Interaural Cross-Correlation Affects Perceived Elevation of Auditory Images Presented via Height Channels in Multichannel Audio Reproduction Systems

David Stepanavicius[1] and William L. Martens[1]

[1]Faculty of Architecture, Design and Planning, The University of Sydney, Sydney, Australia

ABSTRACT

With 'height channels' in multichannel audio reproduction systems becoming more commonplace, there is substantial and growing interest in understanding production techniques for spatially enhancing the listening experiences made available by such systems. When a group of loudspeakers well above ear level are included in a multichannel loudspeaker array, these 'height channels' present opportunities for manipulation of many spatial auditory attributes, in addition to the possibility of controlling the incidence angle of single virtual sources in both azimuth and elevation. The opportunities investigated in the current study are those enabled by the use of elevated pairs of loudspeakers that present spatially extended stereophonic images of virtual sources. The interaural cross-correlation correlation (IACC) for each of a number of recorded string instruments was manipulated for reproduction via pairs of loudspeakers, the shared elevation angle of which could be varied in elevation. The results demonstrate that the perceived elevation of spatially extended stereophonic images of string instruments could be manipulated by IACC adjustment even though the actual elevation angle was held constant for the loudspeaker pair reproducing each spatially extended component of a multi-instrument ensemble.

1. INTRODUCTION

It is well known that wide auditory images associated with strong lateral reflections produce preferred listening experiences in concert halls (Barron, 1971), but less is known about control of auditory image width for musical ensembles reproduced via multichannel loudspeaker systems that include two or more 'height channels' for enhancing spatial audio reproduction. Concert halls have been designed to create diffuse sound fields for listeners and stereophonic production techniques such as stereo delay have also been used to create a sense of expanse and breadth, and similarly spatially distributed sound is made possible by these so-called 'with 'height systems. Auditory image width is commonly measured with the interaural cross correlation (IACC). IACC measures derive a value that represents the similarity, or dissimilarity of signals arriving at the ears. Low IACC values indicate spatially diffuse sound fields and IACC high values represent a narrowing auditory image. While it is not possible to recreate the sense of an image appearing inside the head with loudspeakers (as evidenced binaurally) due to cross-talk, highly correlated signals reproduced with inter-channel loudspeaker pairs results in a phantom image and decorrelating these outputs results in a wider auditory image (Kendall, 1995).

While stereophonic technology was introduced in 1931 (Alexander, 2013) the first audio reproduction configuration to include 'height-channels' was not seen popularly until the advent of Varese's 11-channel *Poeme Electronique* in 1958 (Paine, et al, 2007). By 1970, a system that could capture and reproduce spherical sound fields was introduced. The periphonic system, coined by Cooper, describes "such systems of recording both the horizontal and vertical direction effect", required a minimum 4-channel (tetraphonic) configuration to decode the captured event. Cooper anticipated that similar, irregular reproduction configurations would eventually be introduced domestically once the technology became feasible. Now that three-dimensional (3D) arrays of loudspeakers have become more common, it is surprising that research into auditory image width for sources presented from elevated loudspeakers has remained relatively unexplored. Image width plays a role in the design of rooms and reproduction systems and the technology is no longer new. Research with 3D reproduction systems has concluded that directional resolution is proportional to channel resolution (Furness, 1990), that vertical incidence provides little contribution to listener envelopment (LEV) (Furuya, 2001) and that the localization blur of overhead phantom images can be alleviated when sources are located between 60-90° elevation relative to the listener (Barbour, 2003). However, system designers continue to develop 3D audio reproduction configurations in the absence of published research as to how height-channels can be used to effectively provide the impression of source elevation

or vertical spaciousness. In this relation, while IACC has been a useful predictor for spatial attributes such as apparent width, its role as a predictor of apparent source elevation is relatively unknown (cf. Martens and Cabrera, 2012).

With the growth in the number of commercial systems designed to enable listeners to experience audio with height, the question that must be asked is how to control the variety of spatial auditory attributes most effectively. And since the control parameters may interact in typical applications, a number of other practical questions naturally arise. For example, does variation in multichannel signal correlation that is used to control image width also exert a substantial influence on the perceived elevation of spatially extended musical sources? Such an influence might be expected for broad imagery produced by well elevated loudspeaker pairs, since the observed cross correlation of such sources will increase with the elevation angle at which the two loudspeakers are found (Martens and Cabrera, 2012).

When using musical stimuli that vary in pitch, it is also expected that elevation reports will be influenced by the pitch of the sound sources - when two sources of differing pitch radiate from the physical coordinates, it is likely that the phenomenon termed Pratt's Effect will be observed, in which case the higher pitched sound source will be heard as arriving from a higher elevation angle (Pratt, 1930). And as suggested above, it is also anticipated that modulations in auditory width at ear level will have negligible affect on reported height, as ear-level presentation of pairs of signals exhibiting varying correlation are likely to vary only in apparent source width (Kendall, 1995).

2. METHOD

2.1 Multichannel Audio Reproduction System

The 22.2-channel audio reproduction system promoted by NHK (the Japanese Broadcasting Association) was used in this study (International Telecommunication Union, 2015). The system provides three layers of loudspeaker elevation with a varying number of loudspeakers at each level, as illustrated in Figure 1.

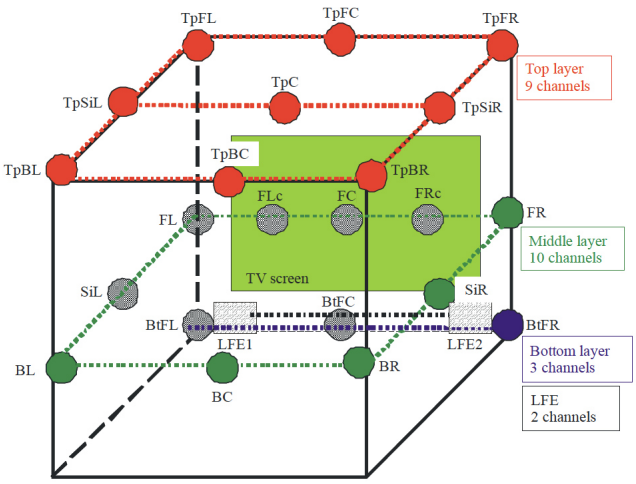

Figure 1. NHK 22.2 Reproduction System Configuration (Image source: International Telecommunication Union, 2015). Recommendations stipulate that from the centre of the listening area, loudspeakers in the bottom layer of the system should be at 15-25° below ear level, the middle layer at 0° elevation, upper front/back loudspeakers at 30-45° elevation and the top middle loudspeaker to be at 90° elevation.

The Spatial Audio Lab at the Wilkinson Building at the University of Sydney contains a multichannel audio reproduction system that is configured to match the 22.2-channel system and also includes an additional layer of loudspeakers (which was not used in this study). It should be noted that the response of two low-frequency drivers provided a roll-off at 2 kHz when used without a cross-over filter. The system was calibrated and level aligned to administer 70 dB(A) of signal to the centre of the listening area. The listening area measures 5 m (L) x 5 m (M) and the loudspeakers were mounted to a frame extending upwards 3 m (H).

2.2 Stimuli

Anechoic recordings of string bass, cello, viola and violin were used to produce two different chords. The first consisted of string bass (A_1 / 55 Hz), cello (D_3 / 146 Hz) and violin (E_5 / 659 Hz). The second chord consisted of string bass (A_1 / 55 Hz), viola (A_4 / 440 Hz) and violin (E_5 / 659 Hz). All recordings were considered complex and contained the frequency bandwidth required for elevation localisation (Blauert, 1969).

Two incoherent, single-channel versions were created from each instrument recording using the *JOLA* function developed in Matlab by the second author. The function windows a selected time region based upon the input signal's amplitude envelope, shifts this windowed portion by a small randomly-selected amount and then 'overlaps and adds' many times, resulting in an ensemble effect sounding output. Figure 2 illustrates the process. The total duration of each signal totalled 6 seconds after processing. String bass and violin signals received a 500 ms fade-in and cello and viola received a 2500 ms fade-in. All signals received an equal 500 ms fade-out. Upon playback of each chord, string bass and violin was heard before cello or viola.

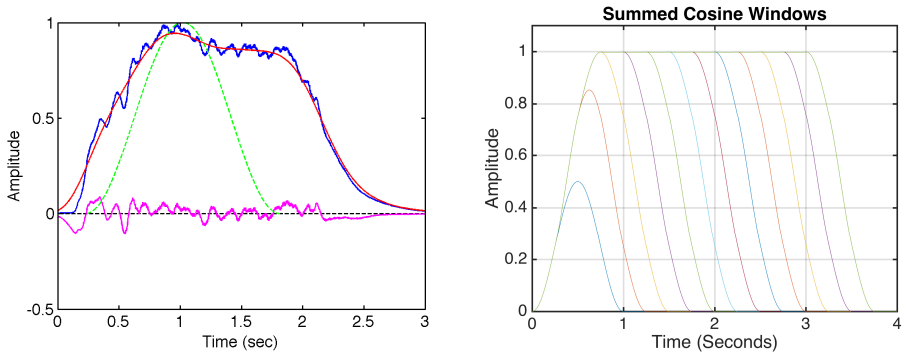

Figure 2. Illustration of the JOLA process. The signal processing algorithm applies a window to an input signal envelope and creates an ensemble sounding output summing multiple copies of the windowed segment that are slightly jittered in time (offset from strict regularity by randomly selected amounts).

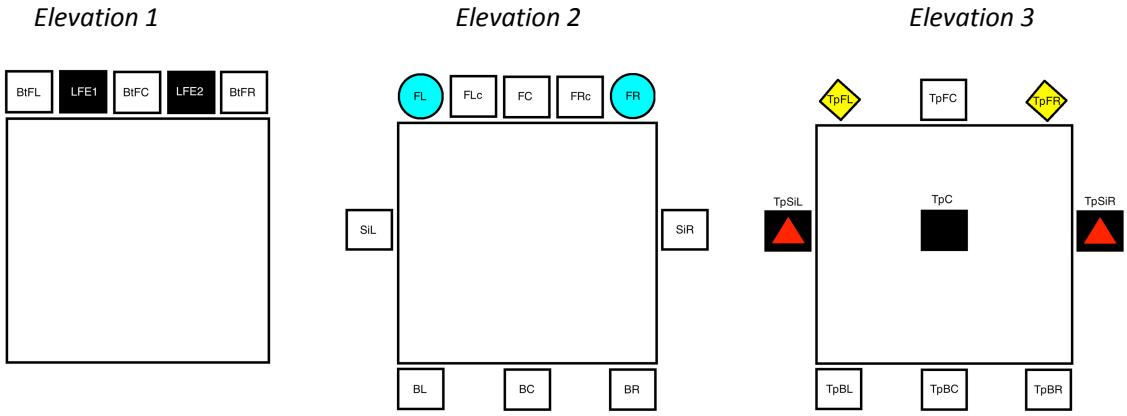

Figure 3: Overhead view of the 22.2 reproduction system in three elevation section layers indicating the loudspeaker pairs employed to reproduce string ensemble components. Black squares indicate the anchor points established by string bass (LFE1 and LFE2) and Violin (TpSiL, TpC and TpSiR). Cyan circles represent the loudspeaker pairs used to deliver stimuli at ear level. Yellow diamonds indicate the loudspeaker pairs used to deliver stimuli at moderate elevation. Red Triangles indicate the loudspeaker pairs used for deliver stimuli near the frontal plane.

Full-band decorrelation was implemented to modulate ensemble stage widths (ESW) of cello and viola to three degrees. Wide (0.1), Narrow (0.4-0.5) and Narrowest (0.9). A Neumann KU100 binaural microphone system was used to record the stimuli at the three ESW modulations at each of the three loudspeaker elevations (for a total of 18 test conditions). The proximal stimulus was analysed with the routine named *IACC_music.m* found within *AARAE*, an open source acoustic analyser that runs within Matlab (Cabrera et al, 2014). Two incoherent string bass signals were assigned to the LFE channels at the lowest layer of elevation and two incoherent violin signals were assigned to TpSiL, TpSiR, with a sum of both signals assigned to TpC. Cello and viola signals were sent to two inter-channel loudspeaker pairs at three elevations between the middle and top layer of the 22.2 system; FL and FR (Ear Level) TpFL and TpFR (Elevated) and TpSiL and TpSiR (Frontal Plane). Figure 3 illustrates the assignment of sources to loudspeakers with the three-dimensional (3D) array.

2.3 Procedure

A total of 8 listeners from the audio and acoustic program at the University of Sydney voluntarily participated in the listening task. Each participant was considered to be an expert listener and reported no hearing deficiencies. Subjects were asked to listen to 18 chords and report on the apparent elevation of the second note within each. It was explained that the string bass and violin that established anchor points of pitch and spatial location would remain static across each condition. The second note however, could be reproduced at any height within these anchor points. Listeners were asked to report each elevation by striking a horizontal line through a 100 mm vertical scale, in reference to the anchor points. String bass was labelled at the bottom of the scale and violin was labelled at the top. Once the task was demonstrated, Matlab generated a randomised listening sequence and each listener completed the 18 tasks from 4 listening locations.

Listening locations were situated along the centre of the listening area. L1 was seated 1.9 behind the front the listening area. Subsequent seat locations provided 0.6 m distance from listener to listener seated front-to-back and each seat was adjusted to provide an approximate 0.4 m incline to reduce head shadowing effects. The ear-level of listeners occupying the middle two seats was in the direct path of loudspeakers on the ear-level layer of elevation. Figure 4 illustrates. Listeners were able to move their heads freely as desired.

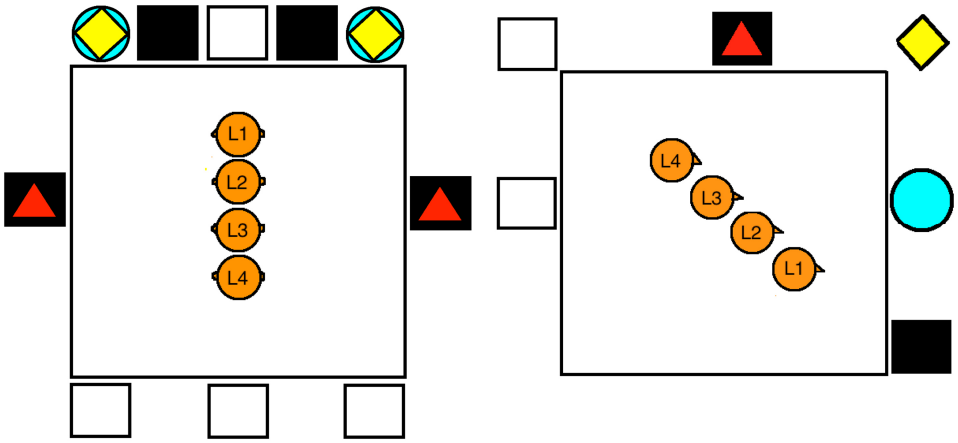

Figure 4. The four listening locations that each subject used to report source elevation. The left image represents an overhead view of the listeners seated front to back. Note that the overhead TpC channel has been discarded from the image for ease of visibility. The right image represents the side elevation view of the four listening positions. L1 was seated cross-legged on the ground. Each ensuing chair was raised at approximately at 0.4 m increased intervals. L2 and L3 were seated in the direct field of the ear-level elevation. (Image not to scale)

A 2-second pause elapsed between conditions. The listening task took less than 15 minutes in total to complete from all four listening positions. Responses were measured from the point at which the vertical line was marked by listeners and rounded to the nearest millimetre. The randomised sequence generated by Matlab was also unshuffled in Matlab and results were collated.

3. RESULTS

The distribution of elevation ratings shown in Figure 5 indicates that the viola tones, the higher pitched of two instrument sounds presented, was typically heard at much higher elevation than the cello tones. The median reported elevation angles for viola tones were consistently higher than those for the cello at all three loudspeaker elevation angles. Changes in IACC had the clearest effect on reported elevation when sources were presented above ear level. Figure 6 shows that the distribution of elevation ratings at ear level was quite broad (quantified here as the Inter-quartile Range, or IQR). Response variation decreased considerably when loudspeakers were elevated either in front of the listener (in the 'Elevated' condition), or overhead (along the frontal-plane). Even the low-elevation outliers in these two conditions were above the median response observed at ear level for all three IACC values. Note however that the only systematic effect on elevation ratings of variation in IACC was observed in the 'Elevated' condition, with variation in the median response along the frontal plane quite minimal. Higher IACC values at the widest ESW conditions are explained by path distance between loudspeakers and the ears. Note that in the 'Elevated' condition, the more highly correlated sources resulted in auditory imagery that was collapsing toward the median sagittal plane as well as being heard to be located at higher elevations.

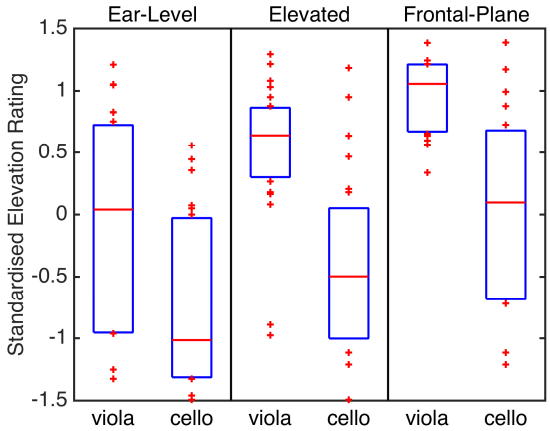

Figure 5. Boxplot showing the distribution of Elevation Ratings observed for a single subject for two instruments (viola versus cello, as written under each box), with separate boxes shown for each of the three speaker elevations (identified by the text above each set of two boxes).

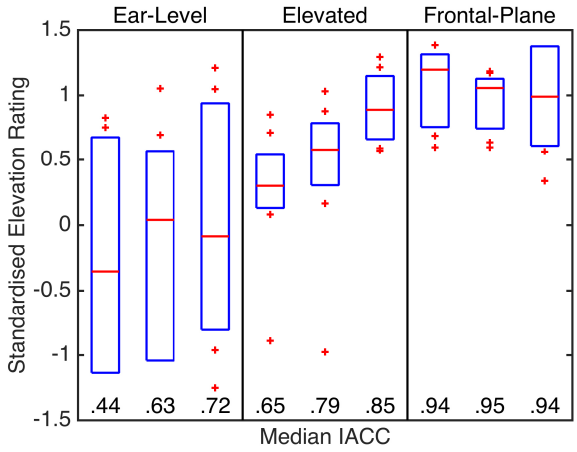

Figure 6. Boxplot showing the distribution of Elevation Ratings observed for a single subject in each of nine conditions, the median Elevation Rating in each case is given by the horizontal red line inside each box. The Median IACC value calculated for each of the nine cases is shown below each box and the speaker elevation is written above each set of three boxes. Results only for the viola are shown here.

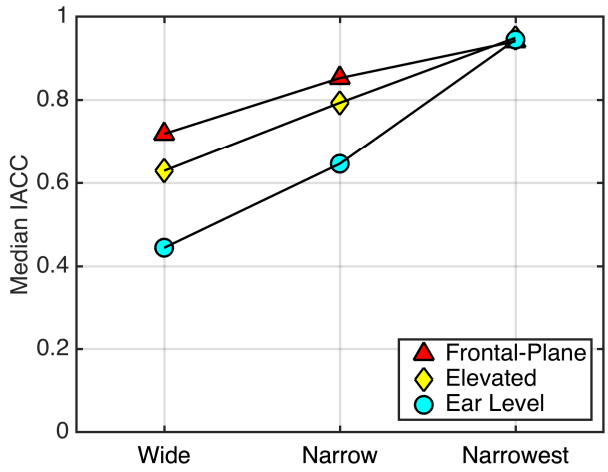

Figure 7. For three different speaker elevations and for three width conditions, the Median IACC value is plotted, with medians calculated for each of the nine cases based upon on a vector of IACC values observed within 10 third-octave bands ranging from 250 Hz to 2000 Hz. Results for the viola only are shown.

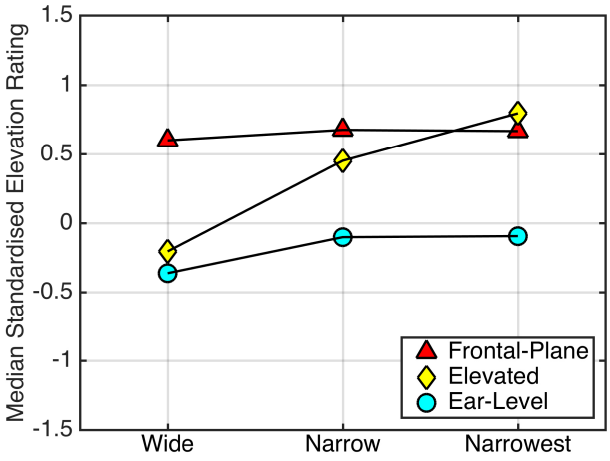

Figure 8. Median Standardised Elevation Ratings calculated for the combined results of 8 subjects in response to the viola stimulus, presented at three different speaker elevations and in three width conditions (Wide, Narrow and Narrowest). Colour coded plotting symbols are the same as those used to plot Median IACC values in Figure 7, which indicate the binaurally measured differences between signals that are identified only nominally on the x-axis.

The elevation ratings for the viola stimulus that were made by 8 subjects (SUB) were submitted to a 3-way ANOVA (results of which are shown in Table 1), with subject (coded as SUB) as a random variable that included the changing listener position as an additional source of random variation. The variance due to subject and position was combined because the results of a preliminary 4-way ANOVA showed extremely small F values for both. The two main variables of interest were those that were manipulated in this experiment, with three levels of each: The three auditory image widths (coded as WID) and the three loudspeaker elevations (coded as ELE). Although IACC values were available for each of the nine viola stimuli presented (as shown in Figure 7), this continuous variable was not included in the analysis of variance. Rather, WID was treated as a categorical variable with the three levels nominally identified as 'Wide', 'Narrow' and 'Narrowest'. The three levels of elevation characterising the loudspeaker variable also could be specified by angular values, but were identified here using the categorical labels 'Frontal-Plane, 'Elevated and 'Ear-Level' (in interaural-polar coordinates, the elevation angles were 90^{o}, 45^{o} and 0^{o}, respectively). By including interaction terms in the 3-way ANOVA, a clear explanation for the pattern of elevation

ratings emerged (which results were presented in Figure 8, with loudspeaker elevation as the parameter of the plot).

Table 1. Results of a 3-way ANOVA performed on the standardised elevation ratings made by 8 subjects (SUB) for the viola stimulus presented at three widths (WID) and three elevations (ELE).

Source	SSQ	d.f.	MSQ	F	Prob>F
ELE	0.78	2	0.391	0.31	0.738
WID	3.97	2	1.987	2.88	0.090
SUB	2.33	7	0.332	0.29	0.939
ELE*WID	13.16	4	3.289	4.01	0.004
ELE*SUB	17.61	14	1.257	1.53	0.099
WID*SUB	9.66	14	0.69	0.84	0.624
Error	200.05	244	0.82	-	-
Total	247.56	287	-	-	-

The main effect of elevation (ELE) looks to be substantial in Figure 8, but is not strong enough to reach statistical significance given the large amount of error variance. The main effect of width (WID) looks more substantial, but again does not reach statistical significance. Rather, it is the interaction between these two factors that is statistically significant due to the manner in which the effect of width on elevation ratings depends upon the elevation of the loudspeakers delivering the stimuli at those varying IACC values. In particular, it is only the moderately elevated loudspeakers that allow the influence of IACC on reported elevation to be observed. When the loudspeakers delivering the viola stimuli were at the listener's ear level, changing image width had no effect on the median of the standardised elevation reports. Likewise, when the loudspeakers delivering the viola stimuli were high above the listener's head, within the frontal plane, changing image width had no apparent effect on reported elevation. The results of this study have stimulated interest in a further investigation that will examine both the perceived elevation and the perceived width of auditory images over a range of loudspeaker elevation angles.

4. DISCUSSION

Results indicate that inter-channel output correlation affects elevation perception most strongly when loudspeakers are located in front of listeners at elevation. Therefore, when doing production for multichannel audio reproduction systems employing 'height channels', understanding this will help to focus efforts upon the most effective control of perceived elevation of spatially extended ensemble components. The results of this experimental study also confirmed that for these spatially extended stereophonic images of string instruments, their perceived elevation could be manipulated by IACC adjustment even though the actual elevation angle was held constant for the loudspeaker pair reproducing each spatially extended component of the ensemble. The results also show that the influence of IACC on perceived elevation is weak for sound reproduced by loudspeakers positioned at ear level. Why is it so important to know about these interactions between parameters that can be used to control elevation of spatially broad auditory images? Rumsey (1998) put it very well in a general statement on the importance of establishing what constitutes subjective 'quality' in research on improving spatial sound reproduction as follows:

> *"Since sound is as much a consumer commodity as any other product these days, there are strong arguments for determining what factors have the greatest effect on consumer preference, and how quality is judged by consumers of sound systems and consumers of recorded material"*
>
> (Rumsey, 1998, p. 123).

To summarize the application of these results then, it should be clear that production techniques that further enhance the listening experience are of substantial interest. When a group of loudspeakers well above ear level are

included in a multichannel loudspeaker array, these 'height channels' present opportunities for manipulation of many spatial auditory attributes, in addition to the possibility of controlling the incidence angle of single virtual sources in both azimuth and elevation. While auditory images presented via loudspeakers at extreme elevation (along the frontal plane) can create images of extreme apparent elevation, adjustments of IACC to control apparent width of those sources will have little effect on perceived elevation in this condition. This informs system designers and mixing engineers that only when using loudspeakers at intermediate elevations, can the adjustment of inter-channel correlation be an effective tool for manipulating the elevation of the region from which ensemble phantom images are perceived to radiate. These results direct attention to the use of a particular subset of loudspeakers positioned on the second and third layer of an NHK 22.2 reproduction configuration to reliably fill space with spatially extended auditory images. The results are also relevant to production techniques used for the Dolby Atmos™ reproduction system. Personal discussions with audio engineers from Dolby Laboratories confirmed that a deeper understanding of these effects would enlighten not only 3D audio reproduction system engineers, but also mixing engineers and researchers alike (Cooper, 2016). Further research will be conducted using a larger group of listeners and a greater variety of stimulus material.

REFERENCES

Alexander, R., 2013. *The Inventor of Stereo: The Life and Works of Alan Dower Blumlein*. CRC Press.

Barbour, J.L., 2003, June. *Elevation perception: Phantom images in the vertical hemi-sphere.* In Audio Engineering Society Conference: 24th International Conference: Multichannel Audio, The New Reality. Audio Engineering Society.

Barron, M., 1971. *The subjective effects of first reflections in concert halls—the need for lateral reflections*. Journal of Sound and Vibration, 15(4), pp.475-494.

Blauert, J., 1969. *Sound localization in the median plane*. Acta Acustica united with Acustica, 22(4), pp.205-213.

Cabrera, D., Jimenez, D. and Martens, W.L., 2014, November. *Audio and Acoustical Response Analysis Environment (AARAE): a tool to support education and research in acoustics*. In Proceedings of Internoise.

Cooper, D., 2016, personal communication, May 27.

Furness, R.K., 1990, May. *Ambisonics - an overview*. In Audio Engineering Society Conference: 8th International Conference: The Sound of Audio. Audio Engineering Society.

Furuya, H., Fujimoto, K., Ji, C.Y. and Higa, N., 2001. *Arrival direction of late sound and listener envelopment.* Applied Acoustics, 62(2), pp.125-136.

International Telecommunication Union, 2015, Multichannel sound technology in home and broadcasting applications. Report ITU-R BS.2159-7, Geneva, Switzerland.

Kendall, G.S., 1995. *The decorrelation of audio signals and its impact on spatial imagery*. Computer Music Journal, 19(4), pp.71-87.

Martens, W. L., and Cabrera, D. A., 2012, Perceived elevation of simultaneously presented sound sources depends upon the correlation between the source signals. The Journal of the Acoustical Society of America, 131(4), 3216.

Paine, G., Sazdov, R. and Stevens, K., 2007. *Perceptual investigation into envelopement, spatial clarity, and engulfment in reproduced multi-channel audio*. In Audio Engineering Society Conference: 31st International Conference: New Directions in High Resolution Audio. Audio Engineering Society.

Pratt, C.C., 1930. *The spatial character of high and low tones.* Journal of Experimental Psychology, 13(3), p.278.

Rumsey, F., 1998, October. *Subjective assessment of the spatial attributes of reproduced sound*. In Audio Engineering Society Conference: 15th International Conference: Audio, Acoustics & Small Spaces. Audio Engineering Society.