applied sciences

MDPI

Article

Localisation of Vertical Auditory Phantom Image with Band-limited Reductions of Vertical Interchannel Crosstalk

Rory Wallis and Hyunkook Lee *

Applied Psychoacoustics Lab (APL), University of Huddersfield, Huddersfield HD1 3DH, UK; rory.wallis@hud.ac.uk

* Correspondence: h.lee@hud.ac.uk; Tel.: +44-1484-471893

Received: 13 January 2020; Accepted: 19 February 2020; Published: 21 February 2020

Abstract: Direct sound that is captured by the upper layer of a three-dimensional (3D) microphone array is typically regarded as vertical interchannel crosstalk (VIC), since it tends to produce an undesired effect of the sound source image being elevated from the ear-level loudspeaker layer position (0°) in reproduction. The present study examined the effectiveness of band-limited VIC attenuation methods on preventing the vertical image shift problem. In a subjective experiment, five natural sound sources were presented as vertically-oriented phantom images while using two stereophonic loudspeaker pairs elevated at 0° and 30° in front of the listener. The upper layer signal (i.e., VIC) was attenuated in various octave-band-dependent conditions that were based on vertical localisation thresholds obtained from previous studies. The results showed that it was possible to achieve the goal of panning the phantom image at the same height as the image produced by the main loudspeaker layer by attenuating only a single octave band with the centre frequency of 4 kHz or 8 kHz or multiple bands at 1 kHz and above. This has a useful practical implication in 3D sound recording and mixing where a vertically oriented phantom image is rendered.

Keywords: vertical interchannel crosstalk; vertical auditory localisation; 3D audio; sound recording; microphone technique; mixing; upmixing; psychoacoustics

1. Introduction

1.1. Background

Audio reproduction systems for surround sound are currently in a state of evolution. Audio engineers and researchers are increasingly looking to improve on the spatial impression and realism that are 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 Dolby Atmos [1] and Auro-3D [2] becoming more widely utilized in film and music productions. Such developments inevitably have implications for the sound recording process, as additional height layers of microphones are required alongside the pre-existing main channel layer to capture the necessary spatial information.

In the context of three-dimensional (3D) sound recording using microphones, it is necessary to consider interchannel crosstalk oriented between vertically arranged microphones in terms of vertical phantom image localisation. Consider a 3D microphone array that consists of two vertical layers of microphones. For most classical music settings, the lower (main) layer would be typically used to mainly capture direct sounds for horizontal source imaging, whilst the upper (height) layer would be used to capture more reflections and reverberation to enhance perceived listener envelopment (LEV). Ideally, in this case the height layer signals should not interfere with the vertical localisation

of source images presented from the main loudspeaker layer. If there were excessive amounts of direct sounds present in the height microphone signals, however the vertically-oriented phantom images may be perceived at intermediate positions between the main and height loudspeakers layers in reproduction. The direct sound present in the height layer is referred to as vertical interchannel crosstalk (VIC) [3]. Conceptually, the phantom image localisation in the presence of VIC can be compared with the precedence effect between the lead and lag signals [4,5]. In horizontal stereophonic reproduction, it is well known that the precedence effect operates when an interchannel time difference (ICTD) is introduced between the loudspeaker signals; a time delay that is greater than around 1 ms would cause the resulting phantom image to be fully localised at the position of the earlier loudspeaker. If the same effects were also effective in the vertical stereophonic reproduction, VIC from a vertically spaced pair of omni-directional microphones would not cause such vertical localisation interference. However, a previous study by the present authors [6] showed that the precedence effect would not be triggered in vertical stereophony, with the influence of ICTD on vertical localisation being mainly dependent on comb-filtering of the ear-input signals. This necessitates the suppression of VIC in the height microphone signals.

From the above background, previous studies [3,7–9] subjectively measured the minimum amount of level reduction that is necessary in the upper loudspeaker for the resultant phantom image to be localised at the position of the sound presented from the lower loudspeaker only, which is commonly referred to the 'vertical localisation threshold' (VLT). The loudspeakers were arranged at 0° (listener's ear height) and 30° elevations in the median plane (i.e., vertical stereophonic setup). Various types of sound sources (cello and bongo [3], conga, quartet, speech, guitar, and oboe [9], broadband and octave-band noises [8], as well as a range of ICTDs, were tested in those studies. The results from the studies largely agreed, in that the VLT for natural sources were commonly around −7 dB for the ICTDs of 1 to 10 ms, and −9.5 dB for the ICTD of 0 ms [9]. In [8], a vertical quadraphonic loudspeaker setup was also tested (main layer loudspeakers at ±30° azimuth and 0° elevation, height layer loudspeakers at ±30° azimuth and 30° elevation, based on the Auro-3D format [2]), but this did not produce significantly different results from the vertical stereophonic setup. For broadband noise, the VLTs were reported to be around −12 dB [7,8], which is considerably lower than the results for the natural sources.

1.2. Aim of the Study

The abovementioned previous studies provide useful practical implications on 3D sound recording and mixing. In terms of 3D microphone array design, the height layer microphones should first be of a directional polar pattern (e.g., cardioid or supercardioid) and angled upwards, so that the level of VIC would be at least −7 dB, relative to the direct sound in the main layer microphones. This VLT can also be a useful reference when a sound engineer aims to vertically pan a phantom image to the main layer position in 3D sound mixing. It is also evident from [10] that a similar amount of ICLD would also cause a phantom image to be fully panned to the physical position of the height layer. In such methods, the amplitude of the signal in the height layer is reduced as a whole (i.e., even attenuation across the frequency spectrum). Henceforth, this method will be referred to as 'blanket VLT'.

However, in [7,8], VLT was also found to have a significant frequency dependency for octave-band noise stimuli, being associated with psychoacoustic phenomena, such as the 'pitch-height' effect [11–13] and the directional bands [14]. In general, the bands centred at 1 kHz and above generally having lower threshold than the lower bands (e.g., −7 dB for the 250 Hz band and −13.5 dB for the 8 kHz band). This formed the initial hypothesis of the current study that the VLT could be applied in a band-dependent manner, rather than as the blanket. The rational for is as follows. As Lee [3] found, the application of VLT does not totally mask the perceived effects of VIC, such as the increase in vertical image spread and tone colouration; further attenuation of VIC is required to achieve a total masking (i.e., VIC being completely inaudible). However, in practical 3D sound recording, mixing, and upmixing applications, such spatial and timbral effects may be found to be preferable, as found in the case of horizontal interchannel crosstalk [15]. Therefore, by applying the band-limited or single band VLT method,

the spatial and tonal benefits would be preserved, whilst ensuring the desired vertical image position. Furthermore, many professional recording engineers prefer using omni-directional microphones rather than directional ones for the height channels [16,17], even though they are more prone to the undesired vertical localisation shift due to the high level of VIC that is presented in the height channel signals. This seems to be mainly because of their fuller and more spacious characteristics of sound benefitted from the extended lower frequencies and the greater ICTD between direct sounds. In such a case, VIC could be reduced only for specific band(s) in post-processing to avoid the vertical localisation shift issue with minimal influence on the perceived spatial and tonal characteristics. This leads to the following research question for the present study: Is it possible to achieve the goal of avoiding the localisation interference effect of VIC by applying VLTs for a limited frequency range or a single band instead of the blanket VLT method?

The rest of the paper is organised, as follows. The design of the subjective experiment conducted is first described. Statistical analyses of the data obtained from the experiment are presented in Section 3. Following this, discussion on the results and their practical implications are provided in Section 4. Finally, Section 5 concludes the paper.

2. Methods

2.1. Physical Setup

The physical setup of the experiment was identical to the author's previous VLT study while using natural sources [9]. The experiment was conducted in an ITU-R BS.1116 [18]- compliant critical listening room (6.2 × 5.2 × 3.5 m; RT = 0.25 s; NR = 12) at the Applied Psychoacoustics Lab of the University of Huddersfield. Figure 1 shows the loudspeaker configuration for the experiment. The loudspeakers were arranged in a vertical quadraphonic configuration that was based on the front part of the Auro-3D 9.1 loudspeaker format [2] (without the centre channel). The L and R loudspeakers, which had the horizontal base angle of 60°, were raised at the listener's ear height (0° elevation). The HL and HR loudspeakers in the height layer were placed right above the L and R loudspeakers and elevated at 30° from the listener position. The loudspeaker positions also comply with recommendations in ITU-R BS.2152 [19]. The distance between the listening position and each of these loudspeakers was 2 m. The direct-to-reverberant (D/R) energy ratios were 7.7 dB and 6.5 dB for the L and HL loudspeaker, respectively, when measured while using an omni-directional microphone at the listening position. Furthermore, within 15 ms after the direct sound, there was no reflection that exceeded −14 dB when compared to the direct sound level, which meets the requirement of ITU-R BS. 1116. Therefore, it can be considered that the influence of reflections on the perception of VIC would be minimal.

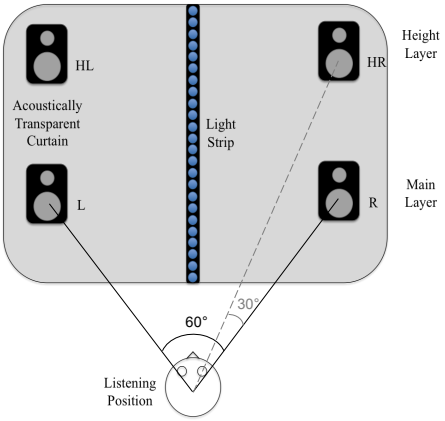

Figure 1. Loudspeaker setup used for the listening test. The main layer loudspeakers are configured with the standard 60° subtended angle from the listening position. The distance between the listener and each of the main layer speakers was 2 m. The height layer loudspeakers are placed right above the main layer ones, with 30° elevation from the listening position, based on the Auro-3D configuration [2].

A light-emitting diode (LED) strip was directly positioned in front of the listening position. This was located behind the acoustically transparent curtain and it was to be used by subjects when making their localisation judgments.

2.2. Test Stimuli

The experiment used five natural sound sources of conga, oboe, quartet, speech, and guitar, all of which were anechoic recording excerpts that were taken from [20]. They were chosen due to their varied spectral and temporal characteristics, as shown in Figure 2.

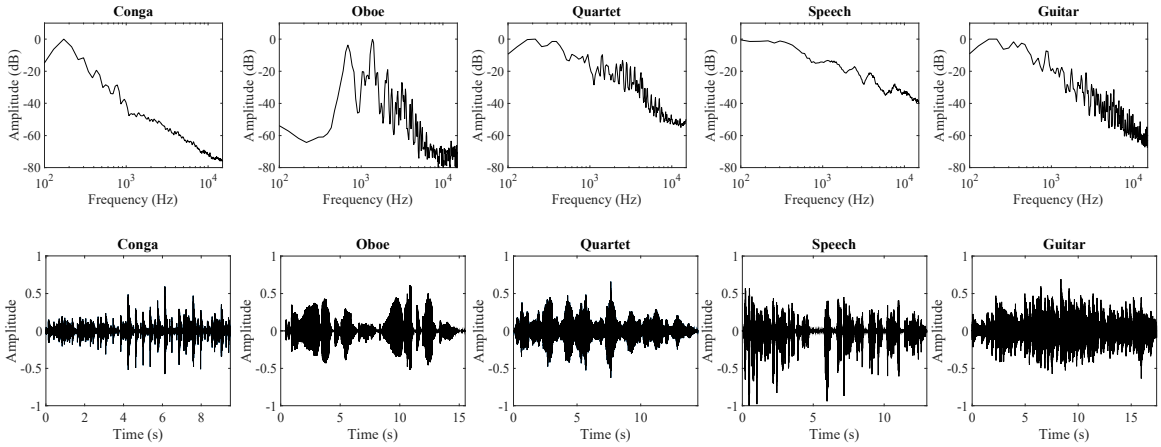

Figure 2. Long term average spectra and waveforms of test stimuli used for Experiment One.

The following eight experimental conditions were applied to the sources to create test stimuli. For the VLT cases, the attenuation was only applied to the direct sound in the height layer, with the ICTDs of 0 and 1 ms.

1. Full band VLT (FB): octave-band-dependent attenuation is applied to the height layer.
2. VLTs for 1 kHz band and above (1 + B): octave-band-dependent attenuation is only applied to the 1 kHz band and above.
3. 4 kHz band VLT (4B): only the band centred at 4 kHz is attenuated.
4. 8 kHz band VLT (8B): only the band centred at 8 kHz is attenuated.
5. Blanket VLT: attenuation is applied to the height layer as a whole.
6. Main layer only: only the main layer is presented.
7. Height layer only: only the height layer is presented.
8. Vertically oriented phantom image with 0 dB ICLD: no level attenuation was applied, thus creating a vertical phantom centre image between the two loudspeaker layers.

The rationales for the 4B and 8B conditions are as follows. Previous work by the present authors [9] suggest that there might be some perceptual dominance of the 7–9 kHz region in determining the VLTs for complex sources. Furthermore, in the literature it has been reported that the primary cues for elevation perception lie in the 4–10 kHz range [21,22]. More specifically, 8 kHz is related to 'above' perception, according to Blauert's directional band theory [14], and therefore it is hypothesised that the band that is centred at this frequency would have a major influence on the perception of image elevation. The 4 kHz band is mapped to 'front' localisation by Blauert. However, in a phantom image condition, which is the case in the present study, Lee [23] showed that, with a pair of loudspeakers elevated at 30° from the listener's ear position, a 4 kHz octave band noise was perceived to be elevated at the same height as the loudspeakers, due to the phantom image elevation effect [24,25]. Therefore, by reducing the energy of this band in the height layer, the perceived elevation of a phantom image that results from the vertical quadraphonic reproduction might be reduced. The 1 + B condition was created

while considering a potential benefit of maintaining the original low frequency energy in the height layer signals captured using omni-directional microphones, whilst reducing all of the high frequencies that are potentially problematic for accurate localisation. In addition, the main and height layer only conditions as well as the blanket and 0 dB ICLD conditions were included to serve as references.

Table 1 presents a summary of the VLTs that were used for each band in the current study This is based on the results from a previous study by the authors [8,9]. For band attenuation, the VLTs that were applied to the speech, oboe, guitar, and quartet sources were the median results obtained for the continuous octave-band noise stimuli in [8]. On the other hand, band reduction for the conga used the VLT obtained for the octave-band noise burst stimuli used in the same study due to its transient nature. It should be noted that, in [8], the thresholds for both the continuous 8 kHz band and the burst 4 kHz band had a significant dependency on the ICTDs, whereas the other octave bands did not, regardless of the transient nature of the sound source. For this reason, in the current study, different thresholds were applied to those two bands, depending on the ICTD and source type. For blanket reduction, the median thresholds that were derived in [9] were used (i.e., −9.5 dB for 0 ms and −7 dB for 1 ms). The 0 ms condition is equivalent to a vertical coincident microphone configuration [26], whereas the 1 ms condition represents the spacing between the main and height microphones being 0.34 m. This would be a practical spacing for a 3D microphone array, since, beyond this spacing, there would be no perceived increase in vertical image spread and the tonal quality would decrease [27]. In addition, the author's previous study on the VLT while using the same sound sources also tested the ICTD of 10 ms, but there was no significant difference found between 1 ms and 10 ms in the results.

Table 1. Vertical localisation threshold (VLT) values applied to different octave bands of the stimuli, based on Wallis and Lee [8].

Centre Frequency (Hz)	Height Layer Attenuation (dB)	
	Guitar, Speech, Oboe and Quartet	Conga
63	−6	−6
125	−7	−6
250	−7	−5
500	−8.5	−6
1 k	−11	−6
2 k	−10	−6
4 k (0 ms)	−9.5	−11
4 k (1 ms)	−9.5	−6
8 k (0 ms)	−13.5	−6
8 k (1 ms)	−10	−6
16 k	−8	−6

The test stimuli were created for each sound source, as follows. For the height channel signal, each source was first broken down into octave bands while using a linear-phase 8th-order Butterworth filter (−48 dB/oct), which was created using the 'forward-backward' filtering method [28] in MATLAB (MathWorks, Natick, Massachusetts, USA). Each octave band then underwent the level reduction according to Table 1. The main channel signal was unaltered. Time delays of 0 and 1 ms were applied to the height channel with respect to the main. During the test, the main channel was routed to L and R, whilst the right was routed HL and HR. For blanket reduction, the amplitude of the height channel was reduced as a whole. From this, a total of 80 stimuli were created (five sources x two ICTDs x eight presentation methods).

The playback levels of the main layers of all stimuli were calibrated at 70 dB LAeq at the listening position. The increase in amplitude when the stimuli were presented as vertically arranged quadraphonic phantom images was dependent on the VLT that was applied to the height layer, as shown in Table 2. It was decided not to match the overall sound pressure level (SPL) of the stimuli as a slight increase in SPL is an inherent result of VIC.

Table 2. A-weighted average sound pressure level (LAeq) for stimuli presented while using each localisation threshold method.

Method	LAeq (dB)
FB	70.0
1 + B	71.2
4B	72.2
8B	72.4
Blanket	70.4
0 dB ICLD	73.1

2.3. Subjects

Ten subjects participated in the listening test. They comprised staff researchers, post-graduate, and final year undergraduate students from the University of Huddersfield's Music Technology (ages that ranged between 21 and 40; Eight males and two females). They were chosen due to their extensive experiences in performing auditory localisation tasks in critical listening test environments. They all reported normal hearing.

2.4. Test Protocol

Localisation judgments were made while using the LED strip located directly in front of the listening position. For each test, the subjects were provided with a handheld rotational knob, which controlled which LED on the strip was turned on. They were required to adjust the knob until the position of the active LED matched the perceived location of the focal point of each stimulus. This method was chosen following research that was conducted by Lee et al. [29], who found that it was faster and produced results with greater accuracy and consistency compared to the numbered scale method, which had been used previously in [9]. The position of the LED selected for each stimulus was converted into an elevation angle within a Max/MSP patch. The heads of subjects were not fixed, but they were instructed to sit up and face forwards at all times, while only using their eyes to look at the LED strip. A small headrest was positioned behind the head of each subject to help maintain the correct seating position. The presentation order of stimuli was randomised for each test.

3. 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 Shapiro–Wilk test showed that not all scores in each condition featured normal distribution, although the results of the Levene test showed homogeneity of variance for all sound sources. For these reasons, non-parametric tests were chosen for the statistical analysis. Wilcoxon signed-rank tests were used to examine the significance of difference between each presentation method for each source and for each ICTD since the experimental design was of a within-subject repeated measure (i.e., all subjects tested all conditions in a randomised order). Boferroni correction was applied to the p values that were obtained from the Wilcoxon tests to reduce a potential Type-I error in multiple comparison.

Figure 3 shows the median perceived elevation of each of the test stimuli, plotted with notch edges (i.e., non-parametric 95% confidence intervals). With respect to the VLTs, it can be observed that each of the proposed band reduction methods, as well as the blanket threshold method, resulted in the perceived location of each stimulus being similar to that for the main layer only condition, suggesting that the proposed methods tend to fulfil their aim. This was the case for all sources, with the median difference in perceived elevation between the main layer only and the VLT conditions ranging between 2.4° and −4.0° for the 0 ms ICTD and between 3.9° and −3.2° for the 1 ms ICTD. Alongside this, the notch edges of the main layer only and VLT conditions all overlap, which suggests no significant differences among the conditions [30].

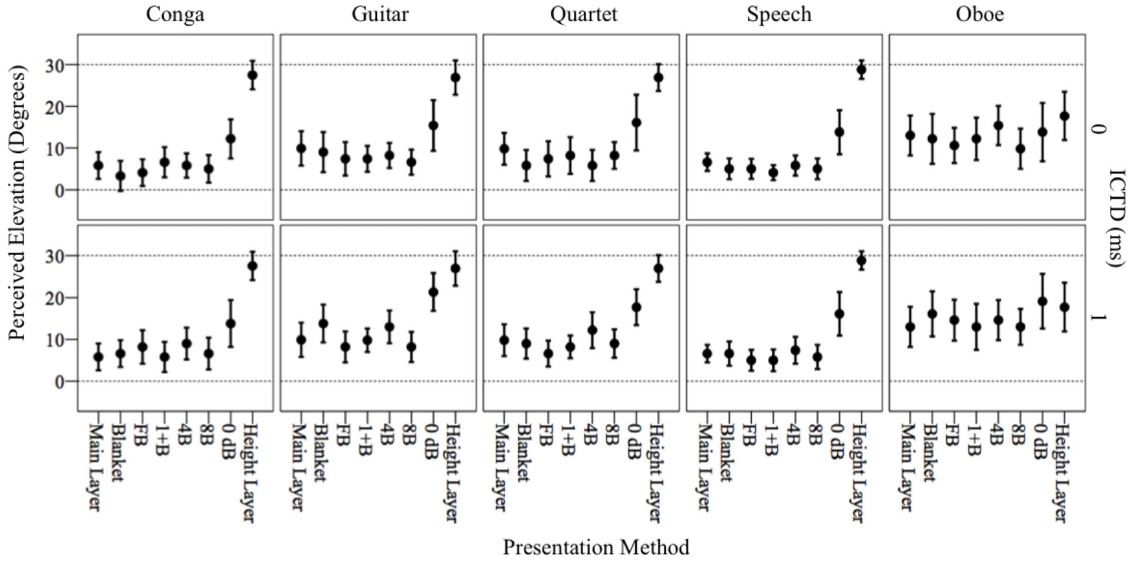

Figure 3. Median perceived elevation and associated notch edges (i.e., non-parametric 95% confidence interval) for each stimulus in the verification experiment. The dotted lines at 0° and 30° represent the positions of the main and height layers, respectively.

The Bonferroni-corrected Wilcoxon tests indicated that, for all sources and ICTDs, none of the VLT methods had a significant difference to the main layer only condition. In addition, the effect size r did not indicate a large effect in any case (<0.5 for all). From this, it can be suggested that all of the proposed VLT methods successfully prevented vertical interchannel crosstalk from affecting the perceived location of the main channel signal.

The medians for the 0 dB ICLD condition were generally higher than those for the main layer only condition ($p < 0.05$), which was expected. However, it is interesting to note that the oboe did not follow this trend; there was no significant difference among all of the conditions ($p > 0.05$). This result might suggest that the application of VLTs would not always be necessary and would have somewhat of a source dependency.

A further result of note can be seen with respect to the main and height layer only conditions. Firstly, for the latter condition, it would appear that perceived elevation judgments were generally accurate for all sources, excluding the oboe, with respect to the physical position of the height layer. Conversely, the judgments were less accurate for the main layer only condition, with perceived source elevation being in the range of 5.8°–13.0° with respect to the main layer's physical position. This elevation of the sound source with respect to the main layer was also maintained for the conditions, whereby a VLT was applied to the height layer. The results of a Wilcoxon signed rank test, which compared the results for the main layer only condition to the physical position of the main layer (0°), showed that each source was perceived to be significantly higher than the physical height from which the source was presented ($p = 0.000$ for all sources).

4. Discussion

4.1. The Effecitveness of the Band-limited VLT Methods

The data that were provided in the present experiment have shown that a shift in the perceived elevation of musical sources is apparent when a sufficient amount of direct sound is present in the height layer. In addition, using any of the band or blanket reduction methods that were tested can prevent this from happening. That both the blanket reduction and FB methods were successful was somewhat expected. For example, the blanket reduction thresholds had already been derived in a previous study [9] and so were reasonably expected to work. Additionally, it can be deduced from Table 1 that the influence of the height channel on the resultant amplitude of each source for the FB

method was low, with the average peak amplitude being equal to that for the main layer only condition. Therefore, it seems reasonable that the perceived location of the main channel signal would not be affected, as the audibility of the direct sound in the height layer was low with respect to that in the main layer. It was the main focus of the current study to investigate whether vertical localisation thresholds (VLTs) could be applied through the selective manipulation of frequency bands within the height layer, with the 1+B, 4B and 8B conditions all being effective.

Consideration was first given to the discussions regarding the mechanisms that might determine the VLTs for complex sources in order to explain the effectiveness of the 1 + B, 4B and 8B methods. The difference in ear-input spectral energy of the height layer to the main layers the vertical quadraphonic condition was measured while using HRIRs that were obtained from the KEMAR dummy head database in order to gain further objective insights into this result [31]. The result is shown in Figure 4, in the range of frequencies covered by the 4 and 8 kHz octave bands (2840–11360 Hz). Any point where the difference is above 0 dB represents the dominance of the height layer over the main layer, and vice versa. From the figure, it can be observed that the height layer has considerably more energy in the 7–9 kHz region when compared to the main layer. Given that 8 kHz 1/3-octave band has an association with above localisation, according to Blauert's directional band theory [14], it is reasonable to suggest that the large energy dominance in the 7–9 kHz region would result in the vertical phantom image condition being elevated with respect to main layer only presentation. By extension, this would also mean that providing sufficient attenuation of the phantom image in this range would result in the VLT being met, which seems to explain why the 8B and 1 + B methods were effective. In either case, the methods would necessitate the attenuation of the 8 kHz octave band by a minimum of 6 dB for transient sources and 10–13.5 dB for continuous sources, based on the listening test results (depending on the ICTD, see Table 1).

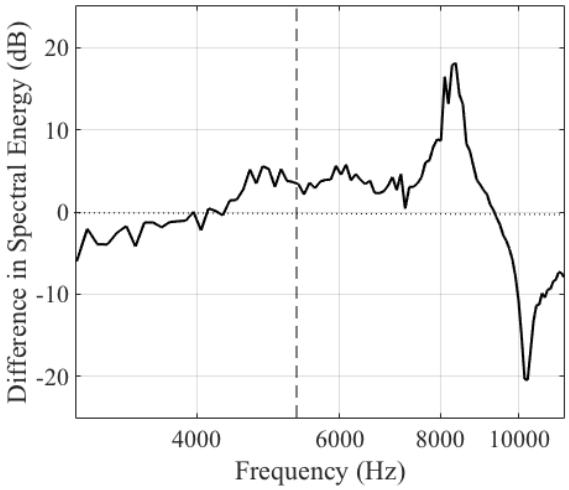

Figure 4. Difference in ear-input spectral energy of the height layer to the main layer of a vertical quadraphonic configuration. The vertical dashed line separates the frequency range for the 4 kHz octave band (left) from that for the 8 kHz octave band (right).

With regard to the 4B method, it can be seen in Figure 4 that the difference in spectral energy between the main and height layers is much smaller for the 4 kHz octave band as compared to the 8 kHz octave band. However, as mentioned in Section 2.2, in [24] the phantom centre image of the 4 kHz octave band was found to be effectively elevated to the physical height of the height loudspeaker layer. This seems to suggest that the attenuation of this band that was tested in the present study was sufficient for suppressing the vertical localisation shift.

4.2. The Relationship Between Perceived Source Elevation and the Vertical Localisation Threshold

The results showed that the difference in perceived elevation between the main layer only and phantom image conditions differed for each source. For example, the speech source that was presented as a vertically oriented phantom image with 0 dB ICLD was perceived as being significantly higher than that for main layer only presentation for both of the ICTDs tested. On the other hand, judgments for the oboe source were generally consistent, irrelevant of how the source was presented to subjects. The remaining sources were more inconsistently affected. The increases in median perceived elevation for the conga, guitar and quartet sources were significant when the ICTD was 1 ms. Despite this, for the 0 ms ICTD the difference was not significant, even though the median perceived elevation notably increased in each case. These results would seemingly indicate that vertical interchannel crosstalk has a source dependent effect with respect to the perceived migration of the main channel signal from the position of the main layer.

With respect to the oboe source, an explanation of the results obtained is thus offered. The literature suggests that narrowband stimuli incident from the median plane are localised on the basis of frequency, with increases in frequency corresponding to increases in perceived elevation (i.e., pitch-height effect) [11–13]. The spectrum for the oboe source was notably narrow, with its predominant energy focused around 1 to 2 kHz, as can be seen from Figure 2. According both to the literature and to the results presented in the present study, band-limited stimuli in this frequency range are localised at a similar vertical position, regardless of which loudspeaker layer presented the source, for both vertical stereophonic [13] and vertical quadraphonic [24] loudspeaker arrangements. Therefore, it might be that the pitch-height effect determined localisation judgments for the oboe, with no relation to the difference in energy in the 7–9 kHz region between the phantom image and main layer only conditions. It might be further suggested that the suppression of vertical interchannel crosstalk is predominantly needed for broadband sources, with less relevance to the sources that are both narrowband and absent in high frequency energy.

4.3. Practical Implications

The results of the present experiment have implications both for microphone techniques for 3D sound recording and for the rendering of 3D images in mixing or upmixing. It was discussed that attributes, such as the tonal colour and spaciousness of the main channel signal, would be affected when the VLT is applied. It is apparent that the perception of such attributes would vary, depending on the VLT method being used. For example, given that the 1 + B, 8B, and 4B methods do not require any attenuation of the low frequencies in the height layer signal, it could be the case that the resultant phantom image would sound fuller compared to the blanket reduction method. In addition, given that Furuya et al. [32] reported that the perception of vertical image spread (VIS) was related to the amplitude of a vertical reflection relative to the direct sound, it could be that the degree of VIS afforded by each method would differ.

Furthermore, as mentioned in Section 1.2, the proposed methods could be used for the post-processing of height layer signals captured using omni-directional microphones. That is, the band-specific VIC reduction could help avoid the potential issue of vertical image shift while maintaining the spatial and tonal benefits of omni-directional microphones.

5. Conclusions

The aim of the study was to investigate the effectiveness of the frequency-band-limited attenuation of vertical interchannel crosstalk (VIC) (i.e., direct sound presented in the height loudspeaker layer) on vertical phantom image localisation in the context of quadraphonic sound reproduction with height channels. The main findings of the study are as follows.

- It is possible to successfully prevent the localisation interference effect of VIC by attenuating only some specific octave bands of the VIC, each by the amount specified Table 1; for example, (i) all of

the bands centred at 1 kHz and above, (ii) only the band centred at 4 kHz only, and (iii) only the band centred at 8 kHz.

- The influence of VIC on vertical localisation has a sound source dependency due to frequency-related psychoacoustic phenomena, such as the pitch-height effect and the phantom image elevation effect. The results suggest that, in vertical quadraphonic reproduction, sound sources with a narrow frequency spectrum (e.g., oboe) would not have a significant VIC influence on perceived vertical image position, since the image would be localised at a vertical position inherent from the frequency content, regardless of the presentation method.
- In addition, the results support a previous finding that neither the precedence effect nor localisation dominance operates vertically. That is, in the vertical domain, time delay between two loudspeakers alone would not cause the auditory image to be localised at the position of the earlier loudspeaker, and a certain level reduction would be necessary for this.

The band-limited VIC attenuation methods might produce a greater magnitude of perceptual effects, such as increases in vertical image spread and fullness when compared to the blanket method (i.e., attenuation as a whole). Their practical application areas include vertical sound image rendering in 3D sound mixing and post-processing of 3D sound recording.

Future works will investigate (i) what the most salient effects of vertical interchannel crosstalk are, (ii) how these vary when the different VLT methods are applied, and (iii) which method is most preferred subjectively.

Author Contributions: R.W. conducted the experiment, analysed the data and wrote majority of the paper. H.L. supervised the project, contributed to the data analysis and co-wrote the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Engineering and Physical Sciences Research Council (EPSRC), UK, Grant Ref. EP/L019906/1.

Acknowledgments: The authors would like to thank all subjects who participated in the listening test of this study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Dolby Atmos. Available online: http://www.dolby.com/us/en/brands/dolby-atmos.html (accessed on 4 January 2020).
2. Listening Formats: Auro 3D. Available online: http://www.auro-3d.com/system/listening-formats (accessed on 4 January 2020).
3. Lee, H. The Relationship between Interchannel Time and Level Differences in Vertical Sound Localisation and Masking. In Proceedings of the Audio Engineering Society 131st Convention, New York, NY, USA, 20–23 October 2011. Preprint 8556.
4. Litovsky, R.Y.; Coulburn, S.H.; Yost, W.A.; Guzman, S.J. The precedence Effect. *J. Acoust. Soc. Am.* **1999**, *106*, 1633–1654. [CrossRef] [PubMed]
5. Blauert, J. *Spatial Hearing: The Psychophysics of Human Sound Localisation*; MIT Press: Cambridge, UK, 1997.
6. Wallis, R.; Lee, H. The Effect of Interchannel Time Difference on Localisation in Vertical Stereophony. *J. Audio. Eng. Soc.* **2015**, *63*, 767–776. [CrossRef]
7. Wallis, R.; Lee, H. Vertical Stereophonic Localisation in the Presence of Interchannel Crosstalk: The Analysis of Frequency-Dependent Localisation Thresholds. *J. Audio. Eng. Soc.* **2016**, *64*, 762–770. [CrossRef]
8. Wallis, R.; Lee, H. The Frequency Dependency of Localisation Thresholds in the Presence of Reflections. In Proceedings of the 29th Tonmeistertagung, Cologne, Germany, 17–20 November 2016; Verbond Deutscher Tonmeister: Cologne, Germany, 2016.
9. Wallis, R.; Lee, H. The Reduction of Vertical Interchannel Crosstalk: The Analysis of Localisation Thresholds for Natural Sound Sources. *Appl. Sci.* **2017**, *7*, 278. [CrossRef]
10. Mironovs, M.; Lee, H. The Influence of Source Spectrum and Loudspeaker Azimuth on Vertical Amplitude Panning. In Proceedings of the Audio Engineering Society 142nd Convention, Berlin, Germany, 20–23 May 2017. Preprint 9782.

11. Pratt, C.C. The Spatial Character of High and Low Tones. *J. Exp. Psychol.* **1930**, *13*, 278–285. [CrossRef]
12. Roffler, S.K.; Butler, R.A. Factors that Influence the Localisation of Sound in the Vertical Plane. *J. Acoust. Soc. Am.* **1968**, *43*, 1255–1259. [CrossRef] [PubMed]
13. Cabrera, D.; Tiley, S. Vertical Localisation and Image Size Effects in Loudspeaker Reproduction. In Proceedings of the AES 24th International Conference on Multichannel Audio, Banff, AB, Canada, 26–28 June 2003.
14. Blauert, J. Sound Localisation in the Median Plane. *Acta Acust. United Acust.* **1969**, *22*, 205–213.
15. Lee, H. Effects of Interchannel Crosstalk in Multichannel Microphone Technique. PhD Thesis, University of Surrey, Gilford, UK, 2006.
16. Lindberg, M. 3D Recording with the "2L Cube". Available online: http://www.2l.no/artikler/2L-VDT.pdf (accessed on 4 January 2020).
17. AMBEO for Loudspeakers. Available online: https://en-uk.sennheiser.com/ambeo-blueprints-loudspeakers (accessed on 4 January 2020).
18. International Telecommunication Union. *Recommendation ITU-R BS.1116-1: Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems*; International Telecommunications Union: Geneva, Switzerland, 1994.
19. International Telecommunication Union. *Advanced Sound System for Programme Production*; International Telecommunications Union: Geneva, Switzerland, 2018.
20. Hansen, V.; Munch, G. Making Recordings for Simulation Tests in the Archimedes Project. *J. Audio. Eng. Soc.* **1991**, *39*, 768–774.
21. Shaw, E.A.G.; Teranishi, R. Sound Pressure Generated in an External-Ear Replica and Real Human Ears by a Nearby Point Source. *J. Acoust. Soc. Am.* **1968**, *44*, 240–249. [CrossRef] [PubMed]
22. Hebrank, J.; Wright, D. Spectral Cues used in the Localisation of Sound Sources on the Median Plane. *J. Acoust. Soc. Am.* **1974**, *56*, 1829–1834. [CrossRef] [PubMed]
23. Lee, H. Perceptual Band Allocation (PBA) for the Rendering of Vertical Image Spread with a Vertical 2D Loudspeaker Array. *J. Audio Eng. Soc.* **2016**, *64*, 1003–1013. [CrossRef]
24. De Boer, K. A Remarkable Phenomenon with Stereophonic Sound Reproduction. *Philips Tech. Rev.* **1947**, *9*, 8–13.
25. Lee, H. Sound Source and Loudspeaker Base Angle Dependency of Phantom Image Elevation Effect. *J. Audio Eng. Soc.* **2018**, *65*, 733–748. [CrossRef]
26. Lee, H.; Gribben, C. Effect of Vertical Microphone Layer Spacing for a 3D Microphone Array. *J. Audio Eng. Soc.* **2014**, *62*, 870–884. [CrossRef]
27. Gribben, C.; Lee, H. The Perception of Band-Limited Decorrelation Between Vertically Oriented Loudspeakers. *IEEE-ACM Trans. Audio Speech Lang. Process.* **2020**. [CrossRef]
28. Forward-Backward Filtering. Available online: https://ccrma.stanford.edu/~{}jos/fp/Forward_Backward_Filtering.html (accessed on 11 February 2020).
29. Lee, H.; Johnson, D.; Mironovs, M. A New Response Method for Auditory Localisation and Spread Tests. In Proceedings of the Audio Engineering Society 140th Convention, Paris, France, 4–7 June 2016.
30. McGill, R.; Tukey, J.W.; Larsen, W.A. Variations of Box Plots. *Am. Stat.* **1978**, *32*, 12–16.
31. Gardener, B.; Martin, K. HRTF Measurements of a KEMAR Dummy-Head Microphone. 2000. Available online: http://sound.media.mit.edu/resources/KEMAR.html (accessed on 4 January 2020).
32. Furuya, H.; Fujimoto, K.; Ji, C.Y.; Higa, N. Arrival Direction of Late Sound and Listener Envelopment. *Appl. Acoust.* **2001**, *62*, 125–136. [CrossRef]

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).