Audio Engineering Society

Convention e-Brief 581

Presented at the 148th Convention
2020 June 2–5, Online

This Engineering Brief was selected on the basis of a submitted synopsis. The author is solely responsible for its presentation, and the AES takes no responsibility for its contents. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Audio Engineering Society.

Directional Dependency of Subjective Sound Pressure Perception on Three-Dimensional Sound

Akihito Nakai[1], Minoru Tsuji[1], and Toru Chinen[1]

[1] *Sony Corporation, R&D Center Tokyo Laboratory 20, Tokyo Japan*

Correspondence should be addressed to Author (Toru.Chinen@sony.com)

ABSTRACT

Sony launched 360 Reality Audio in 2019 that provides a new music experience using object-based spatial audio technology. In this music experience, sounds arrive from various directions. The direction from which the sound arrives affects the subjective sound pressure sensitivity [1, 2, 3]. Sivonen measured the subjective sound pressure sensitivity in seven directions on the left half of the horizontal plane and on the upper front quarter in the median plane [1]. In this e-brief, the number of directions is increased to 31, which includes a whole horizontal plane and the elevations lower than the front center for measurements with three band-limited signals (i.e., 93 conditions in total). As a result, 74 conditions are observed with statistically significant differences.

1 Introduction

Sony announced 360 Reality Audio [4] in CES 2019, and the service was launched in October 2019. The 360 Reality Audio empowers the creation of immersive music experience, such as feeling as if you were at a live concert. The creation of three-dimensional (3D) audio introduces a new process that involves placing audio objects in a 3D space to workflow, which widens the variety of possible artistic expressions for creators. However, such new processes require additional considerations for perceptual differences when listening to 3D sounds. Loudness [5, 6, 7] is an interesting topic in both academic and industrial fields. The desire to better understand how perceptions differ in 3D space have led to investigations on the differences of subjective sound pressure perceptions in 3D space under practical environments in 3D audio content creation.

2 Objectives

In audio content creation, mixing is a common process that adjusts the volumes of multiple audio tracks. The creation of 3D audio content also uses this process, and mixing engineers adjust volumes while they place audio tracks as objects in 3D space. Often, when an object is placed in a position that is different from its original position, the subjective loudness changes, and the mixing engineer must re-adjust the volume of the object.

The idea that changes in subjective loudness depend on the arrival direction of sounds has been considered. Sivonen studied the directional dependency of subjective sound pressure perception in a study in which subjects compared and matched the loudness of sounds by adjusting the volume of the directional sound to a reference sound presented from their front center (Azimuth 0°, Elevation 0°) [1, 2, 3]. The participants used three third-octave-band noises with different center frequencies (400 Hz, and 1 and 5 kHz) in two different playback volumes (45 and 65 dB SPL) as stimuli. The stimuli were presented from a pair of loudspeakers located at the subjects' front and seven others in target directions (five in left-half horizontal plane and two in upper front median plane). The researchers concluded that there was a significant difference in the sound

pressure perception depending on the arrival direction of the sound [1]. This study also utilized head-related transfer function simulated sounds to study the effects of reverberance, the bandwidth of stimuli [2], and individual differences [3].

These studies provide 3D audio content creators with insight into subjective sound pressure perception, which is dependent on the arrival direction of sound. Creators rely on their hearing when positioning objects and adjusting volumes. In this paper, we extend research conducted by Sivonen to include the right half horizontal plane and upper and lower horizontal planes to better understand subjective sound pressure perceptions in 3D space.

3 Experiment

Directions

Subjects listened to and compared two sounds in the center of a loudspeaker system presented from pairs among 32 loudspeakers placed in the positions shown in Table 1. Subjects listened to various conditions that comprised the reference sound at the subject's front and the target sound from a loudspeaker in a different direction. The distances of the loudspeakers should ideally be equal in all directions; however, due to the limitations of the experiment environment, we could not achieve equal distances.

Signals

Three third-octave-band pink noises with a sampling frequency of 48 kHz were used as stimuli. The pink noises had center frequencies (CFs) of 250 Hz and 1 and 8 kHz. Subjects compared only sounds with the same CF in every condition.

Levels

Playback volumes of the reference sounds were measured with a noise meter (Onosokki LA5560) and were set at 71.8, 65.5, and 65.3 dB SPL for 250 Hz and 1 and 8 kHz, respectively. These volumes were determined by considering the calculated loudness levels [8] and the measurement results from the experiment room. Playback volumes of the target sounds were measured with an omnidirectional microphone (Earthworks M50) at the center of the loudspeaker system and were aligned to match the playback volumes of reference sounds for each CF. The playback volumes were determined by using the averaged level of three measurements.

The total number of conditions thus amounted to 93 (three CFs with levels in 31 directions).

Procedures

The subjects sat on a chair placed in the center of the loudspeaker system and were instructed how to use the volume adjustment controls. Subjects were asked to face their front when they compared the loudness of the sounds. The subjects' head and chin were not fixed during the experiment. The subjects pressed the start key to play the reference and target sounds, which originated from two different directions, and compared the loudness. The subjects then adjusted the volume of the target sounds to match their perceived loudness to the reference sound. The adjustment interval was 0.1 dB and the number of comparisons and adjustment controls were unlimited for each condition. The subjects continue with the procedure until they had finished all 93 experiment conditions.

Subjects

There were 29 subjects in total. All participants had previous experience in audio content creation (all subjects are professional musicians) in various genres: eight played classic, four played rock, four played jazz, four played R&B, two played Latin, and there were seven in other areas (including production/direction). All subjects were provided with instructions on the procedure before the experiment. After receiving the instructions, they participated in a practice experiment session.

Sequences

The playback sequence was triggered by the subjects. Reference and target sounds were presented with a 0.6 second pause between the two when the subjects pressed the playback key. The subjects could trigger the sequence again after a 0.6 second pause from the end of the target sound. Initial playback volumes of target sounds were randomized (in the range of ± 3 dB relative to the reference) and the subjects were instructed to avoid hesitating when adjusting volumes. The playback volumes of the reference sounds were kept constant (see *Levels* section). The limits of volume adjustment for the target sounds were set at ±10 dB relative to the reference sound. All subjects' answer range lies within these limits ($-10.0 < \text{volume} < 10$ dB). The sequence of 93

conditions was randomized for each subject; however, a counterbalance of the sequences was not performed. All subjects were offered a break during the experiment after they finish the forty-fifth condition.

Supplemental information

Figure 1 shows the experiment system configuration. Figure 2 shows the experiment setup. The area of the room was 44.52 m^2 (width 5.3 m, length 8.4 m).

4 Results

We obtained the adjusted volumes in 31 target directions for three different CFs from 29 subjects. The arithmetic mean of the adjusted volumes across 29 subjects with 95% confidence intervals in 31 directions along with the loudspeaker positions are listed in Table 1. As can be seen in Table 1, the confidence intervals of the means for 93 conditions (31 directions with 3 CFs) do not cross zero in 74 conditions.

Table 2 shows the number of conditions observed, with significant differences by elevations sorted in three layers by elevations. In the middle layer (ML), the Elevation was 0°. The directions of the upper layer (UL) and bottom layer (BL) were Elevation > 0° and Elevation < 0°, respectively. The denominator indicates the number of conditions (i.e., the number of directions multiplied by three CFs). The numerator indicates the number of conditions that have significant differences.

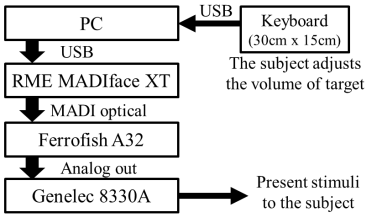


Figure 1 Device connections

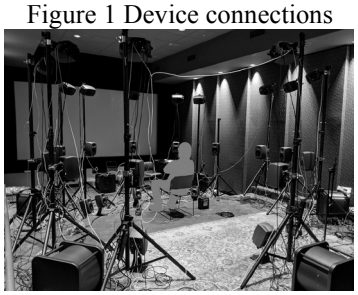


Figure 2 Experiment setup

Per the CF of the signal, there are significant differences in the directions shown in Table 3. The denominator indicates the total number of directions. The numerator indicates the number of directions that have significant differences.

The mean values with confidence intervals are plotted per CF in Figure 3.

Table 2 The number of conditions per layer

	Upper	Middle	Bottom	Total
Conditions	24/36	32/33	18/24	74/93

Table 3 The number of directions per CF

	250 Hz	1 kHz	8 kHz	Total
Directions	21/31	24/31	29/31	74/93

5 Considerations

Comparing the experiment results for each layer of the horizontal plane first, as shown in Table 2, the ML has the highest number of conditions with significant differences (32 out of 33), representing the total number of conditions minus one. There are significant differences for the UL (24 out of 36) and BL (18 out of 24), which represented almost two-thirds of the number of total conditions per each layer. In terms of the adjustment volumes in ML with the highest number of conditions with significant differences, the direction of azimuth 110° in ML (hereafter referred to as ML110°) marks the highest absolute value with −3.42 dB for CF 8 kHz. This implies that if a mixing engineer moves an object from one place to another (e.g., from ML0°, to ML110°), while maintaining the subjective loudness of the object, it may be necessary to lower the volume of the object. On the contrary, for the objects in BL0°, it may be necessary to increase the volume of the object while maintaining its subjective loudness.

Table 3 compares the number of directions with significant differences per CF of signals. The results indicate that CF 8 kHz has the highest number of significant differences (29 out of 31), which represents all the directions except two. The second highest number of directions with significant differences is observed in 1 kHz, with 24 out of 31.

Table 1 Mean values and 95% confidence intervals.
(Angles in the table include errors within ± 1.5°)

Azim. [°]	Elev. [°]	Dist. [m]	CF 250Hz [dB]		CF 1kHz [dB]		CF 8kHz [dB]	
			Mean	Conf. Int. (higher, lower)	Mean	Conf. Int. (higher, lower)	Mean	Conf. Int. (higher, lower)
0	0	2.00	0.00	−.—, −.−−	0.00	−.−−, −.—	0.00	−.−−, −.−−
30	0	2.01	−1.01	−0.61, −1.42	−0.91	−0.50, −1.31	−0.57	−0.02, −1.13
−30	0	1.98	−0.91	−0.50, −1.31	−1.35	−0.91, −1.78	−0.92	−0.37, −1.47
60	0	1.99	−2.51	−2.13, −2.89	−2.28	−1.86, −2.69	−3.09	−2.28, −3.91
−60	0	2.00	−1.45	−0.97, −1.93	−1.74	−1.30, −2.18	−1.52	−0.62, −2.41
90	0	2.00	−0.51	−0.10, −0.92	−2.79	−2.26, −3.31	−2.90	−2.03, −3.76
−90	0	2.02	−1.43	−1.03, −1.82	−2.69	−2.21, −3.17	−3.04	−2.02, −4.06
110	0	2.04	−0.39	−0.01, −0.77	−2.39	−1.98, −2.81	−3.42	−2.46, −4.39
−110	0	2.03	−1.23	−0.80, −1.66	−2.17	−1.70, −2.63	−3.14	−2.18, −4.09
135	0	2.01	0.54	0.92, 0.17	−2.42	−1.92, −2.93	−1.62	−0.57, −2.66
−135	0	2.02	0.04	0.39, −0.31	−2.28	−1.75, −2.81	−1.51	−0.50, −2.51
180	0	2.01	2.25	2.69, 1.81	−1.67	−1.13, −2.22	1.67	2.56, 0.77
0	30	2.36	0.75	1.14, 0.36	1.39	1.82, 0.97	−1.00	−0.32, −1.68
30	30	2.37	−0.69	−0.31, −1.07	−0.14	0.31, −0.59	−2.32	−1.52, −3.11
−30	30	2.34	−1.19	−0.76, −1.62	−0.61	−0.21, −1.02	−1.99	−1.18, −2.80
45	30	2.38	−0.15	0.25, −0.55	0.26	0.61, −0.10	−2.92	−2.06, −3.78
−45	30	2.38	−0.47	0.01, −0.95	−0.06	0.41, −0.53	−3.42	−2.65, −4.20
90	30	2.34	0.39	0.80, −0.03	−1.68	−1.20, −2.16	−4.24	−3.39, −5.09
−90	30	2.42	0.20	0.62, −0.21	−1.40	−0.97, −1.82	−4.14	−3.37, −4.92
110	30	2.42	0.11	0.58, −0.36	−0.38	0.10, −0.85	−3.45	−2.61, −4.28
−110	30	2.40	0.60	1.04, 0.15	−0.38	0.08, −0.83	−2.98	−2.10, −3.86
135	30	2.36	2.63	3.24, 2.03	−0.89	−0.50, −1.28	−1.06	−0.17, −1.94
−135	30	2.37	2.19	2.73, 1.64	−0.83	−0.39, −1.27	−1.54	−0.60, −2.48
180	30	2.36	0.93	1.34, 0.53	−0.12	0.38, −0.62	−0.53	0.56, −1.63
0	−30	2.01	2.36	2.76, 1.97	−1.86	−1.39, −2.33	−0.45	0.32, −1.22
30	−30	2.01	0.66	1.08, 0.24	−1.69	−1.22, −2.15	−1.18	−0.37, −1.99
−30	−30	2.01	1.00	1.38, 0.63	−1.58	−1.10, −2.05	−0.94	−0.34, −1.54
45	−30	2.01	−0.09	0.30, −0.47	−1.94	−1.50, −2.38	−2.06	−1.20, −2.91
−45	−30	2.01	−0.17	0.29, −0.62	−1.83	−1.38, −2.28	−1.01	−0.28, −1.73
135	−20	2.01	0.11	0.60, −0.38	−1.42	−0.98, −1.87	−1.63	−0.77, −2.48
−135	−20	2.00	−0.18	0.25, −0.61	−1.86	−1.38, −2.33	−1.14	−0.43, −1.85
180	−20	2.00	0.62	1.08, 0.16	−0.52	0.03, −1.07	1.16	1.84, 0.47

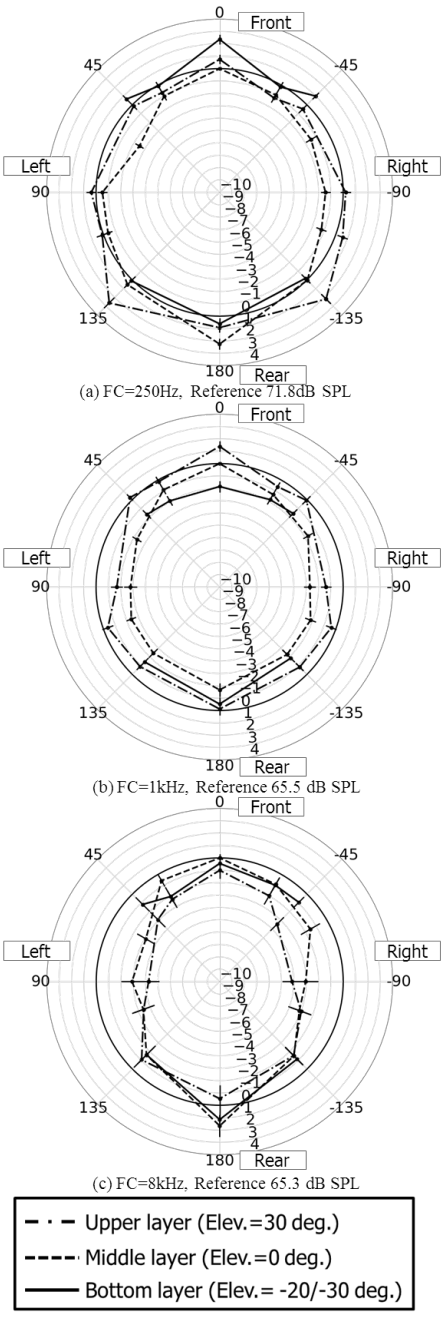


Figure 3 Mean of adjusted volumes across 29 subjects and its 95% confidence intervals per CF. (Each small dot with its confidence interval is the mean of answers in that direction, and the lines connecting the dots are straight line interpolations. The interpolation lines between 45° and 135° in the bottom layer were removed to avoid misleading appearances.)

Table 4. Three most distant volumes per CF

CF [kHz]	Rank	Azim. [°]	Elev. [°]	Mean [dB]
0.25	1	135	30	2.63
	2	60	0	−2.51
	3	0	−30	2.36
1	1	90	0	−2.79
	2	−90	0	−2.69
	3	135	0	−2.42
8	1	90	30	−4.24
	2	−90	30	−4.14
	3	110	30	−3.45

The maximum change in volume observed in all conditions is −4.24 dB in UL90° for 8 kHz. For UL90°, the change for 1kHz is −1.68 dB.

The differences between the CFs indicate that if a mixing engineer places an object from a position to another (e.g., from ML0° to UL90°), it may be necessary to adjust the volume of the object depending on its tone balance. The three maximum absolute values of the adjusted volumes for each CF are listed in Table 4, including CF 8 kHz. As Table 4 shows, the maximum volumes rise as the CF increases. It should also be noted that there were no target directions without any significant differences for all three CFs; here, the directions with a significant difference were only observed in one CF, where they were UL ± 45°, 110°, and 180°. Thus far in this study, we have analyzed the major changes based on the mean values across conditions; however, it is also important to consider the widths of the confidence intervals. The confidence intervals in 30 conditions for CF 8 kHz were counted in one-third of the widest confidence intervals (i.e., in 31 of 93 conditions). The wider confidence intervals suggest that there are more variances in the perceptional differences per subject. Therefore, we suggested volume adjustments to maintain loudness using the mean values in Table 3, although it may differ from person to person and using the values in the confidence intervals that are closer to 0 dB would suppress unintended excess changes. The higher variance observed in the higher CFs compared to other CFs is worthy of further study to unravel the perceptional differences associated with 3D sound.

6 Conclusion

We investigated the differences in subjective sound pressure perception in a 3D audio content creation environment using 32 loudspeakers. The experiment was conducted with 29 subjects who matched the loudness of sounds from reference and target directions in 93 conditions (i.e., three CFs in 31 directions). As a result, we obtained 74 conditions, each of which evidenced a significant difference.

The changes observed in the volumes in the target directions are suggestive of the importance of volume adjustments in audio objects that are to be moved from one position to another to maintain the subjective loudness.

References

[1] V. P. Sivonen, W. Ellermeier, “Directional Loudness in an Anechoic Sound Field, Head-Related Transfer Functions, and Binaural Summation” *Journal of the Acoustical Society of America* vol. 119, pp.2965–2980 (2006).

[2] V. P. Sivonen, “Directional Loudness and Binaural Summation for Wideband and Reverberant Sounds,” *Journal of the Acoustical Society of America* vol. 121, no. 5, pp. 2852–2861 (2007).

[3] V. P. Sivonen, W. Ellermeier, “Binaural Loudness for Artificial-Head Measurements in Directional Sound Fields,” *Journal of the Audio Engineering Society* vol. 56, no. 6, pp. 452–461 (2008).

[4] 360 Reality Audio Official Web Site https://www.sony.com/electronics/360-reality-audio

[5] H. Fletcher, W. A. Munson, “Loudness, Its Definition, Measurement and Calculation,” *Journal of the Acoustical Society of America* vol. 5, pp. 82–108 (1933).

[6] B. Sharf, “Dichotic Summation of Loudness,” *Journal of the Acoustical Society of America* vol. 45, pp. 1193–1205 (1969).

[7] E. Zwicker, U. T. Zwicker, “Dependence of Binaural Loudness Summation on Interaural Level Differences, Spectral Distribution, and Temporal Distribution,” *Journal of the Acoustical Society of America* vol. 89, pp. 756–764 (1991).

[8] ISO 532 B, “Acoustics―Method for calculating loudness level” (1975).