Head Movement and Gaze Analysis in Examining the Influence of Surround Sound on People Watching Images

Yasuyoshi Kobayashi[1], Shinya Mochiduki[2], Mitsuho Yamada[2]

[1]Tokai University, Graduate School of Information and Communication Studies Department, 〒108-8619
[2]Toaki University, Graduate School of Science and Engineering Graduate School of Science and Engineering, 〒259-1292

Keywords: surround sound, eye movement, head movement, objective evaluation

ABSTRACT

Recent advances in sound technology have been remarkable in conjunction with high definition images. The possibility of an objective evaluation of synergistic effect of image and two kinds of sound was examined based on head and eye movement.

1 INTRODUCTION

In recent years high-definition 4K and 8K video technologies have advanced, and the broadcasting of such images began in earnest in Japan in December 2018. These ultra-high-definition images have a fine pixel structure, so you can enjoy images with a wide field of view even at short distances. We have analyzed the effect of short-distance viewing on human psychological characteristics based on eye movements and head movement [1,2]. Broadcasting of 4K and 8K images is expected to be enhanced by the high presence feeling of 22.2-channel surround sound. However, the ability of sound to induce a sense of "presence" or three-dimensional reality is assumed to be impacted by the individual's bias and other senses; a method for its stable evaluation has not yet been established. [3] For this reason, we considered the possibility of using the responses of physiological phenomena such as eye movements to evaluate the acoustic effect. In this study, we measured and compared such data from subjects viewing movies accompanied by 2ch stereo and 5.1ch surround sound: namely, head movement, convergence eye movements and gaze point distributions. In order to clarify the basic characteristics of head and eye movements, we devised an experiment that combined images with three types of sounds: 2-channel, 5.1-channel, and 7.1-channel.

2 EXPERIMENT

In our previous research, an experiment was conducted using a movie (*James Cameron's Avatar*, 2009) capable of 5.1-ch output.[4] The sound pressure was set at 65 ± 5 db at the ear, and the image was projected with a 4k projector (SONY SX3D). The subjects were allowed to watch 10 minutes of 6 chapters in a random combination of 5.1-ch and 2-ch surround sound. We selected the chapters in advance. Each subject watched each scene only once with 2-ch or 5.1-ch audio. The configuration of the experiment is shown in Fig. 1. An EMR-8B (NAC image technology, Tokyo) was used for eye movement measurement, and an acceleration sensor (3DM-GX4-25, LORD MicroStrain, VT) was used for head movement measurement. We used the measurement protocol developed in our laboratory [4]. Each subject wore a cap with an EMR-8B field-of-view camera to capture the entire screen. Data from the left and right eyes were registered and a 9-point calibration was carried out. Next, an acceleration sensor was attached to the EMR-8B cap, the subject was asked to face the screen directly, and the offset of eye movement was measured. Experiments were conducted in a room with soundproof specifications at our university. The maximum luminance was 104 cd/m^2 and the minimum luminance was 0.04 cd/ m^2 from the screen. The subjects were ten students of both genders of our university in their twenties.

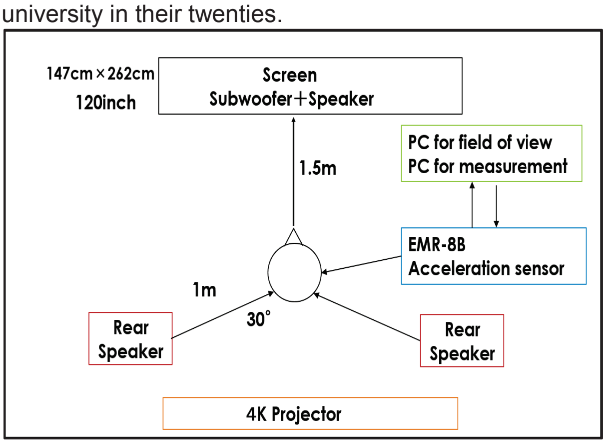

Fig.1 Experiment configuration (previous research)

In our previous studies[5], movies were used and results could thus be dependent to some extent on content. Therefore, we devised an experiment that uses an edited image in combination with 2-ch, 5.1-ch, or 7.1-ch audio. The purpose of this experiment was to create an experimental environment that allows users to view images as easily as they do at home without requiring a large screen or speakers. The image used was an airplane moving from left to right, downloaded from the internet. For sound editing, the DaVinci Resolve 15 video editing software (Blackmagic Design, Melbourne, Australia) was used. The same image was synthesized for 3 seconds with 2-ch, 5.1-ch and 7.1-ch sound, and was watched for 15 seconds. We showed a fixed point for 5 seconds during the advance interval to the next image. The sound was synthesized for those with normal hearing by appropriately arranging tracks and setting the volume so that sound image localization would be easy. The subject was seated in front of a 21.5-inch display (Dell, E2210H) and fitted with an eye movement measurement device (EMR-9, NAC Image Technology, Tokyo) and a wireless acceleration sensor (TSND151, ATR Promotions Inc., Kyoto). The calibration procedure was similar to that used in our previous studies.[5] The luminance was 250 cd/m^2 from the display. The subject was put on headphones incorporating five real speakers per side (Tiamt7.1V2 Razer) as shown in Fig.2. At that time, he/she was instructed to adjust the volume properly with the volume controller at hand. Since it is assumed that the user will be viewing in a relaxed posture at home, the chair was chosen to be variable and the subject was instructed to set it at a comfortable height. The experiment was conducted in our laboratory. The configuration of the experiment is shown in Fig. 3.

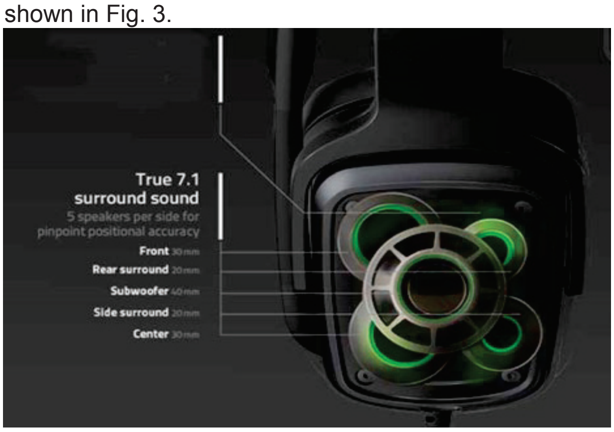

Fig.2 Side view of the headphone which embedded five speakers each side

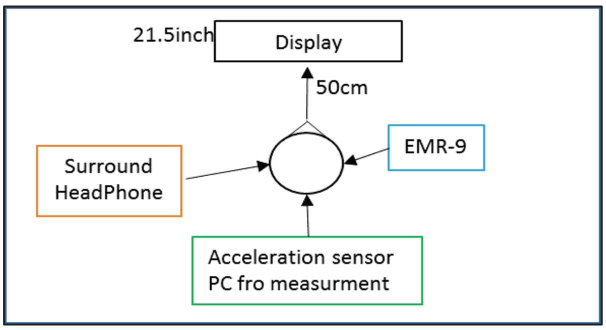

Fig.3 Experiment configuration (Newly proposed research)

3 RESULTS

3.1 Eye movement

First let us consider the results from the previous research.[5] The results of Chapters 1 and 4 where the data has changed significantly are shown below. In Chapter 1, the sound of the aircraft includes surround sound and characteristic sounds. The definition of characteristic sound is that anyone can recognize the change in the left-right direction clearly. On the other hand, 4 is a scene that includes only human conversation and no other features. The data for 10 subjects shown in the Fig.4,5 shows the results of Chapter 1 including characteristic sounds. The orange dot is the subject who watched with 5.1ch, and the green dot shows the subject who watched with 2ch. The 0 point in the figure indicates the center of the screen. The y axis indicates vertical direction of the screen, and the x axis indicates the horizontal direction. A gaze point is a point that occurs when a person reads a sentence or looks at an image and stops the movement of the eye for a short while. Generally, a state in which the eye movement speed of 5 deg / s or less continues for 150ms or more is defined as one gaze point. A t-test was conducted to see if there was a difference in perspective when watching the video whether there was a characteristic sound or not. Fig.4 shows the results of the subjects compared in the t test. A red dot indicates a subject who has watched chapter 1 and a purple dot has watched chapter 4.For both chapters 1 and 4, a two-sided t-test (α=5%) was performed between 5 subjects who watched on 5.1ch. As a result, p = 0.0044 was found, showing a significant difference. From the above, it was found that the gaze point spreads laterally when the surround system includes characteristic sounds.

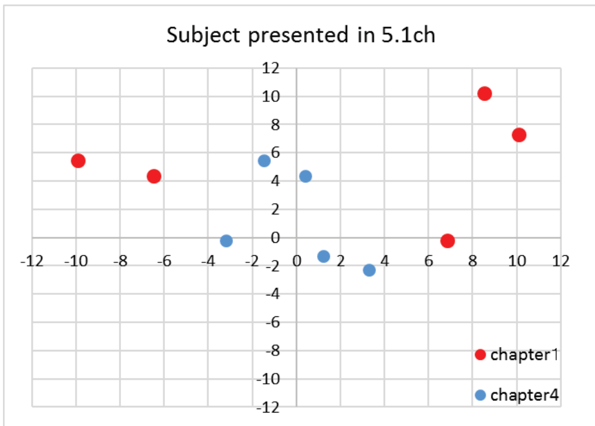

Fig.4 Average gaze position with 5.1ch

Similar t-tests were performed on subjects presented in 2ch for both chapters 1 and 4. The result is shown in Fig.5. As a result, no significant difference was found for p = 0.325. From the result of the t-test, it was found that the gazing point spreads in the lateral direction of the screen in the case of the surround sound, including the characteristic sound.

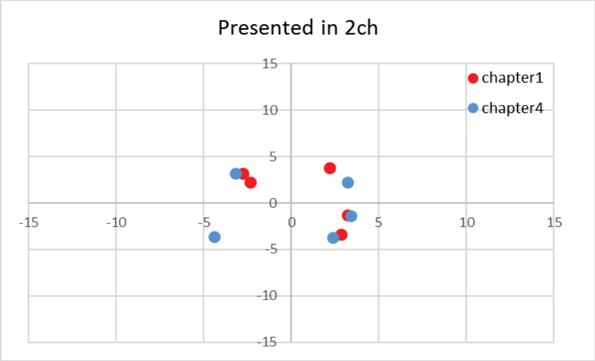

Fig.5. Average gaze position with 2ch

Next, we describe the results of eye movements obtained in the newly proposed experiment. Figure 6 is a summary of the average gaze positions. The central black dot indicates the center of the display. The red dot indicates the average gaze position when viewing with 5.1ch, the green dot corresponds to viewing with 2ch, and the blue dot to viewing with 7.1ch.

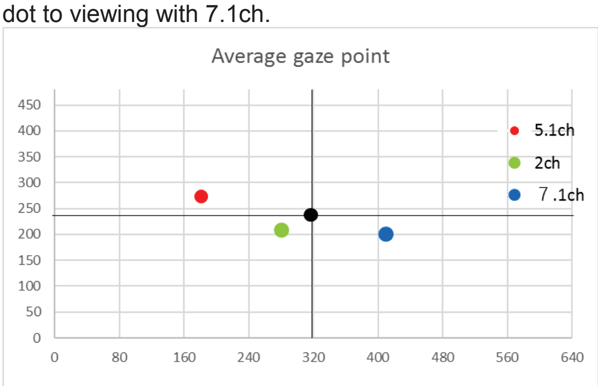

Fig.6 Average gaze position (Newly proposed research)

These results show that the gaze position tended to expand in the left-right direction as the number of channels increased, as in previous studies.

3.2 head movement

A comparison of two subjects is shown for head rotation in previous studies. Fig.7 shows the subject KSM who watched Chapter 1 in 5.1ch, and Fig.8 shows the subject MRT who watched in 2ch. The black arrow shows a 2-minute scene where the sound of the airplane spreads in the left and right directions in surround. From the point of the black arrow, the head movement of 5.1ch KSM showed a change in the horizontal and vertical directions, but no change was seen in 2ch MRT.

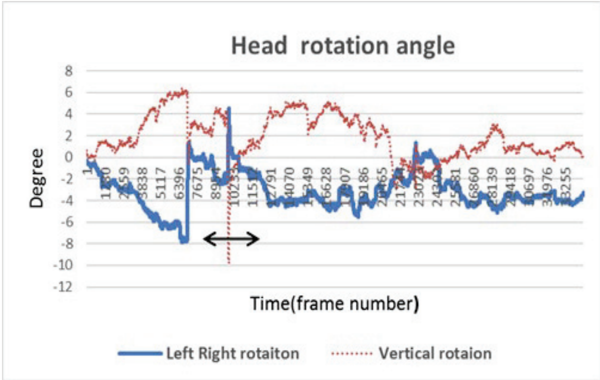

Fig.7 Chapter 1 KSM 5.1ch case

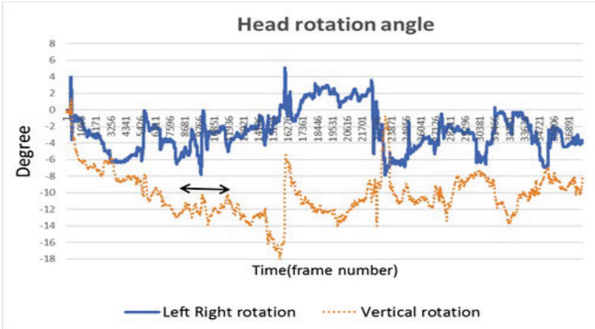

Fig.8 Chapter 1 MRT 2ch case

Now let us consider the present experiment. Figure 9 shows horizontal movement according to the number of channels, and Fig. 10 shows the vertical rotation. In both figures, the green, orange, and blue lines correspond to 2, 5.1, and 7.1-ch, respectively.

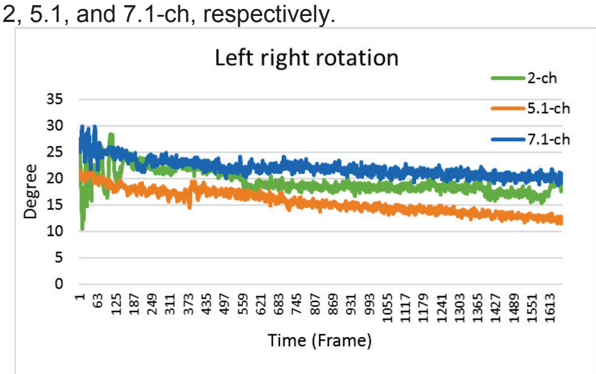

Fig. 9 Left-right rotation

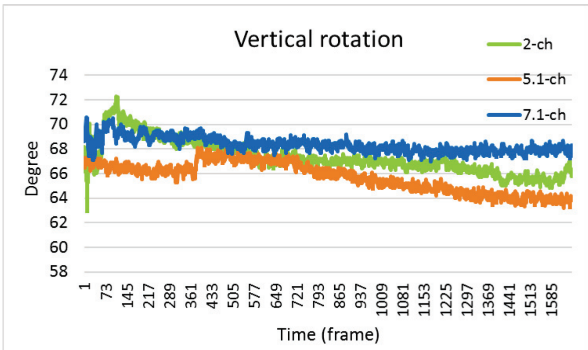

Fig. 10 Vertical rotation

Abrupt changes were seen in previous studies[5], but in the present experiment, moderate changes were seen. In a previous study, lateral movement involved head movement of 10 degrees or more, while in the present experiment, we saw a range of about 5 degrees that gradually changed from left to right regardless of the number of channels. It was found that the vertical movement also changed smoothly from the front to the bottom. Both horizontal and vertical movement showed the tendency that the movement of the 5.1ch was bigger than those of 7.1ch and 2ch.

4 DISCUSSION

When presented in 5.1ch in the previous research, more head movements and changes in the gazing point occurred than when combined with 2ch sound. This result suggests that the effects caused by 5.1ch surround can be evaluated by these parameters. Since 5.1ch surround is excellent in sound reproduction in the left-right direction, we think that the average value of the gazing point may vary in the left-right direction. As for the head movement, the subjects presented in 5.1ch also had a sudden change that was drawn in the direction of the sound, so the reproducibility of the sound increased the attention of the image and moved the head. In the previous research, it was thought that the greater the number of channels, the higher the reproducibility of the sound, and the greater the sense of reality, the greater the gaze position and the greater the movement of the head. However, in the new experimental method used in the present study, the line of sight was the same as that in the previous study in the case of 2-ch stereo sound, but there was a gradual change in head movement. Given that the 3 types of sound were presented at random, we believe that the head may have been moved based on preconceptions about the sound. Regarding the fact that there were no sudden changes in the present experiment, the size of the screen may be related. This idea is supported by results reported by Hatada et al., [6] who found that when the viewing angle at which the left and right edges of the screen are viewed exceeds 20 degrees, subjective coordinates are gradually guided to the image space, and the viewer experiences a sense of reality and high resolution. A viewing angle of 30 degrees has been established as a desirable viewing condition for this effect to become noticeable [6].

5 CONCLUSIONS

In order to elucidate the basic characteristics of gaze and head movement by acoustic effects, we compared subjects' behavior under two kinds of sound systems. We analyzed the head rotation angle and the distribution of gaze points. As a result, it was shown that the fundamental characteristics of the line of sight and head movements during image viewing can be elucidated using the sound effects produced by 5.1ch surround sound using these data. When 5.1ch sound accompanies the movie, changes in head movement and gaze distribution were observed.

However, this phenomenon was limited to chapters with characteristic sounds such as aircraft movement and live noise. In addition, the t-test results show that the condition that changes the line of sight is that the sound is surround sound and includes characteristic sounds. Based on this result, a new experimental method was considered based on the premise of viewing video at home. We found that average gaze positions with 2ch clustered near the center of the display when viewing, but spread to the left and right as the number of channels increased. Head movement of 5.1ch showed relatively big movement. About the reason why the head movement while watching movie with 7.1ch became smaller than the head movement when watching the 5.1ch movie, a problem of the editing of the sound and/or the characteristic problem of headphones which reproduces surround sound expression are considered. A further examination will be necessary to analyze effect of multi-channel surround on views. In order to obtain a synergistic effect between sound and image, a certain field of view is required.

Since only one subject participated in the present experiment, statistical analysis on the distribution of gazing points requires further study with larger numbers of subjects. In future research, we would like to clarify the basic characteristics of eye and head movement based on the synergy of sound and image.

REFERENCES

[1] H. Takahira, S. Mochiduki, M. Yamada, "Analysis of Gaze Movement When Changing Viewing Distance and Viewing Position" Trans. Inst. Electron. Inform. Communi. Engnr. Jpn. (D), vol.J99-D, no.3, pp.293-306, 2016.

[2] S.Mochiduki, H.Takahira, M.Yamada, "Analysis of Head movement viewing the display with wide viewing angle,"IMQ, vol.115, no.133, pp.15-20, 2015.

[3] T.Ito, A.Ando, "High realism sound system and its evaluation", No.108 ,NHK STRL R&D,3,2008, (in Japanese).

[4] H. Takahira, K. Kikuchi and M. Yamada, "A System for Measuring Gaze Movement and Hand Movement Simultaneously for Hand-Held Devices", IEICE Trans. Commun, E98-B, 1, pp.51-61, 2015.

[5] Y.Kobayashi, S.Mochiduki, M.Yamada "The effect of sound effects on Eye movements and head movement And examination of subjective evaluation" , Y.Kobayashi, S.Mochiduki, M.Yamada, Kobe University, 28th september IMQA 2018

[6] T.Hatada, H.Sakata, H.Kusaka "Research on high-definition screen effects that create a high sense of presence" ITE,1979