Vertical Amplitude Panning for Various Types of Sound Sources

MAKSIMS MIRONOVS AND HYUNKOOK LEE

Applied Psychoacoustics Lab, University of Huddersfield, Huddersfield, UK
e-mail: m.mironovs@hud.ac.uk, h.lee@hud.ac.uk

September 23rd 2016

Abstract

It is widely known that the vertical localisation of a real source relies on spectral cues. In order to examine the influence of source spectrum on the accuracy of vertical amplitude panning, the present study conducted subjective localisation tests using six sound sources with different spectral characteristics: broadband, low-passed and high-passed pink noises as well as speech, bird and tank shot recordings. Results generally indicated that the localisation accuracy of vertical amplitude panning was poor regardless of the source type. However, lower or upper response biases observed in the results were found to be significantly dependent on the target panning angle and the type of sound source. The bird and tank shot sources tended to have upper biases regardless of the target angle. The so-called 'pitch-height' effect was observed for low-passed and high-passed noise sources, but this was not consistent with the target angle. Overall, the results suggest that the localisation of elevated phantom source is significantly frequency-dependant.

1 Introduction

Localisation process of the real sound source in the vertical plane differs from that in the horizontal plane. On the horizontal plane, perceived position heavily relies on the combination of interaural time differences (ITD) and level differences (ILD) detected by listener's each ear (binaural cues), as well as sound source directional filtering, caused by pinnae (spectral cues) [1]. In terms of the elevated source localisation, ITD and ILD cues become ambiguous and in the case of the median plane absent, as the sound source is equidistant and arrives at both ears simultaneously, leaving pinna filtering as the main localisation cue [3].

Vertical localisation has received a lot of attention in the literature, however, median plane has been prioritised the most. Additionally, an extensive amount of research has been done using only real sources, i.e., the stimuli were presented using only one loudspeaker to produce the required target location [4, 5, 6]. However, as spatial audio systems incorporate elevated loudspeakers [7, 8], it is necessary to study the localisation of phantom sources produced by the loudspeaker pairs, i.e., vertical stereophony.

Moreover, to support the interactive audio systems, the vertical stereophony must produce high localisation accuracy. As an example of that would be audio systems, where the user controls the perceived position of sound using a 3D graphical user interface or a motion tracker. User experience will suffer, if panning algorithm will not produce consistent and accurate results.

Despite being the most commonly used panning method in audio systems, amplitude panning is based only on the level differences, resulting a poor localisation performance. Barbour [9] showed this in his research showing that vertical phantom image localisation was unstable and had a large inter-subject variability. However, only pink noise and male speech were used in this study and further research on the effect of stimuli's frequency spectra was suggested.

From the above background, the present study investigates the influence of the spectral characteristics of sound source on the localisation accuracy of vertical amplitude panning using various sound sources. Furthermore, additional vertical loudspeaker pair located 30° of the centre is used to provide more data since only the median and frontal planes were examined in the previous study. In the next section, the details of the subjective experiment conducted are presented, followed by the statistical analysis of data collected from the experiment. Then, results are discussed, focusing on the evaluation of the amplitude panning method in the spatial audio systems.

2 Experiment

2.1 Physical Setup

The listening test was conducted in the critical listening room at the University of Huddersfield, compliant with the ITU-R BS.1116-2 international standard for listening rooms [10]. The room's dimensions were 6.2m x 5.6m x 3.8m and the reverberation time (RT60) comprised of 0.25s. The test involved the usage of 4 Genelec 8040A loudspeakers, divided into two vertical pairs with 0° and 30° elevation and arranged horizontally at 30° interval, as shown in Figure 1. The distance between the listener's position and the loudspeakers on the horizontal plane was 2m. It was assumed that 0° elevation was at the loudspeaker's centre point, located 1.29m from the floor. Height layer was located 2.44m from the floor. In order to reproduce phantom images, tangent panning law was used, which is the basis of the VBAP algorithm [11]. Additionally, as the tangent law works only for equidistant loudspeaker arrangements, time and level alignments were necessary. 1.5dB level attenuation and 0.82ms delay were applied to the lower loudspeaker in order to ensure that loudspeakers are equidistant. Lastly, in order to eliminate any visual bias, all loudspeakers were hidden from the subjects by using acoustically transparent curtains around the listener, combined with low lighting conditions.

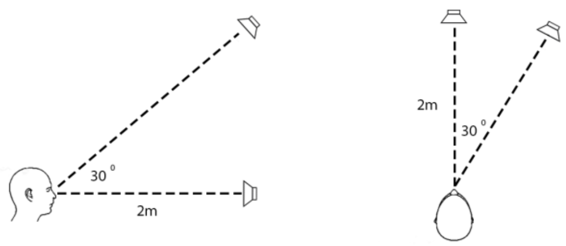

Figure 1: *Loudspeaker configuration used for the listening test.*

2.2 Test stimuli

The experiment involved three real-life, as well as three generated noise sources. The natural sound sources included recordings of the birds and tank shots taken from the BBC Sound Effects Library [12]. These stimuli were chosen due to their varying spectral and temporal content, as well the extensive use in multichannel audio panning. Obtained data provided information on the accuracy of panning, which can be useful for the spatial audio production for films and other media. Additionally, a dry solo vocal track from the Free Vocals recording studio [13] was used as a vocal sample. Although the voice is mainly reproduced using the centre loudspeaker and is panned only on rare occasions, this sample was used to see if there was any psychological bias, as it is unusual to hear elevated voices. Furthermore, noise sources were used to inspect the effect of frequency on perceived elevation in a more controlled fashion. They consisted of broadband, low-pass and high-pass filtered pink noise created using a 16th order Butterworth filter at 3 kHz cut-off frequency. All three noises were 1.4s long with 0.25s linear fade in and fade out, repeated with the interval of 0.7s. This envelope was recreated, as in experiment by Pulkki and Karjalainen [11]. Furthermore, the playback level of each stimulus was adjusted to the average of 75dB (A) at the listening position and all sound sources were recorded at the sampling frequency of 44.1 kHz and 16-bit resolution.

2.3 Subject

The listening test participants comprised a mix of 14 male and female undergraduate and post-graduate students, as well as studio technicians and academic staff from Music Technology courses at the University of Huddersfield. All subjects reported normal hearing and had an extensive experience in critical listening with many previously participating in tests involving the vertical localisation.

2.4 Test procedure

The listening tests were conducted using a novel method for auditory localisation and spread tests [14]. It used a flexible strip containing 160 multicolour light emitting diodes (LED), arranged with the 3cm intervals. The strip was vertically positioned behind the curtains, 2m from the listening position less than 1° difference between each LED. Arduino Micro microcontroller was used to provide communication between the strip and computer. Then, the LED system was controlled using a specifically designed Max 7 patch, providing control over LED position and colour, corresponding to the perceived elevation.

Amplitude panning in 5° steps was applied, resulting in seven amplitude-panned stimuli. The 0° and 30° sources were real, produced only from the bottom and top loudspeakers. Due to the limitation of the project, both vertical loudspeaker pairs were tested separately, resulting in the total of 84 trials (7 x 6 x 2). As all stimuli were tested only once, the consistency of the effect was not investigated. In addition, all stimuli were

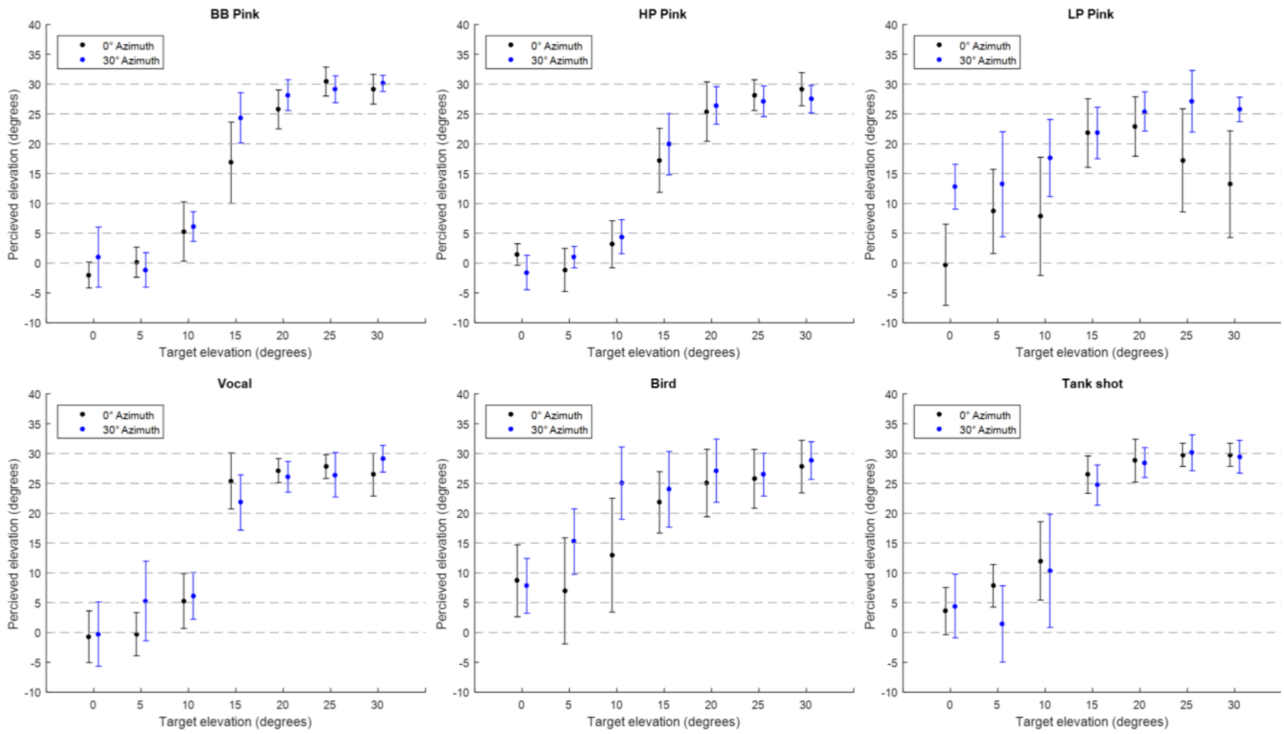

Figure 2: *Subject responses obtained for each source. The data is represented as a median with notch edges. Notches are equivalent to 95% confidence interval. No overlap between notches suggests that there is a significant difference between pairs with 95% confidence. Dashed lines represent amplitude panning target angles used in the experiment.*

randomised and divided into two subsets in order to overcome boredom and ear fatigue.

Each subject sat in a chair located at the listening position. Their height was adjusted using a laser pointer, placed at 90° azimuth from the centre loudspeaker. The subject was asked to adjust his or her position until the laser was pointing directly to the ear centre, ensuring that the head is in the right direction and at the equal distance from the speakers. Before starting the test, subjects were reminded to concentrate only on the source elevation, regardless of the azimuth and spectral characteristics. All subjects were requested to hold his or her head still during the test as this could produce misleading results. Lastly, they were asked if there were any questions and proceeded with the test.

3 Results and Discussion

This section will focus on the discussion of the results obtained from this experiment. The raw data collected from the Max patch were organised in Microsoft Excel and then imported into IBM SPSS and MATLAB software, used for the data analysis and plotting. As all subjects were tested against the same conditions, the within-subject analysis was chosen. In order to determine the correct method, Levene's test of homogeneity variance and Shapiro-Wilks' test of normality were performed first. The results indicated that both tests were not satisfied ($p < 0.001$), violating the assumptions of the Repeated Measure Analysis of Variance (RM ANOVA). Therefore, non-parametric analysis was deemed to be appropriate. Friedman test was used for examining the main effect and Wilcoxon test with the Bonferroni correction was used for pairwise comparisons.

3.1 General trend

The subjects' responses, obtained from the listening tests are presented using median and notch edges in Figure 2. The notch edge is the non-parametric equivalence of 95% confidence interval. As stated by McGill, Tukey and Larsen [15], non-overlapping notch edge ranges between two test conditions suggest that there is a significant difference between them.

The plotted results show that the effect of amplitude panning on the perceived elevation is statistically significant for at least one pair of stimuli for each source type, both on the median and off-centre planes. Friedman test confirms that the main effect of panning was significant at 1% level. Nevertheless, the graphs in Figure 2 indicate that the localisation curves were not linear and they followed a similar trend. As target position increased, phantom sources were localised towards the lower loudspeaker. Near the midpoint, the perceived localisation curves became steeper, producing large localisation blurs. Lastly, the localisation curve became flatter and perceived sources were localised towards the upper loudspeaker, mirroring behaviour of the lower region. This trend can be seen across all presentation methods with the exception of the low-pass filtered pink noise and bird stimuli for both loudspeaker pairs.

Overall, this experiment showed that the amplitude panning cannot produce continuous phantom image elevation, instead effectively resulting in a "3 point panning" (low, mid and high regions). These findings are similar to the Barbour's research [9], where amplitude panning in the vertical stereophony was examined. Furthermore, the study conducted by Baumgartner and Majdak [16] showed that the phantom source localisation is the worst when both loudspeakers have the same gain, i.e. 15° target elevation. This can be caused by the differences in HRTF of the phantom and real life stimuli.

3.2 Source effect on localisation

It was observed, that the stimuli with the different frequency content had an impact on the perceived localisation. It can be seen that there is no significant effect of the frequency at the target elevation of 0°, as there was a considerable overlap of the notch edges for all six stimuli. However, the bird sample, with the median of approximately 8.6°, was considerably higher than the target elevation. As the notch edges were not overlapping with the target elevation, it was assumed that the difference is significant ($p < 0.05$). Similarly, the data was distributed for the 5° target elevation. However, a significant difference was observed for the broadband, high-passed pink noise and vocal stimuli. All three were perceived lower than the target level ($p < 0.05$). Stimuli notch edges continued to overlap for the 10° target level, however, the high-passed pink noise and vocals were lower. At 15°, there was no significant difference between sources, apart from the tank shot and high-passed pink noise ($p < 0.05$). Tank shot was perceived higher than the noise sample but the difference was not significant ($p > 0.05$). Additionally, all sources, apart from the broadband and high-passed pink noises, were perceived higher than the target elevation. There was no difference between sources at 20° target elevation, however, sources were also perceived higher. Exception to that was the low-passed pink noise and the bird sound. Starting from 25° and onwards, there was no significant difference between sources, except for the low-passed pink noise ($p > 0.05$). Large localisation blur and significant lower bias for the low-passed noise above the 25° could be explained by the lack of high frequency cues used for the pinna effect [17]. Furthermore, the so-called "pitch height" effect was observed for the stimuli, containing high frequency content. This can be seen as the notch edges for these stimuli became smaller at the high target elevation, as well as due to the upper bias in the lower region.

Lastly, it can be observed that there is no significant difference in the localisation performance on the off-centre plane, apart from the low-passed noise at 0° and 30° target angle ($p < 0.05$). It is considered that interaural cues introduced in this playback condition might have resulted in an upper bias for the low frequency content. This behaviour is interesting and needs to be studied further as it contradicts the "pitch height" effect. Furthermore, notch edge ranges were reduced in the upper region resulting in smaller inter-subject variability. This result seems to suggest that ITD and ILD play an important role in the vertical stereophonic localisation, specifically for the sources without the dominant high frequency content.

4 Conclusion

The results from the current study show that the vertical localisation of the sound source is dependent on the frequency spectrum of the sound source. It is considered that the localisation performance of a vertical phantom image is generally poor since the spectral energy distribution of the ear input signals does not match real HRTF of the target image position. In interactive spatial audio systems, where localisation accuracy is important or continuous panning is necessary, this can cause problems. Similarly, when new control interfaces are designed for such systems, localisation error needs to be taken into consideration.

The results also suggest that the pitch height effect operates in the vertical stereophonic localisation as well as the vertical localisation of a real source; sources with dominant high frequency content were generally localised higher that the target elevation. However, it was also observed that even with the presence of the pitch height effect, the low-pass filtered pink noise was elevated using the amplitude panning. The effect was stronger on the off-centre plane and was present both for phantom and real sources. This seems to suggest that interaural cues play an important role in the vertical localisation. This needs to be investigated further as there is no theoretical explanation to support this effect yet.

References

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

[2] Jongkees, L. B. W., & Groen, J. J. (1946). On directional hearing. *The Journal of Laryngology and Otology*, *61*(9), 494-504. doi:10.1017/S002221510000832X

[3] 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.

[4] Blauert, J. 1969-70, *Sound Localization in the Median Plane*, Acustica, Volume 22, pp. 205-213

[5] Roffler, S. and Butler, R. 1968 *Factors that Influence the Localization of Sound in the Vertical Plane*, Journal of the Acoustical Society of America, Volume 43, No. 6, pp. 1255-1259

[6] Roffler, S. and Butler, R. 1968 *Localization of Tonal Stimuli in the Vertical Plane*, Journal of the Acoustical Society of America, Volume 43, No. 6, pp. 1260-1266

[7] Dolby Atmos, URL: http://www.dolby.com/us/en/guide/dolby-atmos-speaker-setup/index.html (2016).

[8] Auro Technology, URL: http://www.auro-3d.com/system/listening-formats/ (2016).

[9] Barbour, J.M. (2003). Elevation perception: phantom images in the vertical hemi-sphere. In *AES 24 the International Conference on Multichannel Audio*, Retrieved from http://www.aes.org.libaccess.hud.ac.uk/tmpFiles/elib/20151204/12301.pdf.

[10] International Telecommunications Union. (2014). *RecommendationsITU-RBS.1116-2: Methods for the Subjective Assessment of Small Impairments in Audio Systems including Multichannel Sound Systems*. Retrieved from https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.1116-1-199710-S!!PDF-E.pdf.

[11] Pulkki, V. (2001). Localisation of amplitude-panned virtual sources. II: Two- and three-dimensional panning. *AES: Journal of the Audio Engineering Society*, *49*(9), 753-767. Retrieved from: http://tinyurl.com/p6pu6zn.

[12] Canford Audio PLC. (n.d.). *BBC Sound Effect Library*. Retrieved from http://www.canford.co.uk/BBC-SOUND-EFFECTS-LIBRARY.

[13] Free Vocals Recordings, URL: https://freevocals.com (2016).

[14] Lee, H., Johnson, D., & Mironovs, M. (2016 June). *A New Response Method for Auditory Localisation and Spread Tests*. Convention e-Brief to be presented at AES 140th Convention, Paris.

[15] McGill, R., Tukey, J. W., & Larsen, W. A. (1978). Variations of box plots. *The American Statistician*, *32*(1), 12-16. doi:10.1080/00031305.1978.10479236

[16] Baumgartner, R. & Majdak, P. (2015). Modeling Localisation of Amplitude-Panned Virtual Sources in Sagittal Planes. *Journal of Audio Engineering Society, 63* , 562-569. Retrieved from http://www.aes.org/e-lib/browse.cfm?elib=17842.

[17] Batteau, D. W. (1967). The role of the pinna in human localisation. *Proceedings of the Royal Society of London. Series B. Biological Sciences, 168*(11), 158-180. Retrieved from: http://rspb.royalsocietypublishing.org.libaccess.hud.ac.uk/content/168/1011/158