Subjective assessments for the effect of the number of channel signals on the sound field reproduction used in wavefield synthesis

Toshiyuki Kimura[1], *Kazuhiko Kakehi*[1], *Kazuya Takeda*[1] *and Fumitada Itakura*[2]

[1]Graduate School of Information Science, Nagoya University, Japan
[2]Center for Information Media Studies, Nagoya University, Japan
kimura@cog.human.nagoya-u.ac.jp

Abstract

To study the number of required channel signals is very important to reproduce a realistic sound field based on wavefield synthesis. In this paper, 2 subjective assessments were performed to evaluate the effect of the number of channel signals on the directional perception of a sound source and on the spatial impression. From the results, it was confirmed that a 24-channel system was enough to realize practical directional perception and spatial impression if the control area was limited to a circle of radius 2m.

1. Introduction

Wavefield synthesis method [1]–[3] is a technique to reproduce a sound field as that in other place. Since it needs to transmit a great number of channel signals in order to reproduce the sound field, it is very important to evaluate the number of required channel signals to design a practical reproduction system.

There are 2 approaches in the evaluation of the number of required channel signals, e.g. objective or subjective one. There are many studies [4, 5] due to the objective approach, in which the physical accuracy of synthesized wavefronts is evaluated. However, the relationship between the physical accuracy and a subjective assessment has not been studied enough. Thus, the number of channel signals is evaluated according to the subjective experiment.

Sound field perception has dimensions of directional perception, distant perception and spatial impression according to Morimoto [6]. In this paper, the results of 2 subjective experiments are described to evaluate the effect of the number of channel signals on the directional perception and on the spatial impression.

2. Subjective Assessment for Directional Perception

2.1. Experimental Environment

Subjective assessment was performed in a room (reverberation time: about 80ms). A loudspeaker array was set on the arc of radius 2 meters as shown in Figure 1. Grey circles indicate sound images reproduced by the loudspeaker array in this assessment. A background noise level was 25.0dB(A) and a sound pressure level was set to about 60dB(A) at the position of the subject.

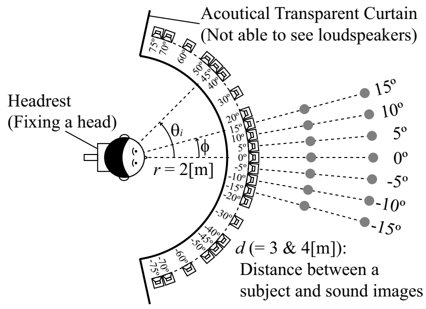

Figure 1: Experimental environment of the subjective assessment for the directional perception.

The directional perception mainly depends on the direct sound from a sound source. When the sound field was assumed to be a free space, the room transfer function $g_i(n)$ is denoted as shown in Eqn. (1), (F_s(= 48[kHz]): Sampling frequency, c(= 340[m/s]): Sound velocity)

$$g_i(n) = \frac{d-r}{d_i}\delta(n - \text{round}\Big(\frac{d_i F_s}{c}\Big)). \tag{1}$$

The d_i (the distance between the sound source and the ith loudspeaker) is calculated as shown in Eqn. (2),

$$d_i = \sqrt{d^2 + r^2 - 2dr\cos(\phi - \theta_i)}. \tag{2}$$

White noise and speech (1 second long) were used as a source signal $s(n)$. Then $x_i(n)$ is calculated from the source signal $s(n)$ and $g_i(n)$ as shown in Eqn. (3),

$$\begin{aligned} x_i(n) &= D(\theta_i,\phi)\{g_i(n) * s(n)\} \\ &= D(\theta_i,\phi)\frac{d-r}{d_i}s(n - \text{round}\Big(\frac{d_i F_s}{c}\Big)). \end{aligned} \tag{3}$$

$D(\theta_i, \phi)$ (directional sensitivity of a microphone) is defined as shown in Eqn. (4),

$$D(\theta_i,\phi) = \begin{cases} 1 & (\cos(\theta_i - \phi) \geq \frac{r}{d}) \\ 0 & (\cos(\theta_i - \phi) < \frac{r}{d}) \end{cases}. \tag{4}$$

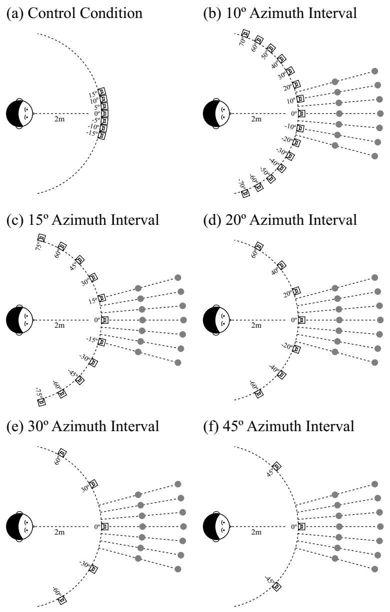

Figure 2: Experimental conditions of the subjective assessment for the directional perception.

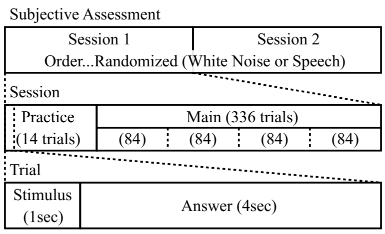

Figure 3: Experimental design of the subjective assessment for the directional perception.

Experimental conditions pertinent to the number of channel signals are shown in Figure 2. Subjects listen to the sound source from each loudspeaker in control condition and the sound image synthesized with the loudspeaker array in other 5 conditions.

2.2. Experimental Design

Subjects were 8 graduate students (4 males and 4 females). The experimental design of the subjective assessment is shown in Figure 3. A rest time was introduced in every 84 trials. The conditions of practice and main trials are shown in Table 1. The subject was instructed to report the direction of sound within 4 seconds after listening the 1 second stimulus. Subjects reported the direction due to a scale which is placed in front of them and marked from -25° to 25° at every 2.5° interval.

2.3. Results and Discussions

Experimental results are shown in Figure 4. Localization results for the 10° and 15° azimuth interval conditions are almost same as those of the control condition. On the other hand, perceived directions tend to be biased towards 0° direction in 20°, 30° and 45° azimuth interval conditions. The reason of the bias is explained by using the case of 45° azimuth interval as shown in Figure 5. The subject localizes the sound image which is synthesized based on $x_0(n)$, $x_{45}(n)$ and $x_{-45}(n)$ (ref. Figure 5a). Because d_0 is the shortest of all distances, $x_0(n)$ is the fastest sound to be heard. On the other hand, $x_{45}(n)$ and $x_{-45}(n)$ are delayed more than 1ms compaired with $x_0(n)$ because d_{45} and d_{-45} are the $0.001c$[m] longer

Table 1: Condition of trials.

	Number	Value
Practice (14)	= 1 distance × 7 directions × 2 conditions	3m 0°, ±5°, ±10 ° & ±15 ° (a) and (b) shown in Figure 2
Main (336)	= 2 distances × 7 directions × 6 conditions × 4 repetitions	3 & 4m 0°, ±5°, ±10 ° & ±15 ° From (a) to (f) shown in Figure 2

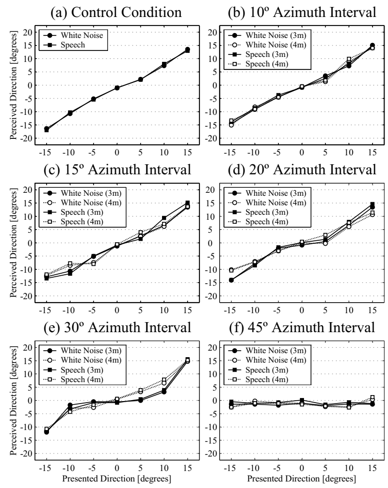

Figure 4: Results of the subjective assessment for the directional perception.

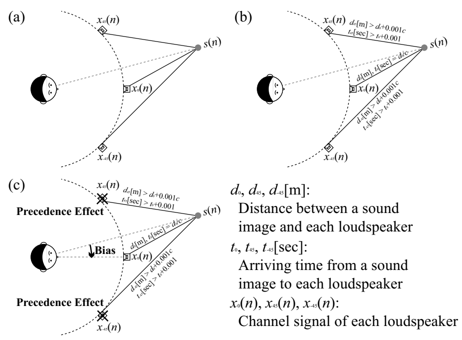

Figure 5: Reason of the bias in the 45° azimuth interval condition.

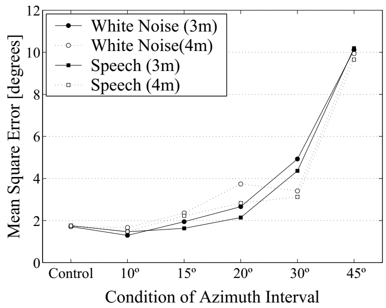

Figure 6: Results of the mean square error for the directional perception.

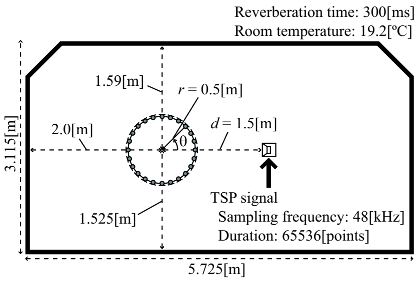

Figure 7: Measurment environment of room transfer functions.

than d_0 (ref. Figure 5b). Due to the precedence effect [7] of $x_0(n)$, only $x_0(n)$ contributes the perceptual localization (ref. Figure 5c).

The accuracy threshold is discussed by calculating the mean square error (MSE) between presented directions y_i and perceived directions y_i' as shown in Eqn. (5),

$$\mathrm{MSE} = \sqrt{\frac{\sum_i (y_i - y_i')^2}{7}} \quad (i = 0, \pm 5, \pm 10, \pm 15). \quad (5)$$

Results are shown in Figure 6. The MSEs of the 10° and 15° azimuth interval conditions are same as that of the control condition (about 2°). Thus, it is considered that the accuracy threshold of interval is 15°, or 24 loudspeakers on the circle of radius 2 meters due to the subjective assessment.

3. Subjective Assessment for Spatial Impression

3.1. Experimental Environment

Room transfer functions used in an experiment were measured in a real room. A microphone array and a loudspeaker were set as shown in Figure 7. Room transfer functions were measured by playing a TSP signal [8] from the loudspeaker. A background noise level was 19.4dB(A) and a sound pressure level of the TSP signal was set to 91.6dB(A) at the 1 meter from the loudspeaker. Because the reverberation time is transformed into that of the realistic concert hall, the measured room transfer

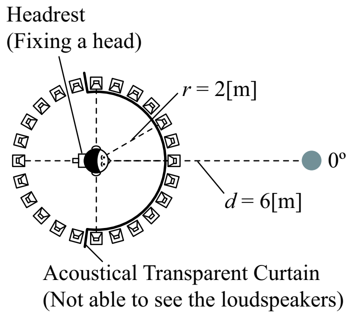

Figure 8: Experimental environment of the subjective assessment for the spatial impression.

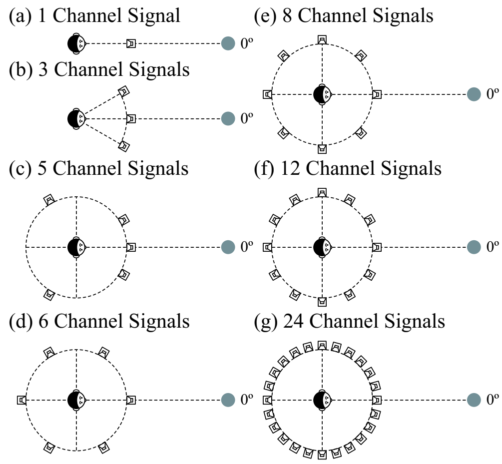

Figure 9: Experimental conditions of the subjective assessment for the spatial impression.

functions were treated as the FIR filters of which the sampling frequency was 12kHz. Thus, the reverberation time becomes 4 times (300[ms]×4=1.2[sec]). Because the diffused time and distance of sound become 4 times, the values of r and d shown in Figure 7 are 0.5[m]×4=2[m] and 1.5[m]×4=6[m]. Sound signals of speech or flute (sampling frequency: 12kHz, duration: 4 seconds) were used as a dry source. Channel signals were synthesized by convolving measured room transfer functions to the dry source.

Subjective assessment was performed in a room (reverberation time: about 80ms). A loudspeaker array was set as shown in Figure 8. A grey circle indicates the sound image reproduced by the loudspeaker array. A background noise level was 25.0dB(A) and a sound pressure level was set to about 70dB(A) at the position of the subject. Experimental conditions pertinent to the number of channel signals are shown in Figure 9.

3.2. Experimental Design

Subjects were 8 males. Scheffé's paired comparison [9] was introduced as an assessment method. The experi-

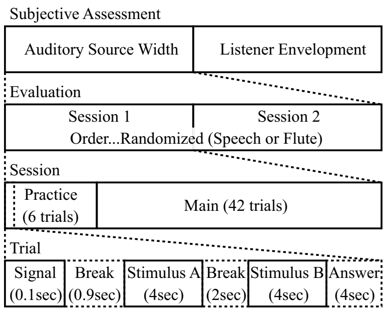

Figure 10: Experimental design of the subjective assessment for the spatial impression.

Table 2: Scale of Scheffé's paired comparison.

Grade	ASW	LEV
3	Very wide	Very enveloped
2	Fairly wide	Fairly enveloped
1	Little wide	Little enveloped
0	The same	The same
-1	Little narrow	Not little enveloped
-2	Fairly narrow	Not fairly enveloped
-3	Vary narrow	Not very enveloped

mental design of the subjective assessment is shown in Figure 10. Morimoto [6] indicates that the spatial impression consists of at least 2 factors, e.g. Auditory Source Width (ASW) and Listener Envelopment (LEV). In this assessment the effect of the number of channel signals on ASW and LEV was evaluated in each evaluation. 6 practice trials are the permutation of 3 conditions ((a), (b), (g) of Figure 9). the permutation of 7 conditions ((a)–(g) of Figure 9) resulted in 42 main trials.

Before the evaluation, the subjects were instructed in the definition of ASW and LEV. The subjects graded the stimulus B in Figure 10 in reference to the stimulus A in Figure 10 according to the 7-step scale shown in Table 2.

3.3. Results and Discussions

The psychological scales of all conditions pertinent to the number of channel signals are shown in Figure 11. Error bars indicate 95% confidence intervals. When the number of channel signals is 1, 3 and 5, psychological scales are significantly lower than those of the 24 channel signals condition which is the nearest one to the original sound field in all conditions. On the other hand, when the number of channel signals is 6, 8 and 12, psychological scales are same as those of the 24 channel signals condition. Thus, it is considered that there is no difference of the spatial impression between the reproduced sound field and the original sound field when the number of channel signals is larger than, or equal to 6. That is the number of required channel signals to realize the spatial impression is 6.

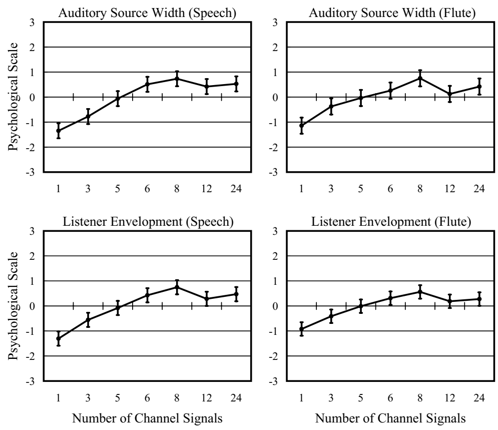

Figure 11: Results of the subjective assessment for the spatial impression.

4. Conclusion

The number of required channel signals in the directional perception and the spatial impression was evaluated for a wavefield synthesis system. It was shown by the subjective assessment for the directional perception that the number of required channel signals was 24. It was also shown by the assessment for the spatial impression that the number of required channel signals was 6. As a result, it was confirmed from the 2 subjective assessments that 24 channels are enough to realize the sound field inside a circle of 2 meters radius.

5. References

[1] Camras, M., "Approach to recreating a sounf field," J. Acoust. Soc. Am., 43(6):1425–1431, 1968.

[2] Berkhout, A. J., Vries, D. de and Vogel, P., "Acoustic control by wave field synthesis," J. Acoust. Soc. Am., 93(5):2764–2778, 1993.

[3] Poletti, M. A., "A Unified Theory of Horizontal Holographic Sound Systems," J. Audio Eng. Soc., 48(12):1155–1182, 2000.

[4] Kirkeby, O. and Nelson, P. A., "Reproduction of plane wave sound fields," J. Acoust. Soc. Am., 94(5):2992–3000, 1993.

[5] Ward, D. B. and Abhayapala, T. D., "Reproduction of a Plane-Wave Sound Field Using an Array of Loudspeakers," IEEE Trans. Speech and Audio Proc., 9(6):697–707, 2001.

[6] Morimoto, M., "The relation between spatial impression and the precedence effect," Proc. of International Conf. on Auditory Display, 297–306, 2002.

[7] Wallach, H., Newman, E. B. and Rosenzweig, M. R., "The precedence effect in sound localization," Am. J. Psychol., 62(3):315–336, 1949.

[8] Suzuki, Y., Asano, F., Kim, H. Y. and Sone, T., "An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses," J. Acoust. Soc. Am., 97(2):1119–1123, 1995.

[9] Scheffé, H., "An analysis of variance for paired comparisons," J. Am. Stat. Ass., 47:381–400, 1952.