PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

A study on the relationship between speed of sound image and evaluation value of moving sound image

Mika KURABAYASHI[1]; Kenji MUTO[2]

[1, 2] Shibaura Institute of Technology, Japan

ABSTRACT

The overall aim of this research is to develop a high-presence sound system for an emergency drill simulator using a virtual reality display system that runs on a personal computer. Previous studies have reported systems that provide high-presence sound, but these systems needed powerful computers to perform the required processing. With our proposed method, the processing is reduced by using a discontinuous moving sound image given by discrete coordinates. We evaluated the synchronization of the moving sound image by conducting an experiment in which pairs of stimuli were compared. In the experiment, moving sound images with different speeds were compared in order to obtain a threshold for the number of discrete coordinates of the moving sound image to realize high presence. We analyzed the relationship between the evaluations and the speed using a paired comparison method. The results showed that the evaluation of the realism depended on the speed of the sound image.

Keywords: Moving sound image, Discontinuous moving sound image, Subjective evaluation, Pair comparison method

1. INTRODUCTION

A number of systems have been developed to reproduce highly realistic sound. For example, Ise et al. developed a high-presence interactive sound field (1) based on the boundary-surface control (BoSC) principle (2). Poletti defined the theory of three-dimensional sound systems on the basis of spherical harmonics (3). Highly realistic sound systems are necessary to produce higher immersion, such as virtual reality content. For example, Hamasaki et al. developed a 22.2 multichannel sound system for ultra-high-definition television systems (4). However, such systems are very expensive because they require a powerful computer to perform a huge amount of calculations.

Our aim is to develop a highly realistic sound system for an emergency drill simulator and to have the system run on a personal computer. Reducing the frame rate enables the size of transmitted video to be reduced (5). We have proposed a method for reducing the amount of required processing by giving discrete coordinates to a sound image and moving it discontinuously. In many cases, sound image was given by the difference of sound pressures and time delay as in stereo system. In our experiments, sound image was given by an acoustical panning technique; the difference of sound pressures, and it was obtained good localization (6).

The moving sound image was shown with a continuous movie. When visual capture occurred, the moving sound image was perceived as being natural (7). Some studies have clarified the conditions under which visual capture occurs for a continuous moving image (8, 9). However, the conditions under which it occurs when a discontinuous moving sound image is displayed remain unclear. Moreover, the effect of the speed of the sound image on the conditions has not been investigated.

In this paper, we propose a moving sound image technique that uses a visual effect to simplify sound images. In this study, we presented moving sounds in conjunction with moving sound icons at fast and slow speeds with stereo speakers. We conducted an experiment to evaluate the synchronization between a sound image moving discontinuously and a movie of icon moving continuously.

[1] ma18038@shibaura-it.ac.jp, [2] k-muto@shibaura-it.ac.jp

2. TECHNIQUE FOR MOVING SOUND IMAGE

We proposed a technique to reduce the amount of processing by using a sound image that moves discontinuously (6). The sound image was moved the same distance and at the same speed by using the difference in sound pressure between left and right channels. Figure 1 shows the concept of the moving sound and movie and figure 2 shows the definition of the parameters for sound image localization. Let D be the length of an arc between two loudspeakers which the center is the observer. The velocity of the sound image v is given by the following equation:

$$v = D/(t_{max} - t_0). \tag{1}$$

When the sound image is given different levels in stereo speakers by using an acoustical panning technique. It was easy for reducing the amount of required processing (10). Our previous research shows that accuracy of the sound image localization was high, and the location identification was around the theoretical value (6).

The amplitudes of the left channel A_L and the right channel A_R are different. Here, $S(t)$ is the input signal, P_L is the sound pressure output from the left, and P_R is the sound pressure output from the right channel, with P_L and P_R thus given by the following equations,

$$P_L = A_L S(t), \tag{2}$$

$$P_R = A_R S(t). \tag{3}$$

The sound pressure at the center of the observer's head is given by the following equation:

$$P = (A_L + A_R)S(t). \tag{4}$$

In Figure 2, x gives the location of the sound image on the arc between loudspeakers centered at the observer. The arc is divided in the number of x_{max}. The amplitudes of the left channel and the right channel at x are given by

$$A_L = C_1(x_{max} - x) \tag{5}$$

$$A_R = C_1 x \tag{6}$$

$$A_L + A_R = C_2 \tag{7}$$

where C_1 and C_2 are constants, and P is always constant because the distance between the sound image and observer is also constant. Figure 3 shows the time waveform of P. The time waveform of P_L and P_R in the case where $x_{max} = 4$ is shown in Figure 4 and the waveform in the case where $x_{max} = 150$ is shown in Figure 5. The icon for movie moved same direction as the sound image at 30fps. The icon size is constant because it moved on the arc.

We presented sound and movie stimuli and evaluate whether they were perceived as being synchronized by using a paired comparison method.

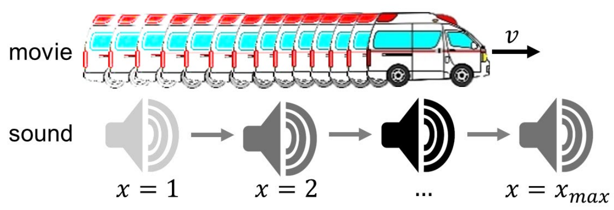

Figure 1 - The concept of the moving sound and movie

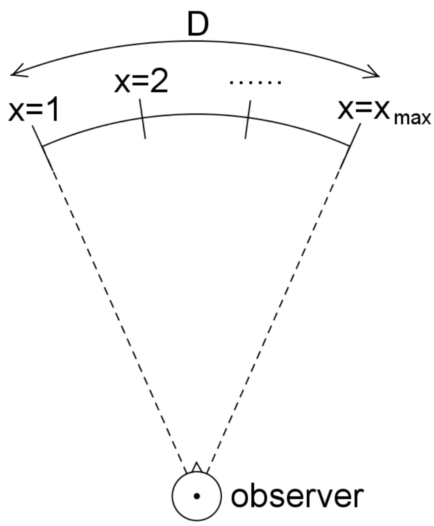

Figure 2 – Location examples of sound image
Each location x_i is of a moving sound image

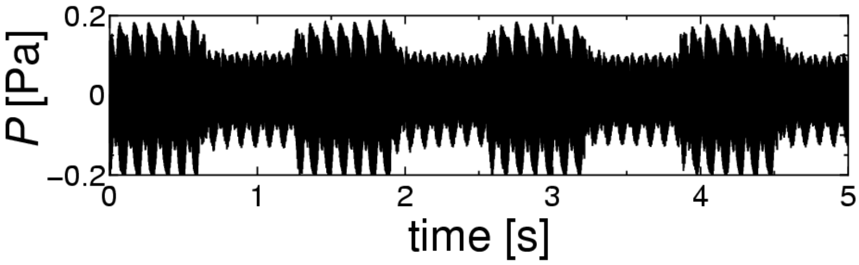

Figure 3 – Time waveform for mixed right channel and left channel

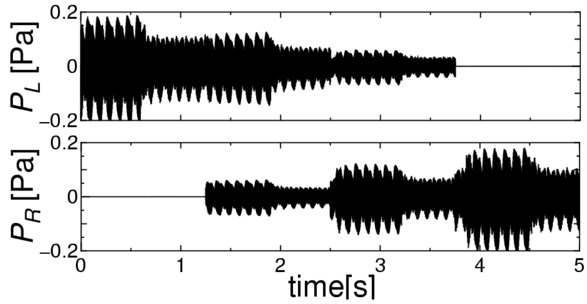

Figure 4 – Time waveforms of sound stimuli in left and right channels, where x_{max}= 4

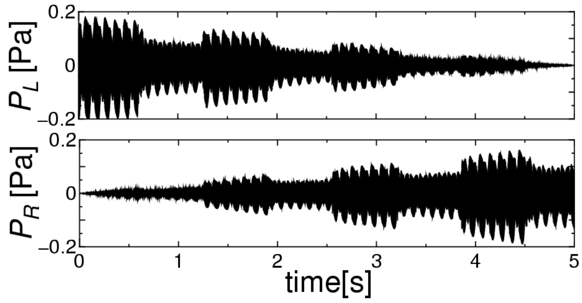

Figure 5 – Time waveforms of sound stimuli in left and right channels, where x_{max}= 150

3. EXPERIMENTS TO EVALUATE SYNCHRONIZATION

We measured the synchronization between a moving sound image and a moving icon. The type of sound used for moving sound is the siren of a Japanese ambulance in both experiment. The Doppler effect didn't include in the experiment because the sound image was moved the same distance and speed from the observer.

3.1 Stimuli Patterns

The length of the stimuli had four patterns: 1.3 s, 2.5 s, 5.0 s, and 10 s; therefore, the speeds of the sound image were 36.8 deg/s, 18.4 deg/s, 9.2 deg/s, and 4.6 deg/s, respectively. The total number of discrete coordinates for the sound images was 150. We used an ambulance as the icon for the moving sound. The movie stimuli were produced at 30 fps. Table 1 shows the four patterns of stimuli used in the subjective experiment.

Table 1 Categories of stimuli

Index	The number of the discrete coordinates	Length of the video [s]	Speed [deg/s]
VF150(Very Fast)	150	1.3	36.8
F150(Fast)	150	2.5	18.4
S150(Slow)	150	5.0	9.2
VS150(Very Slow)	150	10	4.6

3.2 Experimental environment and procedure

The stimuli were played back using 24-inch PC monitors (DELL P2416) and USB loudspeakers (Towa Electronics, Olasonic TW-S7). The distance between the monitor and the observer was 60 cm and the viewing angle 46 degrees. The 11 observers were college students in their twenties.

We presented a pair of stimuli to the observer and asked to him or her to indicate on a questionnaire whether the moving sound was synchronized with the moving icon. Figure 6 shows a conceptual representation of the experiment. Four patterns of stimuli were arranged as six different pairs and we presented these pairs to each observer in a random order. For each pair of stimuli, the stimulus presented first was named A and the stimulus presented second was named B. The interval between A and B was 2 s. After all the pairs were presented, we showed the stimuli in reverse order for each pair. In the questionnaire, the observers entered their evaluations of the synchronization on a 7-point scale. The evaluations were analyzed using Ura's method of paired comparison (11).

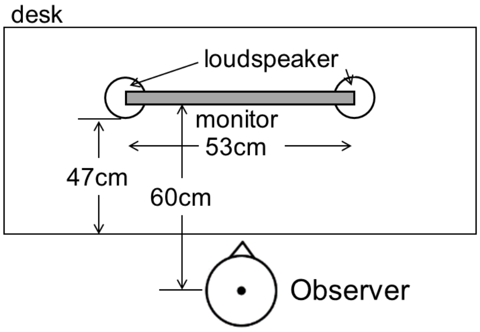

Figure 6 – Experiment composition

4. ANALYSIS OF ACCURACY AND SYNCHRONIZATION MEASUREMENTS

Figure 7 shows the analysis of these evaluations. As shown in the figure, speed affected the evaluation of the synchronization. The arrow in the figure shows the confidence interval used to assess the significance. The case where the speed was 18.4 deg/s is considerably different from the case where the difference was 36.8 deg/s. It indicates that when the speed is faster, the evaluation is lower. Figure 8 shows the analysis results for the relationship between the number of sound images x_{max} and the evaluation of the synchronization and Table 2 shows the ten patterns of stimuli used in the subjective experiment (12). A number next to a letter in the figure indicates the value of x_{max}. In the case where the speed of the sound source is 9.2 deg/s, there is no great difference in the evaluations between S150 to S10, in the result for this speed means that x_{max} = 10 is not over the threshold for being proposed as the number of locations of a continuous moving sound. In contrast, in the case where the speed of the sound source is 18.4 deg/s, there is a considerable difference in the evaluations between F150 to F10, in the result for this speed means that x_{max} = 10 is over the threshold for being proposed as the number of locations of a continuous moving sound. In summary, when the number of sound images was less, the evaluation was lower.

These experiments demonstrated that speed affects the evaluation of synchronization, as shown in Figures 7 and 8. When the speed was faster, the length of video was shorter. We think that the decreasing synchronization was due to the length of time available for presentation of the sound image icon being too short when the speed was fast.

To construct a system meeting the objective of this research, we need to obtain a threshold for the number of discrete coordinates for a moving sound image when the observer moves the head to search for the sound source.

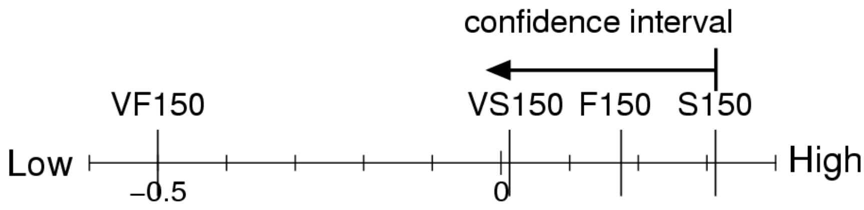

Figure 7 – Scale diagram showing confidence of synchronizing at different speeds of moving sound images and moving icons

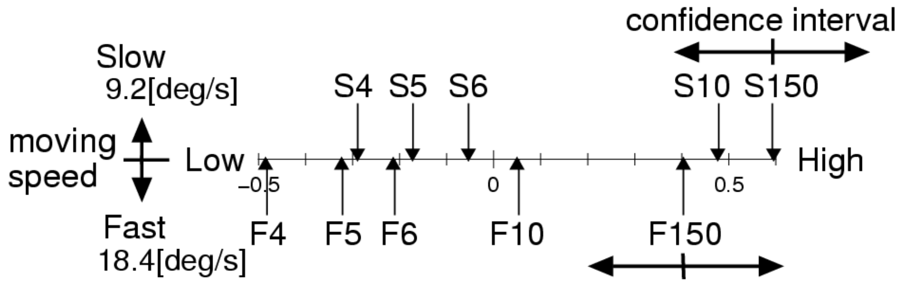

Figure 8 – Scale diagram showing confidence interval of synchronizing for five patterns of x_{max} and two patterns of speed

Table 2 Categories of stimuli used in Figure 8

Index	The number of the discrete coordinates x_{max}	Length of the video [s]	Speed [deg/s]
S4	4	5	9.2
S5	5	5	9.2
S6	6	5	9.2
S10	10	5	9.2
S150	150	5	9.2

Index	The number of the discrete coordinates x_{max}	Length of the video [s]	Speed [deg/s]
F4	4	2.5	18.4
F5	5	2.5	18.4
F6	6	2.5	18.4
F10	10	2.5	18.4
F150	150	2.5	18.4

5. CONCLUSIONS

Our aim was to reduce amount of processing used to generate visual effects for developing a high-presence sound system that runs on a personal computer. In this study, we measured the effect of the speed of a sound image on the synchronization between sound and movie. The analysis of the evaluations of the synchronization shows that the evaluation of realism depends on the speed of the sound image. In future work, we will investigate sound images moving at various speeds.

ACKNOWLEDGEMENTS

This work was supported by JSPS KAKENHI Grant-in-Aid for Scientific Research (C) Numbers 21500922.

REFERENCES

1. Ikeda Y, Ise S. Sound Cask: Music and Voice Communications System with Three-dimensional Sound Reproduction Based on Boundary Surface Control Principle. Proc ICA 19; 2-7 June 2013; Montrial, Canada 2013. pp. 1-8
2. Ise S. A Principle of Active Control of Sound Based on the Kirchhoff-Helmholtz Integral Equation and the inverse (in Japanese). J. Acoust. Soc. Jpn. 1997; 53(9): 706-713

3. Poletti M. A. Three-Dimensional Surround Sound Systems Based on Spherical Harmonics. J. Audio Eng. Soc. 2005; 53(11): 1004-1025
4. Hamasaki K, Nishiguchi T, Okumura R, Nakayama Y, Ando A. 22.2 Multichannel Sound System for Ultra High-Definition TV. SMPRW Motion Imaging Journal. 2007; 117(3): 40-49
5. Takishima Y, Wada M, Murakami H. A Study of Optimal Frame Rate in Low Bit Rate Video Coding. ITEJ technical report 1991; 15(49): 45-52
6. Kurabayashi M, Muto K. A Study on the Impression of Moving Sound Image Influenced by the Speed of the Image of Sound Source (in Japanese). Acoust. Soc. Conf. 2019; 5-7 March 2019; Tokyo Japan 2019. pp.
7. Slutsky D.A., Tecanzon G. H. Temporal and spatial dependency of the ventriloquism effect. NeuroReport. 2001; 12(1): 7-10
8. Kobayashi M, Fujii S, Iwaya Y, Sakamoto S, Suzuki Y. Temporal Characteristics of the Ventrioquism Effect (in Japanese). TVRSJ 2011; 16(1):93-97
9. Kimura M, Kajii H, Takahashi M, Yamamoto K. The Ventriloquism Effect in Peripheral Vision (in Japanese). TVRSJ 1999; 14(1): 253-260
10 Hirahara T, Ashihara K, Ozawa K, Miyasaka E. Sound and humans (in Japanese), Tokyo, Japan: Corona publishing; 2013. p. 183.
11. Amasaka K, Nagasawa S. Basics and Applocations of Sensory Evaluation – For the Kansei Engineering in Automobile Development – (in Japanese), Tokyo, Japan: Japan Standards Association; 2000. pp. 182-193.
12. Kurabayashi M, Muto K. A Study on the Impression of Moving Sound Image Influenced by Presence Order of stimuli (in Japanese). IEICE Gen. Conf. 2019; 19-22 March 2019; Tokyo, Japan 2019. p. 46