A Study on Recording Techniques for Describing Frontal Images with High Separation Between Sound Sources in Multichannel Audio-Focusing on the OCT (Optimized Cardioid Triangle) Method

I. Introduction

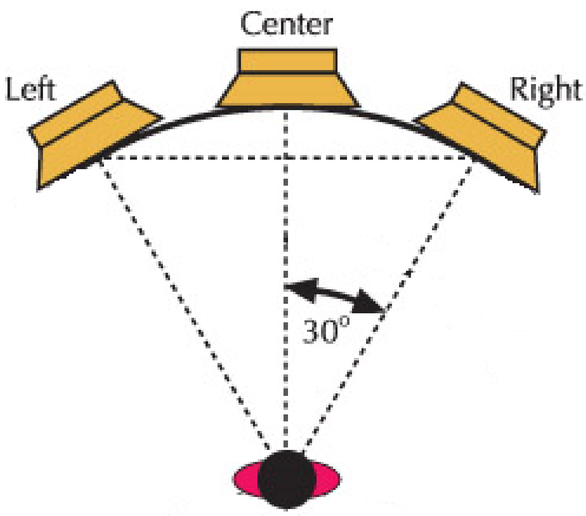
1.1. Research Background

1.1.1. The Role and Correlation of Front Channels in Surround Sound

The range in which a person can accept information composed of sound is from sound coming from all directions surrounding the two ears. Today's advances in audio technology have taken a leap to provide realistic sounds that people can practically accommodate beyond gradual development of Two-channel audio system. With introduction of various playback technologies, and surround and 3D recording technologies that can be integrated with them, there have been sound receiving and methodological researches to realistically contain spatial information without distortion. [1][2]

In 1880s, the Two Channel Audio System were first demonstrated by Clément Ader, and since the announcement of the fact by Carl Stumpf, German philosopher and psychologist in 1916, that both ears of human perceive different information, respectively, the auditory image based on the 'psychological sound' of how human perceives the sound information from today's surround and immersive audio systems has been an important criterion for the development of audio technology to reenact this. According to the R&D engineers of BBC (British Broadcasting Corporation), if there is any omitted or distorted part of the audio during playback that is regarded as important information for human perception of space, it can difficult to experience realistic virtual reality as much as the loss due to this [3]. As such, for surround and 3D audio it is important for playback system and recording to provide auditory experiences similar to actual spaces based on 'psychological sound' beyond two channel audio system.

In a realistic audio playback system, in order to describe the results of recording and playback similar to the auditory experience in real space, the criteria of correlation between the playback system and recording technology must be properly defined and matched. [4] [5] The front channel of the standard monitor of the multi-channel audio according to the ITU-R BS 774-1 scheme, which is the speaker array of the surround system, which is the background of the present study, is building a system with continuous arrangement of three speakers at 30° intervals [Figure 1]. It quotes the stereo main pair angle of a two-channel audio system as it is, but differs from stereophonic due to the addition of a center speaker, and fundamentally changes the perceived characteristics of the frontal sound region, so it should be understood as a continuous space divided by Front L and Center, Center and Front R. Therefore, in order to experience in-phase listening to the listener through the surround system, the microphone array for recording the information on the front face must also be a technique in which the perceived characteristics of the sound region on the front face are considered.



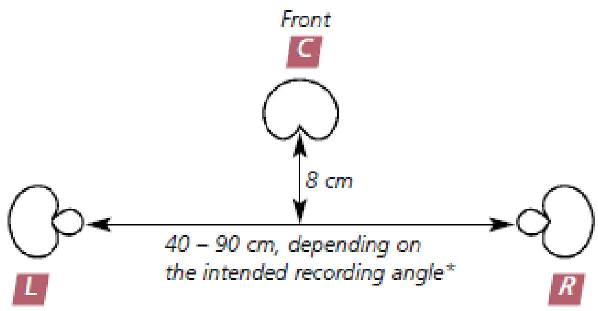
[Figure 1] Front Channel Configuration of ITU-R BS 774-1

1.1.2. Characteristics Discrete of Surround Recording Method OCT

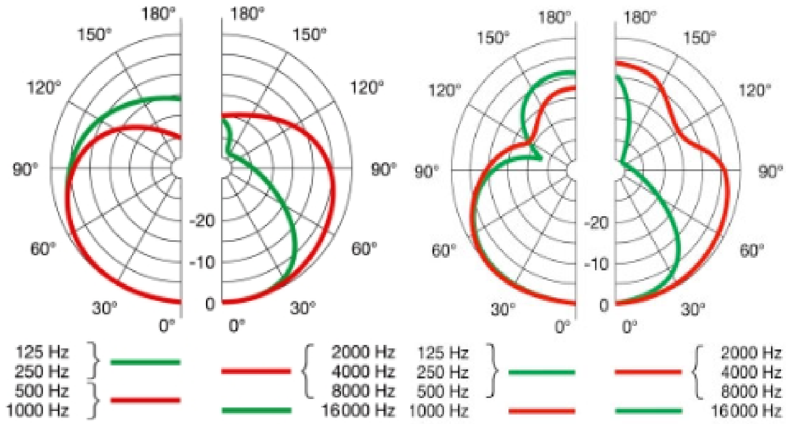
(Optimized Cardioid Triangle)

Using conventional OCT (Optimized Cardioid Triangle) [6] [7] recording methods, the time difference and level difference between the front L and center microphones, and the center and front R microphones can be derived continuously, so this is the ideal recording method for a surround playback system [Figure 2].

When the location information of sound sources located in front of a certain space can be recorded and played back with high separation, a multi-channel audio monitor environment can be provided with an auditory experience similar to that in a real space. Recording and playback using the OCT method create high separation through two main features. Firstly, different polar pattern cardioid and super cardioid microphones are arranged in the center and L-R [Figure 3]. This produces a distinct and continuous level change between center and L or between center and R. Secondly, set the offset angle between the L and R microphones to 180°. This is to record the location information of the sound source in front by using the level attenuation characteristic by the polar pattern from the front lobe to the rear lobe from the super cardioid microphone of L-R forming the level difference with the cardioid microphone of the center [Figure 3], allowing you to clearly record and describe a wider range of content than the normal angle for the front channel.



[Figure 2] OCT setup



[Figure 3] Polar diagram characteristics of AKG C414 xl ii microphone by frequency (left, cardioid right, super cardioid)

1.2. Purpose of Study

The purpose of this study is to propose a recording technique for high-definition sound source description between front channels (L-C-R) in relation to the surround microphone technique for multichannel audio. To do this, the polar pattern and 0° direction of the L and R microphones are changed from the optimized cardioid triangle (OCT) microphone array of existing Scheops company. After that, the level difference between the L and the center channel in each method is analyzed and the hearing evaluation is conducted to evaluate the objectified superiority.

Ⅱ. Main Subject

2.1. Experiment

2.1.1 Experiment Overview

This study tried to find an excellent method in terms of the separation of sound sources among front channels (L-C-R) of multi-channel audio standard monitors, which is the recommended method of ITU-R BS 774-1 with additionally recording through total of three modified methods together with the OCT method as shown in [2]. The method used here is as follows. First, it changed the polar pattern of the L and R microphones of the OCT method from super cardioid to cardioid, and second, it changed the 0° direction of the first method of the L-R microphone to 0° at +/- 90° respectively. AKG C414 ULS cardioids were used for the center microphone, and AKG C414 xl ii cardioid and super cardioid polar patterns were recorded for the L-R microphones.

The sound source used in the experiment was a sine wave of 1 kHz, a clock second hand sound, and a recorded vocal having different characteristics and patterns [Table 1]. These sources were reproduced with 4W output speakers, amplified by Micastasy of RME company, and recorded at AD (Analog to Digital) conversion at 16Bit with a sampling frequency of 44.1kHz through the Protocols HD Native audio interface.

	Type of sound source	Pattern change	Recording length
1.	Sine wave	none	10s
2	Clock second hand	regular	10s
3	Vocal	Irregular	10s

[Table 1] Types and Features of Sound Sources Used in Experiments

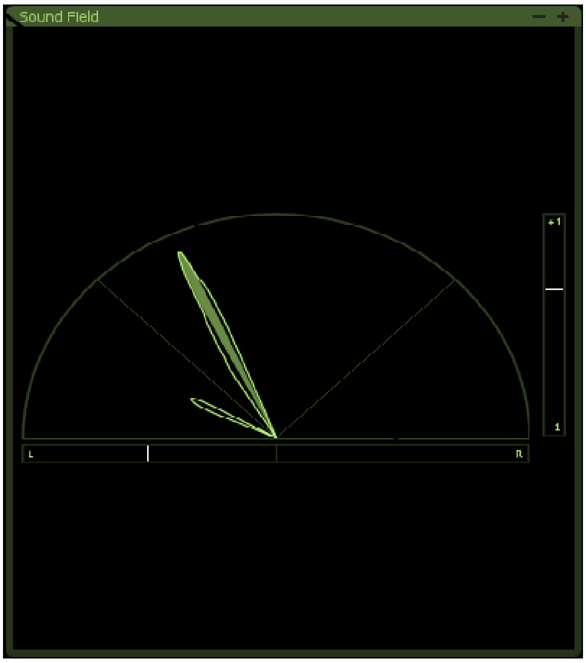
2.1.1 Experimental sequence

Measure the size of the space to avoid highlighting a particular wave reflected by the recording environment as a stationary wave. Several test recordings were made to determine the proper microphone setup position. The microphones were set up at 2.8m width, 3.5m length, and 1.7m in a space of 3.301m width, 6.947m length and 2.508m height.

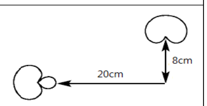
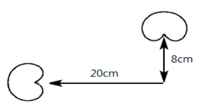
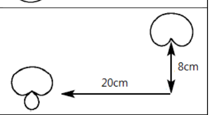
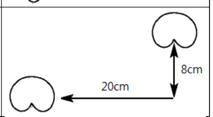
② Determine the amplification value of each microphone. The microphones used in the experiments were cardioid and super cardioid microphones, and the sensitivity was measured by radiating white noise at the size of 96dB SPL at 0° direction, 1m distance of each microphone. Then, determine the amplification value of the preamplifier so that the measured result becomes the output level of -18dB SF of each microphone.

③ Select the position to play the sound sources. In this experiment, the distance between the L-R microphones was set to 40cm, and the position of 45° from the center microphone was selected as the playback position of the sound source. On the same line of 45°, selected the initial point of about 1.2m at the monitor speaker where the sound image becomes 45° by moving the speaker backward as shown in [Figure 4], through the phase scope.

④ The sine wave 1kHz was played and recorded, and the clock second hand sound and the recorded vocal sound were reproduced and recorded at the same time. For each sound source, a total of three modified methods were additionally recorded along with the OCT method. The structure of each method is shown in [Table 2]. First, it is the same as OCT method of [Figure 2] and implemented as [Figure 5] for experiment. Second, the 0° direction of the L and R microphones was kept the same as the OCT method, and the polar pattern was recorded by changing from cardioid to super cardioid. Third, the structure of the existing OCT microphones was maintained, and the 0° direction of the L and R microphones were recorded so as to face the front at +/- 90° [Figure 6]. Fourth, the 0° direction of the L and R microphones was changed to face the front, and the polar pattern was changed from super cardioid to cardioid for recording.



[Figure 4] Check that the sound image of sine wave is 45° through phase scope

	Direction of 0°	Center	L, R	Setup
Method 1	+/- 90°	Cardioid	Super Cardioid	
Method 2	+/- 90°	Cardioid	Cardioid	
Method 3	0°	Cardioid	Super Cardioid	
Method 4	0°	Cardioid	Cardioid	

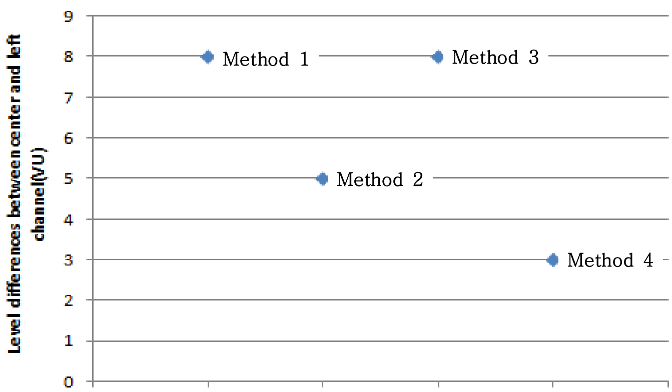
[Table 2] Characteristics of Modified Method for Experiment

[Figure 5] Implementation of OCT method

[Figure 6] OCT type Super cardioid is transformed to face front at +/- 90°, and Super cardioid is transformed to Cardioid in the fourth type

2.1.2. Experiment result

For example, this study compared and analyzed the level difference of the recorded results among L-C channels in total of 4 kinds recording methods, including OCT based on the fact that in the case of stereo microphone technique XY using only the level difference, the greater the level difference between the channels for a sound source, the higher the separation between the sound sources. The level of each channel was measured by VU meter, and the result of level difference between OCT and L-C of each recording method is shown in [Figure 7]. For the OCT method, the sine wave level measured at the center channel is -18dB, the L channel level is -26dB, and the level difference between the center and L channels is 8dB. The levels of the center channels from modified modes 2 to 4 are -16 dB, -16 dB, and -18 dB, respectively, and the levels of the L channels are measured as -21 dB, -24 dB, and -21 dB, respectively, and the level difference due to this between the center and L channel is 5dB, 8dB and 3dB, respectively. Method 1, which is an OCT microphone array type, and Method 3, which changed only the 0° direction of the microphone, showed the same level difference, and Method 2 had a relatively low level difference of 3 dB.



[Figure 7] Variation of level difference of sine wave 1kHz between center and L channel for each method in [Table 2]

2.2. Hearing evaluation

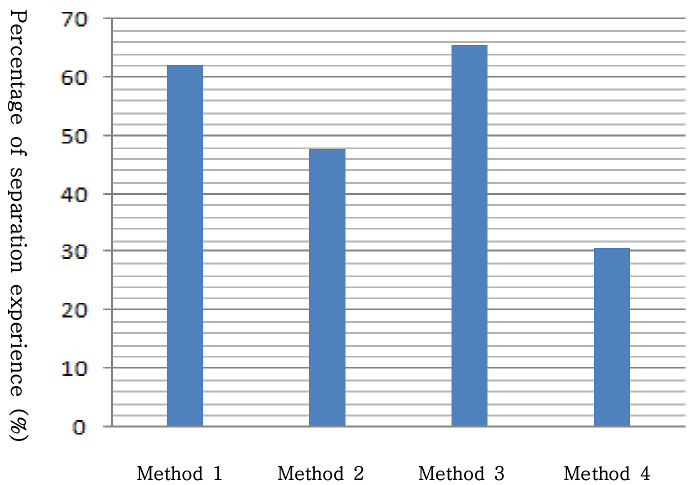
2.2.1 Evaluation method

As for the total of four recording methods of this experiment to describe the separation between the sound sources in the front channels of the surround monitor system hearing evaluation were conducted for the objective evaluation of the superiority. This study used recorded sound source by playing back the speakers at 45° and about 1.2m from the center microphone of the OCT microphone array for the vocal sound source clock second sound whose rhythm and pitch pattern change of sound level is regular and irregular vocal sound source. The sound source was edited with the results of the four recording methods of the experiment in 7-second intervals and arranged at 0.5-second intervals, and the order of each edited sound source was randomly arranged and a total of four sets were used for evaluation.

To assess the degree of separation between sound sources, the subjects were 10 persons who major music. In addition, before the evaluation progressed, the degree of separation of the sound source was fully explained, and the evaluation was conducted to increase the objectivity of the evaluation result. Experiments were carried out to listen to the results of four randomly reproduced recordings of each set and number them in the order of perceived high separation between sources. In addition, the level of individual feeling of the degree of separation of the sound sources was evaluated from 1 to 5, and an absolute evaluation between the recording methods applied in the experiment was attempted.

2.2.2. Evaluation results

[Figure 8] shows the results of evaluation of individual sensory levels as a percentage of the results of the four recording methods applied in the experiment. A relatively large number of subjects experienced a high degree of separation through the method 1 corresponding to the OCT microphone array and the method 3 in which the configuration of the microphone was maintained while the 0° direction was faced in the OCT method. On the other hand, although the sensory level of separation between the results of Method 2 and Method 4 using only cardioid microphones showed a relatively low change of less than 15%, in absolute evaluation, it was confirmed that the sound sources can be experienced with a maximum of 47.7% of separation. In addition, it was confirmed through an interview that three of the subjects who evaluated that the results of Method 1 and Method 3 provided the highest degree of separation were evaluation for the degree of separation experience from 1 to 2 out of 5, which was evaluated by distinguishing only slight differences between the results of each method.



[Figure 8] Percentage of separation experience for the four methods used in the experiment

III. Conclusion

This study was studied for recording techniques in which sound sources are described with high separation when surround sound, which is realistic audio, is played back through multichannel audio. The experiment was conducted by changing the polar pattern of the L-R microphone and the direction of the central axis around the OCT microphone array where the front channels were recorded with high separation, and it is confirmed that the change of the axial direction to the front of the Method 1, OCT microphone array and Method 3, OCT type L-R microphone shows excellent separation in the front channels of the multi-channel audio standard monitor. Method 1 and Method 3 use cardioid microphones for C and super cardioid microphones for LR. This result shows that the level difference between L-C channel and C-R channel can be made more than 3dB than that of using cardioid microphone for L-R, and it is confirmed that the sensory level for the separation in multi-channel audio is also excellent as the level difference is big between channels.

Unlike the OCT method, the actual experiment result showed the same level difference of 8dB in method 3 for the sine wave 1kHz different from the inference that the level difference between L-C or C-R will be smaller than that of the OCT method due to the change in the position of the lobe where the sound source is mainly absorbed because the center axis of the L-R microphone faces the front. On the other hand, despite having the same level difference, the perceived level of separation in multi-channel audio was about 3% higher than that of the OCT method. This result is presumably due to the phase distortion caused by the rear lobe of the super cardioid microphone having a phase characteristic opposite to that of the front lobe.

Based on the above results, it is thought that the recording technique for describing high separation in multi-channel audio is affected not only by the level difference between channels but also by the phase characteristics of the microphone and psychological factors related to hearing, thus the recording technique considering this should be done continuously.

References

[1] Michael Williams, UNIFIED THEORY OF MICROPHONE SYSTEMS for STEREOPHONIC SOUND RECORDING, Presented at AUDIO the 82nd Convention 1987 March 10-13 London

[2] Michael Williams and Guillaume Le Du, MICROPHONE ARRAY ANALYSIS for MULTICHANNEL SOUND RECORDING, Presented at AUDIO the 107th Convention 1999 September 24-27 New York

[3] Spatial Audio for Brodcast.

Helmut Witt Gunther Theile, The recording angle-based on localization curves, Presented at the 112th Convetion 2002 May 10-13 Munich, Germany

[5] Michael Williams, The stereophonic zoom, "Sounds of sootland", BP50, 94364 BRY SUR MARNE CEDEX, France

[6] SCHOEPS_surround_brochure (2006-11)

[7] Schoeps Manual MAB 1000