OPEN ACCESS Freely available online
T. Kamekawa and A. Marui, "How Different Vertical Interchannel Cross-Correlation Affects the Widening of the Listening Area," *J. Audio Eng. Soc.*, vol. 74, no. 9, pp. 612–623 (2026 Sep.).
https://doi.org/10.17743/jaes.2026.0281.

How Different Vertical Interchannel Cross-Correlation Affects the Widening of the Listening Area

TORU KAMEKAWA,* *AES Fellow* , **AND ATSUSHI MARUI,** *AES Member*
(kamekawa@ms.geidai.ac.jp) (marui@ms.geidai.ac.jp)

Musical Creativity and the Environment, Tokyo University of the Arts, Tokyo, Japan

In auditory spatial perception, the degree of correlation between signals reaching the ears is thought to contribute to the sense of horizontal spaciousness. On the other hand, vertical spatial impressions depend on the signal's frequency response rather than on the correlation between the two ears. A previous experiment investigated 3D audio reproduction using pink noise with different interchannel cross-correlation (ICC) values between the top and middle layers of loudspeaker signals (vertical ICC). It found that as vertical correlation decreases, the listening area expands without changing the impression from the central listening position (i.e., sweet spot). This paper investigates in more detail whether the result that lower values of vertical ICC result in a wider sweet spot also applies to real recording situations. The results of an experiment using impulse responses obtained by setting up microphones in an actual concert hall suggest that increasing decorrelation between middle-layer and height-layer microphone signals, achieved by changing the polar pattern and height of the microphones, resulted in lower vertical ICC values, thus widening the sweet spot area.

Keywords: Head Tilt, Sweet Spot, Vertical Spatial Impressions

1 INTRODUCTION

"Main microphone" systems, designed to capture both direct and reverberant sound in a recording venue, are a critical component of classical music recording. Various such recording methods have been proposed, evaluated through numerous formal and informal comparisons. In Wöhr and Nellessen's experiment comparing two-channel stereo using an orchestra as source material, the AB method, which placed omnidirectional microphones at 2-m intervals, received high praise for its overall impression [1]. In a comparison of five-channel surround sound conducted by the AES Japan Section, in which the authors also participated, methods such as the Decca Tree system, employing omnidirectional microphones spaced apart, received high praise for their spatial impression [2].

In the field of 3D audio recording, various formats have also been compared [3–7]. Lee broadly classified current 3D microphone arrays into "horizontally and vertically spaced," "horizontally spaced and vertically coincident," and "horizontally and vertically coincident" arrays [3]. He compared different arrays subjectively or objectively, generally suggesting that significant differences can be perceived between them with respect to both spatial and timbral attributes. It also suggests that the perceived qualities of 3D microphone arrays may depend on both contextual factors and experimental conditions.

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 [7]. Howie et al. proposed another recording technique for 22.2 multichannel sound based on the Decca Tree [8] using omnidirectional microphones and compared their array with Hamasaki et al.'s array and a higher-order Ambisonics capture system [9]. That study found that the two differently spaced arrays showed similar ratings across all subjective attributes 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 [10], the spaced array in which microphones are set apart from each other was also rated as more "realistic" than the coincident array, which is based on Ambisonics. While spaced arrays in concert hall recordings are often employed in current 3D audio recording practices, there is debate over the appropriate positioning and polar

*To whom correspondence should be addressed, email: kamekawa@ms.geidai.ac.jp.

pattern choices for upper-level ("height channel") microphones [3, 4].

In auditory perception of space, horizontal sound-image localization and sense of spaciousness are based on level and time differences between the left and right ears [11–14]. Meanwhile, vertical sound imaging is perceived from differences in frequency response along the direction of sound arrival (spectral cues or direction-determining bandwidth) caused by the shape of the ear [15]. In particular, high-frequency ranges are considered important for median plane localization [16].

Regarding the sense of horizontal spaciousness, the low correlation between the left and right signals contributes to it. In the aforementioned two-channel [1] and five-channel recordings [2], the method of recording with omnidirectional microphones spaced at certain intervals, so-called "spaced omni," likely achieved higher ratings by lowering the interaural cross-correlation (IACC). On the other hand, for the vertical image spread (VIS), spectral cues (frequency that determines vertical direction [15]) are essential, and the change in VIS due to the degree of correlation between the vertical and horizontal signals depends on the frequency response of the contents [17, 18].

Lee recommends placing the microphones in the top layer as close as possible to the middle-layer microphones [3], because the distance between the top-layer and middle-layer microphones causes phase interference at the listening position, resulting in timbre degradation. He proposed a top-layer microphone array with cardioid microphones at the same height as the middle layer, facing upward [4]. On the other hand, in actual recording sites, top-layer microphones are often set 1–2 m above middle-layer microphones [3, 19]. Lee investigated the objective differences between various recording methods for 3D audio [20]. They hypothesized that the arrays would exhibit considerable differences in the perceived magnitudes of apparent source width and the size of the listening area, owing to large differences in interchannel cross-correlation (ICC) across the early segments of the main-layer impulse responses.

1.1 Sweet Spot

Various studies have also examined the so-called "sweet spot," where the listening impression remains unchanged even when listeners move away from the central listening position. Peters et al. investigated how various 5.1 surround recording formats degrade when moved away from the sweet spot [21]. The results showed that the spaced omni method tended to produce less degradation, but this effect varied with the music used and the recording location. In addition, the limitations of binaural headphone monitoring for comparing various listening positions were revealed. There is research by Riedel and Zotter showed that a line source with closely spaced speakers has the effect of widening the sweet spot compared to a spot source with widely spaced speakers [22]. Besides, research by Kiyan et al. showed that lowering the correlation between the left and right channels is related to the expansion of the sweet spot [23].

However, these results depend on the sound source used, and it remains unclear how the left and right ICCs affect the expansion of the sweet spot. In addition, Lee and Gribben found no difference in impression due to differences in the height of the upper microphone. It is suggested that the perceived results were mainly associated with vertical interchannel crosstalk in the signals of each height layer and the magnitude and pattern of spectral change at the listener's ear caused by each layer [4].

1.2 Head Rotation

It is well known that binaural hearing is important for horizontal localization. On the other hand, Hebrank and Wright reported that binaural hearing is not always necessary for vertical localization [24, 25]. They showed that sound image localization in the median plane is possible in one ear (monaural) by providing feedback to the response. This suggests that spectral cues are more important than binaural cues for vertical localization.

Kim et al. have conducted detailed experiments on the effects of head movements on spatial impressions [26]. Their results indicated that listeners moved their heads to a greater extent when judging source width or envelopment than when judging source direction or timbre.

Martens and Han examined the relationship between the auditory spatial diffuseness of upward and head movements [27]. They concluded that listeners could discriminate a "without-height" decrease in auditory spatial diffuseness to a significant extent when they rolled (tilted) their heads from side to side. Those same listeners were unable to discriminate auditory spatial diffuseness when their heads were held stationary.

The authors also investigated the frequency bandwidth required for the reproduction of the top layer within a 3D audio system by comparing the difference between various full-band 22.2 multichannel audio stimuli and those same stimuli filtered to restrict the frequency bandwidth of top layer reproduction [28]. Comparisons of left-ear and right-ear signals measured with a dummy head revealed a significant difference in the high-frequency range above approximately 1 kHz when the head was tilted.

1.3 Structure of This Paper

A previous study by the authors, presented at the 158th AES Convention [29], investigated how differences in the correlation of loudspeaker signals in the vertical direction affect the listening impression, using experiments with pink noise with different correlation values [29]. The current study expands upon that research with additional experiments, new results, and analysis.

This paper is organized as follows: In SEC. 2, the previous experiment using noise will be described, which examines how differences in correlation between upper and lower signals affect the range of the sweet spot. Results obtained with an actual microphone array and a solo-instrument sound source will be presented in SEC. 3. The objective differences between the sweet spot and off-center positions will be discussed in SEC. 4. The relationship between the mea-

surement results of the source signals and perceptual results from the listening test will also be discussed. Finally, SEC. 5 will conclude the paper with a summary.

2 VERTICAL CORRELATION AND LISTENING AREA

In a previous experiment [29], the following two types of pink noise were compared:

- Four-channel uncorrelated pink noise played back from two channels in the middle layer (FL, FR) and two channels in the top layer (TpFL, TpFR). The ICC value r for upper and lower channels (vertical ICC) was almost zero.
- Two-channel uncorrelated pink noise played back from two channels in the middle layer (FL, FR) and simultaneously played back from the top layer (TpFL, TpFR); vertical ICC r was almost 1.

The difference between the two was perceived as a difference in timbre rather than a change in VIS. When the ICC value was low, changes in timbre and sound image were negligible, even when the listener moved to a different position in the room. However, when the ICC value was high, the sound image was felt to change significantly when the listening position was moved only slightly from the sweet spot. In the experiment, therefore, the authors attempted to determine the physical area of listening positions in which the sound impression does not change (i.e., the expansion of the sweet spot).

The ICC value r between two loudspeaker signals (s_1 and s_2) is calculated as the maximum absolute value of the ICC function (ICCF) [14].

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

A lag time (τ) can be set with the ICCF to compensate for any delay between the signals. Considering the distance between the left and right ears, 1-ms lag is used.

2.1 Method for Previous Experiment

The experiment was conducted using a 22.2 multichannel sound system in Studio B at Tokyo University of the Arts. Pink noises with different vertical ICC values, described above, were used as stimuli. 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 vertical ICC values, such as "High" (correlation value $r = 1$), "Mid" ($r \simeq 0.6$), and "Low" ($r \simeq 0.1$) were compared. Fig. 2 shows the ICCs of the middle layer (FL) and top layer (TopFL) per octave band of High, Mid, and Low pink noise. Four different directions of movement were investigated: left, forward, backward, and downward.

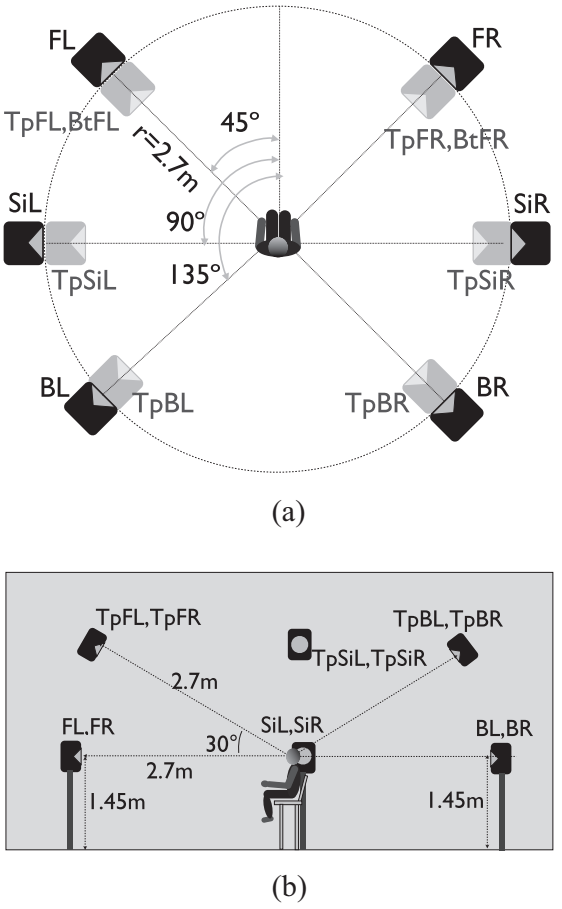


Fig. 1. Layout of the listening experiments. (a) The horizontal direction and (b) section.

Table 1. Combination of middle and top layer loudspeakers used in the experiment.

	Middle	Top	Angle
Front	FL, FR	TpFL, TpFR	±45°
Side	SiL, SiR	TpSiL, TpSiR	±90°
Rear	BL, BR	TpBL, TpBR	±135°

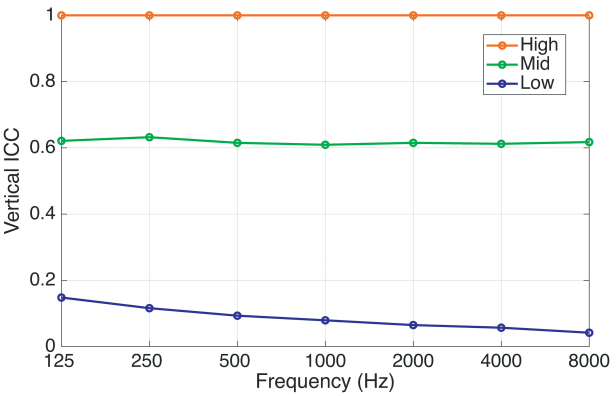


Fig. 2. ICC per octave band between the middle layer (FL) and top layer (TpFL) of High, Mid, and Low pink noise.

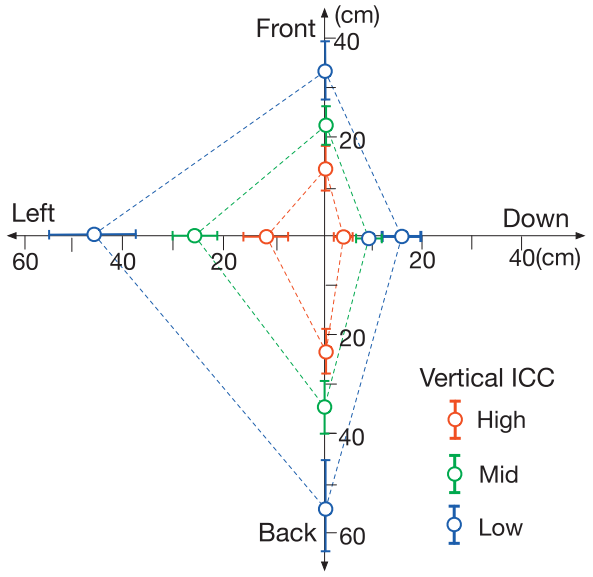


Fig. 3. Differences in perceived distance and vertical ICC values with the central position for each direction of movement. Circles indicate mean values and error bars indicate 95% confidence intervals. ICC values are as follows: High $r = 1$, Mid $r \simeq 0.6$, and Low $r \simeq 0.1$.

The participants were first asked to listen at the central listening position: the location equidistant to all loudspeakers. They then moved gradually to the left and stopped at the point at which their impression shifted from the central listening position. The distance they moved from the center was recorded as d_c. Similarly, the listening position was obtained in four directions (left, forward, backward, and downward while crouching).

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 (mean age = 28.1 years, SD = 16.6 years). They all reported normal hearing and had technical ear training for more than 2 years.

2.2 Results of Previous Experiment

2.2.1 Vertical ICC and the Sweet Spot Size

A three-way analysis of variance (ANOVA) [30] 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 (where the listening impression does not change for 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 summarized in Fig. 3, 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 vertical ICC values (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 Frequency Band Contributing to the Expansion of the Sweet Spot

To find the frequency below which it is acceptable to have correlated signals without reducing the sweet spot size, an additional experiment was executed with the following steps.

1. The low-frequency components below the cut-off frequency f_c in the top layer were duplicated in both the top and middle layers.
2. 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 the previous experiment.
3. The cut-off frequency was initially set at 10 kHz, and if d was less than d_c, f_c was reduced by an octave.
4. When d reached the same position as d_c, f_c was raised by a 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 cutoff filter was created using the Max8 filtergraph function (gain = 1.0, Q = 0.7, −12 dB/octave).

The stimuli and the listener's movements used in the additional experiment were identical to those in a previous experiment. The downward direction, where d_c was expected to be small in the preliminary experiment, was excluded.

2.3.1 Cut-Off Frequency at the Edge of the Sweet Spot

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 edge of the sweet spot d_c (where the listening impression does not change in the case of vertically uncorrelated noise) as the dependent variable. The results showed no significant differences in the direction of movement [$F(2, 63) = 0.78, p = 0.46$] or in the playback speaker position [$F(2, 63) = 0.62, p = 0.54$].

Fig. 4 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 the experiment 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 when the vertical ICC values are high (approaching 1.0) in the frequency band below 150–200 Hz, 75% of listeners will have the same listening area as when the vertical ICC value is almost zero at all frequency bands.

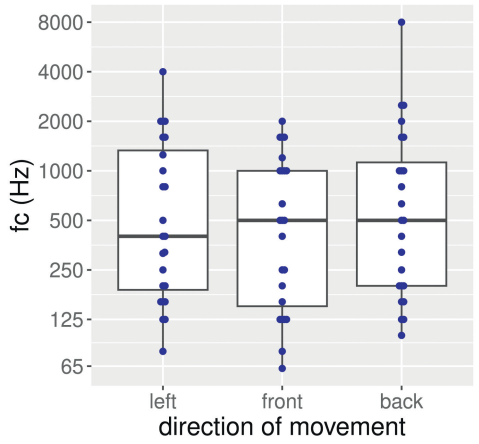


Fig. 4. Scatter plots and box plots of frequency f_c where the difference between the upper and lower uncorrelated cases and the cut case is unclear.

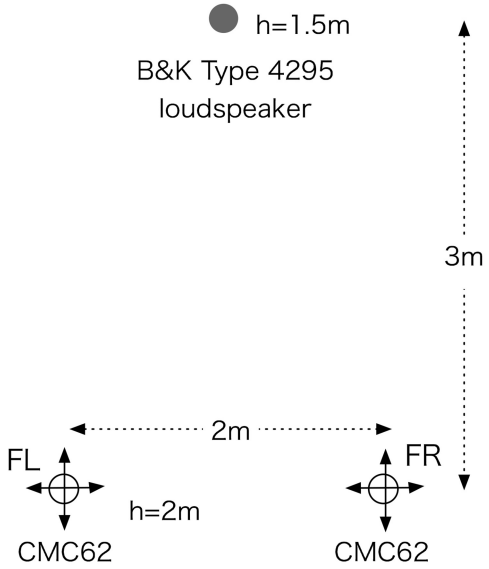


Fig. 5. Layout of microphones with measured impulse responses.

3 EXPERIMENT SIMULATING RECORDING IN AN ACTUAL HALL

A previous experiment showed that lower vertical ICC values resulted in broader impression areas that did not differ from the center listening position. This experiment therefore investigated how the impression of the listening position changes as the height and polar pattern of the top-layer microphones vary.

3.1 Sound Stimuli for the Experiment

The sound sources for the experiment were obtained from impulse responses measured at the university's small concert hall, Hall 6. This hall features 240 seats, a floor space of approximately 450 m^2, a height of about 9 m, and a reverberation time of roughly 1.6 s at 500 Hz, with measurements taken using the setup shown in Fig. 5. The microphones in the top layer were placed at the same position in the horizontal plane as FL and FR in Fig. 5 with the variations in polar pattern and height as shown in Table 2. The cardioid microphones were placed with their directivity axes pointing upward to avoid direct sound from the stage. The other two microphone sets were using omnidirectional microphones (Om0 and Om2). In Om0, the same omnidirectional microphones' signals for the middle layer (2 m high) were used in the top layer. Impulse responses were measured using a 2-s high-to-low (20 kHz–20 Hz) processing sweep sine signal. Using these impulse responses, excerpts of a xylophone and a cello recorded in an anechoic room [31] were convolved.

Table 2. Height and polar pattern of top-layer microphones.

	Model Name	Polar Pattern	Height (m)
Ca0	CMC64	Cardioid	2
Ca1	CMC64	Cardioid	3
Ca2	CMC64	Cardioid	4
Om0	CMC62	Omnidirectional	2
Om2	CMC62	Omnidirectional	4

Note. Om0 used the same signal in the top layer as the omnidirectional microphone used in the middle layer (2 m high).

3.2 Method for the Experiment

When comparing actual studio recordings, the difference due to the height of the upper microphone was not as clear in areas outside the sweet spot as with the pink noise used in the previous experiment. Therefore, in this experiment, two sets of speakers were prepared in advance, and the similarity was compared between a central position and a position away from the central position without the listener moving.

The following two methods were used in the listening experiment.

- Experiment 1: The similarity of the impressions of the three recording techniques at the three listening positions using the multidimensional scaling (MDS) method was examined. MDS is a set of methods for discovering "hidden" structures in multidimensional data without verbal evaluation [32].
- Experiment 2: The impressions of the five types of top-layer microphone arrays at two different listening positions were evaluated using a rating scale method.

Finally, each result was summarized, and the impression of differences from the recording method and the listening position were observed. The listening experiments were conducted using the middle (FL, FR) and upper (TpFL, TpFR) layers of the 22.2 multichannel sound reproduction system in Studio B (the same system used in a previous experiment).

The signals from the aforementioned FL, FR, TpFL, and TpFR microphones were played from the corresponding four loudspeakers. Furthermore, four identical loudspeakers were placed at a position 30 cm to the left of the four loudspeakers as speaker set S (side), so that impressions

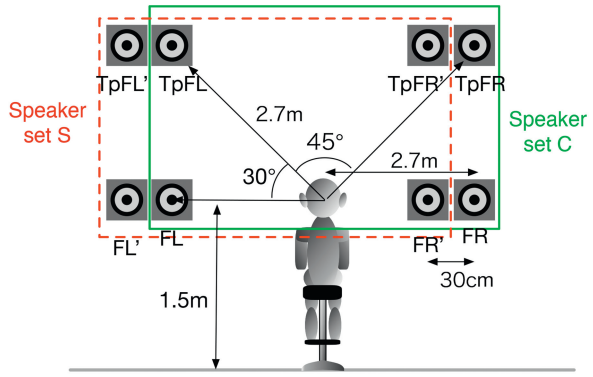


Fig. 6. Layout of speakers for the listening experiments.

at positions displaced from the central listening position (speaker set C) could be compared simultaneously without physically moving (Fig. 6). The value was set to 30 cm because it was the threshold at which a difference was detectable in the authors' preliminary comparison and because it was approximately the same as the vertical ICC of 0.6 in the previous experiment. A black acoustic transparency screen was also installed in front of the loudspeakers to prevent them from being visible to the experimental participants.

A total of ten stimuli (five microphone arrays × two speaker sets) for each xylophone and cello were compared. First, in Experiment 1, participants heard two stimuli randomly selected from the ten stimuli and rated their similarity on a seven-point scale using the GUI created in Max8. The participants were allowed to listen as many times as they wished and compared 45 combinations of xylophone and cello sound sources.

Experiment 2 was then conducted, in which participants listened to ten different stimuli one by one and used a slider to answer whether or not their impressions applied to the following 12 attributes. They were based on evaluation words extracted from a comparative experiment of 3D audio microphones that the authors previously conducted [10]:

- Rich,
- Bright,
- Hard,
- Wide sound image (width),
- Near,
- Clear sound image (clear),
- Listener envelopment (LEV),
- Wide space (broad),
- More reverberation (rev),
- Clear localization (loc),
- The sense of being there (presence), and
- Preference (pref).

The volume of each sound source was adjusted to approximately 70 dB (LAeq) at the central listening position. Fourteen students from the university participated in the experiment (mean age = 22.8 years, SD = 5.8 years). They all reported normal hearing and had technical ear training for more than 6 months. The duration of the experiment varied across participants but averaged approximately 60 min.

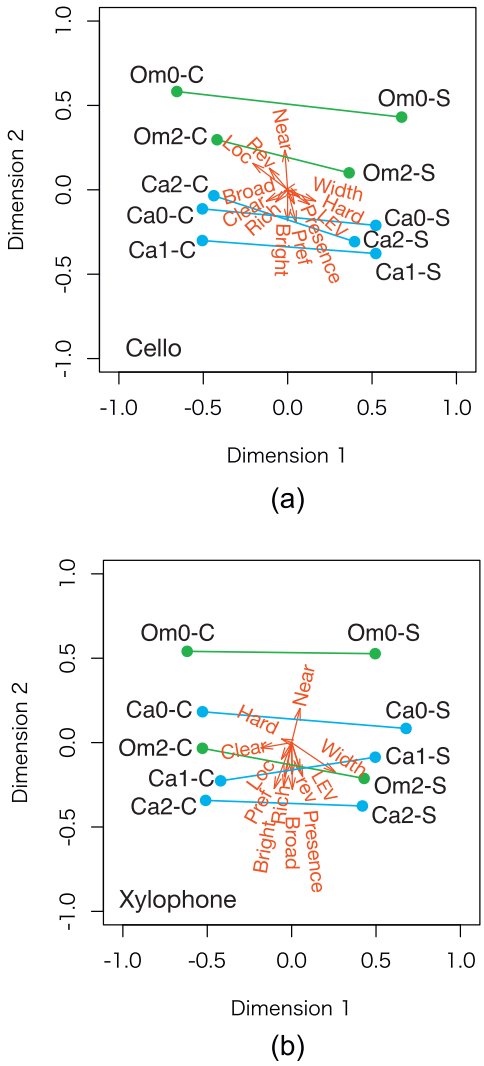


Fig. 7. Result of the rating scale of Experiment 2 superimposed on the figure of distribution of the stimuli obtained from INDSCAL of Experiment 1 in the case of the cello (a) and xylophone (b). The same microphone arrays, with different combinations of listening positions, are connected by lines.

3.3 Result of Experiment

From the results of Experiment 1, the similarity comparison was analyzed using individual differences scaling (INDSCAL) via the smacof package in R [32]. INDSCAL is an MDS method, which is a method of describing the perceptual distance between stimuli as the physical distance within the configuration. Fig. 7 shows the result of the INDSCAL of Experiment 1. The labels in the figures indicate the stimuli (Om0, Om2, Ca0, Ca1, and Ca2) and listening positions (the speaker sets C and S are hereafter referred to as the listening positions). The Stress-1 value did not decrease significantly, even in analyses with three or more dimensions, so a 2D analysis was adopted.

The distance between the symbols in Fig. 7 shows the dissimilarity of the impression for cello (upper) and xylophone (lower). From Fig. 7, it is estimated that the impressions are divided into left and right, depending on the listening po-

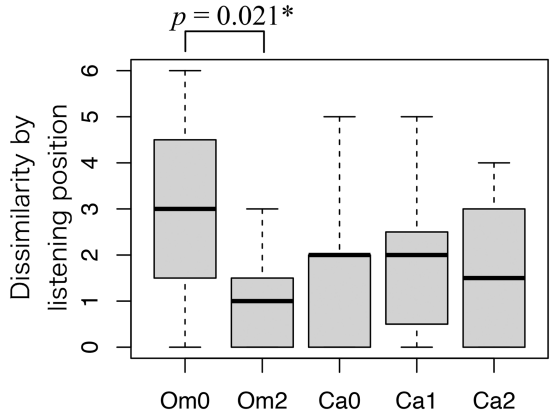


Fig. 8. Box plot of the dissimilarity between the center (C) and the listening position 30 cm to the side (S) in the case of the cello. An asterisk (*) indicates $p < 0.05$.

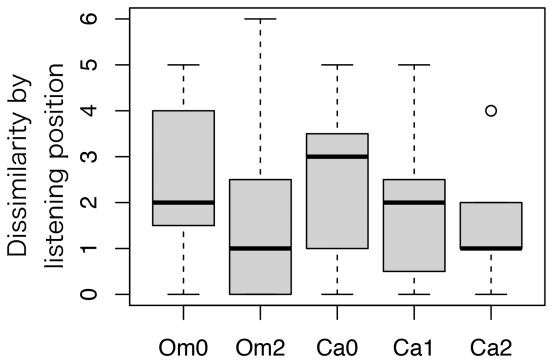


Fig. 9. Box plot of the dissimilarity between the center (C) and the listening position 30 cm to the side (S) in the case of the xylophone.

sition. Especially in the case of the cello, the impressions of Om0-C and Om0-S differ significantly, even when the same microphones are used for the top and middle layers.

The authors then calculated the correlation coefficient between the coordinate values for each dimension of each recording technique, as obtained by INDSCAL, and the scores for each evaluation attribute from Experiment 2. The arrows in Fig. 7 show the common distribution of each evaluation attribute and each stimulus. As it can be seen from the direction of the arrows of the attribute, when the listening position is shifted laterally, the sound image width is perceived as "wide," and the stimuli using the same omnidirectional microphones in the top layer as in the middle layer (Om0-C and Om0-S) are perceived as "near."

The analysis focused on the differences between the central listening position and laterally shifted positions for each microphone array to see the effect of the top microphone height. Statistical significance was then determined using ANOVA and multiple comparison tests. To clarify the difference in impressions between center and right-shifted listening positions, participants who were unable to distinguish between them were excluded. The analysis excluded subjects whose responses had a maximum-to-minimum difference of 2 or less. As a result, two people's responses for the cello stimuli and three for the xylophone stimuli were excluded. The results showed no significant differences for the xylophone. Still, for the cello, the difference between the central and sideways shifted position of Om0 tended to be larger than that of Om2 (Figs. 8 and 9).

An ANOVA was conducted on the responses for each evaluation word between the central and the sideway shifted listening position from the results of Experiment 2, showing significant differences in "rev" [$F(4, 110) = 4.18, p < 0.01$] for cello and "bright" [$F(4, 100) = 3.02, p = 0.021$] for xylophone due to differences in microphone array and "width" [$F(1, 100) = 5.59, p = 0.020$] for xylophone due to differences in listening position. An interaction effect between listening position and microphone was also found for "bright" [$F(4, 100) = 3.37, p = 0.013$] and "presence" [$F(4, 100) = 3.38, p = 0.012$] in the case of the xylophone.

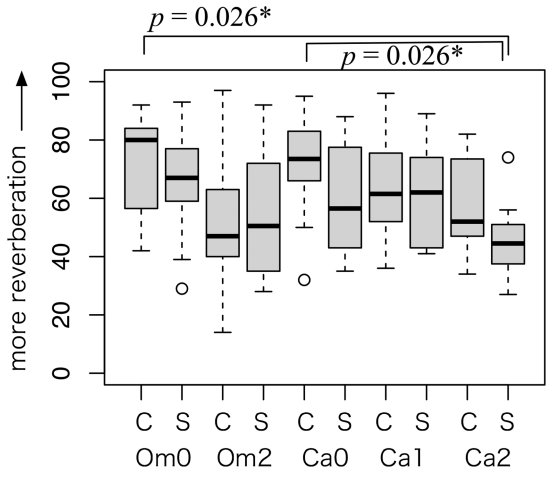


Fig. 10. Box plot of "rev" between the center (C) and the listening position 30 cm to the side (S) in the case of the cello. An asterisk (*) indicates $p < 0.05$.

The box plots for the combinations in which differences in listening positions were observed for each attribute are shown in Figs. 10 and 11. They show the result of the comparison of the "rev" of the cello and the "presence" of the xylophone. Although no statistically significant differences were observed, similar to the differences in microphone arrays, the impressions from listening positions C and S for Om0 tended to be larger than those for the other microphone arrays.

4 OBJECTIVE MEASUREMENT RESULTS AND DISCUSSION

4.1 Vertical Interaural Cross-Correlation

The results of a previous experiment [29] suggest that low vertical ICC widens the sweet spot, the listening area where the impression is not different from that at the central listening position. Furthermore, the results of the additional experiment suggest that regardless of the direction of the playback loudspeaker, if the vertical ICC values are low in the frequency band above 150–200 Hz, 75% of listeners

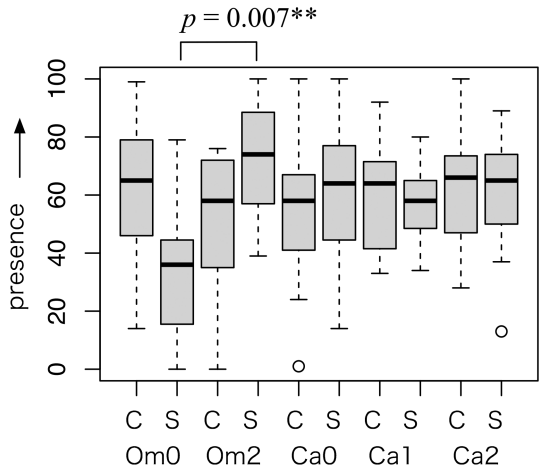


Fig. 11. Box plot of "presence" between the center and the listening position 30 cm to the side in the case of the xylophone. Asterisks (**) indicate $p < 0.01$.

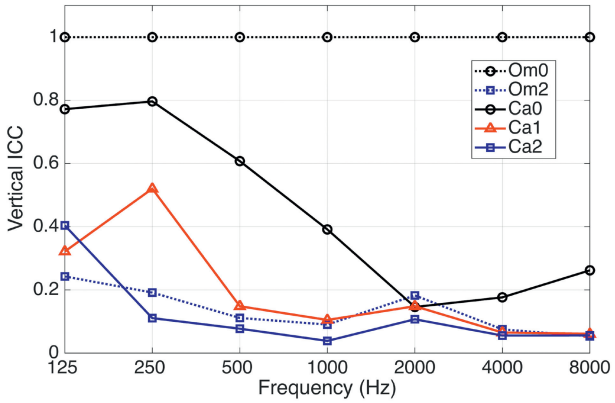


Fig. 12. ICC per octave band between the middle-layer microphone (FL) and the top-layer microphone (TpFL) obtained from the impulse responses.

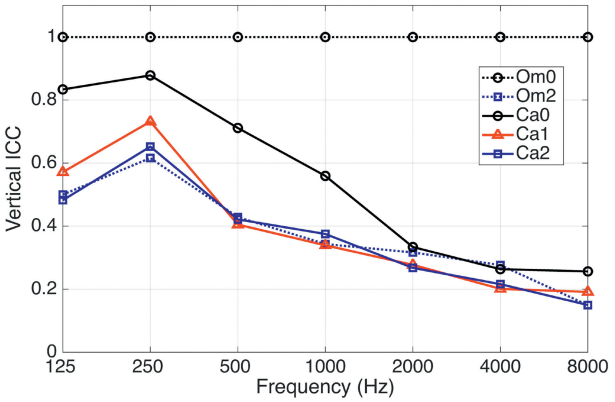


Fig. 13. ICC per octave band between the middle-layer microphone (FL) and the top-layer microphone (TpFL) obtained from the cello.

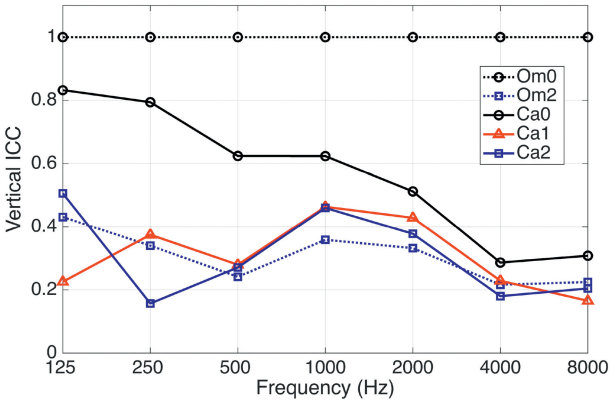


Fig. 14. ICC per octave band between the middle layer microphone (FL) and the top layer microphone (TpFL) obtained from the xylophone.

can obtain a listening area similar to that of the uncorrelated case. Experiment 1, the comparison of similarity, 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 similar listening impressions can be maintained at the off-center listening position when top-layer microphones are placed away (vertically) from the middle layer. Therefore, the vertical ICC of each stimulus used in the experiment was examined.

Fig. 12 shows ICC per octave band obtained from the impulse responses of the microphones in the middle (FL) and the top layer (TopFL). Om0, which uses the same omnidirectional microphones in both the middle and top layers, has a value of 1 across all bands, since 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.

Figs. 13 and 14 show ICC per octave band obtained from the cello 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 the time window by 100 ms within a 300-ms window, excluding values where the level was more than 20 dB below the overall r.m.s. average.

The bandwise ICCs of the sources convolved with cello 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 Ca1, Ca2, and Om2 were particularly close for cello. In all figures, Om2 and Ca2, with microphones placed 2 m higher than the middle layer, exhibited lower ICC values than Ca0 and Ca1 in the mid-low–frequency range below 1 kHz.

When these results are related to the listening experiment in SEC. 3, it is suggested that the high correlation between the upper and lower ICCs is related to the difference in impressions observed when listening from a central listening position and a position shifted to the right (Fig. 8, Fig. 9, Fig. 10, and Fig. 11). On the other hand, the close proximity of ICC values across frequency bands may also be related to the large variability in listening experiment results.

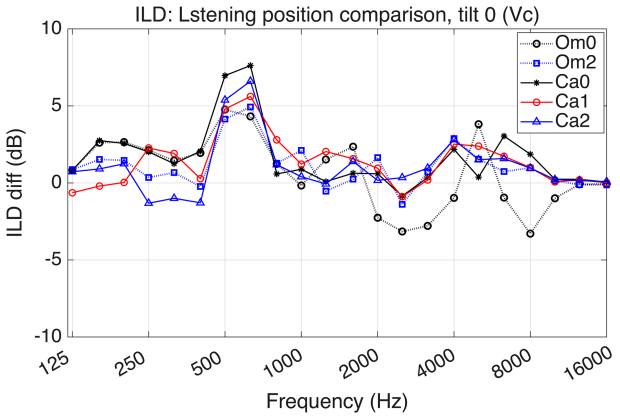


Fig. 15. ILD per 1/3-octave–band differences between the center and right-shifted listening position for each microphone setting obtained from a dummy head recording of the cello.

4.2 The Acoustic Features Corresponding to the Listening Impression

The interaural level difference (ILD), interaural time difference (ITD), and IACC were examined as acoustic features associated with the listening impression. In the two playback speaker sets (C and S) used in the listening experiment, five different sound sources (Om0, Om2, Ca0, Ca1, and Ca2) were played and recorded using a dummy head with miniature microphones (DPA4060) set at the entrance to the ear canal.

For each value, the difference between the central listening position (loudspeaker set C) and the position 30 cm to the right (loudspeaker set S) was calculated. The results showed that for IACC and ITD, there was no apparent difference among the microphone arrays. However, for ILD, there was a relatively significant difference between Om0 (black dotted line) and the other microphone arrays from 2 to 8 kHz (Fig. 15). On the other hand, these differences do not correspond to the band below 1-kHz differences observed in the vertical ICC.

The difference in ICC between the top and middle layers observed here above 1 kHz is consistent with the characteristics reported in Lee and Dale's study using microphone arrays, in which the upper microphones were more than 1 m above the middle level [20]. This is also related to Gribben and Lee's [17, 18] findings that lower ICC between the top and middle layers above 1 kHz contributes to VIS.

Kiyan et al. [23], who examined acoustic features within the sweet spot, found that while there was a relationship between ICC and IACC, this relationship depended on the sound source, and pointed out that it was necessary to examine the relationship with ILD. This is consistent with the present analysis, which found differences in ILD rather than IACC or ITD. They did not investigate the frequency characteristics of ICC and ILD, and it may be key to future investigations of acoustic features in the sweet spot.

4.3 Effects of Head Rotation

As already mentioned, head movement plays an important role in spatial perception. Therefore, assuming that human head motion may act as a cue for perceiving differences in the vertical sound imaging, each stimulus was recorded with a dummy head at the listening position with the head upright (the angle difference between the right and left ears was 0°) and with the right ear tilted upward by 15°. A mannequin with a movable head was used as a dummy head and measured by tilting the head (Fig. 16). Miniature microphones (DPA4060) were set at the entrance to the ear canal of the dummy head.



Fig. 16. A dummy head used in the measurement. Miniature microphones (DPA4060) are set at the entrance to the ear canal. This shows the head tilted at a 15° angle.

The frequency response (1/3-octave power spectrum) of the right ear signal of the dummy head for differences between the center and the off-center position to the right, in the case of the head not tilted [Fig. 17(a)] and the right ear tilted 15° upward [Fig. 17(b)] were compared with the cello recording. From these graphs, when the head is tilted, Om0, which exhibits high vertical correlation, shows significant differences in frequency response due to variations in listening position across the entire bandwidth. It was suggested that the difference in frequency response when the head is tilted may be a clue to the difference in impression depending on the listening position. This trend was also observed in xylophone (Fig. 18). These results also correspond to those obtained in the previous experiment (described in SEC. 2.3), indicating that the frequency band associated with the effect of vertical ICC expanding the sweet spot is above 150–200 Hz.

This can be related to findings from previous studies by Kim et al. [26] and Martens and Han [27], in which head movement improved the perception of vertical sound-image width and envelopment. That is, even when the listening position is shifted from the center, if the vertical ICC is low, the frequency difference caused by head tilt will be smaller, reducing the perceived difference and expanding the sweet spot. However, these results were obtained under limited conditions, and further experiments will be required to draw general conclusions.

5 CONCLUSION

The relationship between the degree of correlation in the vertical direction (vertical ICC) and the listening impression in 3D audio was studied. From the results of a previous experiment using pink noise with different correlation values for the top and middle layers, the lower vertical ICC, the wider the sweet spot, where the impression does not change

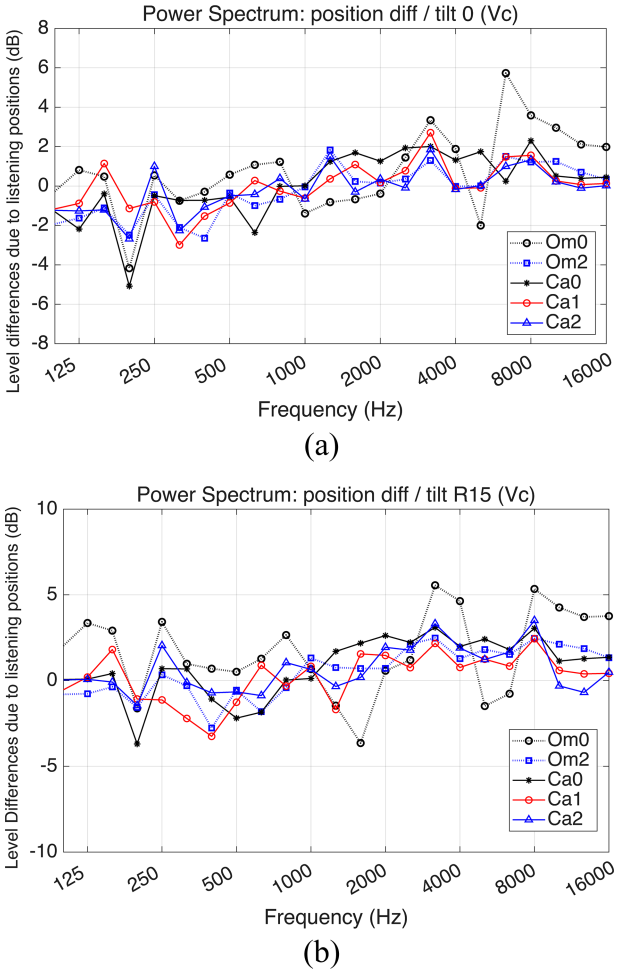


Fig. 17. Power spectrum of the 1/3-octave–band differences between the center and right-shifted listening position for each microphone setting obtained from the right ear of the dummy head recording of the cello. (a) The case of the head not tilted and (b) the case of the head tilted (the right ear is tilted 15° upward).

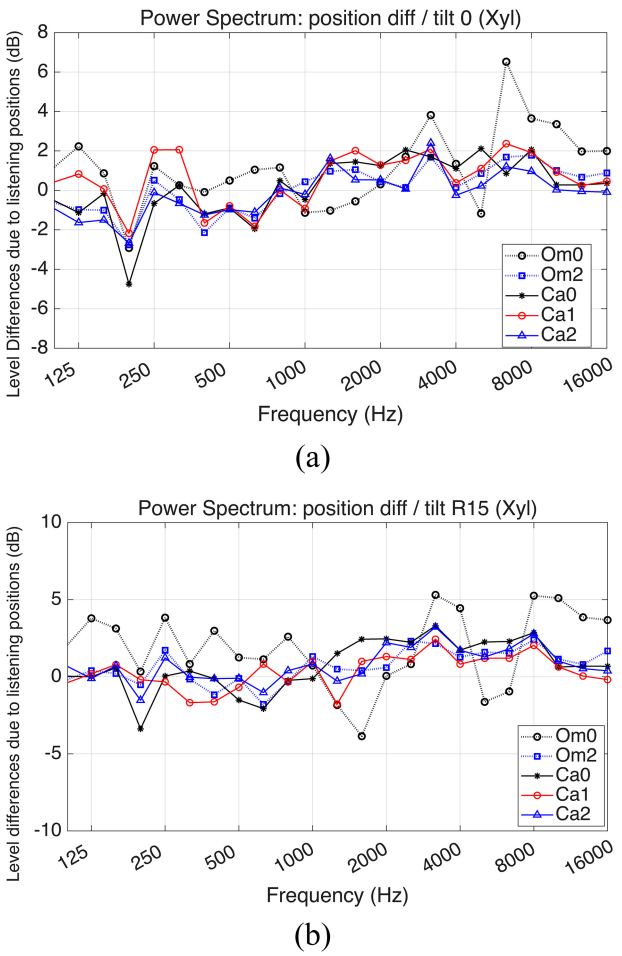


Fig. 18. Power spectrum of the 1/3-octave–band differences between the center and right-shifted listening position for each microphone setting obtained from the right ear of the dummy head recording of the xylophone. (a) The case of the head not tilted and (b) the case of the head tilted (the right ear is tilted 15° upward).

from the central listening position. It was also suggested that the frequency band related to the effect of vertical ICC expanding the sweet spot is above 150–200 Hz.

Therefore, in this paper, two experiments were conducted to investigate whether the sweet spot could be widened by changing the height and polar pattern of the top-layer microphones in an actual hall, thereby changing the vertical ICC. The experiment used instrumental sounds recorded in an anechoic room convolved with impulse responses measured at different directivities and heights of the top-layer microphones as stimuli. INDSCAL and rating scales were used to investigate the differences in impressions between central and off-center listening positions. From the results, when the same microphones used in the middle layer were employed in the top layer to increase the vertical ICC, the difference in impression between the center and off-center listening positions tended to be greater.

The vertical ICCs of the stimulus sounds used in the experiment tended to have lower correlation values in the middle and lower frequency bands when cardioid microphones were used in the upper layers or positioned above the middle layers. It was also shown that when the vertical ICC is high, the difference in the frequency response of the dummy head's single ear is greater at the center and off-center listening positions when the head is tilted. These results suggest that changing the microphone's polar pattern and height on the top layer, and lowering the vertical ICC, would yield a wider sweet spot in which the impression does not differ from that at the central listening position.

6 ACKNOWLEDGMENT

This research was supported by JSPS KAKENHI (Grant-in-Aid for Scientific Research) (B) Grant Number JP21H03761.

7 REFERENCES

[1] M. Wöhr and B. Nellessen, "Untersuchungen zur wahl des hauptmikrofonverfahren," *VDT tmt*, pp. 106–120 (1986).

[2] T. Kamekawa, A. Marui, and H. Irimajiri, "Correspondence Relationship Between Physical Factors and Psychological Impressions of Microphone Arrays for Orchestra

Recording," in Proc. *123rd Conv. Audio Eng. Soc.*, paper 7233:1–14, New York, NY, USA (2007 Oct.).

[3] H. Lee, "Multichannel 3D Microphone Arrays: A Review," *J. Audio Eng. Soc.*, vol. 69, no. 1/2, pp. 5–26 (2021 Jan./Feb.), https://doi.org/10.17743/jaes.2020.0069.

[4] H. Lee and C. Gribben, "Effect of Vertical Microphone Layer Spacing for a 3D Microphone Array," *J. Audio Eng. Soc.*, vol. 62, no. 12, pp. 870–884 (2014 Dec.), https://doi.org/10.17743/jaes.2014.0045.

[5] A. Ryaboy, "Exploring 3D: A Subjective Evaluation of Surround Microphone Arrays Catered for Auro-3D Reproduction," in Proc. *139th Conv. Audio Eng. Soc.*, paper 9431:1–10, New York, NY, USA (2015 Oct.).

[6] W. Howie, D. Martin, D. H. Benson, J. Kelly, and R. King, "Subjective and Objective Evaluation of 9ch Three-Dimensional Acoustic Music Recording Techniques," in Proc. *AES Int. Conf. Spatial Reproduction - Aesthetics and Science*, pp. 1–11, Tokyo, Japan (2018 Aug.).

[7] K. Hamasaki, T. Nishiguchi, K. Hiyama, and K. Ono, "Advanced Multichannel Audio System With Superior Impression of Presence and Reality," in Proc. *116th Conv. Audio Eng. Soc.*, paper 6053:1–12, Berlin, Germany (2004 May).

[8] R. Streicher and F. A. Everest, *The New Stereo Soundbook* (Audio Engineering Associates, Pasadena, CA, USA, 2006), 3rd ed.

[9] W. Howie, R. King, D. Martin, and F. Grond, "Subjective Evaluation of Orchestral Music Recording Techniques for Three-Dimensional Audio," in Proc. *142nd Conv. Audio Eng. Soc.*, paper 9797:1–10, Berlin, Germany (2017 May).

[10] T. Kamekawa and A. Marui, "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. Tech.*, vol. 41, no. 1, pp. 260–268 (2020 Jan.), https://doi.org/10.1250/ast.41.260.

[11] L. Rayleigh, "On Our Perception of Sound Direction," *Lond. Edinb. Dublin Philos. Mag. J. Sci.-Ser.6*, vol. 13, no. 74, pp. 214–232 (1907), https://doi.org/10.1080/14786440709463595.

[12] J. Blauert, *Spatial Hearing* (MIT Press, Cambridge, MA, USA, 1997), 2nd ed, https://doi.org/10.7551/mitpress/6391.001.0001.

[13] T. Hidaka, L. L. Beranek, and T. Okano, "Interaural Cross-Correlation, Lateral Fraction and Low- and High-Frequency Sound Levels as Measures of Acoustical Quality in Concert Halls," *J. Acoust. Soc. Am.*, vol. 98, pp. 988–1007 (1995 Aug.), https://doi.org/10.1121/1.414451.

[14] F. Zotter and M. Frank, "Efficient Phantom Source Widening," *Arch. Acoust.*, vol. 38, no. 1, pp. 27–37 (2013 Mar.), https://doi.org/10.2478/aoa-2013-0004.

[15] J. Blauert, "Sound Localization in the Median Plane," *Acta Acust. un. Acust.*, vol. 22, no. 4, pp. 205–213 (1969 Nov.).

[16] M. Morimoto, M. Yairi, K. Iida, and M. Ito, "The Role of Low Frequency Components in Median Plane Localization," *Acoust. Sci. Tech.*, vol. 24, no. 2, pp. 76–82 (2003 Mar.), https://doi.org/10.1250/ast.24.76.

[17] C. Gribben and H. Lee, "The Frequency and Loudspeaker-Azimuth Dependencies of Vertical Interchannel Decorrelation on the Vertical Spread of an Auditory Image," *J. Audio Eng. Soc.*, vol. 66, no. 7/8, pp. 537–555 (2018 Jul./Aug.), https://doi.org/10.17743/jaes.2018.0040.

[18] C. Gribben and H. Lee, "The Perception of Band-Limited Decorrelation Between Vertically Oriented Loudspeakers," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 28, pp. 876–888 (2020 Jan.). https://doi.org/10.1109/TASLP.2020.2969845.

[19] M. Lindberg, "TUVAYHUN — Beatitudes for a Wounded World (2L-171)," (2021), https://www.lindberg.no/gallery-2l-171/ (accessed 29 Jul. 2026).

[20] H. Lee and J. Dale, "3D Microphone Array Comparison: Objective Measurements," *J. Audio Eng. Soc.*, vol. 69, no. 11, pp. 871–887 (2021 Nov.), https://doi.org/10.17743/jaes.2021.0038.

[21] N. Peters, S. McAdams, and J. Braasch, "Evaluating Off-Center Sound Degradation in Surround Loudspeaker Setups for Various Multichannel Microphone Techniques," in Proc. *123rd Conv. Audio Eng. Soc.*, paper 7197:1–20, New York, NY, USA (2007 Oct.).

[22] S. Riedel and F. Zotter, "Surrounding Line Sources Optimally Reproduce Diffuse Envelopment at Off-Center Listening Positions," *JASA Express Lett.*, vol. 2, no. 9, paper 094404:1–8 (2022 Sep.), https://doi.org/10.1121/10.0014168.

[23] S. Preihs, and J. Peissig R. Kiyan, "Signal Dependencies of the Immersion Sweet Area in Multichannel Loudspeaker Reproduction," *J. Audio Eng. Soc.*, vol. 73, no. 5, pp. 282–292 (2025 May).

[24] J. Hebrank and D. Wright, "Are Two Ears Necessary for Localization of Sound Sources on the Median Plane?," *J. Acoust. Soc. Am.*, vol. 56, no. 3, pp. 935–938 (1974 Sep.), https://doi.org/10.1121/1.1903351.

[25] J. Hebrank and D. Wright, "Spectral Cues Used in the Localization of Sound Sources on the Median Plane," *J. Acoust. Soc. Am.*, vol. 56, pp. 1829–1834 (1974 Dec.), https://doi.org/10.1121/1.1903520.

[26] C. Kim, R. Mason, and T. Brookes, "Head Movements Made by Listeners in Experimental and Real-Life Listening Activities," *J. Audio Eng. Soc.*, vol. 61, no. 6, pp. 425–438 (2013 Jun.).

[27] W. L. Martens and Y. Han, "Discrimination of Auditory Spatial Diffuseness Facilitated by Head Rolling While Listening to 'With-Height' Versus 'Without-Height' Multichannel Loudspeaker Reproduction," in Proc. *AES Int. Conf. Spatial Reproduction - Aesthetics and Science*, paper P4-3:1–11, Tokyo, Japan (2018 Jul.), https://aes.org/publications/elibrary-page/?id=19608.

[28] T. Kamekawa, A. Marui, and W. Howie, "The Effect of Head Tilt on the Perception of the Top Layer of Three-Dimensional Audio," in Proc. *AES Int. Conf. Spatial and Immersive Audio*, paper 24:1–8, Huddersfield, UK (2023 Aug.).

[29] T. Kamekawa and A. Marui, "Correlation Between Middle and Top Layer Loudspeaker Signals and the Listening Area in 3D Audio Reproduction," in Proc. *158th Conv. Audio Eng. Soc.*,

paper 10219:1–8, Warsaw, Poland (2025 Jun.), https://aes2.org/publications/elibrary-page/?id=22855.

[30] S. Satoh, *Statistical Sensory Testing* (Union of Japanese Scientists and Engineers Publishing, Tokyo, Japan, 1985).

[31] V. Hansen and G. Munch, "Making Recordings for Simulation Tests in the Archimedes Project," *J. Audio Eng. Soc.*, vol. 39, no. 10, pp. 768–774 (1991 Oct.).

[32] I. Borg, P. J. F. Groenen, and P. Mair, *Applied Multidimensional Scaling* (Springer, Cham, Switzerland, 2013). https://doi.org/10.1007/978-3-319-73471-2.

THE AUTHORS



Toru Kamekawa



Atsushi Marui

Toru Kamekawa received his Bachelor of Acoustics Design from the Kyushu Institute of Design (currently Kyushu University), Fukuoka, Japan, in 1983. He also received a Ph.D. in design from Kyushu University in 2015. He joined the Japan Broadcasting Corporation (NHK) in 1983 and gained extensive experience as a recording engineer. He participated in various experiments on surround sound production for high-definition television, achieving great success. In 2002, he joined Tokyo University of the Arts as an associate professor in the Department of Musical Creativity and the Environment, and he has been a full professor since 2010. His research focuses on spatial audio and covers auditory perception and musical acoustics. Currently, he is conducting research on archiving acoustic spaces using 3D audio. In 2024, he received the Fellowship Award from the AES for diverse research on immersive audio production and reproduction and for years of outstanding contributions to AES activities.

•

Atsushi Marui received an M.Sci. degree from Pennsylvania State University and pursued his doctoral studies at The University of Aizu, Japan, from which he received a Ph.D. degree in computer science and engineering. He was also a doctoral student in sound recording at the Schulich School of Music, McGill University, where he finished all but the dissertation. He is currently a professor at the Faculty of Music, Tokyo University of the Arts, Japan. His research interests include signal processing algorithms and their psychological evaluations of digital audio effect processors, especially on nonlinear distortion processors for musical instruments, artificial reverberation, human–computer interaction, and auditory display.