SS1-3 2022 International Workshop on Smart Info-Media Systems in Asia (SISA 2022), Sep. 15–16, 2022

The 3D Sound Localization Platform with Reflected Sounds

Yoshinori Kamizono*, Wataru Kobayashi*, Ittetsu Taniguchi*, Hiroki Nishikawa*, and Takao Onoye*
*Graduate School of Information Science and Technology, Osaka University, Japan

***Abstract*—This study proposes a 3D sound localization platform allowing users to freely set each sound reflection from their surroundings based on Head-Related Transfer Functions which hold spatial information. Reflected sounds seem to affect sound localization in a virtual space, however, not so many detailed studies have been conducted at this present, which seems necessary to have a platform that allows a detailed evaluation of the effects of reflected sounds. On the other hand, such a platform has not been reported, which is the motivation to propose the platform. The platform enables the evaluation of 3D sounds in a virtual space with freely parameter settings of reflected sounds. Experimental results show influences of reflected sounds on sound localization of sound sources and contributions of this platform to creating various 3D sound environments according to situations.**

I. INTRODUCTION

Technologies reproducing a physical space in a virtual space, such as XR and Metaverse, have been developed in recent years. A variety of mechanisms have been devised that work on the senses to create more realistic environments. In the field of hearing, it is focused on the experience of the 3D sound field in a virtual space [1]. For instance, 3D sound localization is used to create an appropriate 3D sound field according to situations adding information on direction and distance to sounds. In particular, various research is conducted to localize sounds virtually with high accuracy in affordable environments which consist of small speakers and headphones because the 3D sound fields are gradually becoming more common in general.

In addition, some studies have evaluated the effects of reflected sounds from surrounding environments to create more realistic 3D sound fields [2], [3]. It is said that sound reflections influence 3D sound localization accuracy, which represents the importance to analyze the detailed effects of reflected sounds from the surroundings. However, almost all studies have been conducted regarding the limited effects of reflected sounds. It seems necessary to have a platform that allows the free setting of parameters for each reflected sound and a detailed evaluation of the effects of reflected sounds, however, such a platform has not been reported.

Therefore, this study proposes a 3D sound localization platform allowing users to freely set each sound reflection from their surroundings. The platform enables a detailed evaluation of the effects of sound reflections on the perception of sound direction. The usefulness of this platform is demonstrated by evaluating the effective ways of setting sound reflections.

II. FACTORS RELATED TO 3D SOUND LOCALIZATION

A. HRTF

It is often adopted that inter-aural time difference (ITD) and inter-aural level difference (ILD) are the main factors in a listener's perception of sound direction. The left and the right directions from sound sources can be perceived, however, the front-back and the top-bottom directions are not clear with the above information alone. This phenomenon is called cone phenomenon and front-back confusion [4]. In human binaural ears, a sound is heard through reflections and diffraction with a head, auricle, and torso. Therefore, 3D sound localization can be improved by using detailed information on the influence of human bodies in addition to ITD and ILD. This information is expressed as a frequency characteristic, which is Head-Related Transfer Functions (HRTFs) [5].

Since HRTFs show the effects of human body shapes on the perception of sound direction, individual differences occur in the shape of peaks and dips. Various methods have been proposed to reproduce these individual differences with high accuracy [6]. However, the preparation of individual HRTFs requires a considerable amount of computation costs, which make it difficult to implement in mobile terminals. In particular, since the use of virtual spaces become familiar to the general public, it is important to create 3D sounds on mobile terminals with limited computational resources. Therefore, various attempts have been conducted to generate general HRTFs based on the characteristics of the shape of HRTFs [7]. On the other hand, problems of inaccuracy by the use of general HRTFs still exist in directional perception in the front-back and the vertical directions.

B. The effect of sound reflections

Normally, when humans listen to a sound, it is heard not only as a sound reaching human ears directly from a sound source but also as sounds reflected from walls. Therefore, it is important to analyze the effects of reflected sounds to reproduce a sound field space that more closely resembles reality. Reflected sounds also affect the perception of the direction of a sound source. The precedence effect is widely known as one of these effects [8]. The precedence effect is that when a series of sounds arrive at the ears within a certain period, the sounds are heard as a single sound, and the position of the sounds is mostly determined by the position of the first arrived sound. When a subsequent sound such as the reflected sound is further away from the preceding sound, the sound image is perceived to be expanded after the sounds are perceived at the position of the prior sound, because of a factor of difference in sound arrival delay and volume. Thus, the exact location of the sound source comes to be blurred. On the other hand, it is said that reflected sounds seem to assist to sense the perception of the direction of the sounds. It is reported that the use of reflected sounds can solve the problem of front-back confusion in HRTFs [2]. However, the detailed effects of sound reflections have not been clarified on the solution to the problem of front-back confusion of 3D sound localization in a virtual space. Almost all studies have been conducted regarding limited parameters of reflected sounds in surroundings and remained to evaluate the limited effects of reflected sounds. Moreover, it is required to evaluate the detailed effects of reflected sounds with a platform enabling to set the parameters of each reflection sound freely, however, there are no reports on such platforms.

Therefore, this study proposed a platform that generates multiple reflected sounds and allows detailed evaluation of the effects of each reflection sound by freely setting the parameters of each reflection sound.

III. THE 3D SOUND LOCALIZATION PLATFORM WITH REFLECTED SOUNDS

A. Overviews

The proposed platform provides a rectangular room composed of six reflective materials on the ceiling, floor, front, back, left, and right walls (Fig. 1). One sound source exists in the rectangular room. One early reflected sound happens on each reflective material. The sound arriving directly from the sound source and the reflected sounds from the six reflectors are performed with the 3D sound localization process using HRTFs that add direction and distance information to each sound. The 3D sound localization process uses an embedded 3D sound localization processing method that reduces the amount of computation and generates general HRTFs for use in the development of various applications in the future [7]. In the 3D sound localization process, spherical coordinates are used, in which parameters are distances (r) between listeners and sound sources, horizontal angles (θ), and vertical angles (ϕ). By adding direction and distance information to reflected sounds using HRTFs, various parameters of the reflected sounds can be set freely, and the effects of the reflected sounds can be evaluated from their frequency characteristics. The parameters of type, gain, distance, horizontal angle, and vertical angle can be freely set for a sound source. The parameters of distance and reflection coefficient (h) of reflective materials can be freely set for each reflected sound, which allows the delay and the gain of the reflected sounds to be freely set.

B. Implementation methods

This platform uses the mirror image method to determine the location of the reflected sounds [9]. When a sound is reflected from a wall and reaches a listener, the sound can be regarded to have reached the listener from an imaginary source located at the opposite position of the sound source with the wall as the axis of symmetry (Fig. 2). Therefore, when the position of the listener is an origin of the cartesian coordinates, the distance between the listener and the wall perpendicular to the x axis is defined as x_w, and the position of the sound source is defined as x_s, y_s, z_s, the position of the imaginary reflected sound image is given by Eq. 1. The n in Eq. 1 denotes the type of reflective walls in Fig. 1, corresponding to the positive and the negative symbols depending on the positional relationship with the listener and walls. Similarly, when y_w and z_w are defined as distances between the listener and the wall perpendicular to the y and the z axes respectively, the positions of the imaginary reflected sounds are shown in Eq. 2 and Eq. 3.

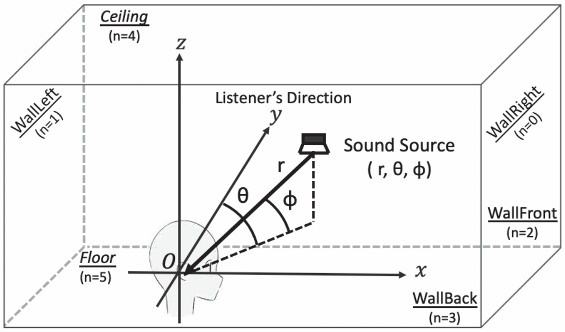

Fig. 1. The sound room of the proposed platform

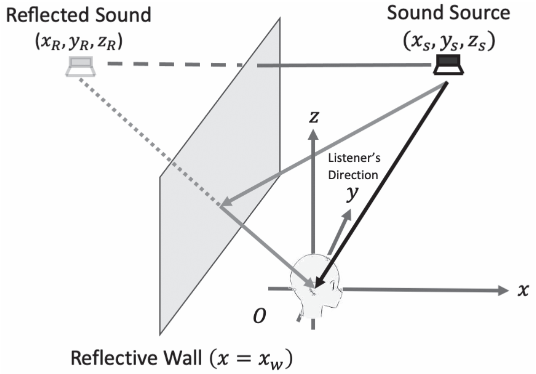

Fig. 2. Image source model

$$\begin{cases} x_R = 2(-1)^n x_w - x_s \\ y_R = y_s \\ z_R = z_s \end{cases} \tag{1}$$

$$\begin{cases} x_R = x_s \\ y_R = 2(-1)^n y_w - y_s \\ z_R = z_s \end{cases} \tag{2}$$

$$\begin{cases} x_R = x_s \\ y_R = y_s \\ z_R = 2(-1)^n z_w - z_s \end{cases} \tag{3}$$

Once the positions of the reflected sounds are determined, the arrival time differences between the direct sound and each reflected sound to the listener are calculated based on the determined position. After each reflected sound is delayed according to the arrival time differences, the 3D sound localization process is performed for each sound using HRTFs which enable to set detailed parameters of each reflected sound. After the 3D sound localization process, the direct sound and all reflected sounds are mixed and output, which prevents other delay differences from occurring except for the arrival time differences.

IV. EXPERIMENTS

A. Overviews

This experiment evaluated the effect of reflected sounds by focusing on the precedence effect, in which reflected sounds do not significantly change the position of a sound source but broaden a sound image. It is reported that humans perceive the front-back position of a sound source more easily when the sound source is set obliquely than in front or side [10]. Therefore, it is expected that the front-back perception of sounds can be improved by stretching a sound image of a sound source near in front or side to the position where the front-back perception of sounds is easier to perceive due to the precedence effect of reflected sounds. Based on the above prediction, this experiment shows the usefulness of the proposed platform allowing the user to set the sound reflections according to situations.

B. Environments

The experiment was conducted with 20s five male subjects under the following three conditions. First (Exp. 1), when the sound source is placed in front (0° in horizontal angle) and back (180° in horizontal angle), it is evaluated whether the reflected sounds are effective in judging the front and the back. Reflected sounds from the left and the right walls are used, which is expected to improve the judgment of the front-back position of the sound source by stretching the sound images to the left-right direction where it is easier to perceive the front-back perception. Second (Exp. 2), the reflected sounds are evaluated for judging the front-back position of a sound source positioned close to the side of listeners. The sound source is placed at horizontal angles of 85° and 95°, and reflected sounds from the front and the back walls are used, which is expected to improve the front-back judgment of the sound source position because the reflected sounds stretch the sound image to a position where the front-back perception is easier to perceive. Third (Exp. 3), the position of a sound source is the same as in the second condition. However, the position of reflected sounds is different, that only the sound reflected from the right wall is used. In this condition, it is expected that the reflected sounds do not have a stretching effect on the sound image and that the effect of the reflected sounds is not so apparent in converse with the other conditions.

The six types of sound sources are used, which are 1/3-octave band noise at 100 Hz (low-frequency narrow-band noise sound source), 1/3-octave band noise at 1 kHz (mid-frequency narrow-band noise sound source), 1/3-octave band noise at 10 kHz (high-frequency narrow-band noise sound source), pink noise (wide-band noise sound source with many low-frequency components), blue noise (wide-band noise sound source with many high-frequency components), and white noise (wide-band noise sound source with full coverage over the entire audible range). The effect of each frequency band on the localization of reflected sounds is evaluated. For other parameters, the distance of the sound source from a listener is fixed at 1.0 m, the reflection walls are all fixed at 2.0 m from a listener, and the reflection coefficient is fixed at $h = 0.9$ which is slightly lower in volume assumed as natural sounds than the direct arrival sound.

TABLE I
THE RESPONSE ITEMS AND THE CORRESPONDING SCORES

Items	Scores (pt.)
Felt a difference	2
Relatively felt a difference	1
Not sure	0
Relatively not felt a difference	-1
Not felt any difference	-2

TABLE II
RESULTS OF THE EVALUATION OF THE FRONT AND THE BACK SOUND AUDIBILITY(NARROW-BAND NOISE SOUND SOURCES)

	100 Hz		
	Non-Reflection	Reflection	P Value
Exp. 1	-1.000	0.000	0.394
Exp. 2	-0.400	-0.400	1.000
Exp. 3	-0.600	0.109	0.600

	1 kHz		
	Non-Reflection	Reflection	P Value
Exp. 1	-0.200	-0.200	1.000
Exp. 2	0.000	-1.600	0.140
Exp. 3	-0.600	0.400	0.142

	10 kHz		
	Non-Reflection	Reflection	P Value
Exp. 1	0.000	1.000	**0.034**
Exp. 2	0.400	0.800	0.587
Exp. 3	1.200	-0.800	**0.047**

The evaluation is conducted using questionnaires. Participants are asked whether or not they perceived different senses when the same type of sound source is played from the front and the back positions respectively. The response items and the corresponding scores are summarized in Table I. The scores are averaged to compare the level of different senses of the front and the back by the presence or the absence of reflected sounds.

C. Results

The results are evaluated using a two-tailed paired-samples t-test that shows statistically significant differences for some sound sources. The results are shown in Table II and Table III divided according to ranges of frequency band.

Bolded numbers indicate the sound sources that showed statistically significant differences in senses when the same type of sound source was played from the front and the back positions respectively. For narrow-band noise sound sources,

TABLE III
RESULTS OF THE EVALUATION OF THE FRONT AND THE BACK SOUND AUDIBILITY(WIDE-BAND NOISE SOURCES)

	Pink Noise		
	Non-Reflection	Reflection	P Value
Exp. 1	1.600	0.800	0.456
Exp. 2	0.000	1.200	0.070
Exp. 3	0.400	0.000	0.772

	Blue Noise		
	Non-Reflection	Reflection	P Value
Exp. 1	0.600	-0.400	0.351
Exp. 2	0.000	0.600	0.529
Exp. 3	1.600	-0.200	0.070

	White Noise		
	Non-Reflection	Reflection	P Value
Exp. 1	1.200	0.000	0.305
Exp. 2	-0.200	1.800	**0.011**
Exp. 3	0.400	-0.200	0.553

a significant difference of $p < 0.05$ was found for 1/3-octave band noise at 10 kHz (high-frequency narrow-band noise sound source) in Exp. 1 and Exp. 3. In the presence of reflected sounds from the right and the left wall, a different sense was caused by the differences in the position of the front and the back sound sources compared to no reflected sounds. On the other hand, when the sound source was positioned at 85° and 95° horizontal angles and the reflected sounds from the right wall were used, which was almost the same direction as the sound source, a different sense was less noticeable with the reflected sounds. It seemed that the different senses in the position of the front and the back sound sources are not improved with reflected sounds located at the almost same position as the sound source location. For narrow-band noise sound sources, left and right reflections far from the sound source seem to have a greater effect. In addition, the high-frequency sound source seems to be comparably easy to be affected by reflected sounds.

For wide-band noise sound sources, when the sound source was positioned at 85° and 95° in horizontal angles and the reflected sounds from the front and the back walls were used, a significant difference of $p < 0.05$ was found for white noise (wide-band noise source with full coverage over the entire audible range). On the other hand, when the left and the right reflections were applied, no effect of the reflections was found on any of the three types of sound sources at the horizontal angles of 0° and 180°, and the horizontal angles of 85° and 95°. For broadband noise sources that extend from low to high frequencies, front and rear reflections seem to be effective for the sense of the difference between front and back.

As described above, some different senses exist in the perception of front and back sounds according to the position of the sound reflections. In addition, the reflected sounds, which affected the front-back perception, differed according to the positions and the types of the sound source. Therefore, it is important to evaluate the effects of each individual reflected sound according to the surrounding environments of listeners with settings of appropriate parameters for the reflected sounds. Other parameters such as delays and gains of the reflected sounds and the effect of the upper and the lower reflected sounds on the localization of sound sources can also be evaluated in this platform. Therefore, the proposed platform can contribute to the evaluation of various sound reflection effects in the creation of a 3D sound virtual environment.

V. CONCLUSIONS

This study proposed a 3D sound localization platform that allows users to freely set individual reflected sounds according to situations. The proposed platform allows users to freely set parameters of reflected sounds that seem to affect the directional perception of sounds and evaluate the effects of sound reflections in detail.

An experiment was conducted to test the difference in the sense of front-back of a sound source with and without the reflected sounds to show the usefulness of the platform. This experiment discovered that the differences in the position of the reflected sound affected the front-back sense of the location of the sound source. Therefore, the requirement to evaluate the effects of individual sound reflections according to the listener’s surroundings was demonstrated in the experiment. In addition, it was shown this platform contributes to the requirement with the ability to enable the evaluation of various parameters such as the positions of sound reflections.

In the future, not only early reflections but also late reverberations will be added to allow a more detailed evaluation of sound reflections. In addition, the shape and material of the reflective walls can be freely set to further evaluate the effects of sound reflections in detail, creating an ideal sound field in the virtual space according to a required environment.

REFERENCES

[1] C. Wang and C. W. Cheng, “Perceiving the world with hearing in virtual reality,” *International Conference on Consumer Electronics - Taiwan (ICCE-TW)*, pp. 1–2, 2019.

[2] D. Poirier-Quinot and B. Katz, “On the improvement of accommodation to non-individual HRTFs via VR active learning and inclusion of a 3D room response,” *Acta Acustica*, vol. 5, p. 25, 2021.

[3] D. K. Reed and R. C. Maher, “An investigation of early reflection ’ s effect on front-back localization in spatial audio,” *Audio Engineering Society Convention 127*, 2009.

[4] T. Fischer, M. Caversaccio, and W. Wimmer, “A front-back confusion metric in horizontal sound localization: The FBC score,” *ACM Symposium on Applied Perception*, 2020.

[5] J. R. Clarke and H. Lee, “The effects of decreasing the magnitude of elevation-dependent notches in HRTFs on median plane localization,” *The Journal of the Audio Engineering Society*, 2017.

[6] K. Iida, Y. Ishii, and S. Nishioka, “Personalization of head-related transfer functions in the median plane based on the anthropometry of the listener’s pinnae,” *The Journal of the Acoustical Society of America*, vol. 136, no. 1, pp. 317–333, 2014.

[7] T. Matsumura, N. Iwanaga, W. Kobayashi, T. Onoye, and I. Shirakawa, “Embedded 3D sound movement system based on feature extraction of head-related transfer function,” *IEEE Transactions on Consumer Electronics*, vol. 51, no. 1, pp. 262–267, 2005.

[8] R. Litovsky, H. S. Colburn, W. Yost, and S. Guzman, “The precedence effect,” *The Journal of the Acoustical Society of America*, vol. 106, pp. 1633–1654, 1999.

[9] S. Yao, “Headphone-based immersive audio for virtual reality headsets,” *IEEE Transactions on Consumer Electronics*, vol. 63, no. 3, pp. 300–308, 2017.

[10] T. Usagawa, A. Saho, K. Imamura, and Y. Chisaki, “A solution of front-back confusion within binaural processing by an estimation method of sound source direction on sagittal coordinate,” *IEEE Region 10 Conference*, pp. 1–4, 2011.