

Auditory and Visual Properties in the Virtual Reality Using Haptic Device

Manabu Ishihara^(✉) and Takafumi Komori

National Institute of Technology, Oyama College, Oyama-City, Tochigi 323-0806, Japan
ishihara@m.iece.org

Abstract. The advantages of a system that uses virtual reality (VR) are that the software program can be changed to permit various types of technical training to be performed with a single device, and that the work environment can also be changed easily. Another advantage is that a network can be used to allow multiple users to train at different remote locations. An example of this phenomenon is the lag between video and sound in a network teleconference system. In an environment where latency exists, such a system cannot be said to be suitable as a technical training system, and this is a problem when using a VR system.

Keywords: Virtual Reality (VR) · Network · Visual · Auditory · Haptic

1 Introduction

The advantages of a system that uses virtual reality (VR) are that the software program can be changed to permit various types of technical training to be performed with a single device, and that the work environment can also be changed easily. Another advantage is that a network can be used to allow multiple users to train at different remote locations. With a standalone system, an independent device that does not interface with a network will experience no latency, and no time difference Δt will be generated for various sensory operations. However, with a VR space connected via the Internet or other network, as a result of network latency and packet loss, as well as differing amounts of information, the data transmission times for various sensory operations will not necessarily be the same. An example of this phenomenon is the lag between video and sound in a network teleconference system. In an environment where latency exists, such a system cannot be said to be suitable as a technical training system, and this is a problem when using a VR system.

Concurrent with these advances has been a wealth of research on haptic interface technology [1], and educators have begun exploring ways to incorporate teaching tools utilizing touch properties in their curriculums [2, 3]. This will make replacing familiar teaching tools with digital media incorporating VR seem more attractive. For example, various learning support systems that utilize virtually reality (VR) technology [4] are being studied. Examples include a system that utilizes a stereoscopic image and writing brush display to teach the brush strokes used in calligraphy [5, 6], the utilization of a robot arm with the same calligraphy learning system [7], a system that uses a “SPIDAR” haptic device to enable remote calligraphy instruction [8], and systems that analyze the

learning process involved in piano instruction [9] or in the use of virtual chopsticks [10]. However, with a VR space connected via the Internet or other network, as a result of network latency and packet loss [11, 12, 13], as well as differing amounts of information, the data transmission times for various sensory operations will not necessarily be the same.

In this study, we created a drum performance system in a VR space to investigate the effect of visual, auditory, and haptic sensation time differences generated while performing an operation in a VR space. As a result, we clarified the impact that delays of various information in a VR space have upon a user, and the impact of delays upon a user when that system is adjusted to more closely approximate an actual network.

2 Experiment Using Ball Striking Action (Preliminary Experiment)

2.1 Description of the Experiment

First, in order to investigate the linkage between cooperative characteristics of haptic, visual, and auditory senses in a VR space, we used PHANToM to create a program that will strike a ball within a VR space. Striking the ball causes a reaction force to be returned, and the user is thus able to sense the striking of the ball in the VR space.

In this study, we used a PHANToM Omni Device (Sensable Technologies) as our haptic device. It was attached to a control computer (CPU: Intel® Core™ i7-2600[3.00 GHz], RAM:4.00 GB, OS:Windows7Pro.,64bit) running Open-Haptics™ toolkit v3.0 as the control program. The participants were seven male (17-20 years old).

The action of striking a ball with a pole causes a collision that returns a reaction force and exhibits basic characteristics. A sound is generated when the ball is struck, and the collision causes the ball to begin to move with respect to the pole. The sound of a billiard ball being hit was used as the sound generated when the ball is struck. A lag between the generation of sound and the start of motion imparted a feeling of sensory discomfort to test subjects when using this system. Therefore, we investigated human reactions and the acceptable range for deviation arising from differences in visual, auditory, and haptic operations occurring in the ball strike timing.

Figure 1 shows the execution screen. The blue sphere in the screen represents the tip of the pole, and PHANToM has been used to create a program in which the tip of the pole is moved to strike the white sphere in the center of the screen. As shown in the figure, by generating delays between the sound and haptic sense generated when striking the ball, and between the haptic sense and the video of the ball beginning to move, we tested the acceptable ranges of time difference for various senses.

The test subjects compared various measurement values for which the video and sound were delayed to a non-delayed reference of 0 ms, and selected one of three criteria corresponding to whether the generated video and sound seemed to be earlier than, later than, or simultaneous with the haptic reference.

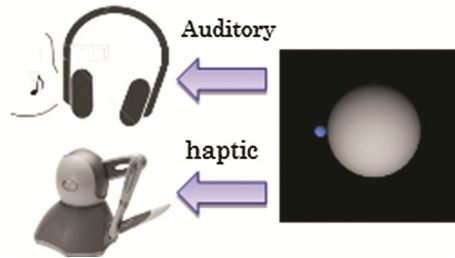


Fig. 1. Overview of the experiment

2.2 Experimental Results

Figures 2 and 3 show the experimental results of the visual and haptic characteristics and the auditory and haptic characteristics, respectively. The horizontal axes indicate the time difference until the respective video or sound is generated with respect to the haptic reference, and the vertical axes indicate the number of people who felt a delay with that time difference. The test subjects were men and women in their teens or twenties; 15 people participated in the visual–haptic experiment and 7 people participated in the auditory–haptic experiment. The time difference settings in these experiments ranged from 0 to ± 150 ms in 10 ms intervals (31 points) and ± 200 ms for a total of 33 points for the visual-haptic experiments and from 0 to ± 200 ms in 10 ms intervals for the auditory–haptic experiments for a total of 41 points.

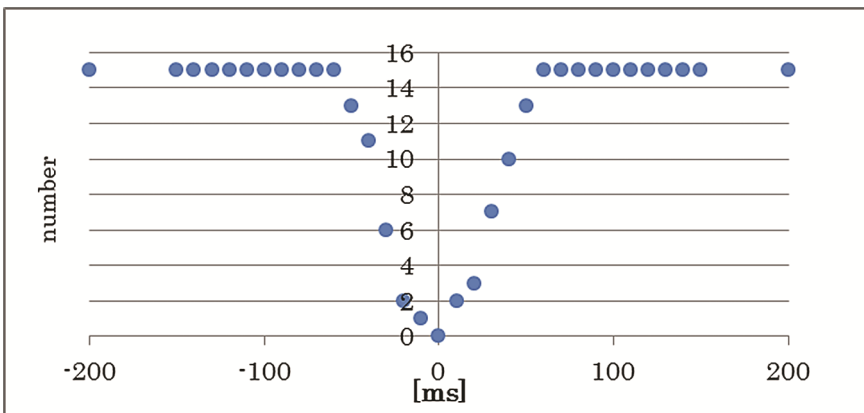


Fig. 2. Visual–haptic sense coordination characteristics

From Fig. 2 it can be seen, for example, that when a 30 ms time difference is generated between the video of the start of ball motion and the haptic reaction force when the ball is struck, 7 people were able to distinguish that difference.

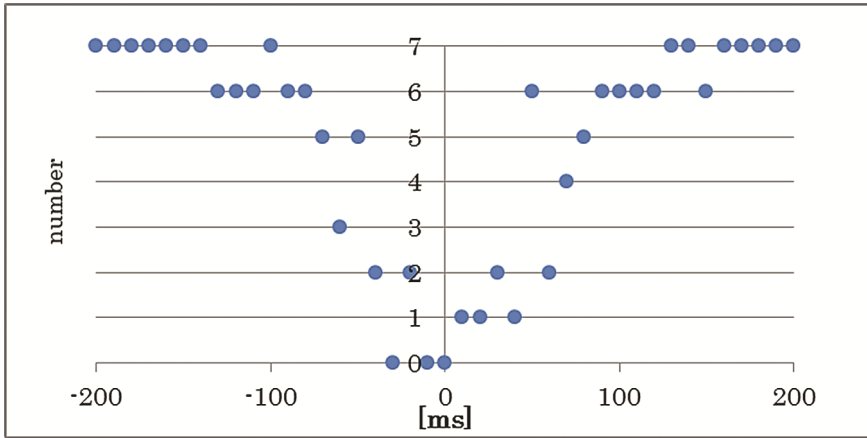
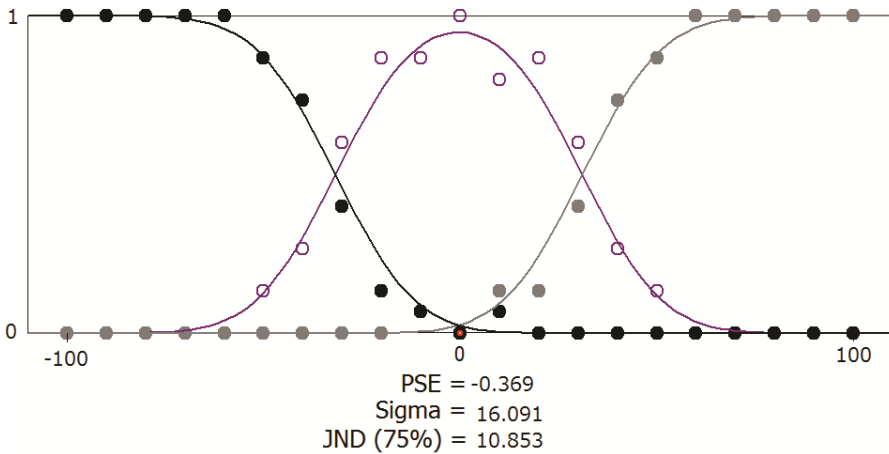


Fig. 3. Auditory-haptic sense coordination characteristics

From Figs. 2 and 3, it can be seen that as the delay time increases, the number of people capable of feeling a difference from the reference also increases. Furthermore, because the increase in people who can feel a difference occurs at a gentler rate in Fig. 3 than in Fig. 2, we can conclude that the effect of an auditory delay has less of an impact than the effect of a visual delay.



Lower difference threshold = -11.222 Upper difference threshold = 10.48

Fig. 4. Visual-haptic sense characteristics

From Fig. 2, in comparing visual and haptic senses, it can be seen that at a time difference of ± 30 ms, approximately 50 % of the people began to feel a difference, and at about ± 50 ms and larger, nearly all the people could feel a difference. From Fig. 3, in comparing auditory and haptic senses, it can be seen that beyond a time difference of

about ± 50 ms, more than 50 % of the total people began to feel a difference, and that from differences of about ± 80 ms, nearly all of the test subjects could feel a difference.

From the obtained results, the method of constant stimuli was used to analyze the time difference threshold, and Figs. 4 and 5 show those findings.

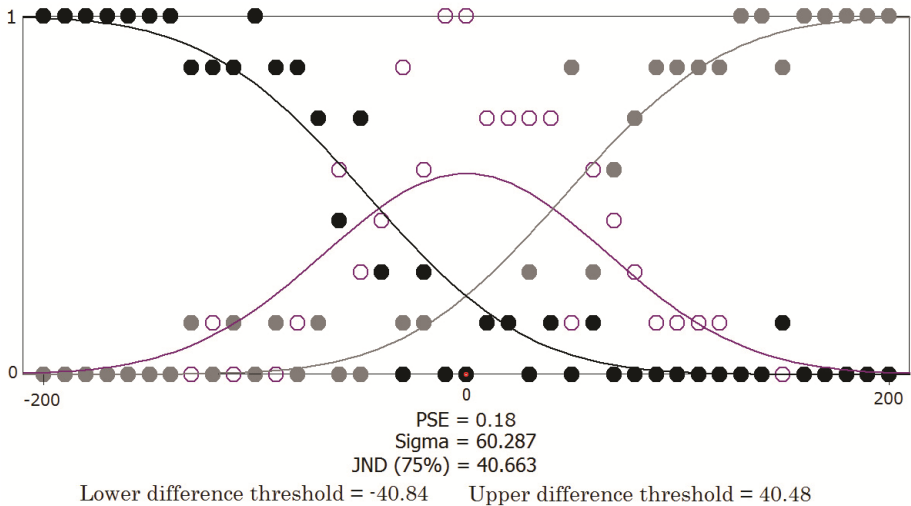


Fig. 5. Auditory–haptic sense characteristics

From Fig. 4, in a comparison of auditory and haptic senses, with respect to a standard stimulus of 0 ms, an upper limen of 10.48 ms and a lower limen of -11.222 ms were obtained. This finding indicates that for time differences of up to approximately ± 10 ms, 75 % of the test subjects did not distinguish any difference from 0 ms. Similarly, from Fig. 5, with respect to a standard stimulus of 0 ms, an upper limen of 40.48 ms and a lower limen of -40.84 ms were obtained. This finding indicates that in a comparison of auditory and haptic senses, for time differences of up to approximately ± 40 ms, 75 % of the test subjects did not distinguish any difference from 0 ms.

It can be seen that the number of users who feel a difference in the program begins to increase when, due to delays, the haptic–visual data time difference begins to exceed 10 ms or the haptic–auditory data time difference begins to exceed 40 ms. Visual sense is said [14] to have a greater impact than auditory sense, and that statement is consistent with these findings. Moreover, for visual sense, there exists research showing that people begin to sense network latency when the delay reaches approximately 30 ms, and this is consistent with the finding that 50 % of the test subjects began to feel a difference at this level.

3 QoE Measurement with Drum Performance System in a VR Space

3.1 Drum Performance System

In this experiment, we created a drum performance system in a VR space to investigate the impact of delay on a person who is performing a more realistic task and the associated quality of experience (QoE). In addition, we changed the way in which the delay was generated to investigate the acceptable range of delays for a person in a state that more closely approximates network latency than in the previous experiment. Figure 6 shows the execution screen of the created drum performance system. The stick on the screen is moved using a PHANTOM haptic device so as to strike a drum. At the timing when the stick strikes a drum, a sound and reaction force are generated, providing the user with a feeling of actually playing the drums. When the stick strikes a drum, we applied delays to the video in the screen and to the sound in order to investigate human reactions to latency in visual, auditory, and haptic operations and the acceptable ranges of those delays.

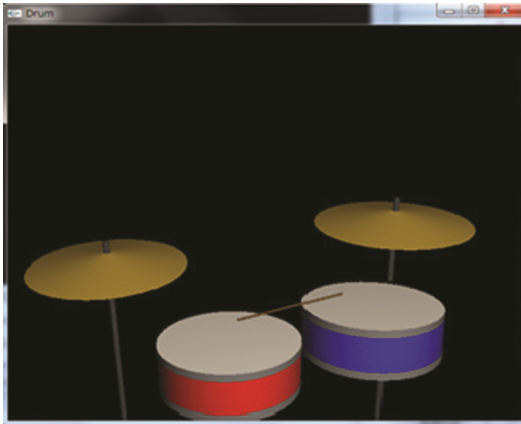


Fig. 6. Program execution screen

3.2 Measurement with Constant Delay and Random Delay

In this experiment, we took measurements using two types of delays: constant and random. Each type is explained below.

3.2.1 Constant Delay

When a drum is struck, a predetermined single type of delay time only is generated. Since a constant delay will always be generated, this differs from an actual network environment, but it is thought that basic human responses to delays could be measured in this manner.

3.2.2 Random Delay

If the same delay time always occurred for all communication, then the delay could be predicted and managed accordingly, but delay times are not necessarily constant. Especially in the case of the Internet, since in principle, the location through which a packet will be transferred is not known, predicting the arrival time of a packet is extremely difficult. In consideration of the fluctuation in such a network, we conducted experiments in which random delays of two or more preset delay times would occur when a drum is struck. Accordingly, the delay format more closely resembled the state when using an actual network than in the case of a constant delay. For this experiment, two types of delays (Delay 1 and Delay 2) were preset in advance. We measured the rates of occurrence of the two delays and set them to occur with nearly equal probability.

3.3 QoE Measurement of Drum Performance System with Constant Delay

Video and sound delays were generated when a drum is struck, and we investigated the impact on people at that time. In the previous experiment, measurements were made using three criteria, but in this experiment, the users performed evaluations to rank each delay on a five-step scale from 1 to 5. The criteria for the five evaluations are shown in Table 1.

Table 1. Experimental evaluation ranking

evaluated value	quality	operability of system
5	Excellent	to have good operability
4	Good	v
3	Fair	v
2	Poor	v
1	Bad	to have not good operability

The experiment was conducted for both sound and video with measurements taken in the range of 0 to 150 ms in 10 ms intervals for a total of 16 measurement points. The delays were presented to the test subjects at random, without prior notification. The experiment was conducted multiple times for each test subject, and measurements were taken a total of 20 times for 7 test subjects.

Figure 7 shows the experimental results. From this figure, it can be seen that the users evaluated the delay at lower values as the delay time increased. It can also be seen that a video delay has a greater impact than a sound delay on the evaluation results.

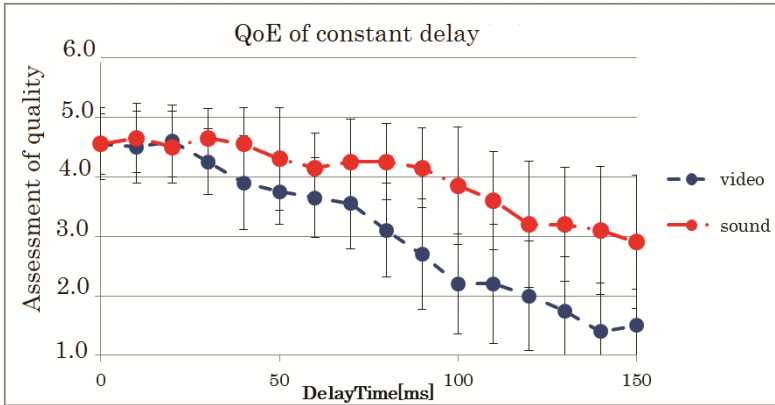


Fig. 7. QoE measurement results with constant delay

In addition, for sound delays greater than the standard deviation, the evaluation value increased gradually as the delay time increased. For video delays, however, it can be seen that the evaluation values became smaller if the delay time became excessively large. This is because users are thought to concentrate their evaluations at the “poor” level when the video delay time becomes large. On the other hand, for sound delays, there was a variety of evaluations even when the delay time was 150 ms. There is thought to be a delay time zone during which individual differences appear in the impact of a delay.

For video delays, evaluation values of 3 or below for “normal” or lower quality were recorded at 80 ms, but for sound delays, there were no evaluation values of 3 or less until 150 ms. From the results of the previous experiment in which the number of users who feel a difference in the program begin to increase when the haptic-visual data time difference begins to exceed 10 ms or the haptic-auditory data time difference begins to exceed 40 ms, there is thought to be a discrepancy between the time difference value at which a difference is felt due to delay and the time difference value at which quality is affected significantly.

3.4 QoE Measurement of Drum Performance System with Random Delay

The measurement range consists of Delay 1 fixed at a single value and Delay 2 varied from 0 to 150 ms in 10 ms intervals, for a total of 16 points. Here, the experiment was performed for two random delays, with the value of Delay 1 being fixed at 0 ms and fixed at 30 ms.

3.4.1 QoE Measurement with Random Delay (Delay 1 = 0 Ms)

Figure 8 shows the experimental results. From Fig. 8, it can be seen that for a random delay in which Delay 1 is fixed at 0 ms, the video quality is evaluated at an overall higher value than in the case of a constant delay, and the evaluation values have risen to a level comparable to those of a sound delay.

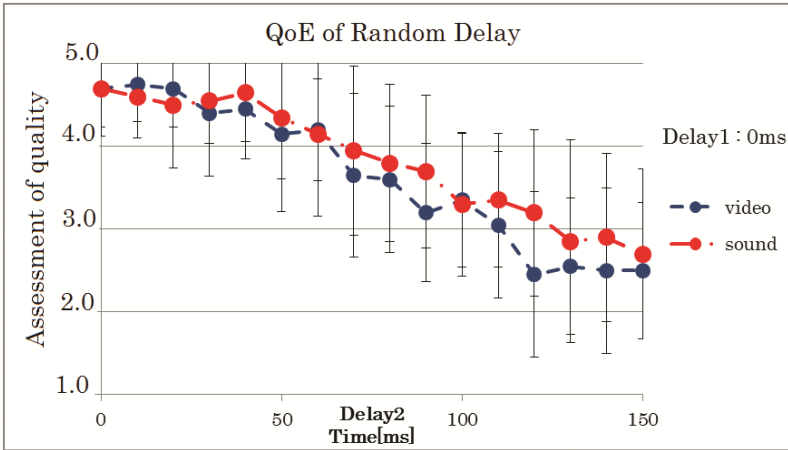


Fig. 8. QoE measurement results with random delay (Delay 1 = 0 ms)

When a video with a constant delay is evaluated, the quality drops below the “normal” level at evaluation value 3 in the vicinity of 80 ms, but with a random delay, the quality evaluation maintains a value of 3 or higher up to 110 ms. Therefore, we found that even if the delay time is not stable, as long as one of the delay times is short, QoE will be able to maintain a high value. Furthermore, in the case of a sound delay, no significant changes with respect to a constant delay were seen. As for the standard deviation, in contrast to a constant delay, we confirmed that the standard deviation would spread out as the delay time increased, even with a video delay. In the case of a sound delay, the spreading out of the standard deviation is more pronounced for a random delay than for a constant delay. From the above, we found that increasing the number of delays presented and more closely approximating the state of using a network causes the determination of quality by users to deviate even further.

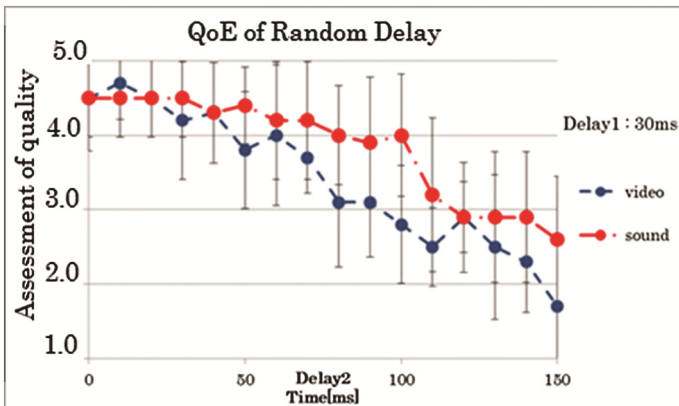


Fig. 9. QoE measurement results with random delay (Delay 1 = 30 ms)

3.4.2 QoE Measurement with Random Delay (Delay 1 = 30 Ms)

Compared to the case when the value of Delay 1 was fixed at 0 ms, the evaluation of random video delays results in low overall values, and this can be seen in Fig. 9. Accordingly, for similar random delays, it is thought that the evaluation values will be lower when the longer delay time is longer.

This result also shows that, for a video delay, people are able to discriminate time differences when the time difference is about 30 ms.

Additionally, Figs. 10 and 11 show the results of Figs. 7 through 9 compiled into graphs for video delays and sound delays, respectively.

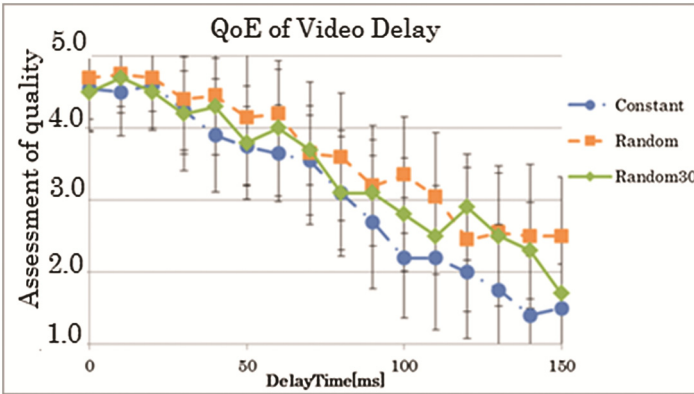


Fig. 10. QoE evaluation values with video delays

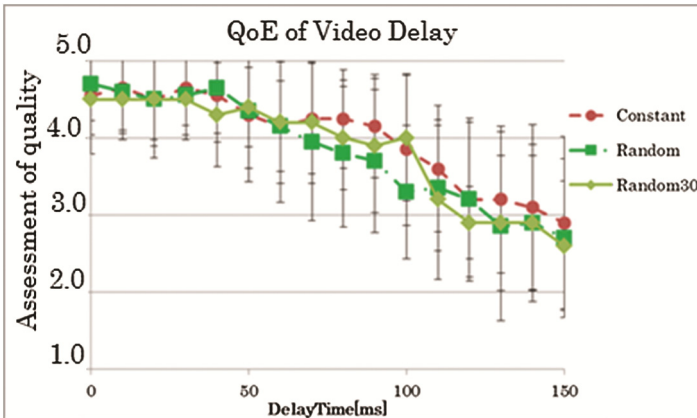


Fig. 11. QoE evaluation values for sound delays

From Figs. 10 and 11, it can be seen that for a video delay, there are no large changes in the evaluation values while the delay time is small, but when the delay time reaches about 80 ms, most evaluation values are in agreement with Constant Delay < Random

Delay (Delay 1 = 30 ms) < Random Delay (Delay 1 = 0 ms). For sound delays, no large change in QoE was observed for either constant delays or random delays.

4 Concluding Remarks

In this study, it can be seen that the number of users who feel a difference in the program begins to increase when, due to delays, the haptic–visual data time difference begins to exceed 10 ms or the haptic–auditory data time difference begins to exceed 40 ms. Visual sense is said to have a greater impact than auditory sense, and that statement is consistent with these findings. Moreover, for visual sense, there exists research showing that people begin to sense network latency when the delay reaches approximately 30 ms, and this is consistent with the finding that 50 % of the test subjects began to feel a difference at this level.

When a video with a constant delay is evaluated, the quality drops below the “normal” level at evaluation value 3 in the vicinity of 80 ms, but with a random de-lay, the quality evaluation maintains a value of 3 or higher up to 110 ms. Therefore, we found that even if the delay time is not stable, as long as one of the delay times is short, QoE will be able to maintain a high value. Furthermore, in the case of a sound delay, no significant changes with respect to a constant delay were seen. As for the standard deviation, in contrast to a constant delay, we confirmed that the standard deviation would spread out as the delay time increased, even with a video delay. In the case of a sound delay, the spreading out of the standard deviation is more pronounced for a random delay than for a constant delay. From the above, we found that increasing the number of delays presented and more closely approximating the state of using a network causes the determination of quality by users to deviate even further.

This result also shows that, for a video delay, people are able to discriminate time differences when the time difference is about 30 ms.

It can be seen that for a video delay, there are no large changes in the evaluation values while the delay time is small, but when the delay time reaches about 80 ms, most evaluation values are in agreement with Constant Delay < Random Delay (De-lay 1 = 30 ms) < Random Delay (Delay 1 = 0 ms). For sound delays, no large change in QoE was observed for either constant delays or random delays.

Acknowledgements. This work was supported by JSPS KAKENHI Grant Number 25350369.

References

1. Ohnishi, H., Mochizuki, K.: Effect of delay of feedback force on perception of elastic force: a psychophysical approach. *IEICE Trans. Commun.* **E90-B**(1), 12–20 (2007)
2. Ishihara, M.: On first impression of the teaching materials which used haptic display. *IEEE J. Trans. Fundam. Mater.* **129**(7), 490–491 (2009). (in Japanese)
3. Ishihara, M.: Haptic device using a soldering test system. In: Tino, A., Stephanidis, C. (eds.) *HCI 2015 Posters. CCIS*, vol. 528, pp. 190–195. Springer, Heidelberg (2015). doi: [10.1007/978-3-319-21380-4_34](https://doi.org/10.1007/978-3-319-21380-4_34)

4. Hirose, M., et al.: *Virtual Reality*, Sangyo Tosho (1993) (in Japanese)
5. Yoshida, T., Muranaka, N., Imanishi, S.: A construction of educational application system for calligraphy master based on virtual reality. *IEEE J. Trans. Electron. Inf. Syst.* **117-C**(11), 1629–1634 (1997). (in Japanese)
6. Yoshida, T., Yamamoto, T., Imanishi, S.: A calligraphy mastering support system using virtual reality technology and its learning effects. *IEEE J. Trans. Fundam. Mater.* **123-A**(12), 1206–1216 (2003). (in Japanese)
7. Henmi, K., Yoshikawa, T.: Virtual lesson and its application to virtual calligraphy system. *TVRSJ* **3**(1), 13–19 (1983). (in Japanese)
8. Sakuma, M., Masamori, S., Harada, T., Hirata, Y., Satou, M.: A Remote Lesson System for Japanese Calligraphy using SPIDAR, IEICE of Jpn., Technical Report, MVE99–52, pp. 27–32, Oct. 1999 (in Japanese)
9. Otsuka, G., Sodeyama, G., Muranaka, N., Imanishi, S.: A construction of a piano training system based on virtual reality. *IEEE J. Trans. Electron. Inf. Syst.* **116-C**(11), 1288–1294 (1996). (in Japanese)
10. Yamaguchi, Y., Kitamura, Y., Kishino, F.: Analysis of learning process of virtual chopsticks. IEICE of Jpn., Technical Report, MVE2001-3, pp. 11–16, June 2001 (in Japanese)
11. Ishihara, M.: Empirical study regarding representing roughness with haptic devices. In: *Proceedings of 2013 IEEE 2nd GCCE*, pp. 471–473, Chiba, Japan, Oct. 2013
12. Sithu, M., Ishibashi, Y., Fukushima, N.: Effects of dynamic local lag control on sound synchronization and interactivity in joint musical performance. *ITE Trans. Media Technol. Appl.* **2**(4), 299–309 (2014)
13. Ishihara, M., Komori, T.: Characteristic evaluation of the audio and visual system of haptic device. In: *Proceedings of 2014 IEEE 3rd GCCE, CFP14GCB-USB*, pp. 356–357, Chiba, Japan, Oct. 2014
14. Ishihara, M., Negishi, N.: Effect of feedback force delays on the operation of haptic displays. *IEEJ Trans. Electr. Electron. Eng.* **3**(1), 151–153 (2008)