Journal of the Audio Engineering Society
Vol. 65, No. 9, September 2017 (© 2017)
DOI: https://doi.org/10.17743/jaes.2017.0028

Sound Source and Loudspeaker Base Angle Dependency of Phantom Image Elevation Effect

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

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

Early studies found that, when identical signals were presented from two loudspeakers equidistant from the listener, the resulting phantom image was elevated in the median plane and the degree of the elevation increased with the loudspeaker base angle. However, sound sources used in such studies were either unknown or limited to noise signals. In order to investigate the dependencies of the elevation effect on sound source and loudspeaker base angle in details, the present study conducted listening tests using 11 natural sources and 4 noise sources with different spectral and temporal characteristics for 7 loudspeaker base angles between 0° and 360°. The elevation effect was found to be significantly dependent on the sound source and base angle. Results generally suggest that the effect is stronger for sources with transient nature and a flat frequency spectrum than for continuous and low-frequency-dominant sources. Theoretical reasons for the effect are also discussed based on head-related transfer function measurements. It is proposed that the perceived degree of elevation would be determined by a relative cue related to the spectral energy distribution at high frequencies, but by an absolute cue associated with the acoustic crosstalk and torso reflections at low frequencies.

0 INTRODUCTION

It is widely known that the localization of sound source placed in the median plane is governed by spectral cues. Studies confirmed that the head-related transfer function (HRTF) above about 3 kHz plays the main role in the median plane localization of broadband signals [1–3]. Research further suggests that notch frequencies between 6 and 12 kHz in the HRTF have particular importance for vertical localization [4–6]. On the other hand, the role of low frequencies in vertical localization has also been reported by Morimoto et al. [7]. Gardner [8] and Algazi et al. [9] showed that torso reflections produce spectral notches in the HRTF below 3 kHz, which are additional localization cues for an elevated source.

With regards to the vertical localization of individual frequencies, Pratt [10] reported that a higher frequency tone was localized at a higher position than a lower frequency one when it was presented from a single loudspeaker in the median plane. A number of studies have confirmed the validity of this so-called "pitch height" or "Pratt's" effect not only for tones [11,12] but also for band-limited noise signals [13–15]. Blauert [16] found, from an experiment using 1/3-octave band noise signals reproduced from loudspeakers placed at the front, rear, side, and overhead positions that frontal localization was mainly associated with the 4 kHz band, back with the 1 kHz band, and above with the 8 kHz band; these bands were referred to as "directional bands." Hebrank and Wright [17] observed similar results for band-passed noise signals.

The aforementioned studies investigated the localization of "real" source image produced from a single elevated loudspeaker. However, early research in stereophony found that there exists the elevation of "phantom" images for two identical signals presented from a pair of loudspeakers arranged symmetrically in the horizontal plane. In 1947 de Boer [18] reported that, when the listener was equidistant from both loudspeakers, the perceived position of the phantom image changed from the front to around above of the listener as the loudspeaker base angle increased from 0° to 180°. However, the sound source used in his experiment was not reported. Damaske and Mellert [19] obtained similar results to de Boer's in their experiment using two identical noise signals ranging from 650 Hz to 4.5 kHz with the loudspeaker base angle varied between 0° and 360°. They found that the 180° angle gave rise to the resulting image being elevated at around 120° from the front. Leakey [20] also observed a similar elevation phenomenon with a speech source in his study on horizontal stereophonic panning. He suggested a linear relationship between loudspeaker base angle and perceived image elevation based on a hypothesis about the role of head movement on the elevation effect, which is discussed in Sec. 3.3, but no systematic perceptual experiment was conducted on this.

The elevation effect has not been exclusively investigated between 1970 and 2010. However, several recent studies on multichannel audio reproduction reported the existence of the effect, although the researchers were unaware of the aforementioned earlier studies. In Jo et al.'s study [21], using a white noise signal ranging from 1 to 16 kHz, it was found that, depending on the subject, the phantom image created by a loudspeaker pair placed at ±110° was elevated to around 45° to 60° from the front in the median plane. Frank [22] showed that the phantom image of broadband pink noise was slightly elevated with the loudspeaker base angle of 40°. The present author [23] showed the elevation effect for the phantom images of octave-band pink noises as well as for broadband pink noise for the 60° base angle. The effect was found for frequency bands centered at 250 Hz and 500 Hz as well as for those centered at 4 kHz and 8 kHz, suggesting the validity of the elevation effect for low frequencies where the HRTF is not relevant.

It is considered that the phantom image elevation effect would be exploited usefully in three-dimensional (3D) multichannel audio applications, such as 3D sound panning, recording, upmixing, and downmixing using horizontal loudspeakers, which will be discussed in Sec. 3.5. Furthermore, it might be relevant to the perception of spatial impression in acoustic spaces. For instance, lateral reflections simultaneously arriving at the ears with the same level might contribute to the perception of a vertically spread auditory image. To date, however, only noise or speech sources have been used in the previous studies mentioned above. Therefore, it is not clear how the effect would be perceived for natural sound sources, especially those that would be perceived to be elevated in real life and therefore be rendered to be elevated in stereophonic reproduction. Furthermore, the perceptual mechanism of the effect has not been fully explored.

From the above background, the present study aimed to investigate the sound source dependency of the phantom image elevation effect. To this end, a series of listening tests were conducted using a wide range of ecologically valid sources with different spectral and temporal characteristics as well as pink and white noise sources in transient and continuous conditions, with seven loudspeaker base angles between 0° and 360°. In this paper methods used for the subjective experiment are first described in detail. The results of statistical analyses conducted for data collected from the tests are then presented, followed by discussion on the effects of sound source and loudspeaker base angle on the perceived magnitude of elevation. Further discussion on the potential theoretical reasons for the elevation effect and the practical implications of the findings is also provided.

1 EXPERIMENTAL DESIGN

1.1 Physical Setup

The listening tests were conducted in an ITU-R BS. 1116-2-compliant [24] listening room (6.2m × 5.6m × 3.8m; RT = 0.25s; NR14) at the University of Huddersfield. Fig. 1 depicts the loudspeaker arrangement used. A total of 12 Genelec 8040A loudspeakers were arranged horizontally at intervals of 30°. The distance between the listening position and each loudspeaker was 2 m. The middle position between the woofer and tweeter of each loudspeaker was 1.28 m high from the floor. The single loudspeakers at the 0° and 180° azimuths were to produce "real" front and back center images, respectively, whereas each of the symmetrically arranged loudspeaker pairs (±30°, ±60°, ±90°, ±120°, and ±150°) was used to create a "phantom" center image. This gave a total of seven loudspeaker base angles: 0° (real front center), 60°, 120°, 180°, 240°, 300°, and 360° (real back center). The loudspeaker setup was hidden to subjects by acoustically transparent curtains placed around and above them in order to avoid a potential visual bias. The subjects were not given any information about the number and positions of the loudspeakers, apart from the fact that there was no loudspeaker below the floor.

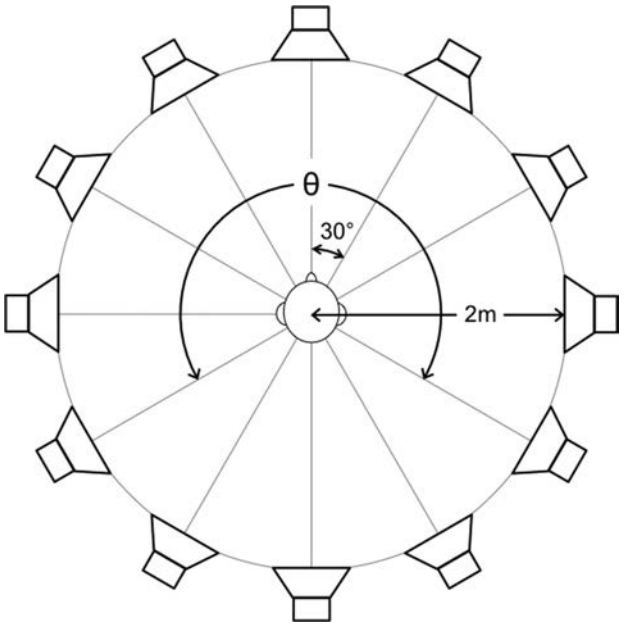

Fig. 1. Loudspeaker setup used for the experiment: stereophonic base angle θ = 0° (front center), 60°, 120°, 180°, 240°, 300°, and 360° (back center). All loudspeakers were placed at the subject's ear level.

Fig. 2 shows operational room response curves measured at the listening position using pink noise reproduced from loudspeakers at each base angle. The ear-input spectrum at the listening position resulting from each loudspeaker base angle is also plotted in Fig. 3.

1.2 Stimuli

Seven natural and four noise sources were used for the experiment. Six of the natural sources comprised the recordings of an airplane, a helicopter, rain, thunder, a bird, and a bell, which were taken from the BBC Sound Effects Library[1]. These sources were chosen not only because they have different temporal and spectral characteristics, but also

[1] http://www.canford.co.uk/Products/81-076_BBC-SOUND-EFFECTS-LIBRARY-Set-of-discs-1-60

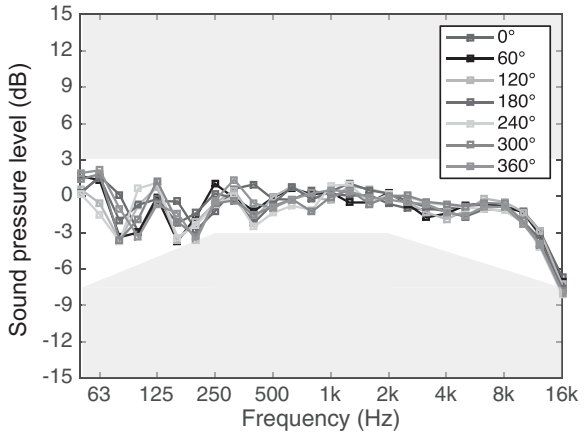

Fig. 2. Operational room responses measured at the listening position for a loudspeaker pair at each base angle tested, using a DPA 4006 omni-directional microphone. The plots show the differences of the sound pressure levels (SPLs) for the 1/3-octave bands of pink noise (over the center frequencies of 50 Hz to 16 kHz) to the average SPL for the 250 Hz to the 2 kHz band. The white area represents the tolerance limits specified in [24].

they would be heard from elevated positions in real life, thus being ecologically valid for practical 3D sound applications. The other natural source was an anechoically recoded male speech, taken from the Bang and Olufsen's Archimedes CD [25]. This was considered to be useful to test the possibility of providing a virtual "Voice of God" effect. The noise sources were chosen in order to examine the effects of temporal and spectral characteristics on perceived elevation in a controlled manner. They comprised 10-second-long broadband pink and white noise signals with 1 second of fade-in and fade-out applied and 200 ms-long broadband pink and white noise bursts repeated with the interval of 500 ms. The onset and offset times for the burst were 5 ms. All signals described above had the sampling frequency of 44.1 kHz and the bit resolution of 16 bits. The long term average spectra (LTAS) of the natural sound sources are shown in Fig. 4.

1.3 Subjects

Twenty-five subjects (24 male and 1 female) participated in the listening tests. They were post-graduate research students, second and final year undergraduate students, and academic staff members from the University of Huddersfield's music technology courses. All of them had previous experiences in localization tests but were not trained particularly for the purpose of the current study. The ages of the subjects ranged from 22 to 38. All subjects reported normal hearing.

1.4 Test Procedure

Listening tests were conducted using a custom-made graphical user interface (GUI) written using the Max 7 software. The total number of stimuli to be tested was 77 (11 sound sources $\times$ 7 loudspeaker base angles). The playback level of each stimulus was calibrated at the average A-weighted sound pressure level (LAeq) of 75 dB at the listening position. Each trial contained a play/stop button for a single stimulus, which was presented in loop and had a side-view circle that was intersected into 12 regions with each covering 30° elevation angle (see Fig. 5). The subject's task was to mark one region in the circle where the sound image was perceived. Each stimulus was presented in a random order for each subject. Prior to the main test, each subject was given a familiarization trial where he or she could listen to all stimuli to be tested.

The present response method was inspired by the method used by Blauert [16], which tested three regions of front, above, and back separated at 90° intervals. In studies where the accuracy of localization is examined by comparing actual and perceived source positions in the median plane, it is a typical response method such that a point on a circle

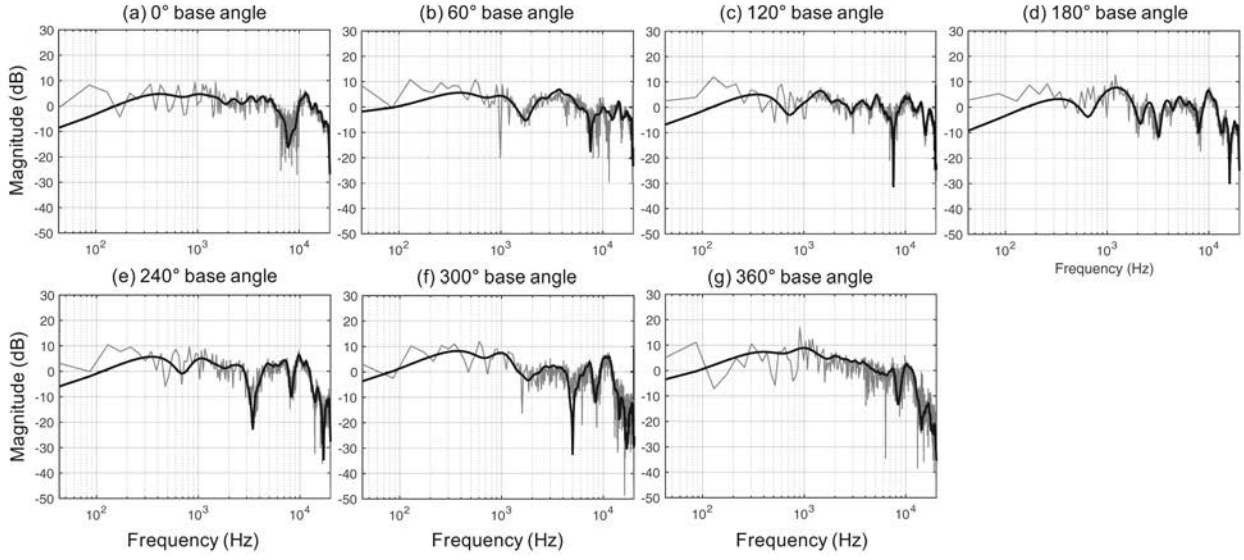

Fig. 3. The spectra of the right ear-input signal for different loudspeaker base angles, measured from binaural room impulse responses captured using the Neumann KU100 dummy head placed at the listening position; the thick black lines are for measurements within the first 2 ms of the BRIRs (i.e., anechoic responses), and the thin grey lines are for those up to 500 ms.

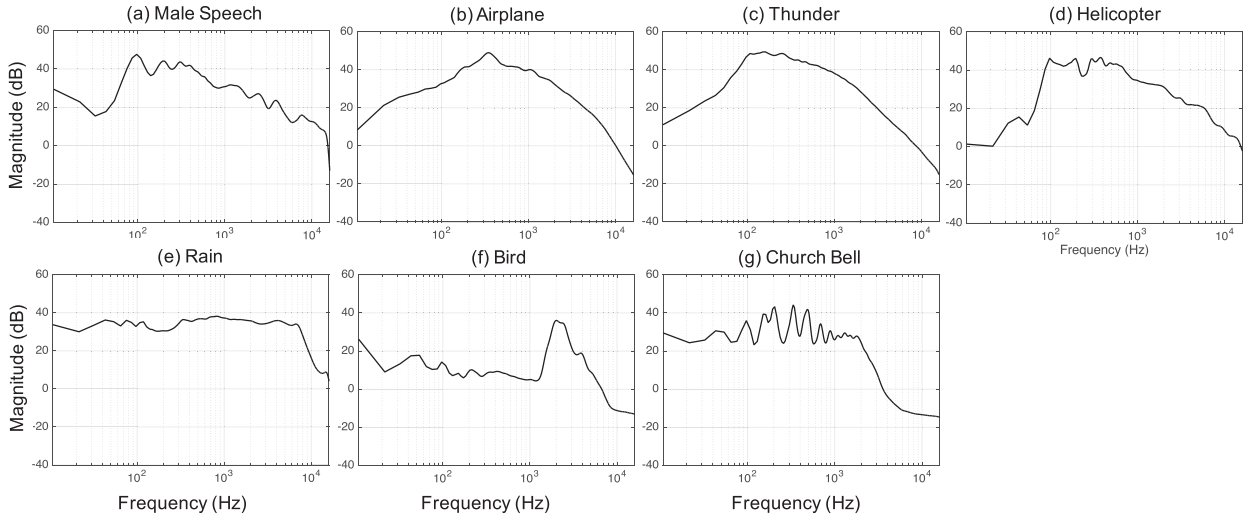

Fig. 4. Long term average spectra (LTAS) of the natural sound sources used for the experiment; the frame length was 4096 samples with a Hanning window and 50% overlap, the FFT point was 4096, and a 1/6-octave Gaussian smoothing was applied to the resulting spectral magnitude.

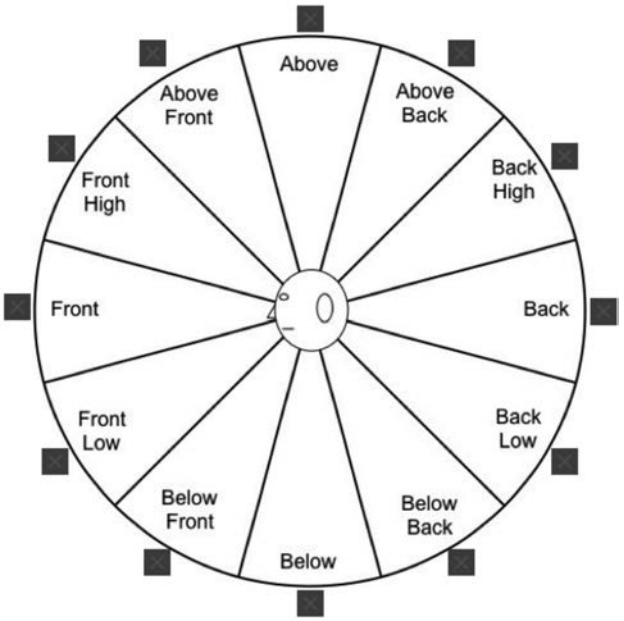

Fig. 5. Graphical user interface used for the listening test.

that corresponds to the perceived position is selected (e.g., Morimoto et al. [7], Algazi et al. [9]). However, from a pilot test it was recognized to be a challenging and time-consuming task to precisely localize perceived image position in the median plane, especially when the image appeared above or behind the listener. Since the aim of the current study was to identify the differences in perceived image elevation among different loudspeaker base angles, the region selection method was considered to be more suitable than the pointing method. Moreover, the 12 sub-regions separated at 30° intervals were considered to produce responses with a sufficiently high resolution for the purpose of this study.

Each subject was positioned so that the ear height was set to 1.28 m, which was also the height of the loudspeaker's acoustic center. The subjects were strictly instructed to face straight ahead and not to move their head while listening and making localization judgments. A small headrest was placed at the back of the subject's head to help them maintain the correct listening position.

2 RESULTS

Data collected from the listening tests were statistically analyzed using the IBM SPSS Statistics 20 software. Since the scale used had an ordinal nature, appropriate non-parametric tests comprising Friedman tests, Wilcoxon tests, Spearman correlation tests, and a correspondence analysis have been performed. The results are described in Sec. 2.1 and Sec. 2.2.

The bubble plots of subject responses obtained from the listening tests are presented in Fig. 6 and Fig. 7. Fig. 6 compares data for different loudspeaker base angles for each sound source, while Fig. 7 data for different sound sources for each base angle. The diameter of each filled circle is proportional to the percentage of responses for the given condition. Table 1 lists the median perceived regions for each pair of sound source and loudspeaker base angles. The figures and the table are used for the discussion of the data in the following sections.

2.1 Relationship between the Loudspeaker Base Angle and the Perceived Elevation for Each Sound Source

Friedman tests were carried out to examine the main effect of the loudspeaker base angle on the perceived image region for each sound source tested. The results presented in Table 2 suggest that there were significant differences among the base angles for every sound source ($X^2(6) > 86$, $p < .01$). However, the effect sizes, represented by Kendall's W coefficient of concordance, vary for different

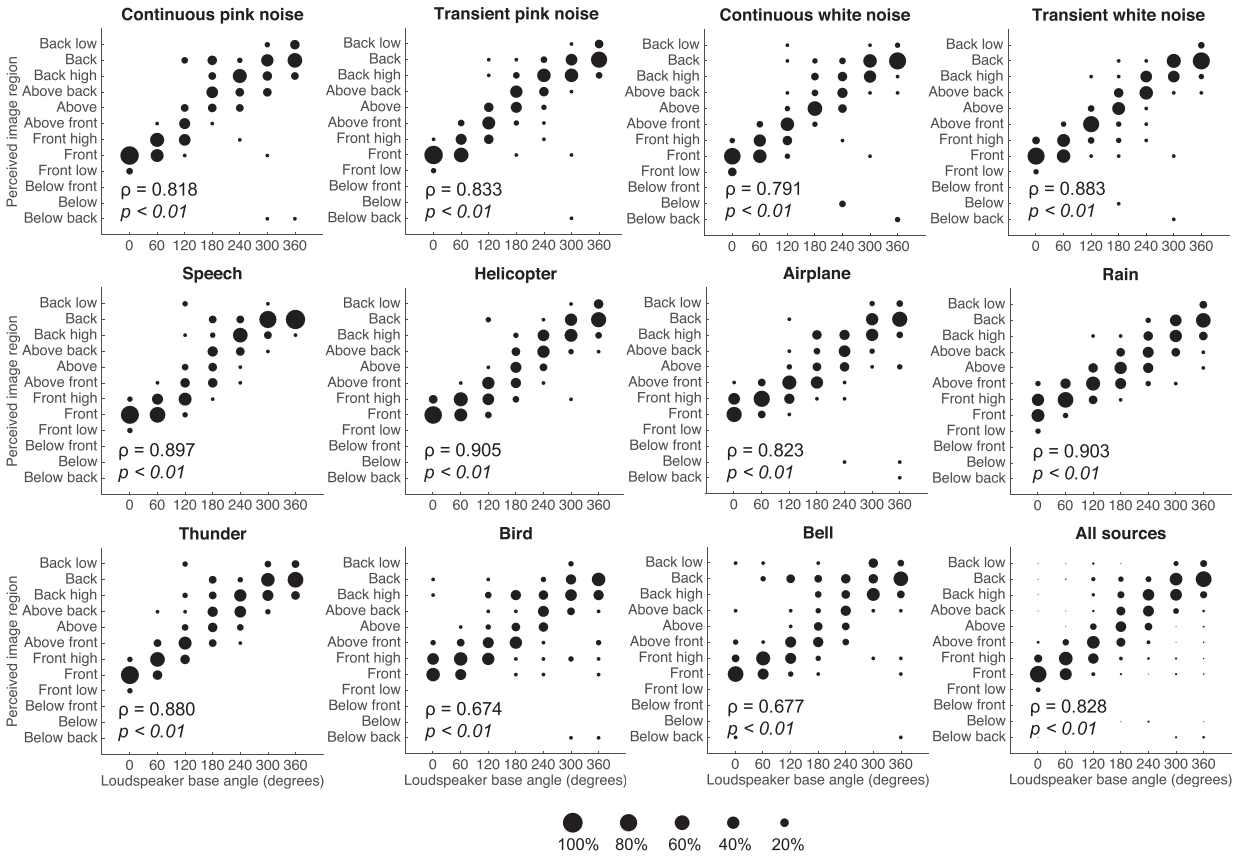

Fig. 6. Bubble plots of subject responses obtained for each loudspeaker base angle, separated for each sound source; the diameter of each filled circle represents the percentage of responses produced for each condition. The ρ values are Spearman correlation coefficients.

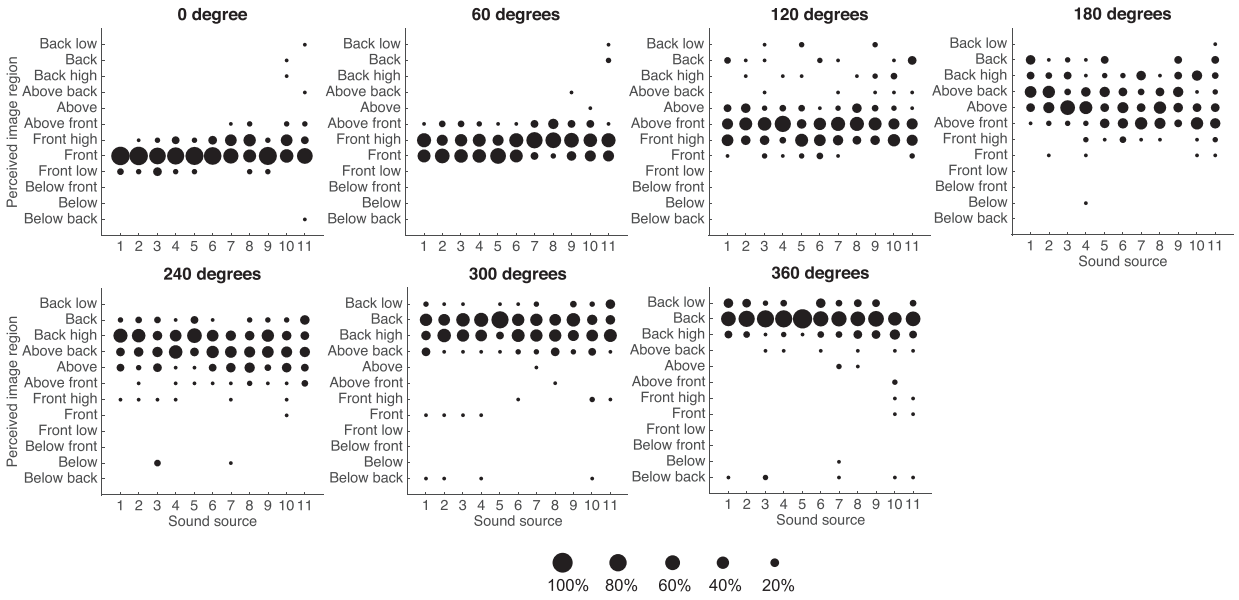

Fig. 7. Bubble plots of subject responses obtained for each sound source, separated for each loudspeaker angle; the diameter of each filled circle represents the percentage of responses produced for each condition. Sound source labels: 1 = continuous pink noise, 2 = transient pink noise, 3 = continuous white noise, 4 = transient white noise, 5 = speech, 6 = helicopter, 7 = airplane, 8 = rain, 9 = thunder, 10 = bird, 11 = bell.

Table 1. Median perceived regions for each pair of sound source and loudspeaker base angle: F (Front), FH (Front High), AF (Above Front), A (Above), AB (Above Back), BH (Back High), B (Back).

	0°	60°	120°	180°	240°	300°	360°
Pink noise – continuous	F	FH	AF	AB	BH	BH	B
Pink noise – transient	F	F	AF	AB	BH	BH	B
White noise – continuous	F	FH	AF	A	AB	B	B
White noise – transient	F	FH	AF	A	AB	B	B
Speech	F	F	FH	AB	BH	B	B
Helicopter	F	FH	AF	A	AB	BH	B
Airplane	F	FH	AF	A	AB	BH	B
Rain	F	FH	AF	A	AB	BH	B
Thunder	F	FH	AF	AB	AB	B	B
Bird	FH	FH	AF	A	AB	BH	BH
Bell	F	FH	AF	A	AB	BH	B

Table 2. Results of Friedman tests showing the main effect of loudspeaker base angle on perceived image region, conducted for each sound source.

Sound source	X^2	df	p	Kendall's W
Speech	129.976	6	.000	.867
Helicopter	131.811	6	.000	.879
Airplane	111.338	6	.000	.742
Rain	133.199	6	.000	.888
Thunder	127.703	6	.000	.851
Bird	87.760	6	.000	.585
Bell	86.429	6	.000	.576
Pink noise (cont.)	113.855	6	.000	.759
Pink noise (tran.)	123.780	6	.000	.825
White noise (cont.)	102.821	6	.000	.685
White noise (tran.)	123.429	6	.000	.823

sources. The bird and bell had medium effect sizes (W = .585 and .576, respectively), whereas such sources as the rain, speech, and helicopter having large effect sizes (W > .86). This indicates that the bird and bell had smaller differences among the base angles in the subjective responses. Among the noise sources, it is noticeable that the transient sources had slightly greater effect sizes than the continuous ones for both the white and pink noises.

The significant effect of the base angle can also be visually observed in Fig. 6. Overall, there appears to be a general tendency that the perceived image region varied from the "front" to the "above" as the loudspeaker base angle increased from 0° to 180°. The 240° angle generally produced the "above back" or "back high" localization, depending on the sound source. Sounds presented from the loudspeaker pairs with the 300° and 360° base angles were generally localized at the "back high" or "back" regions, respectively. From Table 1 it can be observed that, for such sources as the helicopter, airplane, rain, and bell, the median perceived region increased in a regular step as the loudspeaker base angle increased. However, for the speech source, the median response for each base angle was consistently biased towards lower regions than those for the aforementioned sources. For example, 60° had the "front" rather than the "front high," 120° the "front high" rather than the "above front," and 180° the "above back" rather than the "above." It is noticeable that the pink noise sources and the thunder also had the "above back" median response rather than the "above" for the base 180° angle. Moreover, the median responses for the pink noise sources for 240° were the "back high" rather than the "above back." Despite such biases, Spearman's rank-order correlation tests suggest that the base angle and the perceived region had a significant and strong monotonic relationship ($\rho > .79$, $p < .01$) for all sound sources except for the bird and bell, which had moderate correlations ($\rho \approx .67$, $p < .01$).

In order to examine which pairs of base angles were statistically significant for each sound source, Wilcoxon signed rank tests were performed with a Bonferroni correction applied. Cohen's effect size r was also taken into consideration together with p value in judging the significance of difference between base angles, as suggested by a number of researchers [26–28]. An r value greater than 0.5 suggests a large effect, and that greater than 0.4 a medium to large effect [26]. In the current analysis, all pairs with $r > 0.4$ were considered to have a significant difference even though their p values were larger than the standard cutoff of .05. Most pairs of base angles were found to have a significant difference in perceived region ($p < .05$, $r > .45$). However, the number and conditions of significant pairs varied depending on the sound source. For example, for the rain and the helicopter, all pairs of angles were found to have significant differences ($r > .4$), whereas the bell had a number of non-significant pairs ($p < .05$, $r < .4$; 0°-360°, 60°–120°, 120°–180°, 180°–360°, 240°–360°, and 300°–360°). In addition, for a number of sources, several adjacent angles were found to have non-significant differences ($p > .05$, $r < .4$): 0°–60° for the speech, bird, and bell; 120°–180° for the transient white noise, airplane, bird, and bell; 180°–240° for the continuous pink noise, continuous white noise, airplane, and thunder; and 300°–360° for the continuous pink noise, continuous white noise, speech, airplane, thunder, bird, and bell.

Last, a correspondence analysis was performed to provide a perceptual mapping between loudspeaker base angles and perceived image region for the overall responses. The result suggests that there were two effective perceptual dimensions that accounted for 48.7% and 27.8% of the total inertia, respectively. From the two-dimensional correspondence map shown in Fig. 8, it can be observed that the horizontal dimension had a sequential positioning of the loudspeaker base angles from 0° to 360°, divided between

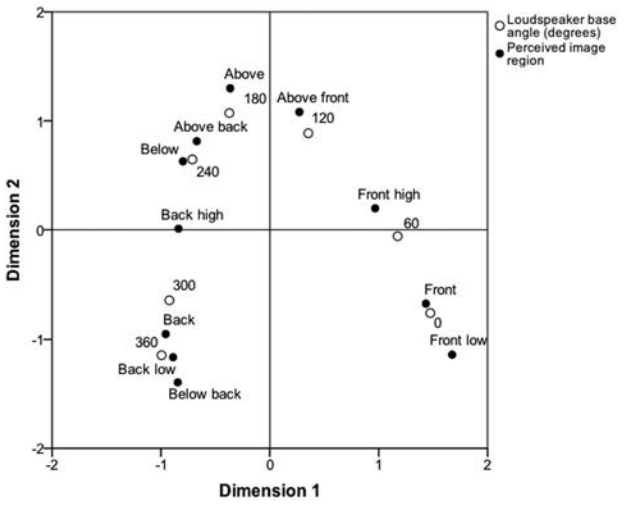

Fig. 8. Correspondence map showing scores on two effective dimensions for loudspeaker base angle and perceived image region.

Table 3. Results of Friedman tests showing the significance of association between sound source and perceived image region, conducted for each loudspeaker base angle.

Loudspeaker base angle	X^2	df	p	Kendall's W
0°	60.993	10	.000	.244
60°	37.743	10	.000	.151
120°	14.123	10	.167	.056
180°	39.693	10	.000	.159
240°	30.098	10	.001	.120
300°	17.645	10	.061	.071
360°	32.340	10	.000	.129

front (0°, 60°, and 120°) and side/rear (180°, 240°, 300°, and 360°) angles. The vertical dimension was divided between base angles closer to the sagittal plane (0°, 60°, 300°, and 360°) and those closer to the frontal plane (120°, 180°, 240°). The sequential pattern is also observed for the perceived image regions, with each loudspeaker base angle clustering with one or more perceived image regions. For example, the 0°, 60°, and 120° are closely located to the front, front high, and above front regions, respectively. The 180° is located in between the above or above back, while the 240° is closest to the above back. The 300° and 360° cluster with the back region.

2.2 Relationship between the Sound Source and the Perceived Elevation for Each Base Angle

From Fig. 7 it can be generally seen that, depending on the loudspeaker base angle, some sound sources had different response patterns compared to others. Friedman tests suggest that the main effect of sound source was significant for all base angles ($X^2(10) > 30$, $p < .01$) apart from 120° and 300°, although the effect sizes (W) were small for all base angles (see Table 3). Based on the results of Wilcoxon signed-rank tests with the Bonferroni procedure and Cohen's effect sizes, only a small number of pairs of sound sources were found to have a statistical significance ($p < .05$, $r > .04$).

For 0°, most of the significant pairs were associated with the bird or the airplane. Especially, the median response for the bird was the "front high," whereas that for the other sources was the "front," as shown in Table 1. This difference between the bird (MED = "front high") and the continuous pink noise (MED = "front") was significant at the 1% level ($p < .01$, $r > .5$). The airplane, which had a slightly response bias towards the "front high," was significantly different from the continuous pink and white noises, which were biased towards the 'front low' ($p < .01$, $r > .5$), although their median response was commonly the "front." On the other hand, all of the significant pairs for 60° were associated with the rain or the speech, with the most significant difference found between the two sources ($p < .01$, $r > .5$); the median regions for the rain and the speech were the "front high" and "front," respectively. For 180°, the continuous pink noise (MED = "above back") was found to be significantly different from the transient white noise, helicopter, rain, and bird (MED = "above") ($p < .01$, $r > .5$). The thunder (MED = "above back") was also found to be significantly different from the rain and the bird (MED = "above") with a medium to large effect size ($r > .4$). Although the 240° and the 360° base angles did not have any sound sources that were significantly different at the 5% level, the speech (MED = "back") was found to be different from the helicopter, airplane, rain, and bird (MED = "back high") with a medium to large effect size ($r > .4$) for 240°, and the bird (MED = "back high") from all sources except the continuous pink and white noises and airplane (MED = "back") for 360°.

3 DISCUSSION

Overall, the current results presented above seem to be in agreement with those found in early studies, e.g., de Boer (1947) reported that the phantom images created by loudspeakers placed directly at the listener's sides were perceived at an elevation angle of around 100°; Damaske and Mellert (1969/70) found that the mean perceived elevation angles for the loudspeaker base angles of 180° and 240° were around 120°. As mentioned earlier, however, the sound source(s) used for de Boer's experiment were not reported, while Damaske and Mellert used only a single source of a white noise ranging between 650 Hz and 4.5 kHz. The current study used 11 sound sources of various types. The results exhibited that there were significant differences between different sound sources in the pattern of elevation perception depending on the loudspeaker base angle. This leads to discussion on what cue(s) trigger the phantom image elevation effect. In the following subsections, the spectral and temporal characteristics of the sound sources, the spectral energy distribution of ear-input signals and potential psychological factors that might have affected the results are considered. Furthermore, a new hypothesis to explain the effect is provided from a cognitive perspective.

3.1 Sound Source Effect

The results for the noise stimuli are first discussed. The white noise has a perfectly flat frequency spectrum, whereas the spectrum of the pink noise is weighted towards lower frequencies. The median responses for all pink and white noise stimuli were "above back" and "above," respectively (Table 1). This seems to suggest that the perceived degree of phantom image elevation is associated with the spectral balance of sound source. The largest difference was found between the continuous pink noise and the transient white noise. The transient white noise was also found to have a slightly larger monotonic correlation (ρ) between the base angles and the perceived image elevation compared to the continuous pink noise (Fig. 6). This seems to be associated with the "pitch-height" effect, which was introduced in Sec. 0, i.e., a lower frequency tends to have a bias towards a lower perceived position, while a higher frequency towards a higher perceived position. It is possible that the pink noise was perceived to be less elevated than the white noise due to the low frequency dominance in the spectrum. Another potential explanation for the result might be the dependency of the directional bands theory on the sound source spectrum, which will be discussed in detail in Sec. 3.3.2. Additionally, despite the statistical non-significance, the median response for the continuous pink noise with the 240° angle was "back high," whereas that for the transient white noise was "above back." From these results, it might be further suggested that the transient nature of sound contributes to the perceived elevation of phantom image.

The dependency of the elevation effect on the spectral and temporal characteristics of the signal is further demonstrated by the results for the natural sources. The response pattern for the rain appeared to be notably similar to that of the transient white noise; both sources had the same median perception region for each base angle apart from the 300°. As can be seen in Fig. 4, the rain had a broad and relatively flat frequency spectrum, which was similar to the spectrum of the white noise. The rain also had transient temporal characteristics. On the other hand, the spectra of the speech, helicopter, airplane, and thunder were similar to that of the pink noise, and the response patterns for these sources were also similar. That is, they had no strong "above" localization for 180°, and the responses for the other angles tended to be more spread toward the front and back regions compared to those for the rain and the transient white noise.

The bird and bell had the most spread responses among all sources. Especially, from Fig. 6 or 7 it can be seen that these sources had the largest number of responses affected by "front-back" confusion. The bird had a narrow spectrum mainly ranging between 1 and 4 kHz. This seems to support Asano et al. (1990) who suggest that frequencies below 2 kHz are important for front-back discrimination. On the other hand, the bell had inharmonic strike tones between 100 Hz and 2 kHz; each partial produces a pitch whose amplitude decays slowly. The onset of the source was about 200 ms, thus having no strong transient cue. It is widely known that steady-state sound is more difficult to localize than transient sound [29, 30], and this might be the reason for the largely inconsistent subject responses for the bell.

Overall, the results generally suggest that a transient source with a broader and flatter frequency spectrum would produce a more effective phantom image elevation effect than a more continuous source with low-frequency-dominant or narrow-band characteristics.

3.2 Loudspeaker Base Angle Effect

Blauert's "directional bands" hypothesis [16], which was described earlier, were derived from subjective localization responses given to three broad regions in the median plane: "front" ($-45° < \varphi \leq 45°$), "above" ($45° < \varphi \leq 135°$), and "back" ($135° < \varphi \leq +225°$), where φ is the elevation angle. Blauert referred to any specific 1/3-octave frequency band as a directional band if the number of responses given to one region for the band was judged to be significantly larger than the total number of responses given to the other regions at the 5% level by binomial tests.

A similar attempt was made for the current results in order to map loudspeaker base angles with the broad "front," "above," and "back" regions, with each covering the 90° span of three corresponding sub-regions (e.g., front = front low + front + front high). For each base angle, the total number of responses given to the three sub-regions of each broad region was compared against that given to the other nine sub-regions through a binomial test. The base angles with a statistical significance ($p < .05$) for any specific region with the probability proportion of total responses being greater than 50% are referred to as "directional base angles" here. Given the number of observations for each base angle condition for each source being 25, the statistical chance levels to achieve the 50% proportion with $p < .05$ and $p < .01$ were 66.9% and 73.8%, respectively. For all data included for each base angle condition (a total of 275 observations), the chance levels for the 5% and 1% significance levels were 55.4% and 57.5%, respectively.

Table 4 shows the results from the analysis. 0° and 60° were found to be the directional base angles for the "front" perception for every source, whereas 300° and 360° were for the "back." However, the 120°, 180°, and 240° had an obvious source dependency. While 180° was the "above" angle for all sources except the bird, bell, and continuous pink noise, 120° and 240° were directional base angles only for a few sources. That is, 120° was the "above" angle for the rain, transient pink noise, and transient white noise. 240° was the "back" angle for the speech, whereas it was the "above" angle for the rain. This analysis further supports the discussion on the dependency of the elevation effect on sound source characteristics. Meanwhile, including data for all sources, the 240° was shown to be the only base angle without a statistical significance for any region.

3.3 Theoretical Explanation

This section first discusses existing theoretical explanations on the phantom image elevation effect based on the role of head rotation and the pinnae-related spectral energy distribution. A novel hypothesis about the role of low

Table 4. Directional base angles derived from binomial tests, after Blauert's method [16]; each region is categorized by gradation and the percentage value indicates the proportion of total responses given to the corresponding region. ** $p < .01$; * $p < .05$.

					Front	Above	Back
	Loudspeaker base angle						
Source	0°	60°	120°	180°	240°	300°	360°
Speech	100%**	96%**	–	72%*	72%*	96%**	100%**
Helicopter	100%**	96%**	–	80%**	–	88%**	96%**
Airplane	96%**	84%**	–	72%*	–	88%**	84%**
Rain	92%**	72%*	76%**	92%**	68%*	76%**	92%**
Thunder	100%**	80%**	–	68%*	–	92%**	100%**
Bird	84%**	80%**	–	–	–	72%*	76%**
Bell	80%**	84%**	–	–	–	92%**	84%**
Pink cont.	100%**	96%**	–	–	–	72%*	96%**
Pink tran.	100%**	88%**	68%*	80%**	–	88%**	100%**
White cont.	100%**	88%**	–	76%**	–	92%**	88%**
White tran.	100%**	92%**	80%**	76%**	–	88%**	96%**
All	95.6%**	86.9%**	56.4%**	72.4%**	–	85.8%**	92%**

frequency acoustic crosstalk on the effect is then proposed based on the evidence of matching between the torso-related spectral notch of a real elevated source and the acoustic-crosstalk-related spectral notch of a phantom source.

3.3.1 The Role of Head Rotation

Early studies by de Boer [18] and Leakey [20] attempted to explain the phantom image elevation effect by the role of head movement on vertical localization. The basic ideas for their explanations were commonly that, when a listener horizontally rotates his or her head by a certain degree in front of a stereophonic pair of loudspeakers presenting coherent signals, the resulting phantom image would be elevated to the position of a real source in the median plane that causes the same amount of interaural time difference (ITD) with the same degree of head rotation. They suggested that this condition would be met if the median plane elevation angle were half the loudspeaker base angle.

However, this explanation can be challenged as follows. First, a real source that is negatively elevated in the median plane can also produce the same ITD as the phantom source with a head rotation. Similarly, the same ITD can be produced between two loudspeaker pairs with the same azimuth from the sagittal plane (e.g., 120° and 240°). Last but more importantly, the elevation effect can be clearly perceived without any head rotation, as observed in the current experiment as well as in Blauert's [16].

3.3.2 The Role of Pinnae-Related Ear-Input Spectrum

The complex spectral shape of the the pinnae-related part (i.e., above 3 kHz) of the head-related transfer function (HRTF) is known to be the main cue for median plane localization [2–6,16]. Existing localization prediction models in the literature [31–33] generally assume that the human auditory system determines the direction of sound source by comparing the ear-input spectrum of the source with a set of template spectra for target source directions that are stored in the system; the source will be most likely to be perceived at the direction where the similarity of the template spectrum to the ear-input spectrum is the highest. In the context of phantom source localization in the median plane, Baumgartner and Madjak [34] claims that localization inaccuracy in the vector base amplitude panning (VBAP) [35] is caused by the discrepancy between the HRTF of a phantom source and that of a real source at the target direction.

From the above, it was of interest to examine if there would be a high similarity between a phantom source resulting from a particular loudspeaker base angle and a real source elevated at the perceived position of the phantom source. To this end, the spectral magnitudes of the left ear signals of the phantom sources resulting from the 60°, 120°, 180°, 240°, and 300° base angles were compared against those of real sources that are positively elevated at 30°, 60°, 90°, 120°, and 150° in the median plane, each of which lies within the median perceived region for each base angle condition for most sound sources (see Table 1). For this analysis, the MIT's KEMAR head-related impulse response (HRIR) database[2] was used. The measurement results are plotted in the panels on the first and second rows of Fig. 9. In general, for any pair of base angle and compared real elevation (e.g., Fig. 9(a1) and Fig. 9(a2)), the overall shapes of the pinnae-related spectra (i.e., above 3 kHz) do not appear to be similar; the positions and magnitudes of the peaks and notches are noticeably different. Having measured the HRTFs of all other possible median elevation angles from the KEMAR database with a 10° resolution, no evidence for a high similarity in the pianne-related ear-input spectrum was found between the real and phantom sources. This suggest that the spectral similarity assumption of the conventional localization models, which were mentioned earlier in this section, could not be applied in explaining the phantom image elevation effect.

In contrast to the viewpoint of the ear-input spectrum matching between the phantom and real source, Blauert [36] asserts that the phantom image elevation effect is the

[2] http://sound.media.mit.edu/resources/KEMAR.html

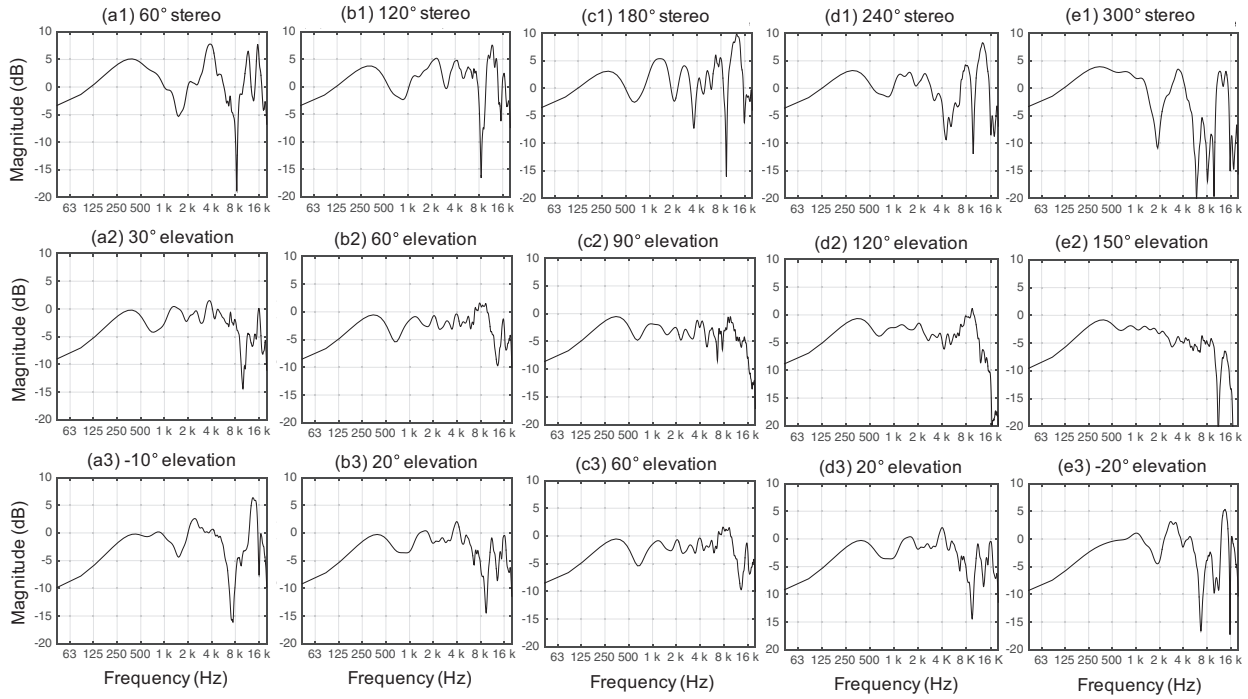

Fig. 9. The ear-input frequency spectra of phantom and real source signals measured using the MIT KEMAR dummy head database[b]. Each panel on the first row shows the spectrum of the left ear signal of a phantom center source resulting from each of the five loudspeaker base angles tested. Each panel on the second row shows the spectrum of the left ear signal for a real single source elevated in the median plane at half the base angle presented in the same column. Each panel on the bottom row presents the ear-input spectrum for a median plane elevation that produces the same first notch frequency as the base angle of the same column.

relative magnitude weightings of the directional bands [16] in the ear-input spectrum. For example, the dominances of frequencies around 1 kHz, 4 kHz, and 8 kHz in the spectrum would respectively determine the perceptual weightings of the "backness," "frontness," and "aboveness" of the resulting image. This appears to be demonstrated in the phantom source ear-input spectra plotted in the first row of Fig. 9. That is, as the loudspeaker base angle increases from 60° to 180°, frequencies around 4 kHz tend to decrease in level, whereas those around 8 kHz tend to increase. Based on Blauert's explanation, this would mean that the perceived image has less "frontness" and more "aboveness" as the base angle increased from 60° to 180°, which agrees with the subjective results shown in Fig. 6. From 180° to 300°, the levels at frequencies around 4 kHz appear to be decreased further, which suggests that the image further loses the "frontness," thus being perceived in the "back" regions. The levels in the 8 kHz region at 240° appear to be similar to those at 180°. Together with the reduced levels in the 4 kHz region, this seems to explain the dominant "above back" perception at 240° found in the subjective results. At 300°, there is a dramatic level decrease in the 8 kHz region, which would further decrease the "aboveness" of the perceived image. A similar trend could be observed in the ear-input spectra of the direct sound part of the BRIR measured in the listening room using the KU100 dummy head (Fig. 3), although the relative magnitudes of the spectra is different to the results obtained from the KEMAR dummy head.

Although the above observation initially seems to validate Blauert's explanation on the elevation effect based on the directional bands, it should be considered that the perceptual influence of the directional band weighting would fundamentally depend on the spectrum of the sound source. For example, a number of natural sound sources tend to have high-frequency roll-off characteristics and therefore the relative effect of the 8 kHz band for "aboveness" would be smaller for such sources. This might be a potential explanation for the result from the current study showing that the low-frequency-biased sources such as the pink noise, the speech, and the helicopter were perceived to be slightly less elevated than the white noise and the rain, which had a flatter spectrum.

3.3.3 A New Hypothesis on the Role of Acoustic Crosstalk and Torso Reflections at Low Frequencies

The previous section attempted to explain the subjective results based on theories focusing on the pinnae-related ear-input spectrum. In this section a new hypothesis on the phantom image elevation effect is provided from a viewpoint of spectral notch produced at a frequency below 3 kHz. The basic idea is that for frequencies below 3 kHz a horizontally oriented phantom center source may be perceived to be elevated to a position of a real source in the median plane where the frequencies of the first notches in the ear-input spectra of the real and phantom sources match. In other words, there may be a cognitive association by the brain between the horizontal phantom and vertical

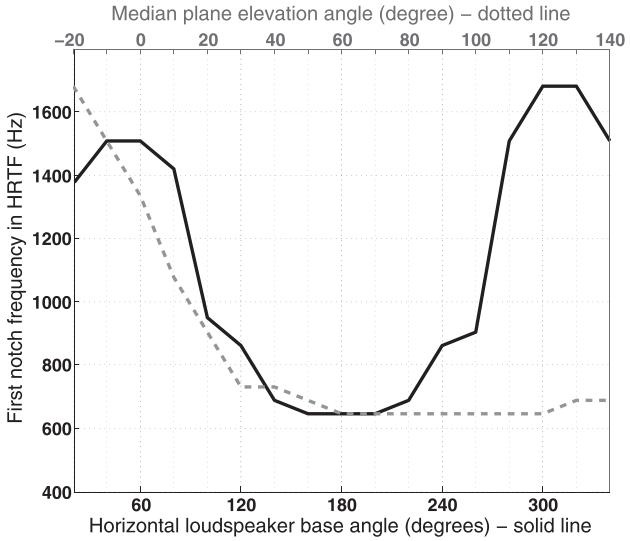

Fig. 10. First notch frequency in the ear-input spectrum of the left ear signal measured as a function of the median plane elevation angle (dotted grey line) and as a function of the horizontal stereophonic base angle (solid black line). The MIT's KEMAR dummy head HRIR database was used for the measurements. The notch frequency is determined with the FFT bin resolution of 43 Hz.

real sources when they produce a spectral notch at an identical frequency below 3 kHz.

The basis for this hypothesis is as follows. As mentioned in Sec. 0, for a sound source in the median plane, a torso reflection produces a spectral notch in the HRTF below 3 kHz when it is combined with the direct sound at the ear [8, 9]. Algazi et al. [9] showed that such a notch varies in frequency as a function of torso reflection delay, which is elevation-dependent. The torso reflection delay tends to increase as the elevation angle increases, and reaches its maximum when the source is in the "above" region. For example, according to Algazi et al.'s measurement data using the KEMAR dummy head, the torso reflection delay is around 0.5 ms at 0° median plane elevation, whereas that reaches the maximum around 0.75 ms at about 60° median plane elevation. A similar relationship can be also observed between the loudspeaker base angle in horizontal stereophonic reproduction and the delay of the acoustic crosstalk (contralateral) signal. That is, the acoustic crosstalk delay increases as the loudspeaker base angle increases, with the 180° base angle producing the maximum delay. In this case, however, the notch produced in the resulting ear-input spectrum is due to the combination of the acoustic crosstalk delay and the torso reflection delays of both ipsilateral and acoustic crosstalk signals.

In order to show the relationship between the loudspeaker base angle of a phantom source and the median plane elevation of a real source with respect to the first notch frequency (f_N) produced in the ear-input spectrum, Fig. 10 plots f_N measured as a function of the median plane elevation angle (dotted grey line) and that as a function of the horizontal stereophonic base angle (solid black line). The MIT's KEMAR HRIR database was used for the measurements, and the FFT bin resolution was 43 Hz. Table 5 presents the median plane elevation angles at a 10° resolution that best match the five loudspeaker base angles of 60°, 120°, 180°, 240°, and 300° with respect to the f_N. The HRTF of the best matching elevation for each base angle is also shown in the bottom row of Fig. 9 for a visual comparison.

Table 5. The angle of the median plane elevation that best matches each of the horizontal loudspeaker base angles of 60°, 120°, 180°, 240°, and 300° with respect to the first notch frequency in the ear-input spectrum, measured using the MIT's KEMAR dummy head HRIR database. The notch frequency is determined with the FFT bin resolution of 43 Hz.

Phantom source		Real source	
Loudspeaker base angle	f_N	Median plane elevation angle	f_N
60°	1507 Hz	−10°	1507 Hz
120°	861Hz	20°	904 Hz
180°	646 Hz	60° – 120°	646 Hz
240°	861 Hz	20°	904 Hz
300°	1680 Hz	−20°	1680 Hz

It was found that the f_N for the 180° base angle condition (646 Hz) matched that for multiple median plane elevation angles ranging between 60° and 120°. This tends to agree with the subjective results showing that the 180° base angle produced the median perceived region of "above" with some spreads from "above front" to "above back." For the other base angles, however, the elevation angles predicted based on f_N did not match the subjective results. For both the 60° and 300° base angles, the best f_N-matching median plane elevation was at a negative angle (–10° elevation for the 60° base angle; $f_N = 1507$ Hz, and –20° for the 300°; $f_N = 1680$ Hz). For both the 120° and 240° base angles ($f_N = 861$ Hz), the best matching elevation was 20° ($f_N = 904$ Hz). From this, it might be further hypothesized that the total degree of elevation for a broadband phantom image would be determined based on the perceptually weighted combination of the crosstalk-related spectral notch cue at low frequencies and the spectral magnitude distribution cue at high frequencies (i.e., the directional bands), depending on the frequency spectrum of the sound source as well as the loudspeaker base angle. For instance, the high-frequency cue was perhaps more responsible for the subjective results showing that the 60° and 120° base angles respectively produced the "front high" and "above front" median responses. This is because the notch cue alone may only have elevated the phantom image to around –10° and 20° for the two base angles, respectively, based on the current hypothesis. On the other hand, for the 180° base angle, the f_N cue might have a similarly or more important contribution to the perceived elevation than for the other base angles since the notch-predicted elevation angle (between 60° and 120°) agrees with the perceived region of "above."

A verification for the above hypothesis is currently underway by the author. As a first step, the role of acoustic crosstalk for the elevation effect at low and high frequencies was examined in an individualized binaural listening environment using headphones. This allows the acoustic

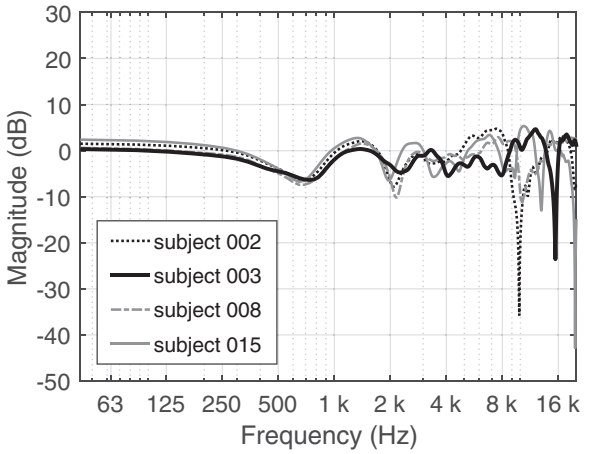

Fig. 11. The ear-input spectra of a phantom center source from the 90° loudspeaker base angle for four different subjects, taken from the SADIE HRIR database.

crosstalk elements in the ear-input signals to be easily removed or manipulated. For an initial listening test, five subjects judged the perceived position of a phantom center image resulting from loudspeakers at the 180° base angle, which was binauralized using their own BRIRs measured in the same listening room that was used in the current study. The binaural stimuli created using white noise, rain, and thunder were manipulated such that the crosstalk signals were totally removed, low-pass filtered or high-pass filtered at 3 kHz. Each subject repeated the test five times in a randomized order. Preliminary results from the test [37] showed that the crosstalk below 3 kHz was necessary for producing a statistically significant "above" and "outside-the-head" perception, whereas the crosstalk above 3 kHz produced a significant "inside-the-head" perception in the "above" or other regions. Similar results were obtained for octave band pink noise signals centerd at 500 Hz and 8 kHz; the crosstalk was necessary for the 500 Hz band to be externalized in the above region, whereas the responses for the 8 kHz band were equally split between "above inside the head" and "above outside the head" regardless of the presence of the crosstalk. This supports the currently proposed hypothesis on the role of crosstalk-related low-frequency cue on the elevation effect. This study requires further tests with more subjects and also for an anechoic condition in order to examine the externalization effect observed with low frequencies. Full results from this study will be presented in a future publication.

3.3.4 Inter-Subject Variability in Ear-Input Spectrum

It is worth noting that pinnae-related ear-input spectrum can have a substantial inter-subject variability due to different pinnae sizes and shapes. This is demonstrated in Fig. 11, which plots the ear-input spectra of a phantom source resulting from the 90° loudspeaker base angle for four different subjects taken from the SADIE HRIR database[3]. The large individual differences observed at frequencies above 3 kHz indicate that the effect of pinnae-related cue on the phantom image elevation effect might be subject-dependent. For example, the spectrum for the subject 003 appears to have a minimal magnitude difference between 4 kHz and 8 kHz compared to the other subjects. This seems to suggest that for this subject the directional band weighting claimed by Blauert would not be an effective cue for the elevation effect. In addition, the subjective result showing the response spread across "above front," "above," and "above back" might also be associated with the inter-subject variability in pinnae-related spectrum.

On the other hand, the crosstalk-related notch frequencies in the low frequency region appear to be relatively consistent across all of the four subjects in Fig. 11. This seems to be due to similar ear-to-ear distances of the subjects. In general, it could be said that the ear-to-ear and ear-to-torso distances would not have dramatic individual differences compared to the differences in the external ear shape. In this respect, the low-frequency notch cue might be suggested to be more predictable than the high-frequency spectral balance cue. However, based on the new hypothesis discussed in the previous section, it is also considered that individual differences in the crosstalk and torso reflection delays due to different ear-to-ear and ear-to-torso distances would likely produce an inter-subject inconsistency in the perceived elevation of a phantom source resulting from a given base angle. The subject-dependency of the phantom image elevation effect will be investigated exclusively in a future study by testing a number of subjects who have substantially different ear-to-ear and ear-to-shoulder distances as well as different external ear shapes.

Additionally, a study by Katz and Parseihian [38] suggests that an accurate presentation of the HRTF for the target source positon might not always be necessary for an accurate localization. In their binaural listening experiment, subjects were asked to choose their most and least "preferred" HRTFs from 46 different individual HRTFs. It was found that the most preferred HRTFs that the subjects chose were not necessarily their own HRTFs and that their localization accuracies improved when they used the most preferred HRTFs rather than the least preferred ones. This result seems to indicate complex cognitive nature of the human localization mechanism, which still requires further research. In the context of the current study, it would be interesting to conduct a similar binaural experiment where one compares his or her own and some other people's HRTFs. It would be examined whether the phantom image elevation effect could still be perceived with a pinnae-related spectral balance or crosstalk-related notch that is different to the subject's own. It would also be worth investigating if the elevation perception could even be enhanced by using someone else's HRTF with particular characteristics.

[3] https://www.york.ac.uk/sadie-project/binaural.html

3.4 Expectancy Bias

From a psychological point of view, some of the current results might demonstrate potential subject-expectancy biases on elevation judgment. That is, subjects' responses might have been affected by the auditory or visual positions of the sound sources that are likely in real life. A good example for this might be the difference found between the speech and some other sources. The speech is most likely to be heard at around the ear height in real life, whereas such sources as the rain, bird, bell, helicopter, and airplane tend to be heard or seen from elevated positions. For the real loudspeakers at the base angle of 0° and 360°, the responses for the speech had strong biases towards the "front" and "back" regions, respectively. On the other hand, the other sources mentioned above had more spread responses for the same loudspeaker positions with a number of responses in elevated regions. Furthermore, the 180° base angle did not produce a strong "above" perception for the speech, whereas it did for the rain.

3.5 Practical Implications for 3D Sound Recording and Rendering

Next generation three-dimensional (3D) multichannel audio formats such as Dolby Atmos and Auro-3D employ additional height and overhead (a.k.a. Voice of God) channels to provide an immersive listening experience. While the VBAP [35] is widely considered as a standard technique for 3D sound panning, new techniques for 3D sound recording [39–41] and upmixing [42–44] are being developed. Furthermore, the recently standardized MPEG-H 3D audio codec [45] offers a highly efficient, object-based coding and transmission of 3D audio for domestic broadcasting.

However, in home environments it is often practically difficult to place loudspeakers at elevated positions or mount them on the ceiling. Several signal processing methods have recently been proposed to create an elevated image using horizontal plane loudspeakers [46–48]. Such methods typically attempt to convey the HRTFs of an elevated source to the listener's ears either by means of acoustic crosstalk cancellation with two frontal loudspeakers [46] or by routing HRTF-filtered signals to rear loudspeakers in the conventional 5.1-channel system [47, 48]. However, results from the current study suggest that such filtering processes would not be necessary for creating the illusion of a virtual overhead image; coherent signals could simply be routed to the side or rear loudspeakers for the same effect, although its effectiveness would depend on the temporal and spectral characteristics of the signals.

From an informal test it was observed that the elevation effect disappeared when the original signals from the two loudspeakers were decorrelated by even a small degree, e.g., lowering the interchannel cross-correlation coefficient (ICCC) from 1 to 0.8; two separate images were localized at the left and right loudspeaker positions. From this, a practical method to capture and render ambient sound over the virtual hemi-sphere without using physically elevated loudspeakers is suggested as follows. Three microphones are placed in line at a large distance from the sound source in a recording venue, with the spacing between each other being 2 to 3 m. Based on the diffuse field correlation model by Cook et al. [49], this range of spacing is considered large enough to produce low cross-correlation between two omni-directional microphone signals above 100 Hz. Further decorrelation could be achieved by using directional microphones [50], e.g., sideward-facing figure-of-eight or backward-facing cardioid microphones for the left and right channels and an upward-facing figure-of-eight or backward-facing cardioid microphone for the center for a maximal attenuation of the direct sound. As illustrated in Fig. 12, the outer microphone signals are then routed to the left and right loudspeakers at the listener's sides in order to produce widely spread ambient sound images laterally. Meanwhile, the center signal is fed into both the left and right loudspeakers in order to create a phantom overhead image. This consequently creates ambient images over the virtual upper-hemisphere, thus enhancing the perception of listener envelopment (LEV).

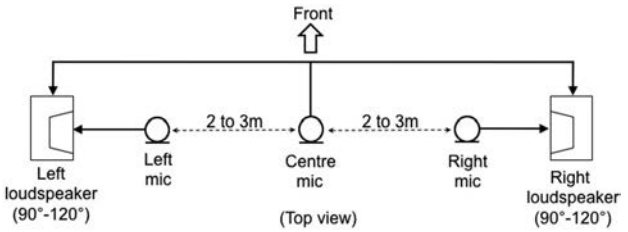

Fig. 12. A new three-channel ambience microphone technique proposed for creating ambient images over the virtual upper-hemisphere.

4 CONCLUSION

This study conducted an extensive investigation on the elevation of phantom center image created by horizontally placed stereophonic loudspeaker pairs. The aim was to examine the dependencies of the effect on sound source and loudspeaker base angle. Eleven sound sources comprising seven natural sources and four noise sources were tested. Twelve loudspeakers were arranged in a circle at 30° intervals. Each source was presented in seven different loudspeaker base angle conditions, comprising 0°, 60°, 120°, 180°, 240°, 300°, and 360°. Listening tests have been carried out to elicit the perceived image region for each experimental condition.

The test results confirmed the general relationship between the phantom image elevation and the loudspeaker base angle reported in early studies; as the base angle increased from 0° to 180° the perceived image was elevated from the front to the above region. However, this tendency was found to have a significant dependency on the spectral and temporal characteristics of the sound source. In general, sources with a broad and flat spectrum, such as the white noise and rain, had the most linear mapping between the base angle and the perceived image elevation. Such sources were also found to produce the strongest

"above" perception for the 180° base angle. Sources with a low frequency weight (e.g., pink noise, speech, thunder, airplane, and helicopter) tended to be less elevated than sources that contained more high frequencies (e.g., white noise and rain). This was particularly true for 120°, 180°, and 240°. The bird, which had a narrow spectrum around 2 to 4 kHz, also produced spread responses for those base angles. The bell, which had tone-like temporal characteristics, produced most inconsistent elevation responses overall.

This paper also provided theoretical explanations for the perceived results. The analyses of the spectra of ear input signals for all base angle conditions showed that the variation of spectral balance above 3 kHz depending on the loudspeaker base angle might be related to the perceived results, although the effectiveness of this cue would depend on the spectral weighting of sound source. A novel hypothesis about the role of acoustic crosstalk and torso reflection at low frequencies was also established. At frequencies below 3 kHz, the brain might use the first notch in the ear-input spectrum, which is produced by the combination of acoustic crosstalk and torso reflection, as a cue for localizing a phantom source at an elevated position in the median plane. It is further suggested that that the overall perceived degree of elevation for the phantom source might be determined by some perceptual weighting between the low and high frequency cues. In addition, an expectancy bias on the perceived position of a certain sound source was discussed as a potential factor that affected the subjective results (e.g., the speech and rain).

Future works will include the testing of the phantom image elevation effect in different acoustical environments such as an anechoic chamber and a reverberant concert hall, the formal verification of the new hypothesis that was proposed in Sec. 3.3.3, and the practical applications of the effect in 3D sound panning, recording, and mixing without elevated loudspeakers.

5 ACKNOWLEDGMENTS

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC), UK, Grant Ref. EP/L019906/1. The author thanks Chris Gribben for the creation of the graphical user interface used for the listening test and the staff members and students of the University of Huddersfield's music technology courses who participated in the listening test. He is also grateful to the editor and two anonymous reviewers of this paper for their constructive and insightful comments to improve the initial manuscript.

6 REFERENCES

[1] R. A. Butler and K. Belendiuk, "Spectral Cues Utilized in the Localization of Sound in the Median Sagittal Plane," *J. Acoust. Soc. Am.*, vol. 61, pp. 1264–1269 (1977). https://doi.org/10.1121/1.381427

[2] J. C. Middlebrooks and D. M. Green, "Sound Localization by Human Listeners," *Annu. Rev. Psychol.*, vol. 42, pp. 135–159 (1991). http://doi.org/10.1146/annurev.ps.42.020191.001031

[3] F. L. Wightman and D. L. Kistler, "Factors Effecting the Relative Salience of Sound Localization Cues," in *Binaural and Spatial Hearing in Real and Phantom Environments*, edited by R. H. Gilkey and T. R. Anderson (Erlbaum, Mahwah, NJ, 1997), pp. 1–23.

[4] F. Asano, Y. Suzuki, and T. Sone, "Role of Spectral Cues in Median Plane Localization," *J. Acoust. Soc. Am.*, vol. 88, pp. 159–168 (1990). https://doi.org/10.1121/1.399963

[5] E. A. G. Shaw, "Acoustical Features of the Human External Ear," in *Binaural and Spatial Hearing in Real and Phantom Environments*, edited by R. H. Gilkey and T. R. Anderson (Erlbaum, Mahwah, NJ, 1997), pp. 25–47.

[6] V. C. Raykar, R. Duraiswami, and B. Yegnanarayana, "Extracting the Frequencies of the Pinna Spectral Notches in Measured Head Related Impulse Responses," *J. Acoust. Soc. Am.*, vol. 118, pp. 364–374 (2005). https://doi.org/10.1121/1.4785467

[7] M. Morimoto, M. Yairi, K. Iida, and M. Itoh, "The Role of Low Frequency Components in Median Plane Localization," *Acoust. Sci. & Tech.*, vol. 24, pp. 76–82 (2003).

[8] M. B. Gardner, "Some Monaural and Binaural Facets of Median Plane Localization," *J. Acoust. Soc. Am.*, vol. 54, pp. 1489–1495 (1973). https://doi.org/10.1121/1.1914447

[9] 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).

[10] C. C. Pratt, "The Spatial Character of High and Low Tones," *J. Exp. Psychol.*, vol. 13, pp. 278–285 (1930). https://doi.org/10.1121/1.1349185

[11] O. S. Trimble, "Localization of Sound in the Anterior, Posterior, and Vertical Dimensions of Auditory Space," *Brit. J. Psychol.*, vol. 24, pp. 320–334 (1934). https://doi.org/10.1121/1.1910977

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

[13] 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 (1968a). https://doi.org/10.1121/1.1910976

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

[15] S. Ferguson and D. Cabrera, "Vertical Localization of Sound from Multiway Loudspeakers," *J. Audio. Eng. Soc.*, vol. 53, pp. 163–173 (2003 Mar.).

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

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

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

[19] 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).

[20] D. M. Leakey, "Some Measurements on the Effects of Interchannel Intensity and Time Differences in the Two Channel Sound Systems," *J. Acoust. Soc. Am.*, vol. 31, pp. 977–986 (1959). https://doi.org/10.1121/1.1907824

[21] H. Jo, W. Martens, Y. Park, and S. Kim, "Confirming the Perception of Phantom Source Elevation Effects Created Using 5.1 Channel Surround Sound Playback," *Proceedings of the ACM SIGGRAPH 9th Int. Conf. on Phantom Reality Continuum and Its Applications in Industry* (Seoul, Korea, 2010), pp. 103–110.

[22] M. Frank, "Elevation of Horizontal Phantom Sources," *Proceedings of DAGA 2014*, (Oldenburg, Germany, 2014).

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

[24] 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).

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

[26] J. Cohen, "Things that I Have Learned (So Far)," *Am. Psychol.*, vol. 45, pp. 1304–1312 (1990).

[27] T. V. Perneger, "What's Wrong with Bonferroni Adjustments," *Br. Med. J.*, vol. 316, pp. 1236–1238 (1998). https://doi.org/10.1136/bmj.316.7139.1236

[28] G. M. Sullivan and R. Feinn, "Using Effect Size—Or Why the p Value Is Not Enough," *J. Grad. Med. Educ.*, vol. 4, pp. 279–282 (2012). 10.4300/JGME-D-12-00156.1.

[29] W. Yost, F. Wightman, and M Green, "Lateralization of Filtered Clicks," *J. Acoust. Soc. Am.*, vol. 50, pp. 1526–1531 (1971). https://doi.org/10.1121/1.1912806

[30] B. Rakerd and W. Hartmann, "Localization of Sound in Rooms, III: Onset and Duration Effects," *J. Acoust. Soc. Am.*, vol. 80, pp. 1695–1706 (1986). https://doi.org/10.1121/1.392474

[31] J. C. Middlebrooks, "Narrowband Sound Localization Related to External Ear Acoustics," *J. Acoust. Soc. Am.*, vol. 92, pp. 2607–2624 (1992). https://doi.org/10.1121/1.404400

[32] E. H. Langendijk and A. W. Bronkhorst, "Contribution of Spectral Cues to Human Sound Localization," *J. Acoust. Soc. Am.*, vol. 112, pp. 1583–1596 (2002). https://doi.org/10.1121/1.1501901

[33] R. Baumgartner, P. Majdak, and B. Laback, "Modeling Sound Source Localization in Sagittal Planes for Human Listeners," *J. Acoust. Soc. Am.*, vol. 136, pp. 791–802 (2014). https://doi.org/10.1121/1.4887447

[34] R. Baumgartner and P. Madjak, "Modeling Localization of Amplitude-Panned Virtual Sources in Sagittal Planes," *J. Audio Eng. Soc.*, vol. 63, pp. 562–569 (2015 Jul./Aug.). https://doi.org/10.17743/jaes.2015.0063

[35] V. Pulkki, "Virtual Sound Source Positioning Using Vector Base Amplitude Panning," *J. Audio Eng. Soc.*, vol. 45, pp. 456–466 (1997 Jun.).

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

[37] H. Lee, "Phantom Image Elevation Explained," presented at the *141st Convention of the Audio Engineering Society* (2016 Sep.), convention paper 9664.

[38] B. Katz and Parseihian, "Perceptually Based Head-Related Transfer Function Database Optimization," *J. Acoust. Soc. Am. EL*, vol. 131, EL, pp. 99–105 (2012). https://doi.org/10.1121/1.3672641

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

[40] 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.). https://doi.org/10.17743/jaes.2014.0045

[41] K. Hamasaki and W. Van Baelen, "Natural Sound Recording of an Orchestra with Three-Dimensional Sound," presented at the *138th Convention of the Audio Engineering Society* (2015 May), convention paper 9348.

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

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

[44] S. Kraft and U. Zölzer, "Low-Complexity Stereo Signal Decomposition and Source Separation for Application in Stereo to 3D Upmixing," presented at the *140th Convention of the Audio Engineering Society* (2016 May), convention paper 9586.

[45] J. Herre, J. Hilpert, A. Kuntz and J. Plogsties, "MPEG-H Audio—The New Standard for Universal Spatial/3D Audio Coding," *J. Audio Eng. Soc.*, vol. 62, pp. 821–830 (2014 Dec.). https://doi.org/10.17743/jaes.2014.0049

[46] C. J. Chun, H. K. Kim, S. H. Choi, S. Jang and S. Lee, "Sound Source Elevation Using Spectral Notch Filtering and Directional Band Boosting in Stereo Loudspeaker Reproduction," *IEEE Trans. Consum. Electron.*, vol. 57, no. 4, pp. 1915–1920 (2011). https://doi.org/10.1109/TCE.2011.6131171

[47] S. Kim, M Ikeda, and W. Martens, "Reproducing Virtually Elevated Sound via a Conventional Home-Theater Audio System," *J. Audio Eng. Soc.*, vol. 62, pp. 337–344 (2014 May). https://doi.org/10.17743/jaes.2014.0019

[48] S. Chon and S. Kim, "Method and Apparatus for Reproducing Three-Dimensional Audio," U.S. Patent 20160330560 (Nov. 2016).

[49] R. K. Cook, R. V. Waterhouse, R. D. Berendt, S. Edelman, and M. C. Thompson, Jr., "Measurement of Correlation Coefficients in Reverberant Sound Ffields," *J. Acoust. Soc. Am.*, vol. 27, pp. 1071–1077 (1955). https://doi.org/10.1121/1.1908122

[50] M. Kuster, "Spatial Correlation and Coherence in Reverberant Acoustic Fields: Extension to Microphones with Arbitrary First-Order Directivity," *J. Acoust. Soc. Am.*, vol. 123, pp. 154–162 (2008). https://doi.org/10.1121/1.2812592

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.