A study on visual attention for measuring responsiveness in online performance environment

Tayhyeong Kim
School of IT Convergence
University of Ulsan
Ulsan, South Korea
steve181813@gmail.com

Wonkuk Boo
School of IT Convergence
University of Ulsan
Ulsan, South Korea
wkb0307@naver.com

Ryongha Kim
School of IT Convergence
University of Ulsan
Ulsan, South Korea
kimha4153@naver.com

Daehwan Kim
School of IT Convergence
University of Ulsan
Ulsan, South Korea
daehwankim@ulsan.ac.kr

Yongwan Kim
VR/AR Content Research Section, Communications & Media Research Laboratory
Electronics and Telecommunications Research Institute (ETRI)
Daejeon, South Korea
ywkim@etri.re.kr

Jinsung Choi
VR/AR Content Research Section, Communications & Media Research Laboratory
Electronics and Telecommunications Research Institute (ETRI)
Daejeon, South Korea
jin1025@etri.re.kr

Ki-Hong Kim
VR/AR Content Research Section, Communications & Media Research Laboratory
Electronics and Telecommunications Research Institute (ETRI)
Daejeon, South Korea
kimgh@etri.re.kr

***Abstract*— This paper is a study to measure the audience's responsiveness by visual attention in an online performance environment. Due to the nature of online performances, it is not possible to know the direct reaction of the audience, so an indirect method of measuring the level of responsiveness is needed. However, since responsiveness is a complex emotion, it is not easy to measure. Recently, studies on selection and concentration using visual attention are in progress. Therefore, the effect of visual attention on responsiveness was investigated through experiments. The experimental results indicate that visual attention is relatively related to the measure of responsiveness.**

Keywords—responsiveness, visual attention, online performance environment.

I. Introduction

Recently, a lot of online performances have been held through online media[1]. Online performances allow a large audience to participate, and real-time communication with the audience compensates for the monotony of simple video viewing. However, the realism of the performance is still not properly conveyed to the audience, and it is difficult for the performer to conduct a performance that allows mutual communication because it is not possible to know the audience's response. Due to the nature of online performances, it is not easy to determine the level of individual audience response. Since it is an environment where people watch in a personal space like home, the degree of response to watching performances is relatively low compared to offline performances. Also, depending on the cultural characteristics, the reaction to the performance is very different.

To become an online performance capable of real-time interaction, a method of measuring online audience acceptance is needed. Recently, studies to evaluate the concentration of video viewing using visual attention are being actively conducted. Visual attention is the voluntary filtering out of unimportant visual perceptual information and selectively focusing on specific important information. In particular, eye-tracking methods help to characterize individual differences that reflect individual attentional shifts and biases[2]. Therefore, this paper conducted a study on visual attention to measure the level of responsiveness in the online performance environment.

II. Background

Many studies on visual attention based on eye-tracking are in progress [2]. There is visual question answering [3, 4], a field of study that infers the correct answer to a given question about an image. The NiCATS system [5] identifies students' level of interest in instruction in the classroom. Atar-HEAD [6] provides a dataset on how visual attention affects human decision-making performance. There is also a study that the human drone pilot's visual attention helps learn to operate an FPV autonomous drone [7].

From these various studies, it has been found that visual attention is widely used by eye-tracking algorithms. Therefore, we used an eye-tracking algorithm to find out the effect on the responsiveness of online performances. In addition, face pose estimation was also used to increase the accuracy. To reproduce a real situation, an image-based face pose estimation and gaze tracking algorithm was used instead of using expensive gaze tracking equipment.

III. Design of Study on Visual Attention to Measure Responsiveness in an Online Performance Environment

We assumed the following environment to find out the effect of visual attention on the level of responsiveness in an online performance environment.

- Online performance viewing environment: small TV, monitor or smartphone
- Audience video capture device: webcam

- Type of viewing video: music video, large-scale performance, or virtual reality performance

Figure 1 shows examples of viewing images captured in the assumed environment. It is typically a viewing screen in a home environment using a TV or PC. It roughly shows a person's upper body.

Fig. 1. Exmpales of viewing images captured in the online performance environment. The white face mesh points are the result of using the face landmark detection algorithm (mediapipe[8]) to estimate the 3D face pose.

To judge the visual attention of performance content, we used a 3D face landmark detection and pupil detection algorithm [8]. The direction of the gaze of the face was estimated from the found 3D facial landmark points, and the direction of the eyes was also estimated by detecting the center of the pupil, the head of the eye, and the tail of the eye. Figure 2 shows images of the estimated face direction and eye gaze direction.

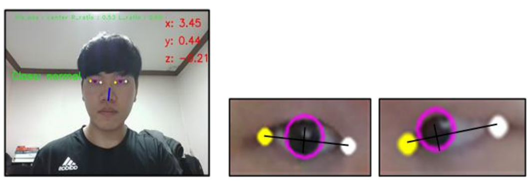

Fig. 2. (Left) The blue line represents the estimated face direction, and (Right) the pink circle represents the estimated eye direction.

We hypothesized that the degree of responsiveness to viewing a performance would be high when the concentration of visual attention on the viewing screen was high. Therefore, it was classified whether the direction of the estimated face and pupil was toward the viewing screen or not. Visual attention was defined as a total of 5 directions (front, up, down, left, right), and it was assumed that if you looked for more than 5 seconds, you were looking in that direction. To recognize the direction of visual attention, it was learned using a convolution neural network (CNN). In addition, long-short term memory (LSTM) was used to determine whether the gaze was correct for more than 5 seconds. Figure 3 shows an example of direction and gaze classification results.

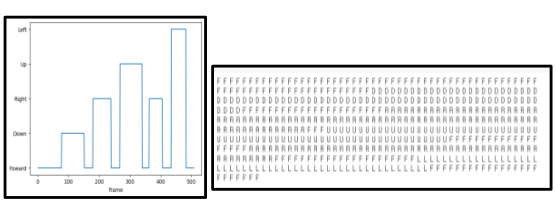

Fig. 3. An example of the gaze direction and gaze classification result for each frame.

To test whether the degree of responsiveness is highly correlated with visual attention, audiences were asked to watch their favorite performance videos and random performance videos. It was configured to measure the direction of visual attention in every frame and store the direction value. The viewing time of favorite performance videos and random performance videos was set to around 15 minutes, and the degree of responsiveness was measured according to the ratio of the direction of visual attention.

IV. Experiment

To measure the effect of visual attention on responsiveness, a total of 10 people participated in the experiment. Each of them chose two of their favorite performance videos, and the other two videos were provided randomly. The experiment participants selected their favorite performance videos from the YouTube site. To remove the quality factor according to the performance video, all 4K video was selected. A total of four videos were arranged in random order, and the interval between viewing the videos was set at 10 minutes. It was recommended to watch it in the most natural position possible and to watch it without any external restrictions. Figure 4 shows an example of storing visual attention direction information and time information for an arbitrary section of one performance video.

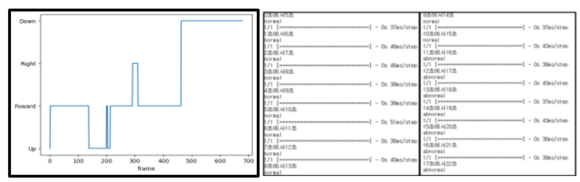

Fig. 4. An example of visual attention directions per frame and their temporal information.

Responsiveness was measured as the ratio of frames viewed from the front of the total number of frames. Table I shows the responsiveness values measured after watching favorite performance videos and random performance videos for 10 experiment participants. The experiment results showed that watching the favorite performance videos showed a 1.4 times higher responsiveness value than watching random performance videos. It was found that the visual attention experiment had a great influence on the degree of responsiveness.

TABLE I. Average of Responsiveness

	Favorite performance videos	Random performance videos
# of videos	20	20
Resoponsivess value	0.85	0.594

V. Conclusion and Future Work

This paper is a study on the effect of visual attention on the measurement of responsiveness in an online performance viewing environment. Visual attention was estimated by the direction of the eyes and pupils, and the experiment was conducted assuming that the subjects were responding to the performance when looking straight ahead. It was found that the responsiveness rate was relatively high when watching a favorite performance. However, responsiveness cannot be measured only with visual attention. Since responsiveness is an expression of complex emotions, we plan to experiment with features such as facial expressions and body posture.

ACKNOWLEDGMENT

This research was supported by Culture, Sports and Tourism R&D Program through the Korea Creative Content Agency grant funded by the Ministry of Culture, Sports and Tourism in 2023. (Project Name: Development of Virtual Reality Performance Platform Supporting Multiuser Participation and Realtime Interaction, Project Number: R2021040046)

REFERENCES

[1] J. W. Jun, “Effects of COVID-19 information and consumer characteristics on untact performance perception,” Studies of Broadcasting Culture, vol. 33, no. 1, pp.193-214, 2021.

[2] O.A. Efeturk, G. Turgut, H. Asadi Dereshgi, and A. Yilmaz, "A review of visual attention research using eye-tracking technologies," Journal of Smart Systems Research, vol.3, no.2, pp. 128-135, 2022.

[3] W. Guo, Y. Zhang, J. Yang, and X. Yuan, “Re-attention for visual question answering,” IEEE Transactions on Image Processing, vol. 30, pp. 6730-6743, 2021.

[4] T. Rahman, S.H. Chou, L. Sigal, and G. Carenini, "An improved attention for visual question answering," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1653-1652, 2021.

[5] A. Sanders, B. Boswell, G. S. Walia, and A. Allen, "Non-intrusive classroom attention tracking system (nicats)," IEEE Frontiers in Education Conference (FIE), pp. 1-9, 2021.

[6] R. Zhang, et. al "Atari-head: Atari human eye-tracking and demonstration dataset." Proceedings of the AAAI Conference on artificial intelligence, vol. 34, pp. 6811-6820, 2020.

[7] C. Pfeiffer, S. Wengeler, A. Loquercio, and D. Scaramuzza, "Visual attention prediction improves performance of autonomous drone racing agents." Plos one, vol. 17, no. 3, 2022.

[8] https://developers.google.com/mediapipe/solutions/guide