Lee, Hyunkook

Evaluation of the Phantom Image Effect for Phantom Images

Original Citation

Lee, Hyunkook (2015) Evaluation of the Phantom Image Effect for Phantom Images. In: 3rd International Conference on Spatial Audio (ICSA), 18-20 Sep 2015, Graz, Austria.

This version is available at http://eprints.hud.ac.uk/25887/

Evaluation of the elevation effect for phantom images

Hyunkook Lee

Applied Psychoacoustics Lab, University of Huddersfield, U.K., Email: h.lee@hud.ac.uk

Abstract

This paper introduces the author's recent research on the elevation effect perceived with horizontal phantom images. Early research in stereophony suggests that a phantom centre image produced by two loudspeakers symmetrically placed from the listener position would be perceived in an elevated position, with its elevation angle increases as the loudspeaker base angle increases. In particular, an image presented from loudspeakers placed around the listener's sides would be perceived overhead. With 3D audio formats employing height and overhead channels in mind, the aforementioned elevation effect is considered to be useful for creating a virtual overhead loudspeaker image, especially for sound effects using just ear-level loudspeakers (e.g. in downmix scenarios). Another important psychoacoustic principle relevant to 3D audio formats is the so-called 'pitch-height' effect, which suggests that the higher the frequency of a sound is the higher its image will be perceived. However, past research in this topic only considered loudspeakers placed in the median plane. From the above background, several subjective experiments have been conducted on the elevation of horizontally oriented phantom image. This paper first presents a vertical localisation test conducted with frontal stereo loudspeakers using octave-band noise stimuli. The results not only confirm the elevation effect for broadband noise, but also show the existence of an elevation effect for middle frequency bands. The second experiment introduced in this paper verifies the existence of the virtual overhead perception depending on loudspeaker base angle but also shows the effect heavily depends on the type of sound source.

1 INTRODUCTION

An important principle for vertical sound localisation is the so-called 'pitch-height' effect, which suggests that the higher the frequency of a tone is the higher the perceived position of the sound is. This effect was first found by Pratt [1] and confirmed by many researchers [2, 3, 4, 5]. In [3, 4], it was shown that the pitch-height effect was valid for band-limited noise signals such as low/high-passed noise and octave-band filtered noise. It was also found in those studies that the perceived positions of low and high frequency images depended on the physical height of the presenting loudspeaker. That is, high frequency images were localised at positions near their presenting loudspeakers, whereas low frequency ones were perceived near the listener's ear height regardless of the loudspeaker's physical height.

In contrast to the above mentioned studies focusing on localisation in the vertical plane in front of the listener, Blauert [6] investigated the relationship between frequency and perceived image position in the median plane. He reported that the 1/3-octave frequency bands centred at 500Hz and 4kHz were localised in front of the listener, 8kHz above and 1kHz at the back, regardless of the physical position of loudspeaker placed in the horizontal plane. These bands are called the 'directional bands'.

These pitch-height studies mentioned above used loudspeakers vertically arranged in the median plane, with each presenting a signal with a different frequency content independently or simultaneously. However, early research in stereophonic sound reproduction suggests that auditory image elevation can be perceived with two loudspeakers placed in the horizontal plane. de Boer [7] was the first who reported that a phantom image produced in stereophonic reproduction would be perceived at an elevated position in the median plane. He found that the perceived elevation angle increased from front to above as the loudspeaker base angle increased from 0° to 180°. Damaske and Mellert [8] obtained similar results in their experiment with the loudspeaker base angle varied between 0° and 360°, which showed that the 180° angle produced an elevated image at around 120°. After almost

40 years of gap since Damaske and Mellert's study, the elevation effect was reported again in several recent studies [9, 10]. Jo et al. [9] showed that a phantom centre image created by the rear loudspeaker pair in a conventional 5-channel loudspeaker setup was elevated to some degrees. Frank [10] found that the elevation angle of the perceived phantom centre image increased as the loudspeaker base angle increased from 0° to 40°.

Except de Boer's paper that did not report the type of sound source used, all of the above mentioned studies used noise sources with different frequency ranges. Damaske and Mellert used a white noise signal ranging from 0.65 to 4.5 kHz, Jo et al. a white noise from 1 to 16 kHz, Frank a broadband pink noise. No natural or practical sound sources have been tested for the elevation effect.

This paper summarises two experiments that the current author recently conducted in order to expand the knowledge about the perception of phantom image elevation. The first one measured the pitch-height effect for the phantom images of octave-band pink noise signals created with each of loudspeaker pairs elevated at 0° and 30° in front. The second experiment investigated into the effect of horizontal loudspeaker base angle on the elevation of perceived phantom image, using a wide range of signals including practical sound sources as well as noises.

2 PITCH-HEIGHT EFFECT FOR PHANTOM IMAGE

2.1 Experiment

An experiment was conducted in a dry listening room (8.3m × 5.4m × 3.4m; RT = 0.2sec) to measure the perceived heights of the horizontal phantom images of octave-band pink noise signals [11]. Four Genelec 8040A loudspeakers were arranged in a frontal two-dimensional (2D) format. Each of the lower and upper loudspeaker pair was to present two coherent signals with no level and time differences applied independently, thus producing a horizontal phantom centre image. The lower loudspeaker pair, which was placed 1.2m high from the floor, had a base angle of 60°. The upper loudspeaker pair was placed directly above the lower pair and was elevated at 30° from the listener's ear position. The distance between the listener and each loudspeaker in the lower layer was 1.78m. The main and height loudspeakers were aligned in terms of arrival time and sound pressure level at the listening position. An acoustically transparent curtain was used to visually hide the loudspeaker setup to the listeners. Vertically oriented number labels ranging from 0 to 300 with the interval of 10, representing the height from the floor in cm, were placed on the curtain.

The sound source used for the experiment was a broadband continuous pink noise. The broadband signal was filtered into nine octave-bands (-48dB/octave) with the centre frequencies ranging from 63Hz to 16kHz. Test stimuli were created in two-channel horizontal stereo for each octave-band as well as for the broadband. The output level of the playback system was calibrated to 70dB LAeq at the listening position using the broadband noise stimulus, and the same output level was maintained for all octave-bands.

12 critical listeners with normal hearing and much experience in spatial audio listening tests participated in this experiment. They were staff researchers, post-graduate students and final year students from the University of Huddersfield's music technology courses.

A graphical user interface (GUI) written using Max was used for the listening tests. There were a total of 20 trials for testing the nine octave-bands and the single broadband stimuli, which were presented from either the main or height loudspeaker layer in a randomised order. In each trial, the subjects were given a continuous grading slider, which had the same scale range as the physical number labels explained above. Their task was to move the slider to a position that corresponded to the perceived image height. The subjects sat on a chair with a small head-rest and their ear height and distance between the ear and one of the main layer loudspeakers were aligned to 1.2m and 1.78m, respectively. They were instructed not to move their head up and down while judging the image height and asked to use their eye movements only.

2.2 Results and Discussion

Figure 1 shows the results obtained from the listening tests. The boxes represents the median values and the associated inter-quartile ranges (IQRs). The white and grey boxes represent the main and height loudspeaker presentations, respectively. The reference lines at 120cm and 214cm indicate the vertical positions of the main and height loudspeakers, respectively.

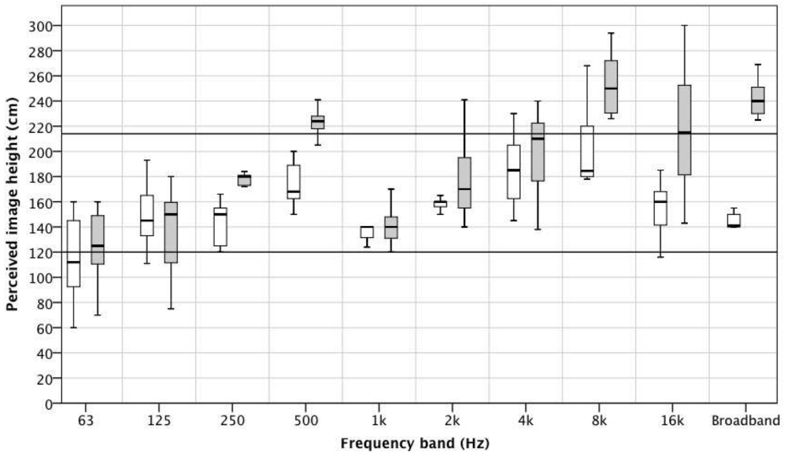


Figure 1: Box plots for the data obtained from the phantom image pitch-height experiment of Lee [11]; the white and grey boxes are for the lower (0° elevation) and upper (30° elevation) loudspeaker presentations, respectively.

The results generally show that the pitch-height effect operated in two separate octave-band regions: 63 to 1kHz, and 1kHz to 8kHz. It can be also seen that the median perceived heights of all bands were higher with the upper loudspeaker layer than with the lower layer. The differences with a statistical significance were found for the 250Hz, 500Hz, 8kHz and 16kHz bands as well as the broadband.

Comparing the current results to those from previous research using real images in the median plane, several unique properties of the localisation of phantom images can be suggested. Firstly, in [3, 4] it was reported that low frequency noise signals were localised at or below the listener's ear height regardless of the physical height of the loudspeaker. However, as can be seen in Figure 2, the phantom images for the 125Hz, 250Hz and 500Hz were perceived to be higher than the loudspeaker height, and this elevation effect was greater with the upper layer than with the lower layer.

Secondly, previous research using median plane loudspeakers generally agreed that broadband noise including frequencies above around 6kHz were accurately localised at the position of the presenting loudspeaker. However, this was not the case with the phantom images for the broadband noise in the current study; the perceived images were perceived to be slightly higher than the loudspeaker positions for both lower and upper loudspeaker layers.

3 ELEVATION PERCEPTION DEPENDING ON LOUDSPEKAER BASE NAGLE AND SOUND SOURCE TYPE

3.1 Experiment

The aim of this experiment was to investigate the loudspeaker base angle and sound source dependency of the phantom image elevation effect. Listening tests were conducted in the University of Huddersfield's ITU-R BS.1116-compliant listening room (6.2m × 5.6m × 3.8m; RT = 0.25s). 12 Genelec 8040A loudspeakers were arranged horizontally at 30° interval and 2m distance from the listening position. The loudspeakers in the front centre (0°) and back centre (180°) 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, the loudspeaker base angle varied from 0° to 360°. The loudspeaker setup was hidden to subjects by placing acoustically transparent curtains around and above the listening position.

Seven natural and four noise sources were used for the experiment. The natural sources comprised the recordings of airplane, helicopter, rain, thunder, bird and church bell, which were taken from the BBC Sound Effects Library, and an anechoic recording of male speech, taken from the Bang and Olufsen's Archimedes CD. The noise sources 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 with the fade-in and fade-out times of 5ms.

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.

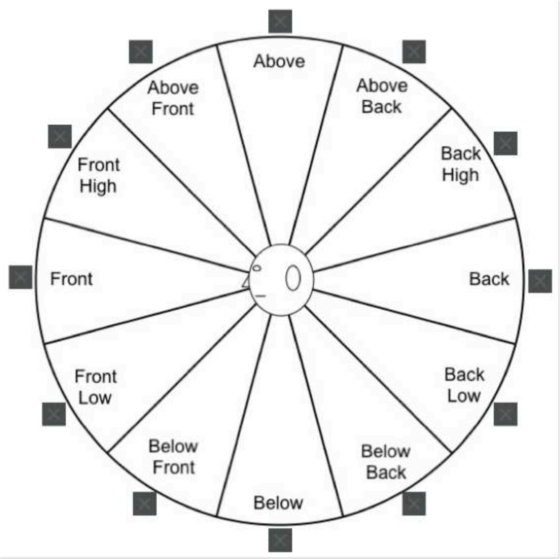


Figure 2: GUI used for the elevation listening test.

A custom-made GUI written using Max was used for the listening tests. The playback level of each source signal was 78dB LAeq at the listening position. Each trial contained a single stimulus and showed a side-view circle that was intersected into 12 regions at 30° interval (Figure 2). The subjects were asked to mark one region in the circle according to the perceived image position. Each stimulus was presented in a random order for each subject.

3.2 Results and Discussion

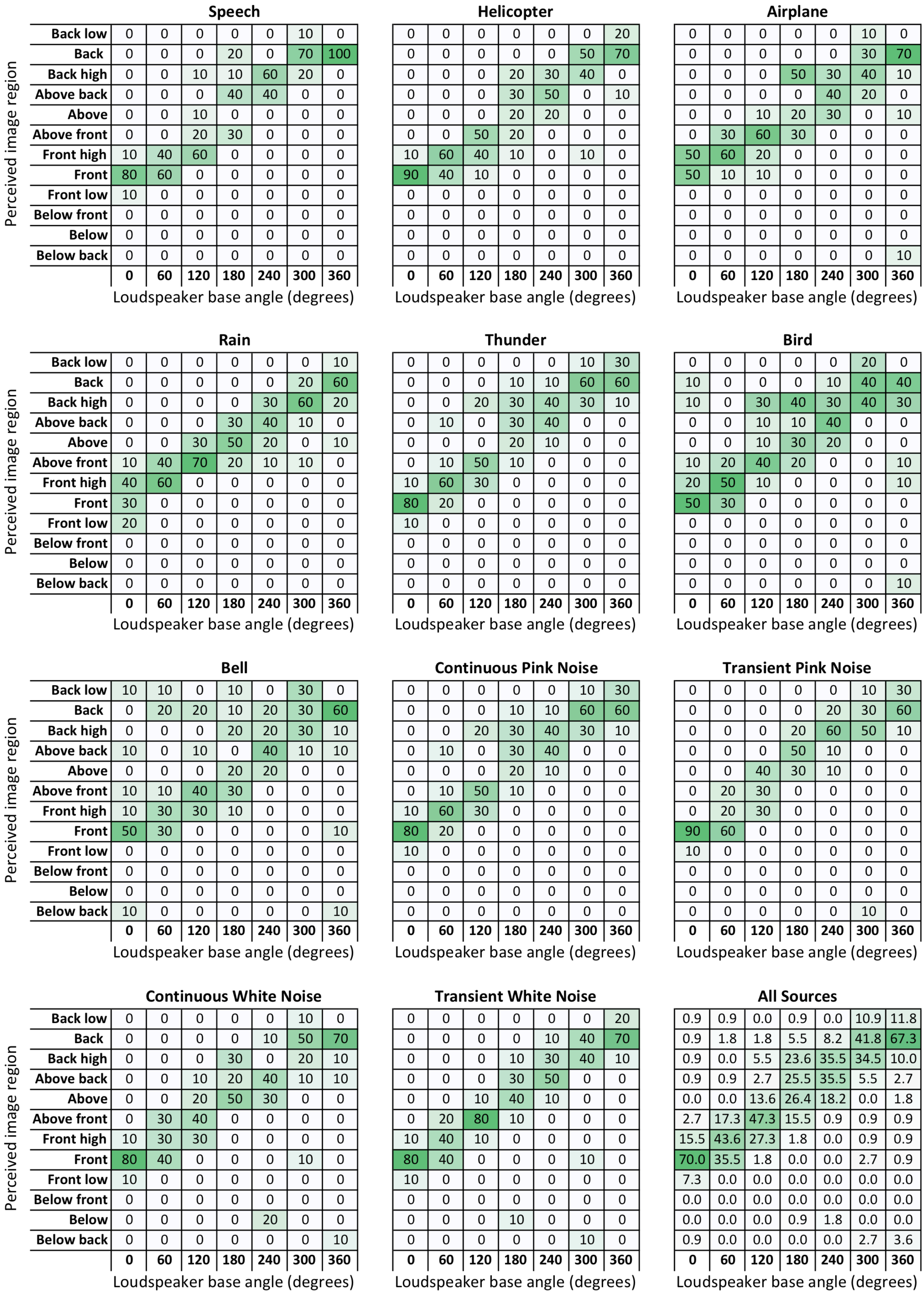
Subject responses from the listening tests are presented as percentage value for each test condition in Figure 3. The colour gradation of each cell is proportional to the percentage of responses.

All sources tend to show a general tendency that the elevation angle of the perceived image increased as the loudspeaker base angle increased from 0° to 240°. The images for the 300° and 360° base angles were generally perceived in 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 pattern seems to be in line with those of de Boer [7] and Damaske and Mellert [8].

The results demonstrate a sound source dependency in the distribution of the responses. For the noise sources, it appears that the white noises were perceived slightly more elevated than the pink noises. Especially, the white noise burst had a more linear relationship between its median response and loudspeaker base angle than the pink noise burst. Furthermore, the white noise burst had the maximal responses for the 180° and 240° base angle in the 'above' and 'above back', respectively, whereas 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. This result seems to be associated with the 'pitch-height' effect mentioned in Section 1. The pink noise has greater low frequency energy than the white noise, and this might have influenced the perceived image to be lower in position with the pink noise than with the white noise.

For the natural sources, the rain source has a similar response pattern to that of the white noise burst. The rain recording used in the current test 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, are more similar to that for the continuous pink noise. In particular, the speech and the continuous pink noise share the same maximal response position for each base angle. The spectra of these sources are also similar in that they have a low frequency dominance.

The bird and bell appear to have the most inconsistent responses, with a tendency of the front and back confusion for the base angles up to 120°. The spectrum of the bird recording is focused around 2 to 4kHz with harmonics up to about 8kHz. This result seems to be explained by the finding of Asano et al. [12] showing that low frequencies below 2kHz are important for resolving the front and back confusion. The bell, on the other hand, has strong harmonic contents between 100Hz and 2kHz. Each of these has a steady state nature similar to a pure tone. The onset of the signal is also slow (e.g. about 200ms). Steady-state sound is known to be more difficult to localise than transient sound in rooms [13, 14].



Speech

Perceived image region

	0	60	120	180	240	300	360
Back low	0	0	0	0	0	10	0
Back	0	0	0	20	0	70	100
Back high	0	0	10	10	60	20	0
Above back	0	0	0	40	40	0	0
Above	0	0	10	0	0	0	0
Above front	0	0	20	30	0	0	0
Front high	10	40	60	0	0	0	0
Front	80	60	0	0	0	0	0
Front low	10	0	0	0	0	0	0
Below front	0	0	0	0	0	0	0
Below	0	0	0	0	0	0	0
Below back	0	0	0	0	0	0	0

Loudspeaker base angle (degrees)

Helicopter

	0	60	120	180	240	300	360
Back low	0	0	0	0	0	0	20
Back	0	0	0	0	0	50	70
Back high	0	0	0	20	30	40	0
Above back	0	0	0	30	50	0	10
Above	0	0	0	20	20	0	0
Above front	0	0	50	20	0	0	0
Front high	10	60	40	10	0	10	0
Front	90	40	10	0	0	0	0
Front low	0	0	0	0	0	0	0
Below front	0	0	0	0	0	0	0
Below	0	0	0	0	0	0	0
Below back	0	0	0	0	0	0	0

Loudspeaker base angle (degrees)

Airplane

	0	60	120	180	240	300	360
Back low	0	0	0	0	0	10	0
Back	0	0	0	0	0	30	70
Back high	0	0	0	50	30	40	10
Above back	0	0	0	0	40	20	0
Above	0	0	10	20	30	0	10
Above front	0	30	60	30	0	0	0
Front high	50	60	20	0	0	0	0
Front	50	10	10	0	0	0	0
Front low	0	0	0	0	0	0	0
Below front	0	0	0	0	0	0	0
Below	0	0	0	0	0	0	0
Below back	0	0	0	0	0	0	10

Loudspeaker base angle (degrees)

Rain

Perceived image region

	0	60	120	180	240	300	360
Back low	0	0	0	0	0	0	10
Back	0	0	0	0	0	20	60
Back high	0	0	0	0	30	60	20
Above back	0	0	0	30	40	10	0
Above	0	0	30	50	20	0	10
Above front	10	40	70	20	10	10	0
Front high	40	60	0	0	0	0	0
Front	30	0	0	0	0	0	0
Front low	20	0	0	0	0	0	0
Below front	0	0	0	0	0	0	0
Below	0	0	0	0	0	0	0
Below back	0	0	0	0	0	0	0

Loudspeaker base angle (degrees)

Thunder

	0	60	120	180	240	300	360
Back low	0	0	0	0	0	10	30
Back	0	0	0	10	10	60	60
Back high	0	0	20	30	40	30	10
Above back	0	10	0	30	40	0	0
Above	0	0	0	20	10	0	0
Above front	0	10	50	10	0	0	0
Front high	10	60	30	0	0	0	0
Front	80	20	0	0	0	0	0
Front low	10	0	0	0	0	0	0
Below front	0	0	0	0	0	0	0
Below	0	0	0	0	0	0	0
Below back	0	0	0	0	0	0	0

Loudspeaker base angle (degrees)

Bird

	0	60	120	180	240	300	360
Back low	0	0	0	0	0	20	0
Back	10	0	0	0	10	40	40
Back high	10	0	30	40	30	40	30
Above back	0	0	10	10	40	0	0
Above	0	0	10	30	20	0	0
Above front	10	20	40	20	0	0	10
Front high	20	50	10	0	0	0	10
Front	50	30	0	0	0	0	0
Front low	0	0	0	0	0	0	0
Below front	0	0	0	0	0	0	0
Below	0	0	0	0	0	0	0
Below back	0	0	0	0	0	0	10

Loudspeaker base angle (degrees)

Bell

Perceived image region

	0	60	120	180	240	300	360
Back low	10	10	0	10	0	30	0
Back	0	20	20	10	20	30	60
Back high	0	0	0	20	20	30	10
Above back	10	0	10	0	40	10	10
Above	0	0	0	20	20	0	0
Above front	10	10	40	30	0	0	0
Front high	10	30	30	10	0	0	0
Front	50	30	0	0	0	0	10
Front low	0	0	0	0	0	0	0
Below front	0	0	0	0	0	0	0
Below	0	0	0	0	0	0	0
Below back	10	0	0	0	0	0	10

Loudspeaker base angle (degrees)

Continuous Pink Noise

	0	60	120	180	240	300	360
Back low	0	0	0	0	0	10	30
Back	0	0	0	10	10	60	60
Back high	0	0	20	30	40	30	10
Above back	0	10	0	30	40	0	0
Above	0	0	0	20	10	0	0
Above front	0	10	50	10	0	0	0
Front high	10	60	30	0	0	0	0
Front	80	20	0	0	0	0	0
Front low	10	0	0	0	0	0	0
Below front	0	0	0	0	0	0	0
Below	0	0	0	0	0	0	0
Below back	0	0	0	0	0	0	0

Loudspeaker base angle (degrees)

Transient Pink Noise

	0	60	120	180	240	300	360
Back low	0	0	0	0	0	10	30
Back	0	0	0	0	20	30	60
Back high	0	0	0	20	60	50	10
Above back	0	0	0	50	10	0	0
Above	0	0	40	30	10	0	0
Above front	0	20	30	0	0	0	0
Front high	0	20	30	0	0	0	0
Front	90	60	0	0	0	0	0
Front low	10	0	0	0	0	0	0
Below front	0	0	0	0	0	0	0
Below	0	0	0	0	0	0	0
Below back	0	0	0	0	0	10	0

Loudspeaker base angle (degrees)

Continuous White Noise

Perceived image region

	0	60	120	180	240	300	360
Back low	0	0	0	0	0	10	0
Back	0	0	0	0	10	50	70
Back high	0	0	0	30	0	20	10
Above back	0	0	10	20	40	10	10
Above	0	0	20	50	30	0	0
Above front	0	30	40	0	0	0	0
Front high	10	30	30	0	0	0	0
Front	80	40	0	0	0	10	0
Front low	10	0	0	0	0	0	0
Below front	0	0	0	0	0	0	0
Below	0	0	0	0	20	0	0
Below back	0	0	0	0	0	0	10

Loudspeaker base angle (degrees)

Transient White Noise

	0	60	120	180	240	300	360
Back low	0	0	0	0	0	0	20
Back	0	0	0	0	10	40	70
Back high	0	0	0	10	30	40	10
Above back	0	0	0	30	50	0	0
Above	0	0	10	40	10	0	0
Above front	0	20	80	10	0	0	0
Front high	10	40	10	0	0	0	0
Front	80	40	0	0	0	10	0
Front low	10	0	0	0	0	0	0
Below front	0	0	0	0	0	0	0
Below	0	0	0	10	0	0	0
Below back	0	0	0	0	0	10	0

Loudspeaker base angle (degrees)

All Sources

	0	60	120	180	240	300	360
Back low	0.9	0.9	0.0	0.9	0.0	10.9	11.8
Back	0.9	1.8	1.8	5.5	8.2	41.8	67.3
Back high	0.9	0.0	5.5	23.6	35.5	34.5	10.0
Above back	0.9	0.9	2.7	25.5	35.5	5.5	2.7
Above	0.0	0.0	13.6	26.4	18.2	0.0	1.8
Above front	2.7	17.3	47.3	15.5	0.9	0.9	0.9
Front high	15.5	43.6	27.3	1.8	0.0	0.9	0.9
Front	70.0	35.5	1.8	0.0	0.0	2.7	0.9
Front low	7.3	0.0	0.0	0.0	0.0	0.0	0.0
Below front	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Below	0.0	0.0	0.0	0.9	1.8	0.0	0.0
Below back	0.9	0.0	0.0	0.0	0.0	2.7	3.6

Loudspeaker base angle (degrees)

Figure 3: Results from the phantom image elevation experiment: the percentages of the responses for each test condition.

4 THEORETICAL EXPLANATIONS FOR PHANTOM IMAGE ELEVATION

Blauert [15] asserts that that the degree of elevation for horizontal phantom image is determined by the spectral energy distribution (i.e., the magnitudes of peaks and dips) of ear input signal. As mentioned in Section 1, his directional bands theory [6] suggests that frontal localisation is associated with 500Hz and 4kHz bands, back with 1kHz band and above with 8kHz band. According to this theory, it can be considered that a greater magnitude of the 8kHz energy in the ear-input signal compared to the other frequencies would increase the "aboveness" of the perceived image. Similarly, more 4kHz energy would mean a stronger "frontness" of the image.

However, the above explanation seems to be only valid for broadband signals containing high frequencies. A number of sound sources tested in the experiment described in Section 3 had high frequency roll-off characteristics (e.g., thunder, airplane) or a narrow bandwidth (e.g., bird), but still produced the elevation effect. Furthermore, from the experiment summarised in Section 2, it was found that low and low-mid frequency bands such as the 250Hz and 500Hz bands were elevated. Since the head-related transfer function (HRTF) is most relevant for high frequencies, the elevation effect for low-frequency sources cannot be explained simply by the directional band theory alone.

In order to explain the elevation of phantom image with no or little high frequency energy, a novel hypothesis is presented from a cognitive perspective as follows. If a centrally placed real loudspeaker is elevated in the median plane, each ear will receive a shoulder reflection after the direct sound. Algazi et al. [16] suggests that the shoulder reflection is the main cue for elevation perception for frequencies up to 3 kHz. The shoulder reflection delay will increase as the source is elevated up to 90° (overhead) in the median plane, as found in [16]. Similarly, the acoustic crosstalk delay produced by a stereophonic loudspeaker pair presenting coherent signals without a time or level difference will also increase as the loudspeaker base angle increases up to 180°. As also shown in [16], the shoulder reflection delay for a source positioned directly above is similar to the acoustic crosstalk delay that could be produced for a loudspeaker pair with the 180° base angle (i.e., around 0.7ms). From this it can be hypothesised that there is a cognitive process involved in the perception of elevation effect, in such a way that the brain interprets the acoustic crosstalk delay as the shoulder reflection delay. A phantom centre image produced with a certain acoustic crosstalk delay would be elevated at an angle where a real source with the corresponding shoulder delay is elevated. This degree of perceived virtual elevation would increase as the loudspeaker angle increase. It is further hypothesised that this time domain explanation for phantom image elevation would be valid for frequencies up to 1kHz due to phase ambiguity problem at higher frequencies, and that the frequency domain explanation described above would be for frequencies above 1kHz. These hypotheses will be investigated in the future study.

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. 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. [16] 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.

5 SUMMARY AND CONCLUSION

This paper summarised experiments that the author has recently conducted on the elevation of horizontally oriented phantom image, and discussed them in comparison with past research results. The role of spectral cues on median plane localisation has been confirmed in the literature, e.g., pitch-height effect and directional bands. The experiments described in this paper differ from conventional median plane localisation studies in that they investigated the elevation of phantom image rather than that of real image. Main conclusions from these studies are as follows:

- The pitch-height effect operates for the horizontal phantom images of octave-band noises, however it is not linear. That is, there is a reset point at 1kHz, which is perceived as low as the 125Hz and 250Hz bands. This is contradictory to the pitch-height effect observed with real images in the literature.
- Phantom image is perceived to be elevated at low-mid frequencies as well as high frequencies. This is not the case with the elevation perception of real image, and also cannot be explained by HRTF alone.
- As the loudspeaker base angle increases, the perceived elevation angle increases. Especially, phantom images are generally localised above with the base angles around 180° (side by side).
- The above effect depends on the type of sound source. Sources with more broad and flatter frequency spectrum produces a more linear increase of elevation angle with increasing base angle. Elevation judgment for sources with a narrow spectrum is more inconsistent.
- It is hypothesised that the phantom image elevation is a cognitive effect at low frequencies – the brain interprets the acoustic crosstalk delay as a shoulder reflection delay, which is important for vertical localisation. On the other hand, it is considered that phantom image elevation at high frequencies is a HRTF-dependent hard-wired effect.

6 ACKNOWLEDGEMENT

The experiments described in this paper were supported by the Engineering and Physical Sciences Research Council (EPSRC), UK, Grant Ref. EP/L019906/1.

REFERENCES

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

[2] O. C. Trimble, “Localization of Sound in the Anterior–Posterior and Vertical Dimensions of ‘Audi- tory’ Space,” *Brit. J. Psychol.,* vol. 24, pp. 320–334 (1934).

[3] 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).

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

[5] S. Ferguson and D. Cabrera, “Vertical Localization of Sound from Multiway Loudspeakers,” *J. Audio Eng. Soc.*, vol. 53, no. 3, pp. 163–173 (2005).

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

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

[8] 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).

[9] 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).

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

[11] 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.

[12] 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).

[13] 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).

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

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

[16] 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).