Author Accepted Manuscript (originally presented at the 139th Audio Engineering Convention in New York, Oct 2015)

Investigation on the Phantom Image Elevation Effect

Hyunkook Lee

Applied Psychoacoustics Lab, University of Huddersfield, West Yorkshire, United Kingdom
h.lee@hud.ac.uk

ABSTRACT

Listening tests have been carried out in order to evaluate the phantom image elevation effect depending on horizontal stereophonic base angle. Seven ecologically valid sound sources as well as four noise sources were tested. Subjects judged the perceived image positions of phantom centre image created with seven loudspeaker base angles. Results generally showed that perceived images were elevated from front to above as the loudspeaker base angle increased up to around 180°. This tendency depended on the spectral characteristics of sound source. The perceived results are explained from both physical and cognitive points of view.

1. INTRODUCTION

In 1947 de Boer [1] found a phenomenon that a phantom image created from two identical sounds radiated by a pair of horizontally and symmetrically arranged loudspeakers tended to be perceived at an elevated position in the median plane. He reported that the elevation angle increased from front to above as the loudspeaker base angle increased from 0° to 180°. However, the sound source used in de Boer's experiment was not mentioned in his paper. Similar results to de Boer's were obtained by Damaske and Mellert [2]. In their experiment using a noise signal ranging from 0.65 to 4.5 kHz with loudspeaker base angle varied between 0° and 360°, it was found that the 180° angle gave rise to the elevation of phantom centre image at around 120°. Several recent studies [3, 4, 5] reported the existence of the elevation effect. In Jo et al. [3] using white noise ranging from 1 to 16 kHz, it was shown that a phantom centre image created by the rear loudspeaker pair in a conventional 5-channel loudspeaker setup were perceived to be elevated to some degrees, but this effect had a high degree of inter-subject variability. Frank [4] and Lee [5] also showed the existence of the elevation effect for broadband pink noise, but only narrow loudspeaker base angles were used in their studies (40° [4] and 60° [5]).

The present study aims to investigate the elevation of horizontally produced phantom image more extensively, testing its effectiveness on a wide range of natural sound sources with different characteristics as well as different types of noise sources. An elevation localisation experiment is performed for phantom centre images created by five stereophonic loudspeaker pairs with the base angle varied between 0° and 360° as well as for real centre loudspeakers in the front and back. It is expected that the experimental data provided in this paper will be useful for practical applications such as 3D to 2D downmix of film sound or object-based rendering of 3D audio where elevation perception is required without physical height or overhead loudspeakers.

2. EXPERIMENTAL METHOD

2.1. Physical Setup

The listening tests were conducted in an ITU-R BS.1116-2 [6]-compliant listening room (6.2m x 5.6m x 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 30° interval from the listening position. The distance between the listening position and each loudspeaker was 2m. The middle position between the woofer and tweeter of each loudspeakers was 1.28m high from the floor. The single loudspeakers in the front centre and back centre were used to produce "real" centre images, whereas each of the symmetrically arranged loudspeaker pairs (±30°, ±60°, ±90°, ±120° and ±150°) was used to create a "phantom" centre image. Therefore, there were a total of seven stereophonic loudspeaker base angles: 0°, 60°, 120°, 180°, 240°, 300° and 360°. The loudspeaker setup was hidden to subjects by placing acoustically transparent curtains around and above the listening position.

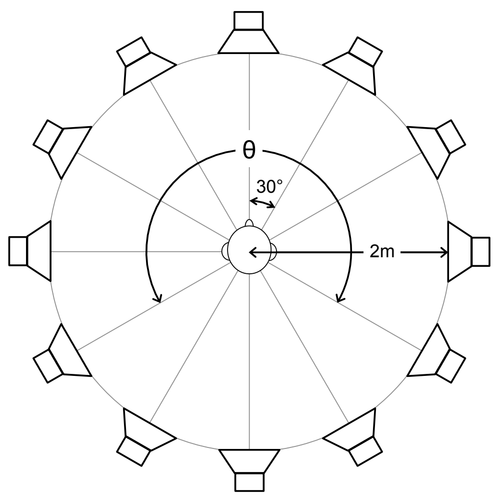

Fig. 1. Loudspeaker setup for the experiment: stereophonic base angle θ = 0° (front centre), 60°, 120°, 180°, 240°, 300° and 360° (back centre).

2.2. Sound Source

Seven natural and four noise sources were used for the experiment. Six of the natural sources comprised the recordings of airplane, helicopter, rain, thunder, bird and church bell, which were taken from the BBC Sound Effects Library. These sources were chosen not only for their different temporal and spectral characteristics, but also for their ecological validity. Since such sources would be heard from elevated positions in real life, experimental data obtained for them would be useful for practical 3D audio applications such as 3D to 2D downmix of film sound or object-based rendering of 3D audio where elevation perception is required without physical height or overhead loudspeakers. The other natural source was an anechoically recoded male speech, taken from the Bang and Olufsen's Archimedes CD [7]. This was considered to be useful to test the possibility of 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 one second of fade-in and fade-out applied, and 200ms-long broadband pink and white noise bursts repeated with the interval of 500ms. The onset and offset times for the burst were 5ms. All signals described above had the sampling frequency of 44.1kHz and the bit resolution of 16 bits.

2.3. Subject

10 male subjects with normal hearing participated in the listening tests. They were post-graduate students, final year undergraduate students and academic staff members from the University of Huddersfield's music technology courses. Each had previous experiences in localisation tests but was not trained particularly for the purpose of the current study. All subjects reported normal hearing.

2.4. 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 × 7 loudspeaker base angles). The playback level of each source signal was calibrated at the average A-weighted SPL (LAeq) of 75dB at the listening position using the front centre loudspeaker. This gave the playback level of 78dB LAeq for the stereophonic stimuli. Each trial contained a single stimulus, which was presented in loop, and had a side-view circle that was intersected into 12 regions at 30° interval (Fig. 2).

The subject's task was to mark one region in the circle where the sound image was perceived. This response method was inspired by Blauert [8], who used a similar region division in the median plane but only for the front, above and back regions at 90° interval. The current method adds the below region for covering the whole median plane and each of the front, above, back and below regions is divided into three sub-regions. Some median plane localisation studies [9, 10] used a free plotting method where subjects mark down the perceived image position on a specific point of a circle. However, from a pilot test it was found to be a challenging and time-consuming task to precisely localise perceived direction in the median plane, especially when the image appears above or behind. Furthermore, the 12 sub-regions at 30° interval were considered to be a high-enough resolution for the purpose of this study, which is to investigate the dependency of perceived elevation on loudspeaker base angle.

The GUI allowed the automatic randomisation of trial order, so each stimulus was presented in a random order for each subject. Prior to the main test, each subject was given a familiarisation trial where he could listen to all stimuli to be tested.

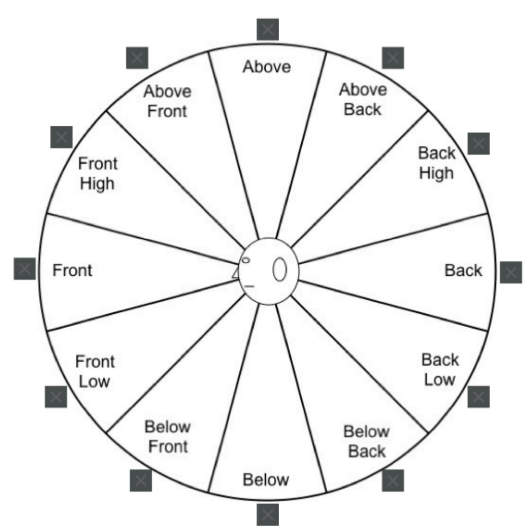

Fig. 2. GUI used for the listening test.

3. RESULTS AND DISCUSSION

3.1. Overall Tendency

Subject responses obtained from the listening tests are presented as bubble plots in Fig. 3. The diameter of each filled circle is proportional to the percentage of responses for each condition. It can be seen that the distribution of responses for each angle varied depending on the sound source. Nevertheless, there appears to be a general tendency for all sources that the elevation angle of the perceived image increased as the loudspeaker base angle increased from 0° to 240°. The images created by the loudspeaker pairs with the 300° and 360° base angles were generally perceived at the back regions. The perceived median elevation linearly increased from the 'front' to the 'above front' as the base angle increased from 0° to 120°. The 180° and 240° conditions had the 'above back' median response, while the 300° and 360° conditions had the 'back' median response. This median response pattern for all sources seems to be in line with those of early studies mentioned in Section 0. Damaske and Mellert [2] found that the mean perceived elevation angles for the loudspeaker base angles of 180° and 240° were around 120°. de Boer [1] also reported that phantom image created by loudspeakers placed directly at the listener's sides was perceived at the elevation angle of around 100° rather than exactly above.

3.2. Sound Source Dependency

The results for individual sources in Fig. 3 show some noticeable differences in terms of the distribution and weight of subject responses for each base angle. Looking at the noise stimuli results first, the continuous and transient characteristics do not seem to have produced dramatic differences in the general elevation pattern for either pink or white noise. Comparing between the pink and white noises, there is a tendency that the perceived positions of the white noise stimuli were slightly more elevated than those of the pink noise ones overall. For example, for the 60° the perceived region with the maximal response was the 'front' with the pink noises, whereas it was the 'front high' for the white noises. Furthermore, the maximal responses for the 180° and 240° conditions were given to the 'above' and 'above back' for the white noises, respectively, while those were the 'above back' and 'back high' for the pink noises. Additionally, the transient white noise had a perfectly linear relationship between median perceived position and loudspeaker base angle, i.e. each base angle is directly mapped to each of the seven perceived regions between the 'front' and 'back'. This result seems to be explained by the so-called 'pitch-height' effect, which suggests that the higher the frequency of a sound is the higher its perceived vertical position tends to be [11, 12, 13]. Since the pink noise has greater low frequency energy than the white noise, it is possible that the influence of low frequencies on the elevation judgment was stronger with the pink noise, thus the pink noise perceived less elevated than the white noise.

The frequency spectrum dependency of the elevation effect is further demonstrated by the results for the natural sources. The response pattern for the rain source appears to be highly similar to that of the white noise burst; both sources had the same median perception region for each base angle apart from the 0°. Furthermore, these sources had the most consistent 'above' responses for the 180° angle amongst all sources.

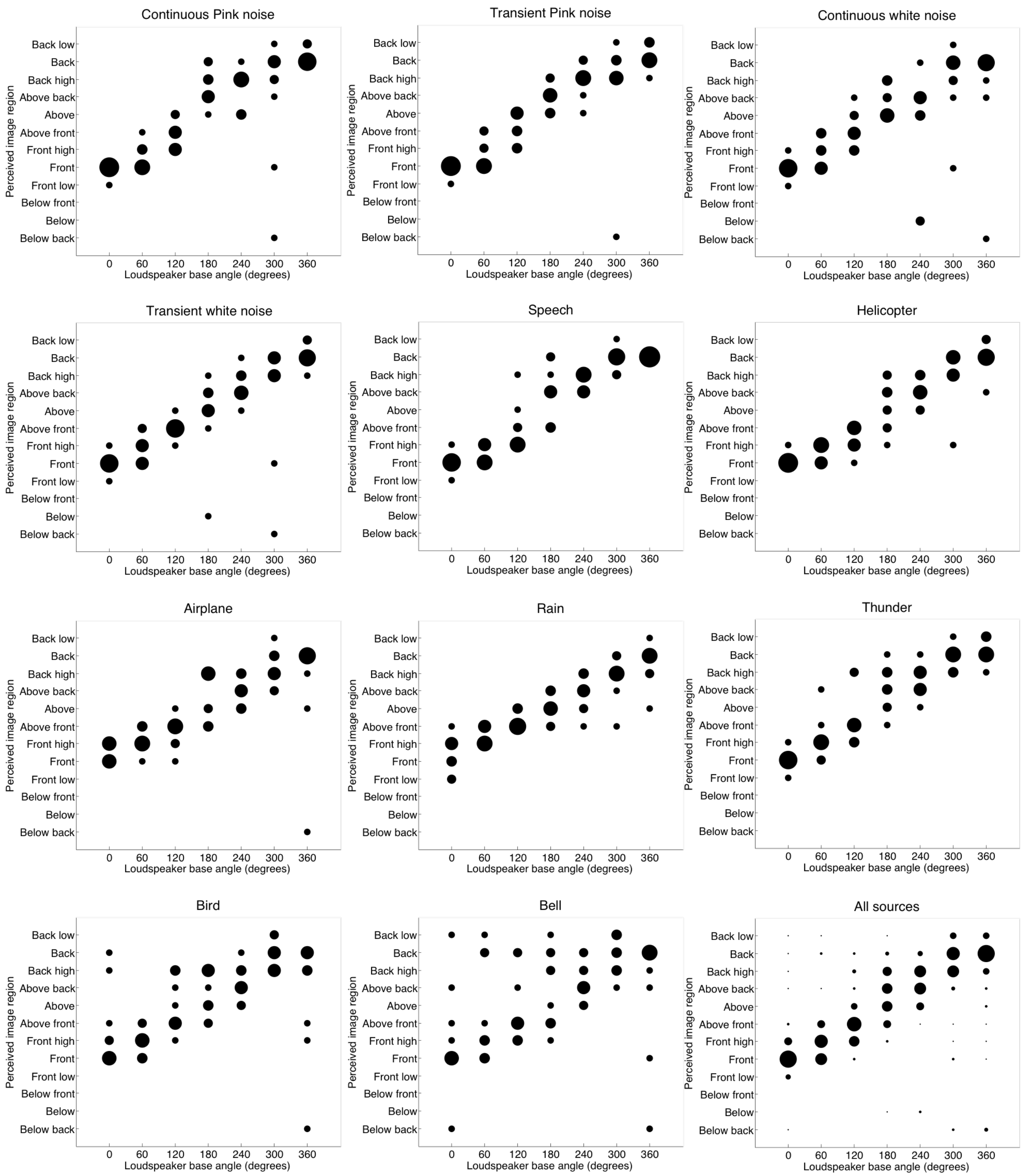

Fig. 3. Subject responses obtained for each source: the diameter of each filled circle represents the percentage of responses produced for each condition.

The rain source tested in the current study has a broad and relatively flat frequency spectrum which is similar to the white noise spectrum. The response patterns for the speech, thunder, helicopter and airplane, on the other hand, appear to be more similar to that for the continuous pink noise. There is no strong 'above' perception with the 180°, and the responses for the other angles tend to be more spread toward the front and back regions compared to the rain. In particular, the speech and the continuous pink noise have the same maximal response position for each base angle. The spectra of these sources have a dominance in a low or low-mid frequency range, and this might have affected the subject responses due to the pitch-height effect.

The bird and bell appear to have the most spread responses among all sources. Especially, a number of responses for the loudspeaker base angles of up to 120° indicate a 'front to back' confusion. The bird source has a sharp spectrum focusing at around 2 to 4 kHz with harmonics up to about 8 kHz. This seems to support the finding of Asano et al. [14] suggesting that low frequencies below 2 kHz are important for the front to back discrimination, although other research suggests the importance of high frequencies [9, 10]. On the other hand, the bell has strong harmonic contents between about 100 Hz and 2 kHz, with each having a tone-like nature; each partial produces a pitch whose amplitude decays slowly. The onset of the source is about 200ms, so there is no strong transient cue. It is widely known that steady-state sound is more difficult to localise than transient sound in rooms [15, 16], and this could be the reason for the largely inconsistent subject responses for the bell.

3.3. Cues for the Elevation Effect

3.3.1. Directional bands

Blauert [17] explains that the phantom image elevation effect is perceived due to the peaks and dips of the spectrum of ear input signal in conjunction with his directional bands theory [8]. He found from an experiment using 1/3-octave band noise signals that frontal localisation was associated with both 500 Hz and 4 kHz bands, back with 1 kHz band and above with 8 kHz band, regardless of the loudspeaker position. According to Blauert [17], the angle of perceived phantom image elevation is determined by the energy distribution in the directional bands. However, this has not been demonstrated experimentally in his report.

In order to examine if this could be the physical explanation for the current results, the spectrum of left ear input signal for each reproduction condition was measured using the MIT's KEMAR head-related impulse response (HRIR) database [18]. In Fig. 4 the measurement results are presented as delta spectra to a reference, which is the ear input spectrum for the front centre loudspeaker, in order to show the spectrum change over loudspeaker base angle more effectively. The plots show relative spectral energy dominance for each condition in a bi-polar way: loudspeaker pair with each base angle (+) vs. front centre loudspeaker (–).

It can be seen that all stereophonic reproduction conditions produce a greater energy at around 8 kHz in the ear spectrum compared to the mono condition. This seems to be an initial explanation as to why phantom image tends to be elevated more than real image. From further observations of the plots, the energy distribution and weighting of the directional bands can be mapped to the pattern of subject responses discussed earlier. The pattern of delta spectrum variation over loudspeaker base angle appears to closely matches the subjective results. The magnitude of the 8 kHz peak increases at about 5dB interval as the loudspeaker base angle increase from 60° to 180°. The difference between 180° and 240° in the 8 kHz peak level is small, although the latter is 1.5dB higher than the former. The peak level is reduced to 14dB at the 300° base angle, which is similar to the peak level at 120°.

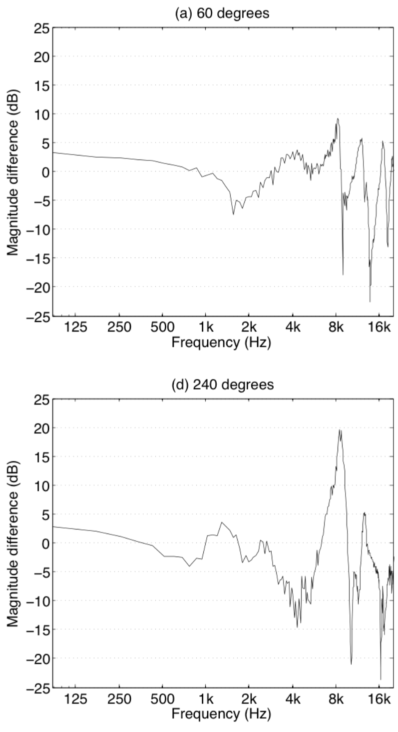

Fig. 4. Spectral magnitude difference of the left ear-input signal of each loudspeaker configuration to that of the front centre loudspeaker.

However, the 8 kHz alone does not explain the front and back biases in perceived elevation. The perceptions of intermediate regions could be explained by the energy-weighted distribution of all directional bands 500 Hz, 1 kHz and 4 kHz and 8 kHz to some extent. For example, the 120° delta spectrum has less energies between 500 Hz and 1 kHz and more energies between 1 kHz and 3 kHz and at around 8 kHz compared to the 60° delta spectrum. This might reduce the 'front-ness' in perceived image while increasing the 'back-ness' and 'above-ness', thus leading to the perception of 'front high' or 'above-front'. The 180° delta spectrum has a dramatically reduced energy at around 4 kHz, which suggests the 'front-ness' of the image might be further reduced, while having a large peak at 8 kHz producing a strong sense of 'above-ness'. It also has a large energy peak around 1.5 kHz compared to the spectra of other angles. This is perhaps the reason why there were a number of responses biased toward back regions for the 180° base angle. If the source signal had a lack of high frequency content or was low frequency dominant, this back-bias might become stronger. In fact, the sources that demonstrated the most spread responses toward the back for the 180° tend to have low frequency dominant characteristics (e.g. speech, thunder and airplane). The 240° delta spectrum appears to be similar to the 180° one, but the bandwidth of the 4kHz reduction is larger. This might suggest that the 'front-ness' of the sound presented by the 240° loudspeaker pair is even weaker than that by the 180°, which might explain why the median response for the 240° was the 'above back'. For the 300°, there appears to be no peak or dip at the directional bands specified by Blauert, other than the 8 kHz peak – this might be associated with the result showing that the noise and helicopter sources had a few front responses for the 300° condition. Lastly, the ear signal spectrum for the back mono loudspeaker (base angle of 360°) appears to have more energies around the back directional band (1 kHz) and less energies around the front directional band (4 kHz) than that for the front mono loudspeaker. From this it might be hypothesised that the hearing system interprets the directions of frequencies around 1 kHz and 4 kHz as back and front because the HRTFs for real front and back sources are most different in those frequency regions.

3.3.2. New hypothesis from a cognitive viewpoint

The above explanation based on the spectral energy distribution seems to be most valid for broadband signals with a flat spectrum or signals containing high frequencies including 8kHz. A number of sound sources used in the current study, however, had high frequency roll-off characteristics (e.g., thunder, airplane) or a narrow bandwidth (e.g., bird), but still their perceived images were elevated. Furthermore, from the author's pilot tests with three subjects using octave-band pink noise signals, low-mid frequency bands such as the 250Hz and 500Hz bands were perceived around the 'above front' or 'above' region for the 180° loudspeaker base angle. This cannot be explained simply by the spectral energy distribution model alone.

A novel hypothesis for the phantom image elevation effect at low frequencies is formulated as follows. If a symmetrically arranged stereophonic loudspeaker pair radiates identical signals at the same time, no interaural differences are produced between the ear-input signals. This is also the case with a real centre loudspeaker. A plausible cue for the brain to tell whether the perceived image is real or phantom might then be the acoustic crosstalk delay occurring between the primary (ipsilateral) and the secondary (contralateral) ear-input signals. This is effectively the same as an interaural time difference (ITD) that would be caused by one of the loudspeakers. Meanwhile, if the real centre loudspeaker is elevated in the median plane, each ear will receive a shoulder reflection after the direct sound. The shoulder reflection delay will increase as the source is elevated up to 90° (overhead) in the median plane, as shown in [9]. Similarly, the acoustic crosstalk delay of the horizontal loudspeaker signals will increase as the loudspeaker base angle increases up to 180°. Algazi et al. [9] suggests that the shoulder reflection is the main cue for elevation perception for frequencies up to 3 kHz. From this it is hypothesised that the brain interprets the acoustic crosstalk delay as a shoulder reflection delay, and that the resulting phantom centre image is elevated at an angle where a real elevated source would produce a shoulder reflection delay corresponding to the acoustic crosstalk delay specific to the given loudspeaker base angle. It is further considered that this time delay explanation would be valid for frequencies below 1 kHz due to phase ambiguity problem at higher frequencies, and that the HRTF-based explanation would be most relevant for frequencies above 1 kHz. This hypothesis will be investigated in a future study.

4. CONCLUSION AND FUTURE WORKS

This study investigated the elevation of phantom centre image created by horizontal loudspeakers. The aim was to examine the dependency of the effect on sound source and loudspeaker base angle. 11 sound sources comprising seven natural sources and four noise sources were tested. 12 loudspeakers were arranged in a circle at 30° interval. Each source was presented in 7 different loudspeaker base angle conditions, comprising 0° (front centre mono), 60°, 120°, 180°, 240°, 300°, 360° (back centre mono). Subject responses were produced from 12 region scale shown on a side-view median plane circle.

The test results confirmed the general relationship between phantom image elevation and loudspeaker base angle found in early studies [1, 2]; as the base angle increased from 0° to 180°, the perceived image was elevated from front to above. This tendency was found to have an obvious sound source dependency. In general, sources with a broad and flat spectrum (e.g., white noise, rain) had the most linear mapping between base angle and elevation. Sources with a low frequency weight or high frequency roll-off (pink noise, speech, thunder, airplane and helicopter) tended to be less elevated than the white noise and rain for the base angles of 120°, 180° and 240°. Particularly with the 180° angle, such sources had a substantially smaller number of responses for the direct 'above' region than the white noise and rain. The bird, which had a narrow spectrum around 2 – 4 kHz, also produced spread responses for the 120°, 180° and 240°, with the presence of ‘front/back’ confusion for the front and back real centre sources. The bell source, which had tone-like characteristics, produced most inconsistent elevation responses overall. From the above results, it could be concluded that the broadness and flatness of source spectrum is an important factor for the perception of phantom image elevation effect. It was also attempted to provide a potential theoretical explanation for the perceived results. The analyses of the spectra of ear input signals for all reproduction conditions revealed that the distribution and magnitude of the peaks and dips of certain high frequency components (i.e., Blauert’s directional bands [8]) in ear signal spectrum could be mapped to the perceived elevation results to some extent. A complementary hypothesis was also formulated to explain the elevation of low frequency dominated phantom sources – the brain cognitively interprets acoustic crosstalk in stereophonic loudspeaker reproduction as should reflection, which is known as the main elevation cue for low frequencies.

The current study led the author to several further research topics. Firstly, as mentioned earlier, it was identified from a pilot test that the phantom image elevation effect could be observed for individual octave bands of pink noise as well as broadband. This will be formally investigated through a listening test similar to the current test. The perceived height of each band for a given elevation region will be also measured. Secondly, the new hypothesis proposed in Section 2.3.2 will be tested by measuring shoulder reflection delay using Algazi et al.’s model [9]. Lastly, the current results are expected to have useful implications for applications such as virtual 3D audio recording/mixing and 3D/2D upmixing and downmixing. New production techniques exploiting the elevation effect will be explored for such applications.

5. ACKNOWLEDGEMENTS

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

6. REFERENCES

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

[2] 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).

[3] H. Jo, W. Martens, Y. Park and S. Kim, “Confirming the Perception of Virtual Source Elevation Effects Created using 5.1 Channel Surround Sound Playback,” *Proc. ACM SIGGRAPH 9th Int. Conf. on Virtual reality Continuum and Its Applications in Industry (VRCAI), Seoul, Korea.* (2010).

[4] M. Frank, “Elevation of Horizontal Phantom Sources”, *Proc. DAGA 2014, Oldenburg (2014 Mar).*

[5] H. Lee, “Perceptual Band Allocation (PBA) for the Rendering of Vertical Image Spread with a Vertical 2D Loudspeaker Array,” *Presented at the 138th International Convention of the Audio Engineering Society* (2015 May), convention paper 9325.

[6] ITU-R,“RecommendationsITU-RBS.1116-2:Methods for the Subjective Assessment of Small Impairments in Audio Systems including Multichannel Sound Systems,” International Telecommunications Union (2014).

[7] 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.).

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

[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] M. Morimoto, M. Yairi, K. Iida, M. Itoh, “The role of low frequency components in median plane localization,” *Acoust. Sci. & Tech.*, vol. 24 (2), pp. 76–82 (2003).

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

[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).

[13] D. Cabrera and S. Tilley, “Vertical Localization and Image Size Effects in Loudspeaker Reproduction,” in *Proc. AES 24th Int. Conf. on Multichannel Audio* (2003).

[14] 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).

[15] B. Rakerd and W. Hartmann, “Localisation of sound in rooms, III: Onset and duration effects,” *J. Acoust. Soc. Am*., vol. 80, pp.1695-1706 (1986).

[16] W. Yost, F. Wightman and M, Green, “Lateralization of filtered clicks”, *J. Acoust. Soc. Am*., vol. 50, pp. 1526-1531 (1971).

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

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