OPEN ACCESS Freely available online
Journal of the Audio Engineering Society Vol. 64, No. 10, October 2016
DOI: https://doi.org/10.17743/jaes.2016.0039

Vertical Stereophonic Localization in the Presence of Interchannel Crosstalk: The Analysis of Frequency-Dependent Localization Thresholds

RORY WALLIS, *AES Student Member,* **AND HYUNKOOK LEE,** *AES Member*
(rory.wallis@hud.ac.uk) (h.lee@hud.ac.uk)

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

Listening tests were conducted in order to investigate the frequency dependency of localization thresholds in relation to vertical interchannel crosstalk. Octave band and broadband pink noise stimuli were presented to subjects as phantom images from vertically arranged stereophonic loudspeakers located directly in front of the listening position. With respect to the listening position the lower loudspeaker was not elevated; the upper loudspeaker was elevated by 30°. Subjects completed a method of adjustment task in which they were required to reduce the amplitude of the upper loudspeaker until the resultant phantom image matched the position of the same stimulus presented from the lower loudspeaker alone. The upper loudspeaker was delayed with respect to the lower by 0, 0.5, 1, 5, and 10 ms. The experimental data demonstrated that the main effect of frequency on the localization threshold was significant, with the low frequency stimuli (125 and 250 Hz) requiring significantly less level reduction (less than 6 dB) than the mid-high (1, 2, and 8 kHz) frequency stimuli (9–10.5 dB reduction). The main effect of interchannel time difference (ICTD) on the localization thresholds for each octave band was found to be non-significant. For all stimuli interchannel level difference (ICLD) was always necessary, indicating that the precedence effect is not a feature of median plane localization.

0 INTRODUCTION

Audio reproduction systems for surround sound are currently in a state of evolution. Engineers are increasingly looking to improve on the spatial impression offered by conventional 5.1 systems through the incorporation of loudspeakers in the vertical domain. The implementation of these so-called "height channels" has seen audio reproduction systems move into the third dimension, with systems such as Auro-3D [1] and Dolby Atmos [2] becoming more widely utilized. Such developments inevitably have implications for the recording process, as additional height layers of microphones are required alongside the pre-existing main channel layer in order to capture the necessary spatial information.

In conventional microphone techniques for horizontal surround sound, pairs of microphones are positioned to capture specific areas of the recording angle [3]; examples of this being the "critical linking" technique developed by Williams and Le Du [4] and the "OCT" technique by Theile [5]. For such techniques the phantom imaging of a given sound source in the reproduction stage is achieved based on the time and level differences between the source signal arriving at each of the microphones covering the recording sector in which the source lies. However, should microphones other than the intended pair pick up the direct sound from a source, which is referred to as interchannel crosstalk, then its phantom imaging at the reproduction stage may be affected [5]. Experiments conducted by Lee [6] showed that the most salient effects of interchannel crosstalk are an increase in source width and a decrease in locatedness.

In the context of microphone techniques for recording three-dimensional (3D) sound in an acoustic space, interchannel crosstalk is also oriented between vertically arranged microphones. Consider a 3D microphone array consisting of two vertical layers of microphones. The lower (main) layer would be typically used for horizontal source imaging, while the upper (height) layer would be used to enhance perceived listener envelopment (LEV). Picking up the direct sound in the height microphones may result in the perceived position of the source image "migrating" vertically from the main channel layer. Additional tonal and spatial effects may also be perceived, depending on the interchannel time and level differences between each layer. Henceforth, in the present paper "vertical" interchannel crosstalk refers to direct sound captured by the height channel microphones.

Lee [7] presented anechoically recorded bongo and cello excerpts to subjects from a pair of vertically arranged loudspeakers directly in front of the listening position. The lower

loudspeaker was not elevated, while the upper loudspeaker was elevated by 30°. Stimuli were presented as vertically oriented phantom images. The experiments subjectively measured the minimum amount of attenuation necessary in the upper loudspeaker for the resultant phantom image to be localized at the position of the lower loudspeaker. Lee referred to this as the "localization threshold." Delays, ranging from 0 to 50 ms were applied to the upper loudspeaker with respect to the lower. The results showed that the localization threshold for both sources was between –6 and –7 dB for interchannel time differences (ICTDs) up to 5 ms. This suggests that, should the upper and lower microphone layers be less than 1.7 m apart (corresponding to an ICTD of 5 ms), vertical interchannel crosstalk would not affect the perceived location of the main channel signal provided the amplitude of the direct sound in the upper layer was reduced by between 6 and 7 dB. Despite this, the influence of the height layer on perceptual attributes such as vertical image spread or timbral coloration would remain audible.

An interesting feature of Lee's [7] results is that ICTD alone was never sufficient to localize the source image at the main channel layer, which suggests that the precedence effect [8] did not operate in the vertical loudspeaker arrangement. This agrees with the results of a more recent localization study conducted by the present authors [9]. It was found that the localization of vertical stereophonic phantom images for octave band pink noise stimuli was governed by the so-called Pratt's effect [10] (also known as the "pitch-height" effect [11]) in general, rather than the precedence effect. According to this phenomenon there exists a correlation between the frequency of a stimulus and the height with which it is localized, with high frequencies being perceived as being physically higher in space than low frequencies. The effect had previously been demonstrated for both tonal and octave band stimuli when presented singularly from vertically arranged loudspeakers [11–14]. The aforementioned study by the authors [9] also demonstrated that the main effects of both frequency and ICTD on vertical localization were statistically significant for octave band noises presented from vertically arranged stereophonic loudspeakers. This leads to the hypothesis that different frequency bands would require different amounts of crosstalk reduction for a vertically oriented phantom image to be localized at the perceived position of the main loudspeaker image.

From the above background, the present study conducts an investigation into the frequency dependency of localization thresholds in relation to vertical interchannel crosstalk. It is also of interest to examine the effect of ICTD on localization threshold for each frequency band. Results from this study provide useful implications for the perceptual rendering of vertical phantom images in 3D sound reproduction as well as the design of microphone array for 3D recording in acoustic environments.

This paper is organized as follows. The first section describes the experimental method used in the study. Following this, the results of the statistical analysis of data obtained from listening tests are presented. Finally, the results are discussed, with a particular focus on the effects of both frequency and ICTD on localization thresholds.

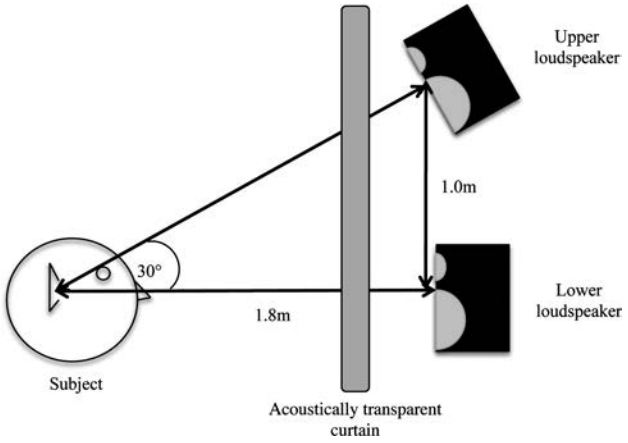

Fig. 1. Physical setup for listening tests.

1 EXPERIMENTAL DESIGN

1.1 Physical Setup

Fig. 1 shows the physical setup used for the experiments, which were conducted in the anechoic chamber at the University of Huddersfield. The experiments utilized two Genelec 8040A loudspeakers, which were positioned as follows. The lower (main) loudspeaker was positioned 1.2 m above the ground, 1.8 m away from, and directly in front of, the listening position. The upper (height) loudspeaker was located 1 m directly above the lower loudspeaker, forming a 30° elevation angle to the listening position. Appropriate time and level alignment was applied to the lower loudspeaker, with respect to the upper, in order to compensate for the differences in distance between each loudspeaker and the listening position. An acoustically transparent curtain was positioned between the listening position and the loudspeakers in order to obscure the nature of the test setup from subjects. The ear height of subjects was aligned to the center point between the woofer and tweeter on the lower loudspeaker using a height-adjustable chair.

1.2 Test Stimuli

The test stimuli used for the experiment were continuous octave bands of pink noise, with center frequencies ranging from 125 Hz to 8 kHz. These were created by brick wall filtering broadband pink noise using an FFT filter. An additional broadband pink noise source was also tested. Each stimulus was ten seconds in duration, which included a one second fade in/out. Stimuli were presented to subjects as vertically oriented phantom images from the loudspeaker pair, with the upper loudspeaker delayed with respect to the lower by 0, 0.5, 1, 5, and 10 ms. The delay times were chosen to simulate differing spacings between the main and height microphone layers; 0 ms is representative of a coincident configuration, while 10 ms corresponds to a spacing of about 3.4 m. In total there were 56 stimuli (eight frequencies with five ICTDs). Each stimulus was calibrated to 75 dB LAeq at the listening position when presented from the lower loudspeaker only. The amplitude of the stimulus when presented as a phantom image was dependent on the amplitude of the upper loudspeaker relative to the

lower, which was to be varied by the subject as described in Sec. 1.4.

1.3 Subjects

Twelve subjects, comprising staff and both postgraduate and final year undergraduate students from the University of Huddersfield's Music Technology courses, participated in the listening tests. These subjects were chosen because of their critical listening experience in spatial audio making them better suited than more naïve subjects to determine the subtle localization differences caused by vertical interchannel crosstalk. They all reported normal hearing.

1.4 Test Method

In order to identify the localization thresholds, subjects were presented with a method of adjustment (MOA) task. This is an indirect scaling method that requires subjects to reduce the amplitude of a stimulus until it is equivalent to that of a reference [15]. Cardozo [16] asserts that the principal application of MOA is in situations whereby stimuli differ from one another by more than one attribute. This is applicable to the present study, as, although subjects were tasked with identifying localization shifts, there would be some timbral changes due to the use of ICTD. Such conditions would make, for example, a two alternative forced choice method [17] inadequate, with Bech [18] reporting subject's difficulty in distinguishing between test and reference stimuli that differed in loudness, timbre, and spaciousness when utilizing an adaptive form of this method.

The graphical user interface for the experiment was created using Max/MSP. The interface split the entire experiment into eight subtests, with each subtest focusing on a single frequency band. Within each subtest was a "reference" and five "test" sounds (labeled A, B, C, D, and E). The reference was the given frequency band played from the lower loudspeaker only. The test stimuli were the same frequency band as the reference presented as vertical phantom images with one of the five test ICTDs applied to the upper loudspeaker.

For each of the test stimuli subjects were presented with a slider with values ranging from 0 to 100 in increments of 1. The slider controlled the amplitude of the upper loudspeaker as follows. Slider values were first divided by 100 to give "x," which lay between 0 and 1. The amplitude of the upper loudspeaker was then multiplied by x. The amplitude of the upper loudspeaker therefore decreased with decreases in the slider value. A slider value of 100 resulted in 0 dB ICLD between the upper and lower loudspeakers. A value of 0 resulted in the upper loudspeaker having zero amplitude ($-\infty$ dB ICLD). Slider values were converted into decibel values internally. The decibel values were not shown to subjects during any part of the test. Subjects were also unaware that they were controlling the amplitude of a loudspeaker. The amplitude of the lower loudspeaker was kept constant throughout each test.

For each test stimulus the subjects' task was to reduce the slider value until the perceived position of the resultant phantom image matched that of the reference. To ensure that the localization threshold was found in each case, subjects were required to set the slider to the highest possible value at which this condition was met. The heads of subjects were not fixed, however they were strictly instructed to face forward, keeping their head still, and using only their eyes to look at the test interface. A guide point for the ear height and distance was placed on the right-hand side of the subject to help maintain the correct listening position throughout the test. Prior to the start of each test, all subjects sat a supervised practice, which utilized a speech source, in order to ensure that the instructions were understood. The order of subtests and the stimuli within each subtest were randomized for each subject.

2 DATA ANALYSIS AND RESULTS

Levene and Shapiro-Wilk tests were first conducted, using the SPSS software, in order to determine the suitability of the collected data for parametric statistical analysis. The results of the Levene test showed homogeneity of variance for all frequencies, while the Shapiro-Wilk test showed that not all scores in each condition featured normal distribution. This therefore meant that the assumptions of Analysis of Variance (ANOVA) were violated. For these reasons, nonparametric tests were chosen for the statistical analysis.

2.1 The Effect of ICTD

Fig. 2 shows the median localization thresholds for each frequency at each ICTD. The medians have been plotted with notch edges. The use of notch edges is a method suggested by McGill et al. [19], who argue that an overlap between notches indicates that pairs of stimuli are not significantly different from one another with 95% confidence.

Based on the notch edges shown in Fig. 2, it appears that changes in ICTD had no significant effect on the localization thresholds obtained for any of the stimuli within the experiment. In order to analyze this further a Friedman test was conducted; the statistical power was judged based on the critical p value of 0.05. The results of this analysis showed that ICTD had no significant effect on the localization thresholds for any stimuli with the exception of 8000 Hz ($p = 0.001$). Additionally the effect size (Kendall's W) was less than 0.5 for all frequency bands including 8000 Hz.

In order to identify which pairs of ICTD were significantly different from one another for the 8000 Hz band a Wilcoxon test was conducted. As such analysis necessitated the performance of multiple pair-wise tests, it was decided to use the Bonferroni correction in a bid to reduce any type 1 errors [20]. The results of this test identified significant differences between the 0 ms and 10 ms ICTDs ($p = 0.03$). Despite this, it is clear from Fig. 2 that there is heavy overlap between the notch edges for this pair of stimuli. When considering this, along with the low effect size (0.413) it seems reasonable to deduce that differences among the different ICTDs for the 8000 Hz band are negligible. Overall it can therefore be concluded that the effect of ICTD on

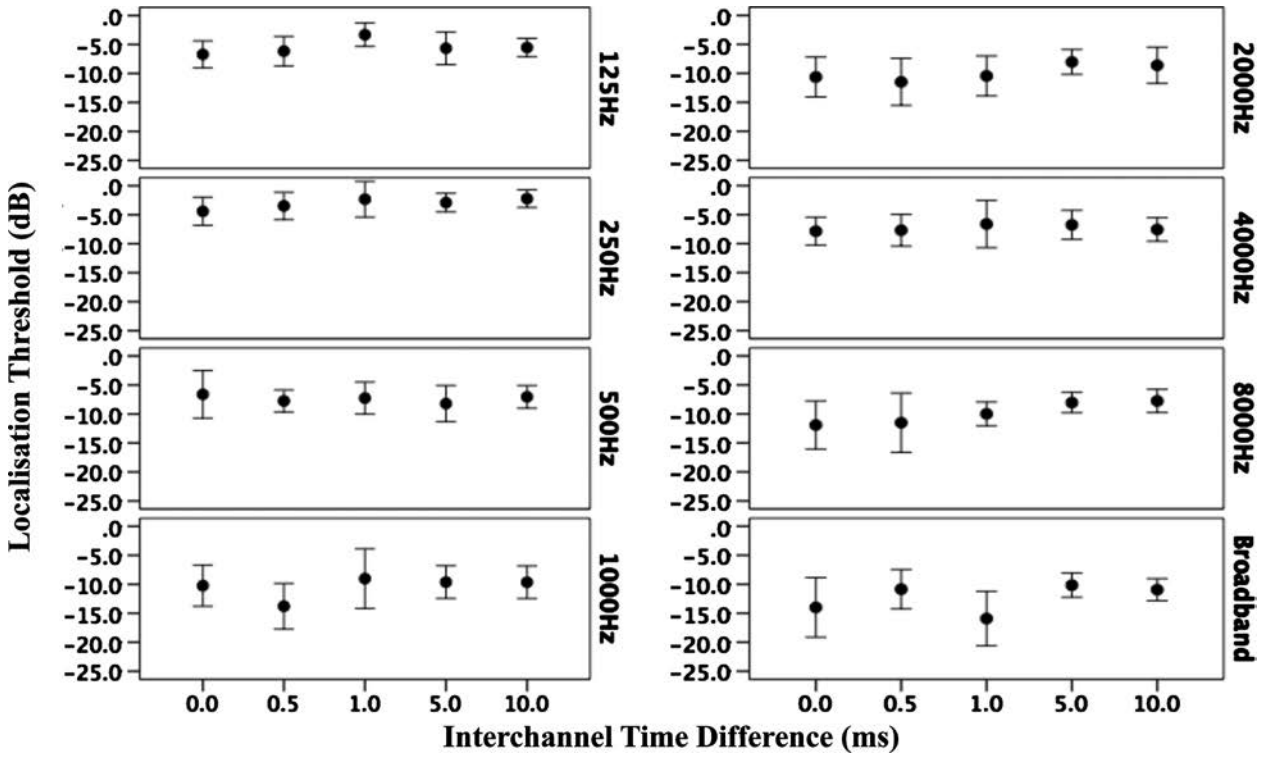

Fig. 2. The effect of ICTD on median localization thresholds for each frequency band, plotted with notch edges. Overlap between notches indicates that pairs of stimuli are not significantly different with 95% confidence.

localization threshold was not significant for any stimulus within the present experiment.

2.2 The Effect of Frequency

In Sec. 2.1 it was shown that ICTD had no significant effect on the localization thresholds obtained for any of the test stimuli. It is therefore possible to combine all the data for each of the frequency bands, rather than consider each ICTD individually. The median localization thresholds for each frequency, with ICTDs amalgamated, are plotted with notch edges in Fig. 3. From Fig. 3 it can be seen that the localization threshold was the highest at low frequencies (–5.3 dB at 125 Hz and –3.03 dB at 250 Hz). This fell gradually to between –9 and –10.5 dB as the frequency increased beyond 1000 Hz. The threshold was the lowest for the broadband source (–11.56 dB), while there was also a small peak for 4000 Hz (–6.96 dB).

Consideration of the notch edges alone suggests that the effect of frequency on localization threshold was significant. A Friedman test was conducted in order to analyze this further (critical p value $= 0.05$). The results of this analysis showed a significant effect of frequency ($p < 0.001$). This result was confirmed with a Wilcoxon test that revealed a large number of significantly different pairs of conditions. Overall, the results of the Friedman and Wilcoxon tests, along with the lack of overlap of notch edges shown in Fig. 4 shows that localization thresholds vary across the frequency spectrum, with the low frequencies needing significantly less level reduction than the mid-high frequencies.

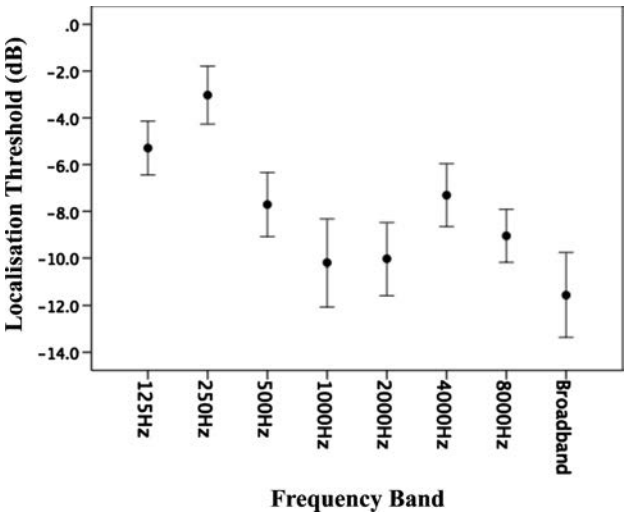

Fig. 3. Median localization thresholds for each frequency band, with results for individual ICTDs amalgamated, plotted with notch edges.

3 DISCUSSION

3.1 Localization Thresholds for Octave Bands

The primary aim of the present study was to analyze how localization thresholds vary across the frequency spectrum and furthermore how they are affected by ICTD. The experimental data obtained demonstrated principally that localization thresholds are not consistent across the full frequency range, with the effect of frequency being

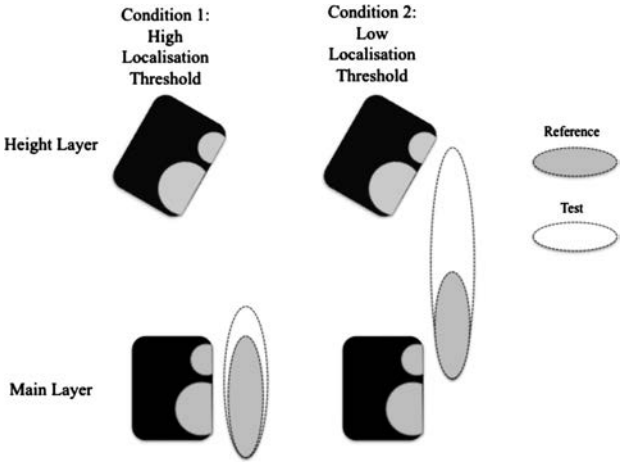

Fig. 4. Example of how differences in the perceived vertical image spread between the "test" and "reference" sounds may have influenced the localization thresholds obtained in the study.

significant. The thresholds were found to be reasonably high for octave bands with center frequencies 125 and 250 Hz (less than –6 dB), with the threshold decreasing to between –9 and –10.5 dB as the frequency increased above 1000 Hz. Within the range of ICTDs tested (0–10 ms), the effect of ICTD on the localization thresholds for each octave band was found to be non-significant. This result is in positive agreement with the localization thresholds obtained by Lee [7] for musical sources with ICTDs up to 5 ms. These results might suggest that, in terms of localization, vertical interchannel crosstalk would be more disturbing to the main channel signal for the mid-high frequencies. Moreover, in the context of microphone array configuration, the amount of attenuation of direct sound necessary in the height microphone layer is consistent irrelevant of the spacing between the upper and lower layers at least up to about 3.4 m, i.e., the ICTD of 10 ms corresponds to a spacing of 3.4 m.

A previous study conducted by the authors [9] found that ICTD generally had a random and inconsistent effect on vertical localization. Perceived median positions for octave band stimuli presented as vertical phantom images were often similar to those for the same stimulus presented from the lower loudspeaker alone. This was the case for middle and high frequency bands tested in the study. However, the results of the current study showed that a large amount of level reduction is necessary for localization thresholds for the 1, 2, and 8 kHz octave bands in particular. This suggests that the difference between the perceived median positions of the test and reference sounds does not directly represent the amount of level reduction required.

In order to address the significant effect of frequency on localization threshold, the authors conducted informal listening exercises during which the perceptual differences between the test and reference stimuli were compared. It was found that the most salient difference, consistent for all stimuli, was a notable increase in the perceived vertical image spread when stimuli were presented as vertical phantom images, compared to lower loudspeaker only presentation. As the upper loudspeaker amplitude was reduced the degree of vertical image spread would decrease leading to the perceived positions of test and reference matching. Based on this, the significant effect of frequency on localization threshold might be explained by the differences in perceived vertical image spread between the test and reference with changes in frequency. This hypothesis is illustrated in Fig. 4. For "Condition 1" the influence of height channel on the increase in perceived vertical image spread is small since the reference inherently has a large spread, necessitating a small amount of reduction in the height channel level (high localization threshold). For "Condition 2," however, the change in vertical spread is considerably larger, requiring an increased amount of level reduction (low localization threshold). From the results of the current study, the following might be inferred. First, based on its non-significant effect on localization threshold, ICTD has little effect on the perceived vertical image spread of octave bands presented from vertically arranged stereophonic loudspeakers in front of the listening position. Additionally, the increase in vertical image spread from single loudspeaker presentation to vertical phantom image presentation is significantly greater for the 1, 2, and 8 kHz octave bands than for the 125 and 250 Hz bands. This hypothesis would require further study.

3.2 Localization Thresholds for Broadband Pink Noise

In [9] it was shown that there was a significant increase in the perceived elevation of broadband pink noise when presented as ICTD-panned phantom images (0.5 and 1 ms) compared to lower loudspeaker only presentation. Therefore, changes in vertical image spread alone are unable to fully explain the localization thresholds observed for the broadband pink noise in the present study. Instead, consideration should be given to how changes in ICLD affect spectral cues, the primary cues used in median plane localization [21].

In order to analyse how changes in ICLD affect the ear-input spectra of broadband stimuli, ear signals for the upper and lower loudspeakers only, as well as stereophonic signals with both 0 and –11.5 dB ICLD (pink noise localization threshold), were simulated using the MIT's KEMAR head related impulse responses (HRIRs) measured at 0° and 30° elevation angles in the median plane [19]. In Fig. 5 the spectra for the upper loudspeaker only, 0 dB ICLD and broadband localization threshold have each been plotted, each with the spectra for the lower loudspeaker subtracted from them (i.e., delta spectrum). For each delta spectrum, any regions where the line is greater than 0 dB represent dominance in the lower loudspeaker. With respect to spectral cues Hebrank and Wright [21] and Asano et al. [23] suggested that key elevation cues exist in the 4–10 kHz region. Additionally, the "above" cue lies in the region between 7 and 9 kHz [21], [24]. This can be seen in the delta spectrum for the upper loudspeaker, which has dominance over the lower loudspeaker at 9 kHz and above. At 0 dB ICLD this dominance is maintained, which would result in the phantom image being perceptually elevated compared

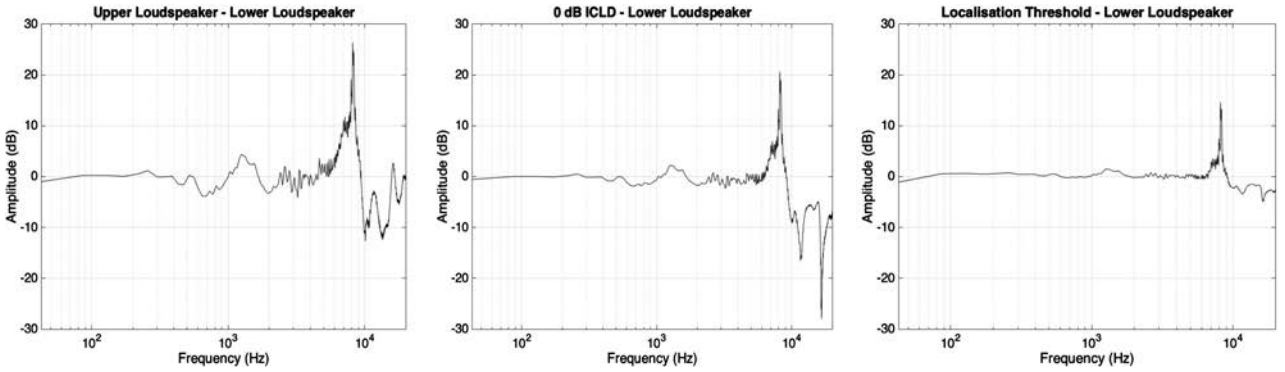

Fig. 5. Difference of the HRTF of (i) upper loudspeaker (30° elevation), (ii) upper and lower (0° elevation) loudspeakers with 0 dB ICLD, and (iii) upper and lower loudspeaker combined with the upper loudspeaker level reduced by 11.5 dB (localization threshold), to that of lower loudspeaker.

to lower loudspeaker presentation. However, with increases in ICLD the influence of the upper loudspeaker on spectral cues would diminish, with the lower loudspeaker becoming more dominant. It can be observed in Fig. 5 that at the localization threshold the dominance above 9 kHz is largely reduced, with the overall spectrum becoming more similar to that for the lower loudspeaker alone (although not identical). It seems reasonable to conclude that the similarity in the spectra for these two conditions is the reason that the two would appear to be co-located.

3.3 The Precedence Effect

The present study suggests the importance of ICLD over ICTD in reaching the localization threshold. For every condition ICTD alone was not sufficient and ICLD was always necessary. This result indirectly suggests that the precedence effect might not be a feature of vertical stereophonic localization; had the effect operated then arguably a sufficient amount of ICTD alone would have been enough for the positions of the test and reference sounds to be perceptually in the same location. This supports the present authors' recent studies reporting that the precedence effect relying on pure ICTD did not operate between vertically arranged loudspeakers [7, 9] for musical and octave band noise stimuli. This might appear to contradict the results of studies conducted by Blauert [25] and Litovsky et al. [26], which suggested that the precedence effect did operate in the median plane. However, it is important to note that both of these studies considered that the precedence effect operated when the position of perceived phantom image was shifted "towards" the earlier loudspeaker, whereas the present authors consider the effect as being valid only if the perceived position of the phantom image exactly "matches" that of the earlier loudspeaker. Despite this, further study, involving a wide range of sound sources, ICTDs and loudspeaker positions, would be necessary to fully investigate whether a vertical precedence effect exists or not. In particular, research is required on the effect of the temporal characteristics of sound source on localization in the presence of a delayed secondary signal that is vertically oriented. In the context of horizontal localization, it is widely known that a strong transient nature of sound is essential for triggering the precedence effect [27]. However, it is not yet clear whether this is still the case for vertical localization. The sound source used in the current study was limited to continuous noise. Hartmann [27] asserts that continuous noise can trigger the precedence effect since it features random amplitude fluctuations that can serve as a series of small impulses. However, it needs to be verified whether the results shown in the current study were obtained due to the nonexistence of the vertical precedence effect itself or due to the small transient cue not being of sufficient strength to trigger the vertical precedence effect. In order to provide more conclusive results on this, sound sources with different temporal characteristics including various natural sources as well as continuous and transient noise signals will be tested in a future study.

3.4 Practical Implications and Future Works

The non-significant effect of ICTD on localization threshold, as well as the absence of the precedence effect in vertical localization, has implications for the design of microphone configurations for recording in 3D audio formats. In the context of preventing vertical interchannel crosstalk from affecting the localization of the main channel signal, it is clear that there should be a focus on the attenuation of direct sounds in the height microphone layer, with the spacing between layers being less of an issue. This would make unidirectional microphones more ideal choice than omnidirectional microphones for the height layer, as the former would be able to provide the necessary attenuation of direct sounds to limit vertical interchannel crosstalk. For example, in the case of a vertically coincident cardioid microphone pair with the main layer microphone pointing directly down towards the sound source and the height microphone pointing away from the source, the localization threshold of about –12 dB, which was obtained for the broadband pink noise in the current study (Fig. 3), could be achieved by applying the subtended angle of about 120° between the microphones. Note, however, that for musical sources the necessary localization threshold is around

–6 dB as found in [7]. For this the subtended angle of the vertical microphone pair needs to be around 90°.

The results demonstrated that the effect of vertical interchannel crosstalk on the localization of the main channel signal is dependent on frequency. It would therefore be of interest to apply the individual localization thresholds obtained for each band to a complex signal, such as music, rather than applying level reduction across the whole frequency spectrum. In this way, any spatial or tonal effects that would potentially be perceived in the presence of the height channel signal could be maintained while achieving localization around the main loudspeaker layer. It may be that this is perceptually more preferable than reducing the amplitude of the full spectrum by a consistent amount. Although this approach may be difficult to execute within a practical recording situation, there would certainly be implications for 3D mixing using discrete sound sources.

In addition to the above, it would be worth examining if the localization shift effect of vertical interchannel crosstalk can be eliminated through the manipulation of selected frequency bands that are perceptually dominant. The delta spectra in Fig. 5 indicate that the localization threshold for complex sounds can be reached by reducing the dominance of the upper loudspeaker on spectral cues. From the delta spectrum for the upper loudspeaker it can be seen that the upper loudspeaker is most dominant over the lower at around 8000 Hz. Chun et al. [28], presented musical sources and speech to subjects from stereophonic loudspeakers arranged on the horizontal plane. The test stimuli first underwent HRTF modeling, followed by spectral notch filtering, directional band boosting, or a combination of both. For the directional band-only condition, the resultant sound sources were perceived as being elevated by up to 20° with respect to the horizontal plane. Based on this, directional band reduction could be applied to perceptually decrease the elevation of sources. This could be an alternative method for preventing the height channel signal from affecting the perceived location of the main channel signal and would have implications for the rendering of 3D images.

It was mentioned in Sec. 3.1 that localization thresholds for octave bands might be as much related to differences in perceived vertical image spread as they are to differences in perceived location. It would therefore be interesting to determine how the thresholds obtained in the present study for band limited stimuli would vary in a room in which reflections are present (i.e., in a more natural listening environment). If reflections are present then this may influence the differences in perceived vertical image spread between test and reference, which in turn may lead to a less strong effect of frequency than was seen in the present study. This would have implications for the application of frequency dependent localization thresholds for complex sources, as the effect would need to be maintained to some extent for the method to have any relevance.

Last, attention should be given to the threshold of acceptability for localization shifts as a result of vertical interchannel crosstalk. Although the present study has considered the amount of attenuation necessary to prevent such a shift, it is not entirely clear yet whether or not complete prevention is desired. It may be the case that small increases in perceptual elevation are acceptable, depending on the type of sound source.

4 CONCLUSION

The present study carried out an analysis of how vertical interchannel crosstalk varies across the frequency spectrum. Seven octave bands of pink noise with center frequencies ranging from 125 Hz to 8000 Hz, as well as broadband pink noise, were presented to experienced subjects as phantom images from vertically arranged stereophonic loudspeakers. The upper loudspeaker was delayed with respect to the lower by 0, 0.5, 1, 5, and 10 ms. Subjects were required to identify the minimum amount of attenuation necessary in the upper loudspeaker for the resultant phantom image position to match that of the same stimulus played from the lower loudspeaker alone (the localization threshold).

The results of the study showed that the main effect of frequency on localization threshold was significant. Thresholds were the highest at low frequencies (–5.3 dB at 125 Hz and –3.03 dB at 250 Hz), falling to between –9 and –10.5 dB as the frequency increased beyond 1000 Hz. It was hypothesized that the primary reason for this was variations in perceived vertical image spread between lower-loudspeaker-only presentation and phantom image presentation with changes in frequency. In addition, the threshold for the broadband pink noise source was the lowest of all thresholds (–11.56 dB). This result was interpreted in terms of the dominance of the upper loudspeaker on spectral cues, with increases in ICLD resulting in a spectrum more similar to that of the stimulus presented from the lower loudspeaker alone.

The main effect of ICTD was not significant on localization threshold for any of the test stimuli. Moreover, ICLD was always necessary for the localization threshold; there was no condition whereby ICTD alone was sufficient. This result suggests that the relative amplitudes between the upper and lower loudspeakers are of greater importance for reducing the localization shift effect of vertical interchannel crosstalk than are the vertically applied time delays. The results also indirectly suggest that the precedence effect does not operate in the median plane, although this requires further study.

The results imply that in configuring 3D microphone array cardioid microphones would be a more appropriate choice for the height layer than would be omnidirectional microphones in terms of localizing source images near the main loudspeaker layer position. In addition, when creating a vertical phantom image different localization thresholds could be applied to different frequency bands of the height channel signal. It is possible that localization thresholds can be achieved by manipulating the levels of single octave bands within the height channel signal rather than by manipulating the signal as a whole.

It should also be noted however that the present study utilized subjects who were experienced in vertical sound localization tests. It is, as yet, unclear how the results would

vary for less experienced subjects and this would require further study. It might be, for example, that the experienced subjects are more sensitive to the effects of ICLD changes on perceived source location and therefore more level reduction might be needed generally compared to if less experiences subjects were tested. As a result of this caution should be exercised when attempting to generalize the results of the present study for all individuals.

5 ACKNOWLEDGMENT

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC), UK, Grant Ref. EP/L019906/1. The authors are grateful to the music technology students and staff members at the University of Huddersfield who participated in the listening tests. They also thank the editor and anonymous reviewers of this paper for their insightful and constructive comments.

6 REFERENCES

[1] Auro Technology, URL: http://www.auro-3d.com/system/listening-formats/ (2014).

[2] Dolby, *URL*: http://www.dolby.com/gb/en/consumer/technology/movie/dolby-atmos-details.html (2014).

[3] F. Rumsey, *Spatial Audio* (Focal Press, Burlington, MA, 2001).

[4] M. Williams and G. Le Du, "Multichannel Microphone Array Design," presented at the *108th Convention of the Audio Engineering Society* (2000 Feb.), convention paper 5157.

[5] G. Theile, "Natural 5.1 Recording Based on Psychoacoustic Principles," presented at the *AES 19th International Conference: Surround Sound—Techniques, Technology, and Perception* (2001 June), conference paper 1904.

[6] H. Lee, "Effects of Interchannel Crosstalk in Multichannel Microphone Technique," Ph.D. Thesis, University of Surrey (2006 Feb.).

[7] H. Lee, "The Relationship between Interchannel Time and Level Differences in Vertical Sound Localization and Masking," presented at the *131st Convention of the Audio Engineering Society* (2011 Oct.), convention paper 8556.

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

[9] R. Wallis and H Lee, "The Effect of Interchannel Time Difference on Localization in Vertical Stereophony," *J. Audio. Eng. Soc.*, vol. 63, pp. 767–776 (2015 Oct.) http://dx.doi.org/10.17743/jaes2015.0069.

[10] D. Cabrera and M. Morimoto, "Influence of Fundamental Frequency and Source Elevation on the Vertical Localization of Complex Tones and Complex Tone Pairs," *J. Acoust. Soc. Am.*, vol. 122, no. 1, pp. 478–488 (2007 Jul.) http://dx.doi.org/10.1121/1.2736782.

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

[12] C. C. Pratt, "The Spatial Character of High and Low Tones," *J. Exp. Psychol.*, vol. 13, no. 3, pp. 278–285 (1930 June).

[13] O. C. Trimble, "Localisation of Sound in the Anterior-Posterior and Vertical Dimensions of Auditory Space," *Brit. J. Psychol.*, vol. 24, no. 3, pp. 320–334 (1930 Jan.), http://dx.doi.org/10.1111/j.2044-8295.1934.tb00706.

[14] S. K. Roffler and R. A. Butler, "Localization of Tonal Stimuli in the Vertical Plane," *J. Acoust. Soc. Am.*, vol. 43, no. 6, pp. 1260–1266 (1968), http://dx.doi.org/10.1121/1.1910977.

[15] S. Bech and N. Zacharov, *Perceptual Audio Evaluation – Theory, Method and Application (*John Wiley and Sons, Chichester, West Sussex, England, 2006).

[16] B. Cardozo, "Adjusting the Method of Adjustment: SD vs DL," *J. Acoust. Soc. Am.*, vol. 37, no. 5. pp. 786–792 (1965 May) http://dx.doi.org/10.1121/1.1909439.

[17] M. Yogan and A. Stocker, "A New Two-Alternative Forced Choice Method For the Unbiased Characterization of Perceptual Bias and Discriminability," *J. Vis.,* vol. 14, no. 3. pp. 1–18 (2014 Mar.) http://dx.doi.org/10.1167/14.3.20.

[18] S. Bech, "Spatial Aspects of Reproduced Sound in Small Rooms," *J. Acoust. Soc. Am.*, vol. 103, no. 1, pp. 434–445 (1998 Jan.) http://dx.doi.org/10.1121/1.421098.

[19] R. McGill, J. W. Turkey and W. A. Larsen "Variations of Box Plots," *Am. Stat.*, vol. 32, no. 1, pp. 12–16 (1978 Feb.) http://dx.doi.org/0.2307/2683468.

[20] R. Simer, "An Improved Bonferroni Procedure for Multiple Tests of Significance," *Biometrika*, vol. 73, no. 3. pp. 751–754 (1986 Dec.) http://dx.doi.org/10.2307/2336545.

[21] J. Hebrank and D. Wright, "Spectral Cues Used in the Localization of Sound Sources on the Median Plane," *J. Acoust. Soc. Am.*, vol. 56, no. 6, pp. 1829–1834 (1974a Dec.) http://dx.doi.org/10.1121/1.1903520.

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

[23] F. Asano, Y. Suzuki and T. Sone, "Role of Spectral Cues in Median Plane Localization," *J. Acoust. Soc. Am.*, vol. 88, no. 1, pp. 159–168 (1990 July) http://dx.doi.org/10.1121/1.399963.

[24] J. Blauert, "Sound Localization in the Median Plane," *Acust.*, vol. 22, pp. 205–213 (1969 Jan.).

[25] J. Blauert, "Localization and the Law of the First Wavefront," *J. Acoust. Soc. Am.*, vol. 50 , no. 2, pp. 466–470 (1971) http://dx.doi.org/10.1121/1.1912663.

[26] R. Y. Litovsky, B. Rakerd, T. C. T. Yin and W. M. Hartmann, "Psychophysical and Physiological Evidence for a Precedence Effect in the Median Sagittal Plane," *J. Neurophysiol.*, vol. 77, pp. 2223–2226 (1997 April).

[27] W. M. Hartmann, "Localization of Sound in Rooms," *J. Acoust. Soc. Am*, vol. 74, no. 5, pp. 1380–1391 (1983 Nov.) http://dx.doi.org/10.1121/1.390163.

[28] C. Chun et al., "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 Nov.) http://10.1109/TCE.2011.6131171.

THE AUTHORS

Rory Wallis

Hyunkook Lee

Rory Wallis is a Ph.D. student and member of the University of Huddersfield's Applied Psychoacoustics Lab (APL). He graduated with a first class degree in music technology with audio systems from Huddersfield and was granted the Vice Chancellor's Scholarship to pursue postgraduate research. The primary focus of his Ph.D. is vertical localization with respect to 3D audio, with a particular interest in the location-based effects of vertical interchannel crosstalk and the development of methods to reduce them. He has published in *JAES* and has also presented research at the 136th, 138th, and 140th AES conventions. Alongside his Ph.D. work he has also taught concert hall recording as part of the University of Huddersfield's Music Technology courses.

•

Dr 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, sound capturing and rendering techniques for 3D and VR audio, 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.