OPEN ACCESS Freely available online

Journal of the Audio Engineering Society
Vol. 63, No. 10, October 2015 (© 2015)
DOI: http://dx.doi.org/10.17743/jaes.2015.0069

The Effect of Interchannel Time Difference on Localization in Vertical Stereophony

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

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

Listening tests were conducted in order to analyze the localization of band-limited stimuli in vertical stereophony. The test stimuli were seven octave bands of pink noise, with center frequencies ranging from 125–8000 Hz, as well as broadband pink noise. Stimuli were presented from vertically arranged loudspeakers either monophonically or as vertical phantom images, created with the upper loudspeaker delayed with respect to the lower by 0, 0.5, 1, 5, and 10 ms (i.e., interchannel time difference). The experimental data obtained showed that localization under the aforementioned conditions is generally governed by the so-called "pitch-height" effect, with the high frequency stimuli generally being localized significantly higher than the low frequency stimuli for all conditions. The effect of interchannel time difference was found to be significant on localization judgments for both the 1000–4000 Hz octave bands and the broadband pink noise; it is suggested that this was related to the effects of comb filtering. Additionally, no evidence could be found to support the existence of the precedence effect in vertical stereophony.

0 INTRODUCTION

The mechanisms used to localize sound sources incident from the median plane are fundamentally different from those used in horizontal plane localization. In the horizontal plane, localization is reliant on a combination of the time and level differences between a given source arriving at each ear (binaural cues) as well as on the directional filtering of the sound source by the pinnae (spectral cues) [1]. However, in the median plane binaural cues are absent as sound sources arrive at each ear simultaneously. As a result, median plane localization relies solely on spectral cues [2].

Median plane localization is a topic that has received much attention in the literature, with numerous studies being particularly concerned with the localization of tonal and band limited stimuli. In early experiments using tonal stimuli presented from vertically arranged loudspeakers, Pratt [3] concluded that localization is governed solely by frequency, with high tones being localized physically higher in space than low tones. A similar observation was made by Trimble [4], who presented tonal stimuli both singularly and in succession to listeners via receiving phones positioned 15 cm from each ear. A more expansive study by Roffler and Butler [5], also using tonal stimuli presented from vertically arranged loudspeakers, affirmed the results presented in [3] and [4], with the authors noting that the effect was maintained irrelevant of listener orientation, visual bias, and whether or not subjects had prior knowledge of the terms "high" and "low" in describing pitch. Subsequent experiments by Roffler and Butler [6] and Cabrera and Tiley [7] demonstrated that the relationship between pitch and height is maintained for the localization of band-passed noise signals and moreover that the perceptual range of pitch-height depends on the physical height of the loudspeaker that presents the signal. In [7] the correlation between pitch and height was referred to as the "pitch-height effect."

Following the Roffler and Butler study [5] it was noted by Blauert [8], from median plane localization experiments using loudspeakers placed in front of, directly above and behind the listener, that frequency also governed the localization of 1/3-octave bands. Under these conditions certain frequency bands were related to specific locations on the median plane, irrelevant of actual loudspeaker position. Blauert called these bands "directional bands." Subsequent studies by Hebrank and Wright [2] and Asano et al. [9] have shown that directional bands are closely related to the spectral cues provided by the pinnae in vertical localization. Additionally, Itoh et al. [10] demonstrated that directional bands are maintained for 1/6-octave bands of noise and that there exist differences in directional bands depending on the listener.

The aforementioned localization studies are similar in that they predominantly considered the localization of stimuli presented from single loudspeakers located on the median plane. However, with the emergence of

three-dimensional (3D) audio systems employing vertically elevated height loudspeakers, such as Dolby Atmos [11] and Auro 3D [12], it becomes necessary to examine localization mechanisms in vertical stereophony, i.e., vertical "panning." Within this context it can be considered that the effect of interchannel level difference (ICLD) is already fairly well understood. Barbour [13] analyzed the effect of ICLD on the localization of male speech and pink noise stimuli with a pair of loudspeakers elevated at various angles in the median plane. The results of this study showed that localization was generally unstable and had a high degree of variance between individual subjects. Despite the inconsistent effects of ICLD all subjects localized the resultant phantom images at a position biased towards the loudspeaker of greater amplitude. Overall, a ±15 dB ICLD was found to be sufficient for a full image shift. Additionally, ICLD serves as the basis for the Vector Base Amplitude Panning (VBAP), which is a 3D panning method proposed by Pulkki [14].

An aspect of median plane stereophony that can be considered as being less understood is the effect of interchannel time difference (ICTD). This is exemplified in the conflicting evidence concerning the operation of the precedence effect, a psychoacoustic phenomenon usually attributed to horizontal stereophony. For stereophonic loudspeakers radiating coherent signals, an ICTD of 1.1 ms is sufficient for the resultant phantom image to be localized at the position of the earlier loudspeaker [15]. Below this threshold the phantom image is perceived at a location biased towards the earlier loudspeaker, a phenomenon known as "summing localization" [15]. Studies conducted by both Blauert [16] and Litovsky et al. [17], using white noise and clicks respectively, provided evidence to support the operation of the precedence effect in the median plane. However, a more recent study conducted by Lee [18], using musical sources, produced results that somewhat counteracted these findings. It should be noted that there were clear differences between the respective studies. With regards to experimental setup, [16] and [17] utilized loudspeakers in front, above, and behind listeners, while [18] utilized vertically arranged stereophonic loudspeakers in front of the listener, with the upper loudspeaker elevated by 30°. Moreover, the respective authors had differing definitions for the precedence effect. Where [18] considered the effect as operating only if the ICTD caused the resultant phantom image to be localized at the position of the earlier loudspeaker, both [16] and [17] considered more of a localization "dominance" of the earlier loudspeaker.

From the above background, the present study conducted subjective experiments in order to investigate the effect of ICTD on vertical stereophonic localization in the median plane and its dependency on frequency. In previous studies using single-loudspeaker presentation it has been demonstrated that localization is governed by frequency, with a correlation between pitch and height. It was therefore of interest to examine whether this correlation is maintained when the stimuli are presented as ICTD-panned phantom images from vertically-arranged stereophonic loudspeakers. With respect to the effect of ICTD, a further aim of the study was to determine whether localization judgments become biased towards the earlier loudspeaker as the ICTD increases and whether or not evidence can be found for the precedence effect in vertical stereophony.

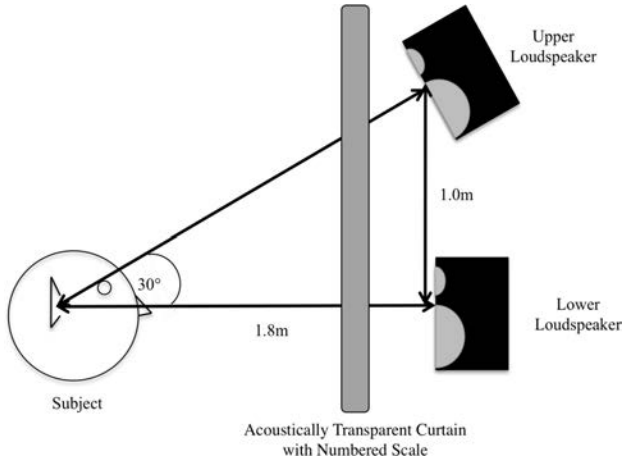

Fig. 1. Loudspeaker setup used for the listening test.

This paper is organized as follows. The first section describes the experimental method used in the study. Following this, the experimental data is presented, with the results analyzed statistically. Finally, the results are discussed, with a particular focus on the implications for median plane localization.

1 EXPERIMENTAL DESIGN

1.1 Physical Setup

The listening tests were conducted in an anechoic chamber at the University of Huddersfield. For the test configuration two Genelec 8040A loudspeakers were arranged vertically in the median plane. The lower loudspeaker was positioned 1.2 m above the ground at a distance of 1.8 m from the listening position. The upper loudspeaker was positioned 1 m above the lower loudspeaker (Fig. 1). The ear height of each subject was aligned to the height of the middle position between the woofer and tweeter on the lower loudspeaker using a height adjustable chair. With respect to the listening position, the upper and lower loudspeakers were elevated by 30° and 0°, respectively. Appropriate time and level alignment was applied to the lower loudspeaker with respect to the upper. An acoustically transparent curtain was placed directly in front of the loudspeakers in order to obscure the test setup from subjects. The curtain featured a numbered scale, ranging from 0 to 100, with a step size of 10, which spanned the entire height of the room. With respect to the listener, the lower loudspeaker was located at '52' on the scale, with the upper loudspeaker at '83'.

1.2 Test Stimuli

The stimuli used in the experiment were created by brick-wall filtering a continuous pink noise source into octave bands using an FFT filter. Seven consecutive octave bands were used, with center frequencies ranging from 125 to 8000 Hz. For comparison, the original broadband pink noise

was also included in the test. Each of these eight signals were reproduced continuously using seven different presentation methods: (i) lower loudspeaker alone; (ii) upper loudspeaker alone; (iii – vii) both loudspeakers together with the upper one delayed with respect to the lower one by 0, 0.5, 1, 5, and 10 ms. There were, therefore, a total of 56 stimuli. The range of ICTDs was chosen for a number of reasons. First, the use of ICTDs below 1.1 ms should allow for any potential summing localization to be identified. Beyond this threshold, the precedence effect should be able to be observed, if it is indeed a mechanism used in median plane localization. Additionally, 10 ms was chosen as a maximum to reflect a likely maximum ICTD in practical recording situations (10 ms corresponds to an upper microphone spaced 3.4 m from the lower). Each stimulus was calibrated to 75 dB LAeq at the listening position and lasted for a total of 10 s (1 s fade in/out, 8 s sustain).

1.3 Subjects

Twelve subjects, comprising of 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 to determine the subtle localization differences between the stimuli than more naïve subjects. They all reported normal hearing.

1.4 Test Method

The graphical user interface used for the experiment was created using Max/MSP. For each test stimulus, subjects were presented with a slider, which had values ranging from 0 to 100, in increments of 1. This slider was to be adjusted until its value matched the perceived location of the stimulus on the scale in front. Each subject's sitting position was adjusted so that the ear height matched the height of the lower loudspeaker (52 on the scale). The distance between the subject's ear and the lower loudspeaker was set to 1.8 m. Although the subjects' heads were not fixed and their movements were not monitored, they were strictly instructed to maintain the set position, facing forward, keeping their head still and using only their eyes if they needed to look at the scale or 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. All subjects sat a supervised practice, which used a speech source, to ensure that they fully understood the test instructions.

Due to the number of stimuli the test was conducted in two parts, containing 28 stimuli each. Subjects were required to wait a minimum of three hours between each half of the test to remove the effects of any fatigue. The stimuli were randomized between the two tests and the test order was randomized for each listener to prevent any psychological bias.

It was decided to present the test results as elevation angles, as opposed to simply showing the gradings given on the scale. To achieve this it was necessary to calculate how many degrees of elevation a step increase of 1 on the scale corresponded to. Given that the height loudspeaker was located at "83" on the scale and was elevated by 30° with respect to the lower loudspeaker, which was located at "52," it was calculated that a step increase of 31 corresponded to an elevation increase of 30°. Therefore, a step increase of 1.03 on the scale was equivalent to 1° of elevation. Consequently, upon completion of each test, all the subjective gradings on the scale were divided by 1.03 in order to present them as elevation angles.

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's test showed homogeneity of variance for all frequencies, while the Shapiro-Wilks 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 this reason, non-parametric tests were chosen for the statistical analysis.

2.1 The Effect of Frequency

A Friedman test was conducted to analyze the main effect of frequency. The results showed that the effect of frequency was significant for all presentation methods at the 1% level. The median perceived elevation angles for each frequency from each presentation method are plotted with notch edges in Fig. 2. 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.

The use of notch edges in Fig. 2 confirms the results of the Friedman test. It can be seen that there is significant difference between the perceived elevations of at least one pair of stimuli for each presentation method. Each graph in Fig. 2 shows that in general the high frequencies (4000 and 8000 Hz) were perceived as being in a perceptually more elevated position than were the low frequencies (125–500 Hz). This difference is significant for lower loudspeaker presentation, as well as for the 0, 1, 5, and 10 ms ICTDs. This being said, there is considerable overlap between the notch edges for the high and low frequencies for 0.5 ms ICTD presentation, as well as between the 4000 and 500 Hz stimuli for upper loudspeaker presentation.

Despite the aforementioned result, it cannot be said that the relationship between pitch and height was linear for any presentation method. Within the low frequency range there was no significant difference between localization judgments for the 250 and 500 Hz bands, neither was there any between the high frequency stimuli.

In addition to this, the mid frequencies (1000 and 2000 Hz) had variable localization depending on the presentation method. The 1000 Hz band, for example, was sometimes localized beneath the position of the lower loudspeaker (lower loudspeaker, 0.5 ms ICTD, 5 ms ICTD),

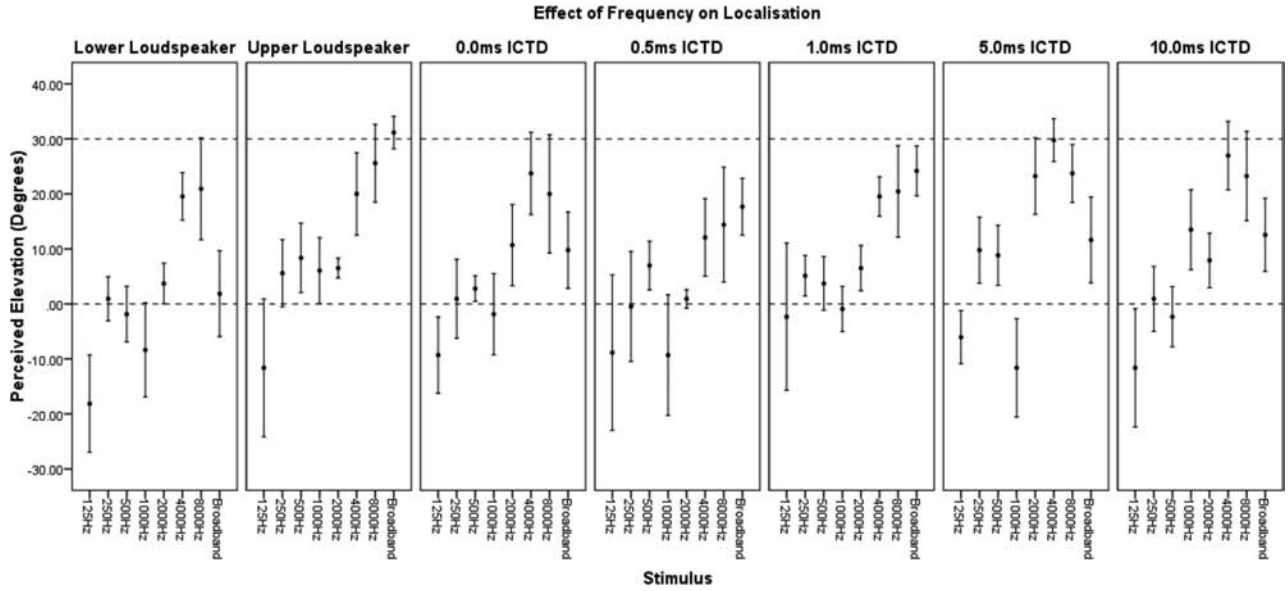

Fig. 2. Median perceived elevation of stimuli with notch edges. The dashed lines at 0° and 30° represent the positions of the upper and lower loudspeakers respectively. Overlap between notches suggests that pairs of stimuli are not significantly different from one another with 95% confidence.

while at other times it was perceived as being above it (10 ms ICTD, upper loudspeaker). Likewise, for the 2000 Hz band localization judgments were sometimes similar to those for the 250 and 500 Hz bands (upper loudspeaker, 1 ms ICTD), whereas for other presentation methods it was perceived as being significantly higher than these stimuli (5 ms ICTD).

2.2 The Effect of Presentation Method

A Friedman test was conducted to analyze the main effect of presentation method on localization. The results showed significance for the 125 Hz ($p < 0.05$), 1000 Hz ($p < 0.01$), 2000 Hz ($p < 0.01$), and 4000 Hz ($p < 0.05$) octave bands as well as for the broadband ($p < 0.001$) stimulus. In Fig. 3 the perceived elevation of each stimulus has been grouped by frequency. The data has again been plotted with notch edges.

From Fig. 3 it can be seen that the effect of presentation method had little or no effect on the localization judgments for the low frequencies. Although the Friedman test results suggested that the effect of presentation method was significant for the 125 Hz band, in Fig. 3 it can be seen that the notch edges for all presentation methods do overlap. It should be noted however that the overlap between lower loudspeaker presentation and the 5.0 ms ICTD is minimal. Regarding the results for the 250 and 500 Hz bands, there is agreement between the results of the Friedman test and Fig. 3 although again the overlap between some pairs of presentation methods is minimal.

As the frequency increased it is clear that the presentation method began to have some influence on localization judgments. Despite this, the significant pairs are not consistent for all stimuli. For example, the results for 1000 Hz show that the 10.0 ms ICTD was localized in a significantly higher position than all other presentation methods, with the exception of upper loudspeaker presentation. However, for 4000 Hz the 10.0 ms ICTD was only localized significantly higher than the 0.5 ms ICTD, while for 2000 Hz this presentation method was localized significantly lower than the 5.0 ms ICTD, with localization being similar to most of the other presentation methods. Presentation method can therefore be said to have had an influence on localization for stimuli in the range of 1000–4000 Hz although the overall effect was somewhat erratic. For the 8000 Hz stimuli, presentation method had no significant effect.

Localization of the broadband stimulus was also influenced by presentation method. There is clear significant difference between localization judgments for upper and lower loudspeaker presentation, where the stimulus was accurately localized at the position of the emitting loudspeaker. Overall, judgments for upper loudspeaker presentation were significantly higher than those for the other presentation methods with the exception of the 1.0 ms ICTD, although notch overlap in this case is minimal. There was no significant difference between localization judgments for the 0.0, 0.5, 5.0, and 10.0 ms ICTDs. The 1.0 ms ICTD was, however, localized significantly higher than the 0.0, 5.0, and 10.0 ms ICTDs.

3 DISCUSSION

3.1 The Pitch-Height Effect

The experimental data obtained in the present study shows that the pitch-height effect governs the median plane localization of octave band stimuli. Moreover, the effect is maintained when the stimuli are presented either monophonically or as ICTD-panned phantom images from vertically arranged loudspeakers. However, despite the fact that for the majority of conditions the high frequencies were localized in a significantly higher position than the low

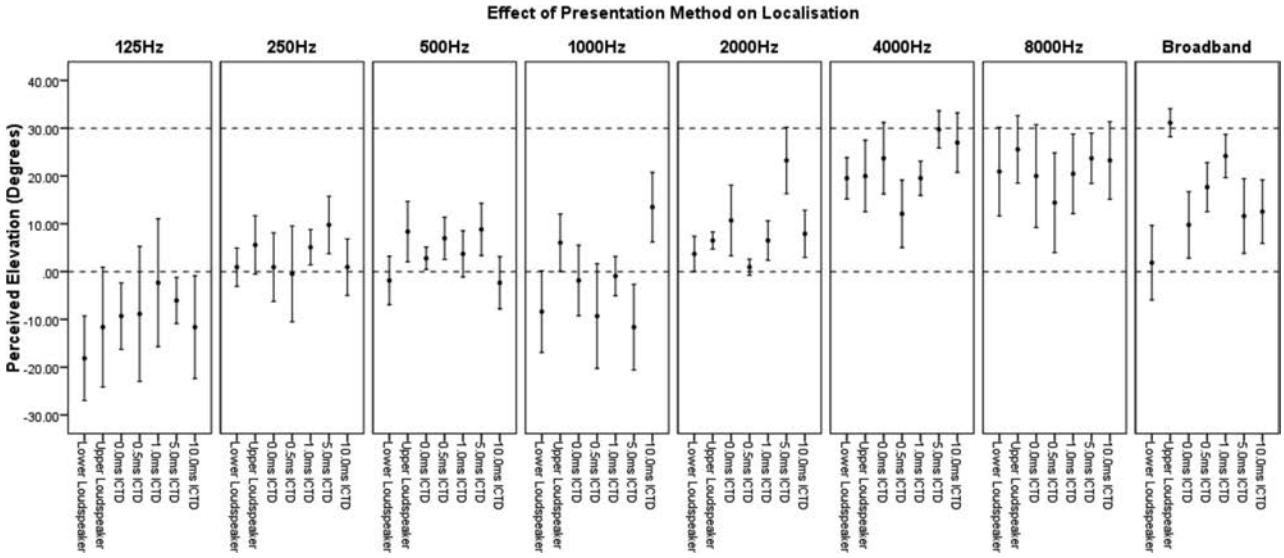

Fig. 3. The effect of presentation method on localization. Overlap between notches suggests that pairs of stimuli are not significantly different from one another with 95% confidence.

frequencies, the correlation between pitch and height was not linear in nature for any presentation method. This is demonstrated in the lack of significant difference between localization judgments for the 4000 and 8000 Hz, as well as for the 250 and 500 Hz, octave bands for all presentation methods. Additionally, localization judgments for the mid frequency stimuli were highly erratic, appearing somewhat random at times, with perceptual elevation certainly not in line with the pitch-height effect.

The result that the pitch-height effect governs the localization of octave bands presented from vertically arranged loudspeakers has been previously demonstrated by Cabrera and Tiley [7]. In their experiment pink noise, filtered pink noise, and octave bands were presented to subjects from either 1, 3 or 5 contiguous loudspeakers arranged vertically in the median plane, with elevation angles of 0°, $\pm 7.9^\circ$, and $\pm 15.6^\circ$. The center frequencies of the octave band stimuli were 125, 500, 2000, and 8000 Hz. Stimuli were presented to listeners in ten 200 ms bursts at loudness levels of 64 and 84 phons. There are a number of similarities between the results obtained in the respective studies. First, when stimulus presentation was from the non-elevated loudspeaker, both studies found that the 125 Hz band was localized beneath the position of the emitting loudspeaker, while 2000 and 8000 Hz were localized above. It should be noted however that in [5], 500 Hz was perceived as being beneath the lower loudspeaker, while in the present study 500 Hz coincided more with the actual loudspeaker position. Additionally, both studies showed that the broadband (pink noise) stimuli were localized accurately at the position of the emitting loudspeaker (this was the case for monophonic presentation in the present study).

However, although for upper loudspeaker presentation both studies showed a relationship between pitch and height, the results in [7] showed that judgments for the 125 and 500 Hz octave bands were almost identical. Conversely, in the present study the 500 Hz band was localized significantly higher than 125 Hz. In this case judgments for 500 Hz were more in line for those with 2000 Hz. Moreover, when the 8000 Hz octave band was presented from the uppermost loudspeaker in [7], localization judgments were significantly higher than for when the same stimulus was presented from the non-elevated loudspeaker. This was the case for both loudness levels. There also appeared to be some correlation between the position of the emitting loudspeaker and the perceived location of the stimulus. No such correlation could be seen in the present study. Additionally, the difference between localization for the 8000 Hz stimuli when presented monophonically from either loudspeaker was not significant. It should be noted that there were a number of differences in the experimental setup that may have contributed to these differences. First, there were differences in upper loudspeaker elevation in the respective studies. In the present study the upper loudspeaker was elevated by 30°, whereas in [7] the uppermost loudspeaker was only elevated by 15.6°. Moreover, the stimuli in [7] were presented as 200 ms bursts; in the present study stimulus presentation was continuous. It is possible that the listeners in [7] were therefore afforded additional localization cues due to the burst nature of the stimuli, leading to more accurate localization of the 8000 Hz octave band. This would require further study.

The localization of octave bands with center frequencies 250, 1000, and 4000 Hz was not considered in [7]. This makes it difficult to analyze whether the lack of pitch-height linearity was as much a feature in that study as it was in the present. The experimental data presented here indicates that localization judgments for the 250 Hz band were similar to those for the 500 Hz band, while the 4000 Hz band was localized similarly to the 8000 Hz band. Localization judgments for the 1000 Hz band were slightly more erratic. Generally this stimulus was localized beneath, or in the position of, the lower loudspeaker. The results for both 1000 and 4000 Hz show some agreement with the

findings of Wallis and Lee [20]. Although the context of the experiments differed somewhat, the experimental data obtained in each study showed that 4000 Hz octave bands are perceptually elevated, while localization judgments for 1000 Hz octave bands are often in positions below the horizontal plane. It should be noted, however, that for both upper loudspeaker and 10 ms ICTD presentation localization judgments for the 1000 Hz were much higher, being above the position of the lower loudspeaker.

3.2 The Effect of Delay Time

The experimental data showed that localization of the low frequency stimuli, being 500 Hz and below, was consistent irrelevant of whether the stimuli were presented monophonically or as ICTD-panned phantom images. For these frequencies the effect of presentation method was found to be non-significant. This was also observed for the 8000 Hz band. Conversely, for the octave bands between 1000 and 4000 Hz, presentation method did have some significance on localization judgments. It can be seen in Fig. 3 that the 4000 Hz octave band was localized particularly low when the ICTD was 0.5 ms. Additionally, the 2000 Hz band had unusually high perceptual elevation when the ICTD was 10 ms, while the 1000 Hz band was localized noticeably low at 5 ms. Moreover, for the broadband stimulus localization judgments were higher for the 1 and 5 ms ICTDs compared to those for the other stereophonic presentation methods.

In order to gain objective insights into the current results, the ICTDs used for the experiment were applied to head related impulse responses (HRIRs) measured for 0° and 30° elevation angles at 0° azimuth, taken from the MIT's KEMAR dummy head database [21]. Fig. 4 shows the spectra of the resulting HRIRs, i.e., head related transfer functions (HRTFs). From this analysis it is clear that comb filtering had a large influence on the frequency content of all stimuli presented stereophonically with a delay in the upper loudspeaker. From these results the significant effect of presentation method on the perceived elevation observed for some of the stimuli in the present study is possible of explanation. It is well established that spectral cues are required for a sound to be localized in a specific region in vertical space. Previous research [2, 9], has shown that the spectral cues for elevation correspond to a notch in the frequency spectrum in the range between 4 and 10 kHz, with an increase in the center frequency of this notch leading to the perception of increased elevation. From Fig. 4 it can be seen that the effect of comb filtering is to introduce additional notches in the frequency spectrum. It is possible that the human auditory system can interpret such notches as spectral cues for vertical localization, therefore affecting perceived elevation.

Consideration of the results for the broadband source, as shown in Fig. 3, somewhat validates the above hypothesis. When the source was presented with ICTDs between 0.0 and 1.0 ms, an increase in ICTD led to the perception of increased elevation. From Fig. 4 it is noticeable that the 0.5 ms ICTD features notches in the region of 3 and 5 kHz, which are not present for the 0.0 ms ICTD. These notches may have been interpreted as elevation cues, leading to the 0.5 ms ICTD being perceptually elevated with respect to the 0.0 ms ICTD. Additionally, the 1.0 ms ICTD features notches in the region of 5, 6, and 9 kHz. This stimulus was perceived to be more elevated than the 0.5 ms ICTD, which may be due to the increased center frequencies of the notches. If this were the case then this would verify the results in [2]; that increased notch center frequency between 4 and 10 kHz leads to the perception of increased elevation. It should be noted, however, that the differences in perceived elevation between 0.0 and 0.5 ms and between 0.5 and 1.0 ms were not significant. It is not entirely clear how the distinct notches affect perceived elevation and this requires further study.

Interestingly, the frequency content for the 5.0 and 10.0 ms ICTDs feature considerably more notches between 4 and 10 kHz than for the other stimuli and yet they were perceptually no more elevated than the stereophonic broadband stimulus with no ICTD applied. This result is again possible of interpretation based on the results in [2], where it suggests that the elevation cue between 4 and 10 kHz is a 1-octave notch. It is well known that as the ICTD increases the bandwidth of the notches due to comb filtering decreases. It may therefore be that the comb filtering notches for the 5.0 and 10.0 ms ICTDs are not of sufficient bandwidth to be interpreted as elevation cues. Additionally, the overall spectral envelope of these stimuli is close to that for the 0 ms ICTD and therefore the similarities in perceptual elevation seem plausible.

With respect to the results obtained for individual octave bands, the aforementioned hypothesis is partially inadequate. It can clearly be seen in Fig. 3 that, even though the effect of presentation method was significant the spread of results for 1000–4000 Hz does not follow the same pattern as for the broadband source. This being said, it is clear that comb filtering would have affected these stimuli and it remains arguable that in some way this would have influenced perceived elevation.

Additionally, despite the presence of comb filtering the 8000 Hz octave band was almost entirely unaffected by changes in presentation method. This might be related to the strength of the relationship between 8000 Hz and above localization as described by Blauert [8], although this would not explain the wide error bars for the 8000 Hz stimuli observed in the present test. It is clear spectral cues in relation to the localization of octave band stimuli requires further study.

3.3 Summing Localization and the Precedence Effect

An additional aim in the present study was to determine whether or not evidence could be found for the precedence effect in median plane stereophony. It initially appears that the precedence effect operates at low frequencies since the perceived image positions of the stereophonic sounds are similar to the physical position of the lower loudspeaker. However, it is important to note that the stimuli presented from the upper loudspeaker were localized at similar positions to those presented from the lower loudspeaker despite

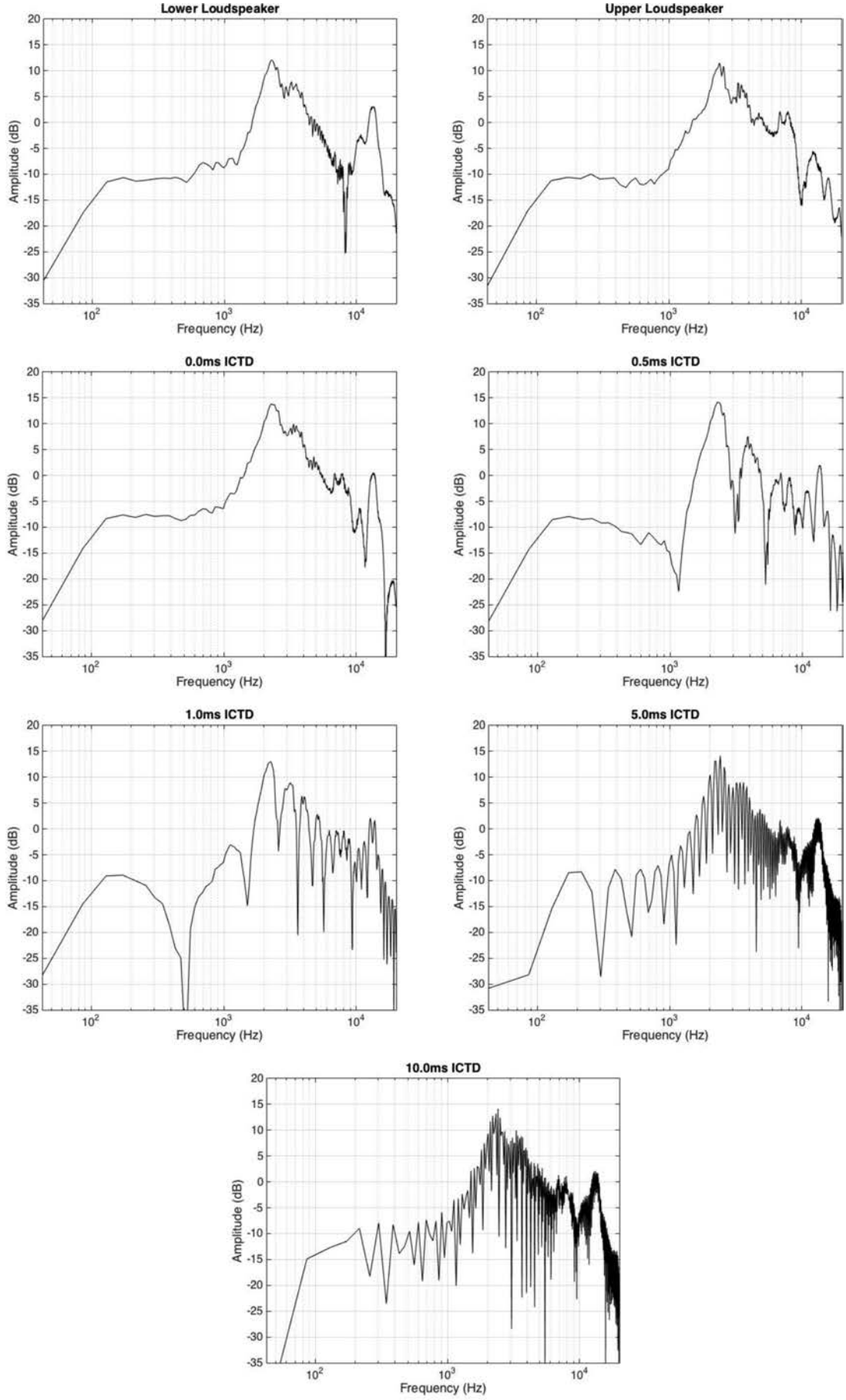

Fig. 4. HRTFs taken from the MIT Database for 0° and 30° elevation, with test ICTDs applied.

differences in physical height. Therefore, it seems logical to consider that the localization results for the low frequencies were due to the inherent pitch-height effect rather than ICTD or the precedence effect. Moreover, localization judgments for the broadband stimulus also showed no evidence of the precedence effect, with judgments for the 5 and 10 ms ICTD stimuli being in a position between the loudspeaker pair. The present experimental data therefore suggests that the precedence effect is not a feature of vertical stereophony, at least in the current experimental setup using the 30° elevation angle. It also seems unlikely that an increase in ICTD beyond 10 ms would result in precedence effect-like localization as Lee [18] tested musical sources with ICTDs up to 50 ms, using an identical test setup to that used here, and likewise found no evidence for the precedence effect. Additionally, the experimental data also appears to suggest that time panning would be ineffective in the median plane. For ICTDs below 1.1 ms, the broadband source was localized progressively closer to the position of the upper loudspeaker as the ICTD increased. For there to be effective time panning the opposite would have to be true.

3.4 Practical Implications

The results obtained in the present study for the broadband stimuli may have some useful practical implications regarding the use of ICTD for vertical panning. Between 0 and 1 ms it is clear that summing localization does not operate in the same way in the vertical domain as it does in the horizontal. This being said, the results identified a somewhat linear pattern for perceptual elevation as the ICTD increased. There is the potential that this result could form the basis of an ICTD-based vertical panning tool. Despite this suggestion, there remains the issue of tone coloration as a result of comb filtering. Any such tool would have to compensate for this in some way or another. There needs to be further research for ICTD-based vertical image panning, although the present results would at least suggest its potential use.

3.5 Future Works

The present study has considered the effects of ICTD on vertical localization primarily with respect to band-limited noise sources. A natural progression of this experiment would be to analyze the effect of ICTD on practical sources such as music and speech. Such an experiment would more clearly determine whether ICTD would be a useful perceptual parameter for vertical image rendering.

Additionally, the present study tested localization for a single loudspeaker positioned directly in front of the listener with an accompanying height channel (real center). It would be of interest to determine whether or not the present results would be maintained for phantom center images formed by stereophonic left and right loudspeakers with accompanying height channels. This would have further implications for 3D audio systems (e.g., Auro-3D [12]), which tend to make use of elevated front left and right loudspeakers.

Finally, Lee [18] conducted an experiment into the relationship between ICTD and ICLD in median plane stereophony. He did this by considering the amount of level reduction necessary in the upper loudspeaker for its influence to be totally masked (masked threshold) as well as for the resultant phantom image to be fully localized at the position of the lower loudspeaker (localization threshold). This could be expanded, making use of band limited noise sources as well as various musical sources. This would be directly applicable to 3D image rendering tools as well as having implications for the design of microphone arrays for 3D audio systems.

4 CONCLUSION

The present study investigated into the effect of ICTD on the vertical stereophonic localization of band-limited stimuli. The study utilized seven octave bands of pink noise, with center frequencies ranging from 125 to 8000 Hz, as well as a broadband pink noise source. Stimuli were presented either monophonically or as stereophonic phantom images, with the upper loudspeaker delayed with respect to the lower. The experiment used ICTDs of 0, 0.5, 1, 5, and 10 ms.

The experimental data obtained from the study showed that localization under the above conditions is governed by the pitch height effect. For the majority of presentation methods, the high frequency stimuli were localized in a significantly higher position than were the low frequency stimuli. Despite this, the relationship between pitch and height was found to be non-linear in all cases. Additionally, localization for the mid frequency stimuli was found to be somewhat erratic, this was likely related to the somewhat "forced" localization of the stimuli in front of the subject.

Localization judgments for the low frequency stimuli were consistent, irrelevant of how the stimuli were presented to subjects. However, as the frequency increased, judgments appeared to become increasingly affected by ICTD. This was found to be the case for the 1000–4000 Hz stimuli, as well as for the broadband source. Despite this, the effect was not consistent, with different frequencies being affected differently by the same ICTDs. It is arguable that the erratic effects of ICTD at these frequencies were the result of comb filtering distorting the spectral cues utilized in vertical localization, particularly for the 4000 Hz and broadband stimuli.

Additionally, no evidence could be found to support the operation of the precedence effect in median plane stereophony. In the present study the only occasions whereby stimuli were localized at the position of the earlier emitting loudspeaker were due to the pitch height effect. There was also no consistent effect of time panning observed, with localization judgments for the broadband source becoming more biased towards the upper loudspeaker as ICTD increased, as opposed to the lower.

5 ACKNOWLEDGMENT

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC), UK, Grant Ref. EP/L019906/1. The authors thank the music technology

students and staff at the University of Huddersfield who participated in the listening tests. They are also grateful to the editor and anonymous reviewers of this paper for their helpful comments.

6 REFERENCES

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

[2] 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 (1974 Dec.), http://dx.doi.org/10.1121/1.1903520.

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

[4] O. Trimble, "Localization of Sound in the Anterior-Posterior and Vertical Dimensions of 'Auditory' Space," *Brit. J. Psychol.*, vol. 24, no. 3, pp. 320–334 (1934 Jan.), http://dx.doi.org/10.1111/j.2044-8295.1934.tb00706.

[5] 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.

[6] S. K. Roffler and R. A. Buttler, "Factors That Influence the Localiaation of Sound in the Vertical Plane," *J. Acoust. Soc. Am.*, vol. 43, no. 6, pp. 1255–1259 (1968). http://dx.doi.org/10.1121/1.1910976.

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

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

[9] 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.

[10] M. Itoh, K. Iida and M. Morimoto, "Individual Differences in Directional Bands in Median Plane Localization," *Appl. Acoust.*, vol. 68, pp. 909–915 (2007 Aug.), http://dx.doi.org/10.1016/j.apacoust.2006.08.001.

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

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

[13] J. Barbour, "Elevation Perception: Phantom Images in the Vertical Hemisphere," presented at the *AES 24th International Conference on Multichannel Audio, The New Reality* (2003, June), conference paper 14.

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

[15] J. Blauert, *Spatial Hearing. The Psychophysics of Human Sound Localization* (MIT Press, Cambridge, MA. 1997).

[16] 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.

[17] 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 Apr.).

[18] 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.

[19] R. McGill, J. W. Tukey 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. Wallis and H. Lee, "Directional Bands Revisited," presented at the *138th Convention of the Audio Engineering Society* (2015 May), convention paper 9278.

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

THE AUTHORS

Rory Wallis

Hyunkook Lee

Rory Wallis is a Ph.D. student and member of the University of Huddersfield's Applied Psychoacoustic Lab (APL). Wallis 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 concern towards the effects of vertical interchannel crosstalk and methods of reducing them. Over the last few years Wallis has presented his research at both the 136th AES Convention in Berlin and the 138th in Warsaw, while also contributing work towards a conference paper presented at the 28th Tonmeistertagung in Cologne.

•

Hyunkook Lee 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. From 2006 to 2010, Dr. Lee was Senior Research Engineer in audio R&D at LG Electronics, South Korea. Currently, he is Senior Lecturer in music technology and the leader of the Applied Psychoacoustics Lab (APL) at the University of Huddersfield, UK. Dr. Lee has also been a freelance recording engineer since 2002. His current research includes spatial audio psychoacoustics, 3D sound recording and image rendering techniques, and interactive virtual acoustics. He is an active member of the Audio Engineering Society and a fellow of the Higher Education Academy, UK.