Audio Engineering Society

Convention Paper 10219

Presented at the AES 158th Convention
2025 May 22–24, Warsaw, Poland

This paper was peer-reviewed as a complete manuscript for presentation at this convention. This paper is available in the AES E-Library (http://www.aes.org/e-lib), all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Correlation between middle and top layer loudspeaker signals and the listening area in 3D audio reproduction

Toru Kamekawa[1] and Atsushi Marui[1]

[1]*Tokyo University of the Art*

Correspondence should be addressed to Toru Kamekawa (kamekawa@ms.geidai.ac.jp)

ABSTRACT

In auditory spatial perception, horizontal sound image localization and a sense of spaciousness are based on level and time differences between the left and right ears as cues, and the degree of correlation between the left and right signals is thought to contribute to the sense of horizontal spaciousness, in particular (see [1, 2]). For the vertical image spread (VIS), spectral cues are necessary. The change in VIS due to the degree of correlation between the vertical and horizontal signals depends on the frequency response [3]. This paper investigated the influence of different correlation values between the top and middle layers of loudspeaker signals within a 3D audio reproduction system on listening impressions through two experiments. The results of experiments using pink noise with different correlation values for the top and middle layers show that the lower the vertical correlation values are, the wider the listening area is, where the impression does not change from the central listening position. From the results of experiments using impulse responses obtained by setting up microphones in an actual concert hall, a tendency to perceive a sense of spaciousness at the off-center listening position was found when cardioid microphones were used for the top layer that were spaced apart from the middle layer. The polar pattern and height of the microphones may have resulted in lower correlation values in the vertical direction, thus widening the listening area of consistent spatial impression outside of the central listening position (i.e., sweet spot).

1 Introduction

Several three-dimensional immersive audio formats including height channels have been proposed, such as NHK's 22.2 channel system, Dolby Atmos, and Auro 3D, all of which are standardized in ITU-R BS.2159 [4].

When recording an orchestra or other instruments in a concert hall, spot microphones are used to capture the details of individual instruments, and main microphones are used to capture the sound of the entire hall. For the main microphone system, various methods have been proposed for 3D audio recording thus far [5, 6, 7, 8, 9]. Many maintain a linear, one-to-one relationship between the number of microphones and the number of corresponding loudspeakers. Hamasaki *et al.* proposed a recording method for 22.2 multichannel sound using super-cardioid microphones for the front and figure of eight microphones for the side, back, and upper directions [9]. Howie *et al.* proposed another recording technique for 22.2 multichannel sound based on the Decca Tree [10] using omnidirectional microphones and compared their array with the Hamasaki's array and Higher-Order-Ambisonics (HOA) [11]. That study found that the two different spaced arrays showed an apparent similarity of ratings for all subjective at-

tributes under investigation: “clarity”, “scene depth”, “naturalness”, “environmental envelopment”, “sound source envelopment”, and “quality of orchestral image”. In a comparison of microphone arrays conducted by the authors, the spaced array was also rated as more “realistic” [12]. Thus, spaced arrays, in which microphones are set apart from each other in concert hall recordings, are often used, but there are various opinions on the position and polar pattern of the upper-level “height channel”) microphones.

Lee recommends that the microphones in the top layer should be placed as close as possible to the middle layer [5], because the distance between the top layer and the middle layer microphones causes phase interference at the listening position, resulting in timbre degradation. Therefore, a top layer microphone array was proposed with cardioid microphones placed at the same height as the middle layer, facing upwards [6]. On the other hand, in actual recording sites, top layer microphones are often set 1-2 m away from middle-layer microphones [5, 13].

In auditory perception of space, horizontal sound image localization and sense of spaciousness are perceived based on level and time differences between the left and right ears [14, 15], while vertical sound imaging is perceived based on the difference in frequency response in the direction of sound arrival (spectral cues or direction-determining bandwidth) caused by the shape of the ear [16]. On the other hand, for the vertical image spread (VIS: Vertical Image Spread), spectral cues are essential, and the change in VIS due to the degree of correlation between the vertical and horizontal signals depends on the frequency response [3].

As mentioned above, lower correlation between the upper and lower layer signals, obtained by separating the top layer microphones from the middle layer may not necessarily contribute to VIS. Therefore, this paper investigates how differences in correlation of loudspeaker signals in the vertical direction affect the listening impression, through experiments using pink noise with different correlation values and impulse responses obtained by setting up a microphone array in an actual concert hall.

Before the experiment, the first author compared the following two types of pink noise:

- four channel uncorrelated pink noise played back from two channels in the middle layer (e.g. FL, FR), and two channels in the top layer corresponded to the middle layer (e.g. TpFL, TpFR). (ICC: the inter-channel cross-correlation value r for upper and lower channels was almost zero.)
- two channel uncorrelated pink noise played back from two channels in the middle layer (e.g. FL, FR), and simultaneously played back from the top layer corresponding to the middle layer (e.g. TpFL, TpFR). (upper and lower ICC r was almost 1.)

The difference between the two was perceived as a difference in timbre but not as a change in VIS. On the other hand, when the ICC value was low, the change in timbre and the sound image was not felt different, even if the listening position was moved, but when the ICC value was high, the change was felt to be significant when the listening position was moved slightly. In the experiment, therefore, it was decided to investigate the area of listening positions where the sound impression does not change.

The ICC coefficient (ICCC) between two loudspeaker signals (s_1 and s_2) is calculated as the maximum absolute value of the inter-channel cross-correlation function (ICCF) [2].

$$ICCF(\tau) = \frac{\int_{-\infty}^{\infty} s_1(t) \cdot s_2(t+\tau)dt}{\sqrt{\int_{-\infty}^{\infty} s_1^2(t)dt \int_{-\infty}^{\infty} s_2^2(t)dt}} \quad (1)$$

$$ICCC = \max|\mathrm{ICCF}(\tau)| \quad (2)$$

A lag time (τ) can be set with the ICCF to compensate for any delay between the signals. In this paper, 1 ms lag is used.

2 Experiment 1

Experiment 1 investigated how different vertical correlation values affect the size of the sweet spot where the impression doesn't differ from that of the center.

Pink noise with different ICC values between the upper and lower layers described in the previous section was used as stimuli.

The experiment was conducted using a 22.2 multichannel sound system in Studio B at Tokyo University of the Arts. The loudspeaker distance was set at 2.7 m from the central listening position, and the elevation angle of the top layer was set 30°(Fig.1). The combination of

playback loudspeakers is described in Table 1. Three types of correlation values between the top and middle layers, such as "High (correlation value r=1)';, "Low ($r \simeq 0$)", and "Mid ($r \simeq 0.5$)" were compared.

To compare the difference between center and off-center listening positions, four types of movement directions such as left, forward, backward, and downward were investigated.

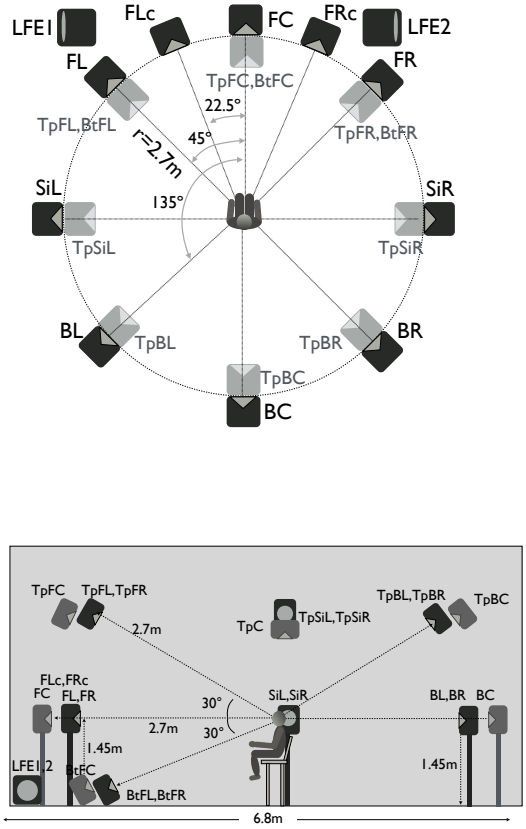


Fig. 1: Layout of the listening experiments. The upper panel indicates the horizontal direction, and the lower panel indicates the section.

2.1 Method for Experiment 1-1

The participants were first asked to listen at the central listening position: the location equidistant to all loudspeakers. Then they moved gradually towards the left and stopped at the position where their impression was changed from the central listening position. The distance they moved from the center was recorded as d_c. In the same way, the listening position was obtained in a total of four directions (left, forward, backward, and downward by crouching).

Table 1: Combination of middle and top layer loudspeakers used in Experiment 1

	middle	top	angle
Front	FL, FR	TpFR, TpFR	$\pm$ 45°
Side	SiL, SiR	TpSiL, TpSiR	$\pm$ 90°
Rear	BL, BR	TpBL, TpBR	$\pm$ 135°

The three playback loudspeaker combinations (front, side, and rear) in Table 1 were also carried out in the four directions of movement mentioned above.

Nine students and faculty staff from the University participated in the experiment.

2.2 Method for Experiment 1-2

Following Experiment 1-1, to find the frequency band of the correlation of vertical direction which contributes to the expansion of the listening position, Experiment 1-2 was executed in the following steps.

1. The low-frequency components below the cut-off frequency f_c in the top layer were cut and mixed into the middle layer.
2. As in Experiment 1, the participants were asked to move from the central listening position to a position d where the listening impression did not change, and compared with the position d_c where the listening impression did not change for the uncorrelated vertical noise obtained in Experiment 1.
3. The initial cut-off frequency is initially set at 10 kHz, and if d is less than d_c, f_c is reduced by 1/3 octave band.
4. When d reached the same position as d_c, f_c is raised by 1/3 octave, and when d became less than d_c, the f_c immediately before that was set as the cut-off frequency and recorded.

The filter for the cut-off was created using the Max8 filtergraph function (gain = 1.0, Q = 0.7, -12dB/oct.).

The stimuli and the listener movements used in Experiment 1-2 were the same as in Experiment 1-1, except for the movement for the downward direction. Experiment 1-2 was conducted following Experiment 1-1, with eight participants from the students who had participated in Experiment 1-1. The duration of the experiment varied between participants, but Experiments 1-1 and 1-2 together lasted approximately 60 minutes.

2.3 Results of Experiment 1

2.3.1 Correlation between upper and lower layer signals and d_c

A three-way ANOVA (analysis of variance) was conducted using the ICC value, the position of the playback loudspeaker and the direction of movement as independent variables, and the position d_c at which the listening impression does not change in the case of uncorrelated vertical noise as the dependent variable. The results showed significant differences in ICC value ($F(2,279) = 102.1, p < 0.001$) and direction of movement ($F(3,279) = 65.2, p < 0.001$), but not in playback speaker position ($F(2,279) = 1.4, p = 0.25$). This suggests that while ICC value and movement direction significantly influence perception, speaker placement does not play a crucial role.

The playback loudspeaker positions for which no significant differences were found are summarised, showing the mean values of d_c and 95% confidence intervals for the perceived difference from the central position for each direction of movement and different ICC values between the upper and lower (Fig.2). For low ICC values, the listening area with an unchanged sound impression extends approximately 45 cm to the sides, 35 cm forward, 55 cm backward, and 15 cm downward.

2.3.2 Cut-off frequency f_c at d_c

A two-way ANOVA was conducted with the playback speaker position and direction of movement as independent variables, with the cut-off frequency f_c at the position d_c where the listening impression does not change in the case of vertically uncorrelated noise as the dependent variable. The results showed that there were no significant differences in both the direction of movement ($F(2,63) = 0.78, p = 0.46$) and the playback speaker position ($F(2,63) = 0.62, p = 0.54$).

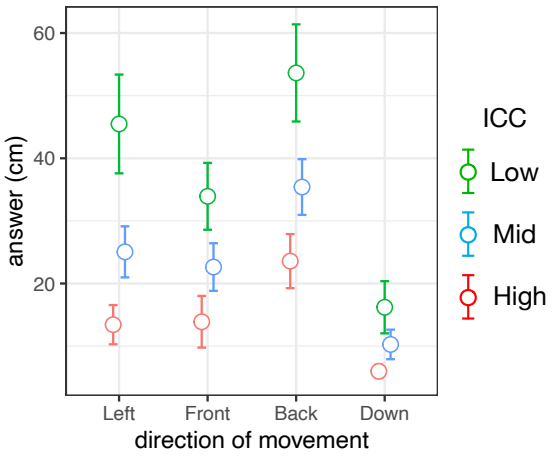


Fig. 2: Differences in perceived distance and vertical ICC values with the central position for each direction of movement. Circles indicate mean values and error bars indicate 95% confidence intervals. ICC values r are as follows; High=1, Low $r \simeq 0$, and Mid $r \simeq 0.5$

Fig.3 shows the scatter plots and box plots for each direction of movement of the frequency f_c (Low cut-off frequencies in d_c obtained in Experiment 1-1 at which participants cannot distinguish the difference from the central listening position). The figure shows that the first quartiles of the cut-off frequency are 190 Hz (left), 151 Hz (front) and 200 Hz (back), suggesting that even if the upper and lower ICC values are high in the frequency band below 150-200 Hz, 75 % of listeners will have the same listening area as when the upper and lower are uncorrelated at all frequency band.

3 Experiment 2

Experiment 1 showed that lower ICC values between the upper and lower layers resulted in a wider area of impressions that did not differ from the center listening position. In Experiment 2, therefore, we decided to investigate how the impression changes with the listening position when the height and polar pattern of the upper microphones are changed and its correlation with the middle microphone is varied.

3.1 Sound stimuli for Experiment 2

The sound sources used in the experiment were obtained from the impulse responses measured at the

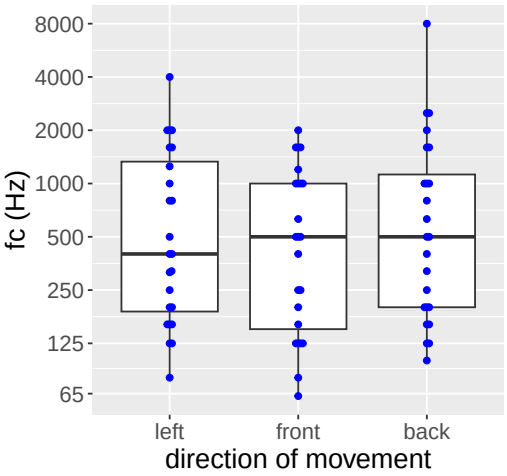


Fig. 3: Scatter plots and box plots of frequency f_c where the difference between the upper and lower uncorrelated cases and the cut case is unclear.

university's small concert hall, Hall 6 (240 seats, floor space approx. 450 m^2, height approx. 9 m, reverberation time approx. 1.6 s@500 Hz), where an omnidirectional loudspeaker and microphone were set at the positions shown in Fig. 4. The microphones in the top layer were placed at the same position in the horizontal plane as FL, FR, BL, and BR in Fig. 4 with the polar pattern and height shown in Table 2. The cardioid microphones were placed with their directivity axes pointing upwards or side to avoid direct sound from the stage. The center microphone (FC) in the middle layer was figure-of-eight, while the other four microphones were omnidirectional at 2 m height. For "O0m", omnidirectional microphones signal with the same height of 2 m as the middle layer was used for the top layer. Using these impulse responses, excerpts of xylophone and solo female singing recorded in an anechoic room were convolved.

3.2 Method for Experiment 2

Listening experiments were conducted using the middle (FL, FC, FR, BL, BR) and upper (TpFL, TpFR, TpBL, TpBR) layers of the 22.2 multi-channel sound reproduction system in Studio B at the university (same as Experimetnt 1: distance from the central listening position to the loudspeakers 2.7 m, elevation angle of the upper layer 30°).

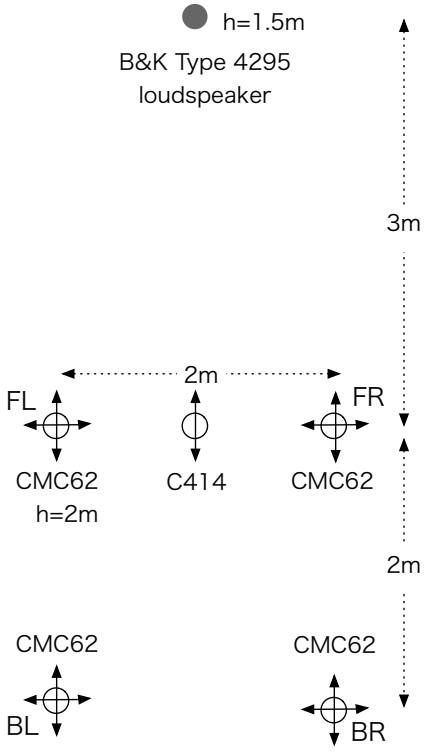


Fig. 4: Layout of microphones with measured impulse responses.

Prior to the experiment, the aforementioned five sound sources were compared, and it was confirmed that there were differences in spatial expanse, distance, and reverberation. Therefore, this time we decided to focus on the impression of spatial expanse and to obtain answers by pairwise comparison, using the GUI created in Max8, with the middle layer of Fig.4 as common, and the five upper layers shown in Table 2. Stimuli "A" and "B" were randomly assigned to the 5+4-channel impulse response convolved sound sources (xylophone, female vocal), which were a combination of microphones, and the participants were asked to answer which one makes the sound scene in which the musicians are playing feel more spacious on a seven-point scale. The participants were allowed to listen as many times as they wished, and the experiment was conducted using Scheffe's pairwise comparison method (Nakaya's variant [17]), which does not take order effects into account. To compare the impressions of different listening areas, it was decided to examine the impressions at three loca-

tions: the center (C), 50 cm to the left of the center (L), and 50 cm behind the center (B), referring to the area where the impression does not differ from that of the center when the ICC between the upper and lower parts is low in the previous report. The listening positions and the order of the materials (xylophone and female solo) were set randomly according to the experimental participants.

The volume of each sound source was adjusted to be approximately 70 dB (LAeq) at the central listening position. Sixteen students and teachers from the University participated in the experiment. The duration of the experiment varied from participant to participant but lasted approximately 60 minutes in total.

3.3 Result of Experiment 2

The results of the pairwise comparisons between the female vocal and xylophone are shown in Fig. 5 and Fig. 6, respectively. The horizontal brackets indicate pairs where significant differences were found for each combination (* $p < 0.05$, ** $p < 0.01$). The figures show that the top layers with cardioid microphones (C1m, C2m) perceived more spatial extent than those with omnidirectional (O0m, O2m). There was also a tendency for the top layer to feel more spacious for higher microphone positions. Furthermore, it can be said that the differences between the microphones in the upper tiers were easier to perceive with the xylophone than with the female vocal. Concerning the differences in listening position, it can be considered that the differences were easier to understand because there were many combinations where significant differences ($p < 0.01$) were found in the backward (B) in the case of female soloists and in the lateral direction (L) in the case of xylophones.

Table 2: Height and polar pattern of top layer microphones. (Height of main layer microphones: 2 m)

	model name	polar pattern	height (m)
C0m	CMC64	cardioid	2
C1m	CMC64	cardioid	3
C2m	CMC64	cardioid	4
O0m	CMC62	omni	2
O2m	CMC62	omni	4

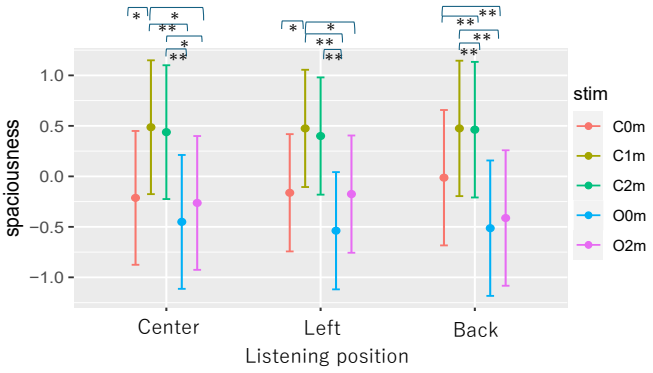


Fig. 5: Means and 95% confidence intervals obtained from pairwise comparisons of the female vocal.

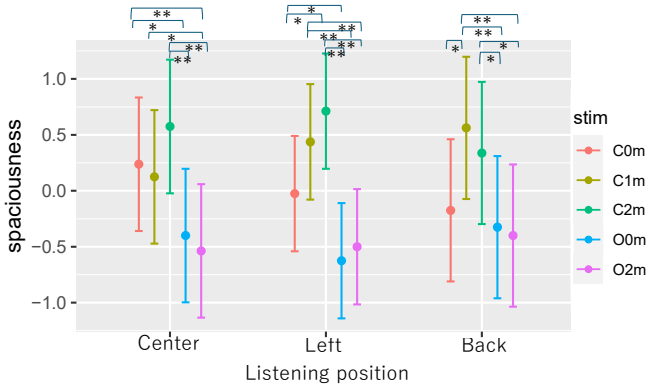


Fig. 6: Means and 95% confidence intervals obtained from pairwise comparisons of the xylophone.

4 Discussion

The results of Experiment 1-1 suggest that low vertical correlation widens the listening area, the so-called sweet spot, where the impression is not different from the central listening position. Furthermore, the results of Experiment 1-2 suggest that, regardless of the direction of the playback loudspeaker, if the upper and lower correlation values are low in the frequency band above 150-200 Hz, 75% of listeners can obtain a listening area similar to that of the uncorrelated case. Experiment 2 investigated the relationship between listening impression and listening position when the height and polar pattern of the top layer microphones were varied. The results show that a sense of spaciousness can be kept at the off-center listening position when cardioid

microphones are placed away from the middle layer. Therefore, the ICC of the upper and middle layers of the stimuli used in Experiment 2 were examined.

Fig. 7 shows ICC per octave band obtained from the impulse responses of the microphones in the middle (FL) and the top layer (TopFL). Although not shown in the figure, the omnidirectional microphones at the same height as the middle layer (O0m) have a value of 1 in all bands as it is the same as the middle layer. The figure shows that when cardioid microphones are used, the ICC in the band below 1 kHz is lower when the height of the microphone in the top layer is raised above the middle layer. This result may correspond to the fact that the "C2m" cardioid microphones placed at a height of 4 m (2 m higher than the middle layer), had a higher sense of spaciousness at the off-center listening position.

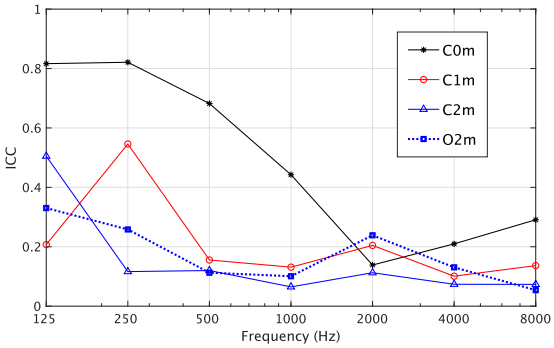


Fig. 7: Inter-channel cross-correlation (ICC) per octave band between the middle layer microphone (FL) and the top layer microphone (TpFL) obtained from the impulse responses.

Fig. 8 and 9 show ICC per octave band obtained from the female vocal and the xylophone in the middle (FL) and the top layer (TopFL). The ICC obtained cross-correlation values for the entire sound source by shifting every 100 ms in a 300 ms time window, excluding values where the volume was less than 20 dB below the average of the overall r.m.s.

The band-wise ICC of the sources convolved with vocal and xylophone are different from the ICCs of the impulse responses, suggesting that the listening impression differs depending on the source, as seen in the results of Experiment 2. The values of "C1m", "C2m", and "O2m" were particularly close for female vocals, corresponding with the impression from the participants that the female vocal was relatively more difficult to judge.

In all figures, the O2m with an omnidirectional microphone placed 2 m higher than the middle layer has a low ICC value in the mid-low frequency range below 1 kHz, but in listening experiments, the impression was different from that of the C2m with a different polar pattern. This suggests that other factors contribute to the perception of spaciousness, not just the vertical ICC. This requires further study in the future.

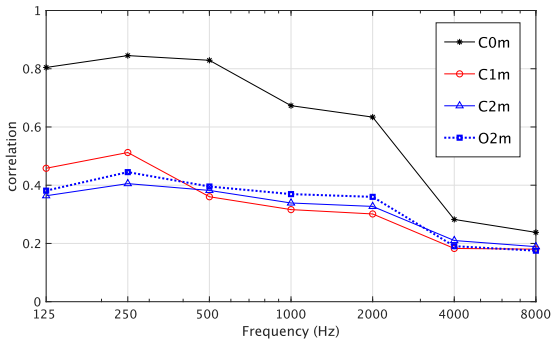


Fig. 8: ICC per octave band between the middle layer microphone (FL) and the top layer microphone (TpFL) obtained from the female vocal.

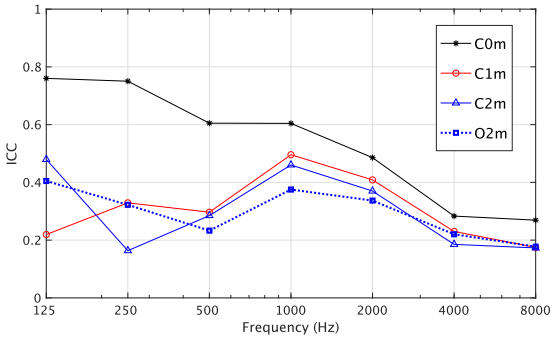


Fig. 9: ICC per octave band between the middle layer microphone (FL) and the top layer microphone (TpFL) obtained from the xylophone.

5 Summary

The relationship between the degree of correlation in the vertical direction and the listening impression in

3D audio was investigated. Two experiments were conducted to investigate the effect of different vertical correlations on the listening position. From the results of Experiment 1 using pink noise with different correlation values for the top and middle layers, the lower the vertical correlation values, the wider the listening area where the impression does not change from the central listening position.

From the results of Experiment 2 by Scheffe's pairwise comparison (Nakaya's variation) using impulse responses obtained by setting up microphones in an actual concert hall, there was a tendency to perceive a sense of spatial spaciousness when the cardioid microphones were positioned away from the middle layer.

ICC between the microphones in the middle and top layers from the impulse responses used in the experiment showed a tendency for the correlation values in the lower-middle frequency band to be close to zero when cardioid microphones were positioned 2 m higher than the middle layer. The results of Experiment 1 showed that the lower the vertical correlation, the wider the sweet spot area where the impression does not differ from the central listening position. This suggests that in Experiment 2, depending on the polar pattern and height of the top-layer microphone, the lower the vertical correlation value, the wider the sweet spot area. This may have led to a clear difference in impression at the off-center listening positions when the omnidirectional microphones were used and the cardioid microphones were set close to the middle layer.

ACKNOWLEDGMENT

JSPS KAKENHI Grant Number 21H03761 supported this research.

References

[1] Hidaka, T., Beranek, L., and Okano, T., "Interaural Cross-Correlation, Lateral Fraction and Low- and High-Frequency Sound Levels as Measures of Acoustical Quality in Concert Halls," *J. ASA*, 98, 1995.

[2] Zotter, F. and Frank, M., "Efficient Phantom Source Widening," *Archives of Acoustics*, 38, 2013.

[3] Gribben, C. and Lee, H., "The Frequency and Loudspeaker-Azimuth Dependencies of Vertical Interchannel Decorrelation on the Vertical Spread of an Auditory Image," *J-AES*, 66(7/8), 2018.

[4] ITU-R-BS.2159-9, "Multichannel sound technology in home and broadcasting applications," *ITU Recommendation*, 2022.

[5] Lee, H., "Multichannel 3D Microphone Arrays: A Review," *J. Audio Eng. Soc.*, 69(1/2,), 2021.

[6] Lee, H. and Gribben, C., "Effect of Vertical Microphone Layer Spacing for a 3D Microphone Array," *Journal of Audio Engineering Society*, 62(12), pp. 870–884, 2014.

[7] Ryaboy, A., "Exploring 3D: A subjective evaluation of surround microphone arrays catered for Auro-3D reproduction," *AES 139th Convention, convention paper 9431*, 2015.

[8] Howie, W., Martin, D., Benson, D. H., Kelly, J., and King, R., "Subjective and objective evaluation of 9ch three-dimensional acoustic music recording techniques," *AES International Conference on Spatial Reproduction - Aesthetics and Science, conference paper*, 2018.

[9] Hamasaki, K., Hiyama, K., and et al., "Advanced Multichannel Audio System with Superior Impression of Presence and Reality," *Audio Engineering Society 116th Convention paper, 6053*, 2004.

[10] Streicher, R. and Everest, F. A., *The New Stereo Soundbook*, Audio Engineering Associates, third edition, 2006.

[11] Howie, W., King, R., Martin, D., and Grond, F., "Subjective Evaluation of Orchestral Music Recording Techniques for Three-Dimensional Audio," *142nd AES Convention, convention paper 9797*, 2017.

[12] Kamekawa, T. and Marui, A., "Evaluation of recording techniques for three-dimensional audio recordings: Comparison of listening impressions based on difference between listening positions and three recording techniques," *Acoust. Sci. and Tech.*, 41(1), pp. 260–268, 2020.

[13] Lindberg, M., "TUVAYHUN, Recording sessions Photo Gallery," *https://www.lindberg.no/gallery-2l-171/*, 2021.

[14] Rayleigh, L., "On our perception of sound direction," *hilosophical Magazin Series*, 6(13), pp. 214–232, 1907.

[15] Blauert, J., *Spatial Hearing, 2nd edition*, MIT Press, 1997.

[16] Blauert, J., "Sound localization in the median plane," *ACOUSTICA*, 22, pp. 205–213, 1969.

[17] Satoh, S., *Statistical Sensory Testing*, Union of Japanese Scientists and Engineers Publishing, (available in japanese language) edition, 1985.