spatial audio evaluation and reported normal hearing.

1.1.4 Test Method

The tests were performed using a graphical user interface (GUI) written using the Max software. There were a total of 20 trials for testing the nine octave-bands and the single broadband stimuli, which were presented from either the main or height loudspeaker layer in a randomized order. In each trial, the subjects were given a slider with which they could change a numerical value between 0 and 300 with the resolution of 1. Their task was to use the slider to report a value that represented the perceived vertical image location according to the number scale in front, which also ranged between 0 and 300. This response method was based on those used in previous vertical localization studies mentioned earlier [11–14].

The directional bands theory [15] suggests that certain frequency bands could be mapped to above or behind localization. Since the current study focused only on the vertical location of perceived image, the subjects were instructed to make their judgments according to the vertical scale in front of them even if any stimulus was perceived from behind. In case a stimulus was perceived to be elevated beyond the scale range (e.g., directly above), the subjects were to locate the slider to the maximum position, 300.

The subjects sat on a chair with a head-rest. The height of the chair was adjusted so that the subjects' ear height was set to 1.2 m, which was also the height of the main layer loudspeaker. They were instructed not to move their head up and down while judging the vertical image location and asked to use their eye movements only, which was monitored by the author.

1.2 Results

Data collected from the localization tests were analyzed statistically using the SPSS software. Shapiro-Wilk and Levene's tests suggested that the data were not suitable for parametric analysis—the normal distribution and equal variance requirements for parametric testing were not satisfied. Therefore, the non-parametric Wilcoxon signed-rank test was used for the analysis of statistical differences between conditions. A Bonferroni adjustment has been applied to the original p values in order to reduce the Type-I error.

Fig. 2 shows box plots for the main and height loudspeaker presentations for each stimulus. The boxes represent the median values and the associated inter-quartile ranges (IQRs), while the whiskers show the range of the highest and lowest data points within 1.5 times IQR. The reference lines at 120 cm and 214 cm represent the visual marker positions on the acoustic curtain that corresponded to the middle positions between the woofer and tweeter for the main and height loudspeakers, respectively. The spacing between the woofer and tweeter of the loudspeaker used in the test was 14 cm, and the cross-over frequency of the loudspeaker was 3 kHz. Therefore, the actual positons of sound radiation on the marker scale were slightly lower or higher than the reference positions of 120 cm and 214 cm, depending on the frequency bands. The woofer and tweeter positions of the main loudspeakers corresponding on the visual scale were 113.2 cm and 126.8 cm, respectively, whereas those of the height loudspeakers were 206.5 cm and 221.5 cm. This has been taken into account in the statistical analysis that examined the significance of difference between the sound radiation position and the perceived position for each octave band. For the broadband signal, the middle positions were used for the statistical analysis.

From the plots, the pitch-height effect can be observed for two separate regions of octave bands independently with both loudspeaker layers: 63 Hz – 500 Hz and 1 kHz – 8 kHz. As the center frequency of octave-band increased from 63 Hz to 500 Hz, the vertical image location tended to increase from the physical height of the main loudspeaker layer towards that of the height layer in general. However, it appears that this effect was "reset" at 1 kHz; the perceived location for this band was similar to those for the 63 Hz and 125 Hz bands. Wilcoxon tests confirmed that there were

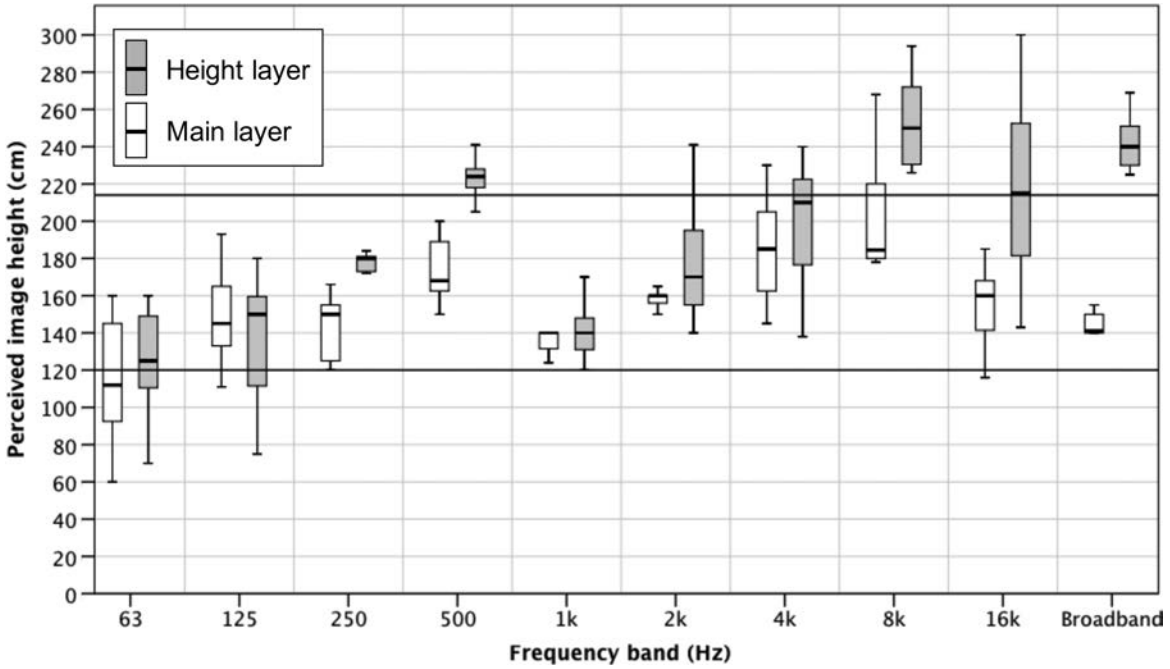

Fig. 2. Localization test results. Each box plots the median (the 2nd quartile) and the interquartile range (IQR: the 1st quartile to the 3rd quartile) of the data for each experimental condition; the white and grey boxes are for the main and height loudspeaker presentations, respectively. The top and bottom whiskers indicate the highest and lowest values in the data within 1.5 times IQR, respectively. Each horizontal line crossing the figure represents the physical height of the middle position between the woofer and tweeter for the main or height loudspeaker layer.

no statistically significant differences among these bands ($p > 0.05$). The pitch-height relationship can be observed again between the 1 kHz and 8 kHz bands, with the perceived location of the 8 kHz band presented from the height loudspeaker layer being the highest among all tested conditions, but the effect broke down with the 16 kHz band. It is interesting that the perceived location of the 16 kHz band with the main layer was at a similar height as those of most of the lower bands. For the height layer, the perceived location of the 16 kHz band was similar to those of the 500 Hz and 4 kHz.

The results generally show that the median vertical locations of the stimuli presented from the height layer were higher than those from the main layer. Significant differences between the loudspeaker layers in vertical image location were observed for the octave-bands with the center frequencies of 250 Hz ($p < 0.01$), 500 Hz ($p < 0.01$), 8 kHz ($p < 0.05$), 16 kHz ($p < 0.05$), and the broadband ($p < 0.01$). On the other hand, the physical height of the loudspeaker layer did not have a significant effect for the 63 Hz, 125 Hz, 1 kHz, 2 kHz, and 4 kHz bands ($p > 0.05$).

It can be also observed that the range of the perceived image location in each of the pitch-height regions was greater with the height loudspeaker layer than with the main layer. For example, although the 63 Hz band was localized near the ear height regardless of its loudspeaker layer, the 500 Hz band presented from the height layer was localized slightly higher than the physical height of the layer, whereas that from the main layer was perceived halfway between the main and height layers. A similar tendency is observed also for the 1 kHz and 8 kHz bands.

Last but not least, for the main loudspeaker layer, one-sampled Wilcoxon tests suggest that the perceived locations of all bands except the 63 Hz band were significantly higher than the sound radiation position of the layer ($p < 0.05$). For the height loudspeaker layer, on the other hand, the 500 Hz and 8 kHz bands as well as the broadband were found to be localized significantly higher than the sound radiation position of the layer ($p < 0.05$).

2 EXPERIMENT 2: RENDERING OF VERTICAL IMAGE SPREAD

The aim of the second experiment was to examine the ability of the PBA to create different degrees of vertical image spread (VIS). A set of stimuli was created using octave-band pink noise signals based on the results of the first experiment. A listening test was conducted to compare the perceived magnitudes of vertical spread for the stimuli.

2.1 Experimental Design

2.1.1 Stimuli

A total of six stimuli were created for the experiment using the same nine octave-band pink noise signals from the previous experiment as described in Table 1. The underlying hypothesis was that different degrees of VIS of a broadband signal could be produced by allocating each sub-band of the signal to its desired median perceptual location obtained from Experiment 1. To this end, each octave band was mapped to only one loudspeaker layer, either the main or height layer, depending on the desired vertical

Table 1. Band allocation schemes for the test stimuli.

	Bands for the main layer	Bands for the height layer
(a) Main only	63, 125, 250, 500, 1 k, 2 k, 4 k, 8 k, 16 kHz	None
(b) Height only	None	63, 125, 250, 500, 1 k, 2 k, 4 k, 8 k, 16 kHz
(c) PBA-1	63, 1, 2, 4 kHz	125, 250, 500, 8 k, 16 kHz
(d) PBA-2	500, 1 k, 8 k, 16 kHz	63, 125, 250, 2 k, 4 kHz
(e) PBA-3	63, 125, 250, 500, 1 kHz	2 k, 4 k, 8 k, 16 kHz
(f) PBA-4	2 k, 4 k, 8 k, 16 kHz	63, 125, 250, 500, 1 kHz

location of its perceived image. The schematic diagrams in Fig. 3 shows how the stimuli were created. The overall vertical span of the frequency bands for each stimulus predict perceived VIS. The white and black circles correspond to the median perceived locations of octave-bands presented from the main (white) and height (black) loudspeaker layers, respectively, which are from Experiment 1. The "main only" and "height only" are conditions where the octave-bands were presented from the main or height layer only. The "PBA-1" and "PBA-2" are PBA-rendered stimuli that aimed for maximum and minimum vertical spreads respectively, with a constraint being that the perceived image location of each octave-band should be distributed as evenly as possible vertically. The "PBA-3" and "PBA-4" were created from simple low-passed and high-passed approaches; in (e) octave-bands with the center frequencies of 2 kHz and higher were routed to the height layer while the lower bands were to the main layer, and vice versa in (f).

2.1.2 Test Method

The loudspeaker and playback system setup used for this experiment was identical to that used in the first experiment (see Sec. 1.1.1). Since this experiment compared the perceived vertical spreads of different stimuli relatively, the vertical number labels used in the previous experiment was not used.

A multiple stimulus comparison test with a reference (REF) and a hidden reference (HR) was conducted. The inclusion of the REF and HR was based on recommendations in ITU-R BS.1524-2 [21]. The "main only" stimulus was chosen as REF and HR for two reasons; (i) it was initially assumed to produce the smallest vertical spread based on the predictions described in Fig. 3; (ii) in a practical application such as vertical ambience upmixing, upmixed signals from both main and height channels would be compared with the original signals from the main channels. The same subjects from Experiment 1 participated in this listening test, and they were provided with a second custom-made Max GUI, on which they could switch between the six stimuli and REF instantaneously. Their task was to grade them on bi-polar continuous scales in terms of the "*perceived magnitude of vertical image spread*," using sliders provided on the GUI. The scale range, which was internally recorded, was –50 to 50, with the middle point 0 representing "no difference." There were no semantic labels used in the scale, but the directions of grading were indicated as "larger" towards 50 and "smaller" towards –50, with each end point implying a perceptually extreme difference.

2.2 Results

Data collected from the listening test were first normalized with respect to mean and standard deviation according to [22]. Since the Shapiro-Wilks and Levene's tests again showed the data were not suitable for parametric analysis, boxplots were used for the visual presentation of the results and the Bonferroni-corrected Wilcoxon test was used for pairwise multiple comparisons. As can be seen from Fig. 4, the "height only" condition was found to be the largest in perceived vertical spread and the second largest was the "PBA-3," but the difference between these two conditions was statistically not significant ($p > 0.05$). The "PBA-1" was slightly less spread than the "PBA-3," but again this difference was non-significant ($p > 0.05$). The perceived vertical spread of the "PBA-4" was ranked between those of the "PBA-1" and the "main only," with its difference to each being significant ($p < 0.01$). The "PBA-2" was found to be the least spread stimulus, although its difference to the "main only" was non-significant ($p > 0.05$).

3 DISCUSSION

This section will discuss the perceived results from the above two experiments, together with the objective analysis of ear-input signal spectra.

3.1 Vertical Localization of Phantom Images

Past pitch-height studies using pure tone [9–13] generally suggest that the perceived vertical image location becomes higher as the frequency increases. However, the current results showed that there were two independent regions where the pitch-height effect operated with the octave noise bands (63 Hz–500 Hz and 1 kHz–8 kHz), with the 1 kHz band being a "reset" point. Furthermore, the 16 kHz band was localized lower than or similar to some of the lower bands, such as the 500 Hz, 4 kHz, and 8 kHz. This suggests, at least in the context of the current experimental condition using octave-band phantom images, that the pitch-height effect does not operate entirely linearly across the whole frequency range. The result showing that the 1 kHz band was perceived almost at the listener's ear level regardless of the height of loudspeaker layer seem to be related to Blauert's directional bands theory [15], which suggests that a 1/3-octave-band centered at 1 kHz tends to be localized behind the listener regardless of the location of

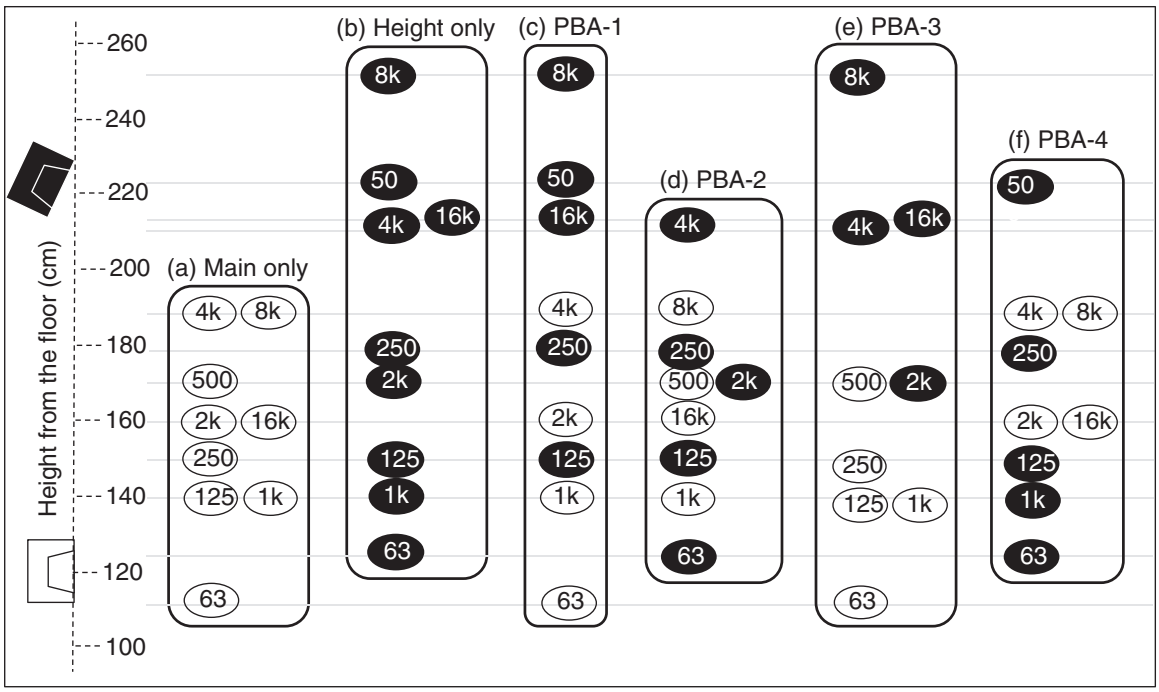

Fig. 3. Schematic diagrams of the stimuli created; the solid and open ellipses represent frequency bands presented from the height and main loudspeaker layer, respectively.

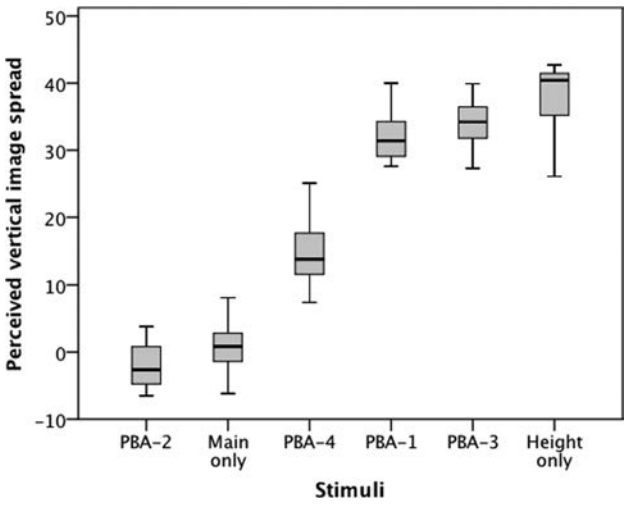

Fig. 4. Results of Experiment 2: The Y-axis scale is arbitrary as a result of the data normalization.

the presenting loudspeaker. In the current experiment, several subjects reported a front-back confusion phenomenon for the 1 kHz band rather than a consistent back perception for the band, and this would have caused the height of the band to be judged to be near the ear level.

It seems worth discussing the current results in comparison with Cabrera and Tiley's previous "pitch-height" results obtained using "real" images of selected octave-band pink noises [13]. They measured perceived vertical image locations of broadband and four octave-band pink noises in an anechoic chamber with five vertically elevated loudspeakers. The loudspeakers were placed vertically in front of the listener, and the elevation angles were 0°, ±7.9°, and ±15.6° with respect to the listener's ear. The center frequencies of the octave-bands tested were 125 Hz, 500 Hz, 2 kHz, and 8 kHz. The current results appear to partly agree with their results in that the perceived location of a sound presented from a physically higher loudspeaker tended to be higher than that from a lower loudspeaker. However, comparing the results for the four octave-band and broadband signals commonly used in the two studies, the perceived image heights in the current study are found to be generally higher than those of Cabrera and Tiley's study. For example, the 125 Hz band was localized to be significantly lower than the ear-level loudspeaker in Cabrera and Tiley's study, whereas the perceived height of the same band was significantly higher than the ear-level (main) loudspeaker layer in the current study. A more radical difference was observed for the 500 Hz band. Cabrera and Tiley's results showed that the perceived image location for the 500 Hz band was lower than the ear height regardless of the physical loudspeaker height. In the current results, however, the same band was localized significantly higher than the presenting loudspeaker layer's height, for both main and height layers. Especially, the perceived location of the band for the main layer was as high as those of the 4 kHz and 8 kHz for the same layer. Furthermore, the current results showed that the broadband noise was localized slightly but significantly higher than the height of the presenting loudspeaker layer, whereas Cabrera and Tiley's studies, as well as Roffler and Butler [12], showed that the broadband pink noise was accurately localized at the physical height of the loudspeaker that presented the signal.

The above-mentioned differences suggest that the relationship between frequency and its perceived image height is associated not only with the physical height of sound source (as already found in previous studies [12, 13]), but also with the nature of the image being real or phantom. The stereophonic loudspeaker configuration used in the current experiment produced a "phantom" center image for each band, whereas Cabrera and Tiley's experiment was conducted with a single loudspeaker placed in the center at each physical height, thus producing a "real" center image.

The elevation of horizontally oriented phantom center image was first reported by de Boer [23] and later confirmed by Damaske and Mellert [24], Frank [25], and Lee [26]. It is generally suggested that as the base angle of a stereophonic loudspeaker pair increases from 0° to between 180° and 240°, the perceived image is elevated from front to overhead. Blauert [27] explains that this effect is caused due to the spectral energy distribution of ear-input signal, which varies depending on the loudspeaker base angle, based on his "directional bands" theory [15]. For example, more energy around 8 kHz and less around 4 kHz in the frequency spectrum of ear-input signal would mean that the resulting phantom image is elevated more towards the directly overhead position according to Blauert suggesting that the 1/3-octave 8 kHz and 4 kHz bands are mapped to above and front perceptions, respectively.

Although Blauert's theory seems to be valid for the elevation of broadband or signals containing frequencies above about 3 kHz, it cannot explain the reason for the elevations of individual low frequency bands such as the 250 Hz and 500 Hz bands, which were found in the current study (Fig. 2). In [26] the current author proposed a new hypothesis suggesting that the low frequency phantom image elevation is perceived due to the brain's cognitive association between the acoustic crosstalks of horizontally arranged loudspeaker signals and the torso reflections of a real source elevated in the median plane. The basis for this is the fact that the acoustic crosstalks and torso reflections have similar natures. As Algazi et al. [28] found, the low frequency component of head-related transfer function (HRTF) for a source elevated in the median plane is a feature of torso reflection, which is the main cue for elevation localization. Acoustic crosstalks also mainly feature low frequencies due to the head-shadowing effect. As the loudspeaker base angle increases, the delay between the ipsilateral and crosstalk signals increases and reaches its maximum of around 0.7ms at the base angle of 180°. Similarly, the maximum torso reflection delay occurs when the source is elevated to directly above and it is also around 0.7 ms according to Algazi et al.'s analysis. Therefore, it could be suggested that the low frequency content of a phantom center image produced with a specific acoustic crosstalk delay would be perceived to be elevated at the position of a real source in the median plane that produces a torso reflection delay corresponding to the crosstalk delay. Further experiments are currently ongoing in order to verify the above hypothesis and the results will be presented in a future paper.

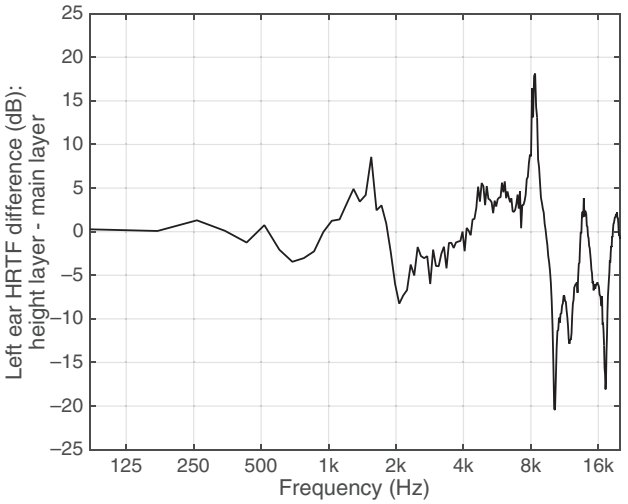

Fig. 5. Spectral magnitude difference of the left-ear head-related transfer function (HRTF) of the height loudspeaker layer signal to the left-ear HRTF of the main layer signal; calculated using MIT's KEMAR head-related impulse response database.

3.2 Vertical Image Spread Rendering by PBA

First, the results from Experiment 2 suggest that the PBA is able to increase the perceived magnitude of vertical image spread (VIS) of a broadband signal presented from the main loudspeaker pair. This is important for vertical stereophonic upmixing, which is the main application of the proposed method. The results also showed that various degrees of VIS could be rendered by applying different band-to-loudspeaker mapping schemes. The stimuli intended for a larger vertical spread were indeed perceived to be have a significantly larger VIS than those for a smaller spread. It is initially considered that this was mainly due to their differences in the upper boundary median location rather than the lower one for the following reason. For all stimuli, the 63 Hz band defined the lower boundary as shown in Fig. 3. Although the median vertical location of the band varied slightly for different loudspeaker layer presentations, this had no statistical significance and therefore all the stimuli would have had similar perceived lower boundary of the image. On the other hand, the "height only," "PBA-3," and "PBA-1," which were the three most spread stimuli in both predicted and perceived results, all had the 8 kHz band presented from the height layer as the upper boundary, whereas the other stimuli had the 4 kHz or 500 Hz upper boundary band. As presented in Fig. 2, the 8 kHz band from the height layer was localized significantly higher than any other bands regardless of their presenting layer.

However, the upper boundary position alone does not seem to explain the reason why the "height only" and "PBA-3" were perceived to be more spread than the "PBA-1." A possible explanation for this can be provided based on the differences between the ear-input spectrum of the main layer signal and that of the height layer signal. Fig. 5 shows the spectral magnitude difference of the height layer to the main layer for the left ear-input signal, measured using the MIT's KEMAR Head-Related Impulse Response database [29]. As can be seen, the height layer HRTF has emphases

between 1 kHz and 2 kHz and between 4 kHz and 8 kHz, compared to the main layer HRTF. On the other hand, the main layer HRTF has more weighting between 2 kHz and 4 kHz than the height layer HRTF. In the current experiment, the "height only" and "PBA-3" presented the 2 kHz and 4 kHz bands from the height loudspeaker layer, while the "PBA-1" presented them from the main layer. Considering the relative weighting of the frequencies within the two bands in the main loudspeaker HRTF, it might be that the two bands were perceptually more dominant with the "PBA-1" than the other two stimuli. This would potentially have produced a distinct image focus around the perceived vertical locations of the 2 kHz and 4 kHz, which were in between the main and height loudspeaker layers (see Fig. 2). Consequently, the subjects might have perceived the "PBA-1" to be vertically narrower than the "height only" or the "PBA-3," which had more 4 kHz to 8 kHz dominance in the height layer.

The "height only" was found to have a slightly larger VIS than the "PBA-3." Although the difference was statistically not significant, this result seems to suggest a potential influence of the elevations of individual bands on the perception of VIS. As can be seen in Fig. 3, the difference between the two stimuli conditions in terms of band allocations lies only in the frequency bands with the center frequencies between 63 Hz and 500 Hz; the "PBA-3" allocates those bands to the main loudspeaker layer, whereas the "height only" to the height layer. From the localization results shown in Fig. 2, it is evident that the 250 Hz and 500 Hz bands were localized at significantly higher positions when they were presented from the height layer than when from the main layer. Especially, the 500 Hz presented from the height layer was localized slightly above the physical height of the height layer, whereas that from the main layer was localized in between the main and height layers. Moreover, as the delta HRTF plot in Fig. 5 indicates, the height layer has more spectral energy than the main layer at those bands. From the above, it might be suggested that the "height only" was perceived to have a greater VIS than the "PBA-3" due to its 250 Hz and 500 Hz bands being more elevated and perceptually emphasized.

The reason why the "PBA-4" was perceived significantly more spread than the "PBA-2" can be explained as follows. The upper boundary for the "PBA-4" was the 500 Hz band presented from the height layer, whereas that for the "PBA-2" was the 4 kHz band from the same layer. From the results of Experiment 1, the 500 Hz band had not only a slightly higher vertical location, but also a much narrower interquartile range (IQR) than the 4 kHz band when they were presented from the height layer. This suggests that the 500 Hz had a greater localization certainty than the 4 kHz one in terms of determining the perceived upper boundary of the broadband image. Furthermore, Fig. 5 shows a slight peak at 500 Hz and a large dip around the 4 kHz band region, which suggests that with the height layer the 500 Hz band would have been perceptually more prominent than the 4 kHz band.

Originally it was assumed that the "main only" (reference) condition would be perceived to have the smallest vertical spread, but this was not found to be the case. A possible explanation for this is as follows. The upper boundary of the "PBA-2" (4 kHz from the height layer) was higher than that of the reference (4 kHz from the main layer). However, the statistical difference between the two bands was not significant as shown in Fig. 2. Furthermore, the 4 kHz band in the "PBA-2" would have been perceived quieter than that that in the reference due to the HRTF difference between the main and height layer shown in Fig. 6.

3.3 Practical Implications

The result showing that the "height only" condition produced the largest VIS initially seems to suggest that the use of main layer loudspeakers would not be necessary for maximally increasing the perceived VIS. In fact, this might have useful implications for 3D recording and loudspeaker arrangement. For example, ambience recordings that were originally made for 2D surround, e.g., 5.1, could be simply allocated to the height channels in order to increase perceived vertical spread in 3D reproduction. However, it is important to note that a larger VIS alone might not necessarily mean a greater magnitude of overall 3D LEV. The lack of horizontally presented signals in the "height only" condition might reduce the perceived magnitude of horizontal image spread, despite the large VIS. This argument is supported by a previous result in [17] showing that the "height only" condition was graded lower than the "PBA-3" condition in perceived 3D LEV for musical ambience signals. The current results showed that the "PBA-1" and "PBA-3" conditions could create a VIS that was comparable to that of the "height only." They distribute frequency contents to both main and height layers, and therefore would be able to produce both the horizontal and vertical senses of LEV, thus potentially producing a greater sense of 3D LEV than the "height only" condition.

It is also worth pointing out that the inherent HRTF difference between the main and height loudspeakers, which was shown in Fig. 6, can change the perceived tonal color of the original broadband signal in the PBA process. However, the tone coloration mentioned here might not necessarily be a negative thing for a subjective tonal quality perception in practical applications. Since each sub-band is allocated to one selected loudspeaker layer only, the signals combined at the ear does not suffer from an audible comb-filtering effect, which might occur when conventional image widening methods are applied vertically. For instance, the prominent frequencies between 4 kHz and 8 kHz in the HRTF of the height layer (Fig. 6) might produce a perceptually pleasing effect (e.g., more "clarity" or "brightness") as well as increasing the perception of elevation, while the reduced response around 2 kHz to 4 kHz in the same signal might reduce any potential "harshness" or "hardness" of the sound.

3.4 Limitations and Future Works

The present study presented experimental data for PBA using the phantom images of octave-band pink noises presented from a vertical 2D loudspeaker array in front. A future study will measure the perceived vertical image

locations of individual frequency bands for each loudspeaker azimuth angle individually in a conventional 3D loudspeaker configuration, e.g., 0°, ±30°, ±90°, and ±120°. From this it is aimed to propose 2D to 3D upmixing methods using PBA that are optimized for each loudspeaker azimuth.

In relation to the above, possible ways to exploit the phantom image elevation effect, which was discussed in Sec. 3.1, in PBA upmixing will be investigated. Further studies to investigate this effect are currently being conducted by the author. Attempts will be made to integrate results from the studies in the PBA-based upmixing method.

The broadband and octave-band pink noise stimuli were used in the present study since the focus of the study was on presenting context-free and controlled experimental data for PBA. However, future works will practically evaluate various PBA schemes derived from the aforementioned further tests for 2D to 3D ambience upmixing tasks using a wide range of musical sources. Conventional image spread rendering techniques such as all-pass filter and complementary comb-filter decorrelators will also be compared against the PBA method. While the PBA allocates individual frequency bands decomposed from a broadband signal to the main and height loudspeaker layers independently, thus no overlapping of frequency content at the ear, the conventional methods feed same frequency content to both loudspeaker layers but with the alteration of phase relationship. It is considered that this would have an effect on perceived timbral quality and subjective preference as well as spatial quality, which will be investigated in the future study.

Last but not least, it was shown in the localization test results (Fig. 3) that different bands had different IQRs depending on which loudspeaker layer presented them. The IQRs seem to represent the locatedness of the image, but this might also be related to the perceived VIS. It was also observed that certain bands had substantially larger IQRs than the broadband signals, suggesting that different bands might have different degrees of vertical locatedness and VIS. In a future study the relationship between vertical image locatedness and VIS for individual bands and its influence on the perceived vertical location and spread of the broadband image will be investigated in a controlled manner.

4 CONCLUSION

This paper described two listening experiments conducted to investigate a novel vertical image rendering method, "Perceptual Band Allocation (PBA)." This method is based on the psychoacoustic principle referred to as the "pitch-height" effect, and aims to create different degrees of vertical image spread (VIS) by allocating each frequency band split from an original broadband signal to either the lower (main) or upper (height) loudspeaker layer depending on its unique vertical location perceived with each layer. In contrast with past vertical localization studies using loudspeakers vertically arranged in front of the listener, the current study used a frontal two-dimensional (2D) stereophonic loudspeaker configuration, thus testing the vertical localization of "phantom" center images rather than "real" center images. A broadband pink noise signal was used as a sound source, and it was filtered into nine octave-bands with the center frequencies ranging from 63 Hz to 16 kHz.

The first experiment measured the perceived vertical location of each octave-band and the original broadband signal, with each presented from the main and height loudspeaker pairs individually. The results generally showed that the pitch-height relationship was not entirely linear across the whole frequency range. All band signals were generally localized higher when they were presented from the height layer than from the main layer, which agrees with the literature. However, in contrast with the results of previous studies obtained from a real image condition, the vertical locations of most bands were found to be higher than the physical height of the presenting loudspeaker layer.

In the second experiment, six different stimuli, which were aimed to produce different degrees of VIS, were created based on the median height of each condition measured from the first experiment. A listening test was conducted to compare the six stimuli in terms of the perceived magnitude of VIS. One PBA condition aimed for a large spread and the condition where all bands were presented from the height layer only were found to produce the largest vertical spread of image. Two other PBA conditions with each aimed for a large and a medium spread were indeed perceived as predicted with a statistical significance. All of the three PBA conditions aimed for large and medium spreads were perceived to be significantly more spread than the reference condition with all bands presented from the main layer. These results generally show that it is possible to effectively render the perceived magnitude of VIS by using different PBA schemes.

5 ACKNOWLEDGMENTS

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC), UK, Grant Ref. EP/L019906/1. The author thanks the staff members and students at the Applied Psychoacoustics Lab of the University of Huddersfield who participated in the listening tests

6 REFERENCES

[1] G. S. Kendall, "The Decorrelation of Audio Signals and Its Impact on Spatial Imagery," *Computer Music J.*, vol. 19, no. 4, pp. 71–87 (1995), http://dx.doi.org/10.2307/3680992

[2] J. Herre, K. Kjorling, J. Breebaart, C. Faller, S. Disch, H. Purnhagen, J. Koppens, J. Hilpert, J. Roden, W. Oomen, K. Lintmeier, and K. S. 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.).

[3] F. Zotter, and M. Frank, "Efficient Phantom Source Widening," *Arch. Acoust.*, vol. 38, pp. 27–37 (2013), http://dx.doi.org/10.14279/depositonce-12

[4] H. Lauridsen, "Experiments Concerning Different Kinds of Room-Acoustics Recording," *Ingenioren*, 47 (1954).

[5] M. A. Gerzon, "Application of Blumlein Shuffling to Stereo Microphone Techniques," *J. Audio Eng. Soc.,* vol. 42, pp. 435–453 (1994 Jun.).

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

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

[8] H. Lee, and C. Gribben, "Effect of Vertical Microphone Layer Spacing for a 3D Microphone Array," *J. Audio Eng. Soc.*, vol. 62, pp. 870–884 (2014 Dec.), http://dx.doi.org/10.17743/jaes.2014.0045

[9] C. C. Pratt, "The Spatial Character of High and Low Tones," *J. Exp. Psychol.*, vol. 13, pp. 278–285 (1930).

[10] O. C. Trimble, "Localization of Sound in the Anterior–Posterior and Vertical Dimensions of 'Auditory' Space," *Brit. J. Psychol.*, vol. 24, pp. 320–334 (1934).

[11] S. K. Roffler, and R. A. Butler, "Localization of Tonal Stimuli in the Vertical Plane," *J. Acoust. Soc. Am.*, vol. 43, pp. 1260–1266 (1968), http://dx.doi.org/10.1121/1.1910977

[12] 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), http://dx.doi.org/10.1121/1.1910976

[13] D. Cabrera, and S. Tilley, "Vertical Localization and Image Size Effects in Loudspeaker Reproduction," presented at the *AES 24th International Conference: Multichannel Audio—The New Reality* (2003 Jun.), conference paper 46.

[14] S. Ferguson, and D. Cabrera, "Vertical Localization of Sound from Multiway Loudspeakers," *J. Audio Eng. Soc.*, vol. 53, pp. 163–173 (2005 Mar.).

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

[16] M. Barron, and A. H. Marshall, "Spatial Impression Due to Early Lateral Reflections in Concert Halls: The Derivation of a Physical Measure," *J. Sound. Vib.*, vol. 77, pp. 211–232 (1981), http://dx.doi.org/10.1016/S0022-460X(81)80020-X

[17] H. Lee, "2D-to-3D Ambience Upmixing Based on Perceptual Band Allocation," *J. Audio Eng. Soc.,* vol. 63, pp. 811–821 (2015 Oct.), http://dx.doi.org/10.17743/jaes.2015.0075

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

[19] MDG, 2+2+2 Recording Technique, URL: http://www.mdg.de/222e.htm2015).

[20] Dolby, Dolby Prologic IIz, URL: http://www.dolby.com/us/en/technologies/dolby-pro-logic-iiz.html2015).

[21] ITU-R, "Recommendations ITU-R BS.1534-2: Method for the Subjective Assessment of Intermediate Quality Level of Audio Systems," International Telecommunications Union (2014).

[22] ITU-R, "Recommendations ITU-R BS.1116-2: Methods for the Subjective Assessment of Small Impairments in Audio Systems including Multichannel Sound Systems," International Telecommunications Union (2014).

[23] K. de Boer, "A Remarkable Phenomenon with Stereophonic Sound Reproduction," *Philips Tech. Rev.*, vol. 9, pp. 8–13 (1947).

[24] P. Damaske, and V. Mellert, "A Procedure for Generating Directionally Accurate Sound Images in the Upper Half-Space Using Two Loudspeakers," *Acustica*, vol. 22, pp. 154–162 (1969/1970).

[25] M. Frank, "Elevation of Horizontal Phantom Sources," *Proc. DAGA 2014*, Oldenburg (2014 Mar).

[26] H. Lee, "Investigation on the Phantom Image Elevation Effect," presented at the *139th Convention of the Audio Engineering Society* (2015 Oct), convention paper 9441.

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

[28] 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), http://dx.doi.org/10.1121/1.1349185

[29] B. Gardner, and K. Martin, URL: http://sound.media.mit.edu/resources/KEMAR.html2000).

THE AUTHOR

Hyunkook Lee

Hyunkook Lee is Senior Lecturer 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 includes spatial audio perception, capturing and rendering techniques for 3D and VR audio, intelligent sound engineering, and interactive virtual acoustics. Hyunkook is an active member of the Audio Engineering Society since 2001 and a fellow of the Higher Education Academy, UK.