The degree of gaze-induced shifts in overt attention explains inter-subject variability in long-term memory performance

Touchai THAWAI[a,b], Sakol TEERAVARUNYOU[a,b], Sirawaj ITTHIPURIPAT[a,c]

[a]*Learning Institute, King Mongkut's University of Technology Thonburi, Bangkok, Thailand*

[b]*School of Architecture and Design, King Mongkut's University of Technology Thonburi, Bangkok, Thailand*

[c]*Psychology Department, Vanderbilt University, Nashville, Tennessee, USA*
itthipuripat.sirawaj@gmail.com

Abstract: Gaze is an important cue thought to facilitate effective social interaction and communication. Previous studies have shown that gaze could induce an attentional shift toward a location that match gaze direction and this attentional shift could in turn enhance sensory information processing in simple perceptual tasks. However, less is known about how gaze cues may influence selective attention and long-term memory in more complex real-world tasks. To examine this issue, we monitored eye movements via an infrared eye-tracking camera in 95 male and female adults, while they were reading and listening to sentences containing autographical information. On some trials, the sentence was presented by itself. On other trials, there was also an animated facial stimulus, which either gazed toward the sentence (congruent), gazed directly to the viewer (neutral), or gazed away from the sentence (incongruent) . We found that the congruent gaze cue effectively induced overt shifts of attention to the sentence as the probability that the eyes landed on the sentence increased compared to the incongruent gaze cue. Moreover, the degree of gaze-induced attentional modulations in the eye movement data positively correlated with the degree of attentional modulations in long-term memory performance. Taken together, these results suggest that gaze could induce overt attentional shifts toward relevant information in a complex behavioral task that requires learning and memory. Moreover, the attentional enhancement of memory performance varies across individuals depending on the degree at which social cues influenced attentional and oculomotor systems.

Keywords: Gaze, Social communication, Social attention, Eye-tracking, Video-based learning

Introduction

Due to the dramatic growth of internet users in past decades, video-based learning such as massive open online course (MOOC) has become increasingly popular and is thought to be an alternative way of effective learning. Accordingly, this learning style has attracted millions of students globally. However, a number of students who take online courses or those who choose video-based learning may not succeed. For example, they may not complete the course or they may not remember what they learn as well as students who learn directly from teachers in the classroom. Due to a lack of face-to-face interaction between students and teachers, it is generally thought that the engagement level during video-based learning may be lower than during in-class learning and thus the former may not be as effective as the latter. According to this idea, recent studies have examined whether the presence of the instructor in video-based lectures could increase the engagement level of

students and could improve their long-term memory performance about learning contents in the video-based lectures (Kizilcec, et al., 2014; Cui, et al., 2012; Mayer 2005). Surprisingly, they found that memory performance was comparable between video-based materials with and without the presence of the instructor (Kizilcec, et al., 2014; Li et al., 2016). This null instructor effect on memory suggests that even though the overall engagement level of students may increase by having the instructor in the video, the presence of the instructor may compete for attentional resources and thus worsen memory encoding processes (i.e., the instructor may distract students, thus reducing their abilities in paying attention and memorizing relevant learning materials). Consistent with this idea, Kizilcec et al. (2014) found that eye movement patterns, which are commonly used to index overt spatial attention (Kizilcec, et al., 2014), were changed by the presence of the instructor. Specifically, they found that the probability at which the eyes fixated within the text area (i.e., relevant information) was decreased by the presence of the instructor and this is due to the fact that students tended to look at the face of the instructor substantially more frequently (Kizilcec et al., 2014).

In the past few years, research has focused on finding key factors that could enhance the instructor effect on information encoding during video-based learning. These include using virtual agents (e.g., social robots and animated cartoon characters) instead of live-video recordings of human instructors (Carter, et al., 2013; Li et al., 2016). These studies have found that using these virtual agents could be less distracting than using human instructors. For example, Carter, et al. (2013) found that fixation time and fixation count within the instructor's face area decreased as the instructor's face became more artificial. However, it is less clear whether the reduction in distractibility of these virtual agents could have a significant impact on the encoding of task-relevant information (Carter, et al., 2013; Li et al., 2016).

In the fields of cognitive psychology and social neuroscience, it is well known that gaze is an important social cue that can enhance joint attention between senders and receivers during social communication (Nummenmaa and Calder., 2016; Birmingham and Kingstone., 2009). Several past studies have demonstrated that gaze could automatically induce covert and overt shifts of attention toward locations in the visual scene, which match gaze direction (Nummenmaa and Calder., 2016; Birmingham and Kingstone., 2009; Caruana et al., 2014; Shepherd et al., 2009; Risko et al., 2012). These gaze-induced shifts in visuospatial attention have been shown to increase the efficiency of sensory information encoding at the behavioral level (i.e., increase accuracy and reduce reaction time during simple visual detection/discrimination tasks) (Nummenmaa and Calder., 2016; Birmingham and Kingstone., 2009; Caruana et al., 2014; Shepherd et al., 2009; Risko et al., 2012). Moreover, gaze stimuli have been found to activate brain regions including intraparietal sulcus and frontal eye field, commonly known to support visuospatial attention and oculomotor functions (Nummenmaa and Calder., 2016; Birmingham and Kingstone., 2009; Caruana et al., 2014; Shepherd et al., 2009; Risko et al., 2012). Taken together, these findings suggest that gaze is a promising element that could potentially enhance information encoding during learning. However, research has not yet investigated whether gaze could be used to increase the quality of video-based learning materials. Therefore, in the present study, we designed different types of videos where gaze directions of animated cartoon characters (i.e., instructors) were manipulated (Figure 1). These included trials (i) without the instructor (text-only), (ii) with the instructor gazing towards the sentence (compatible gaze), (iii) with the instructor gazing directly to the viewer (neutral gaze), and (iv) with the instructor gazing away from the sentence (incompatible gaze). We recorded eye movement patterns of human participants while they were freely watching these videos. We hypothesized that the compatible gaze condition should induce an automatic shift of eye movement patterns towards relevant information (i.e., sentences) compared to the neutral

and incompatible gaze conditions. Importantly, this gaze effect should be automatic and happen without directly instructing participants about the purpose of varying gaze directions.

1. Methods and Materials

1.1 Participants

We recruited 100 volunteers who has normal or corrected-to-normal vision and audition from the community at King Mongkut's University of Technology Thonburi (KMUTT) (age = 18-35, 21 male, 1 left-handed). All participants provided a written consent before participating in the experiment and were compensated. All experimental procedures were approved by the Institutional Review Board at KMUTT.

1.2 Stimuli and Tasks

Participants were seated 72 cm in front of the LCD computer monitor (24inches, 1280x1024 resolution, a 60Hz refreshing rate) in a quiet and dimly lit room. To reduce head movements, we instructed participants to rest their chins on the chin rest. The eye-tracking experiment was performed on a laptop running WindowXP. The stimuli were presented via the Experimental Design software and the eye-tracking data were recorded via an infrared video camera (RED-4.2-913-140, sampling frequency = 50Hz) with the iViewX RED software developed by SensoMotoric Instruments (SMI) (Teltow, Germany).
Each trial started with a fixation for 1000ms located at the center of the screen on the horizontal axis and 2.33 degrees visual angle below the center of the screen on the vertical axis. After the fixation period, the sentence appeared at the upper left corner of the screen (the start of the sentence was at 13.9 and 3.37 degrees to the left and to the top of center, respectively; the length of the sentence ranged from 17-21.5 degrees). The sentence was presented in Thai and contained autographical information about different individuals with different names. There were four different trial types including the text-only, compatible gaze, neutral gaze and incompatible gaze conditions. In the text-only condition, the sentence was presented alone and it stayed on the screen for 6,000ms. For the other three gaze conditions, an animated character (either male or female) was simultaneously presented with the sentence and located at the lower right corner of the video (the center of the character's face was located at 8.65 degrees and 2.33 degrees to the right and to the bottom of center, respectively: the size of the character was 6 x 9.6 degrees square). In the neutral gaze condition, the animated character gazed directly to the viewer for the entire trial. In the compatible gaze condition, the animated character gazed directly to the viewer for 500ms and then gazed toward the sentence for 3,500ms and then returned to the neutral direction and stayed on for 2,000ms. The trial structure for the incompatible gaze condition was the same as that in the compatible gaze condition, except that the animated character gazed away from of the sentence. There were 3 experimental blocks. Each block started with an eye-tracking calibration period followed by 16 trials (4 trials for each experimental condition). There were 48 trials with 48 unique sentences for the entire experiment. For each participant, these 48 unique sentences were pseudo-randomly assigned to different experimental conditions. Block and Trial orders were also randomized. Each block lasted 2 min and the entire experiment including breaks lasted 10 minutes.

1.3 Eye-Tracking Analysis

We performed eye-tracking analysis using the BeGaze software provided by SMI and customized MATLAB-R2017a codes. First, we binned the eye-tracking data, including horizontal (HEP) and vertical eye positions (VEP), into four different bins: text-only, compatible gaze, neutral gaze, and incompatible gaze conditions. Next, we averaged the data in each bin for individual participants and then plotted the data averaged across participants. We also calculated within-subject standard error of the mean (SEM) (Loftus and Masson., 1994). To examine differences between all experimental conditions, we used one-way repeated-measures ANOVAs on HEP and VEP values across all time points (300 time points from 0-6s). Multiple comparisons across time points were then corrected using the false discovery rate (FDR) method (Benjamini and Hochberg., 1995). To test the instructor effect on the HEP and VEP values, post-hoc paired t-tests were used to compare differences between the data in the text-only condition and the data averaged across all gaze conditions. To test the cuing effect on the HEP and VEP values, post-hoc paired t-tests were used to compare differences between the compatible and neutral gaze conditions, between the compatible and incompatible gaze conditions, and between the neutral and incompatible gaze conditions. Multiple comparisons across all post-hoc tests and time points were then FDR-corrected (Benjamini and Hochberg., 1995). We used 2-tailed t-tests to be conservative.

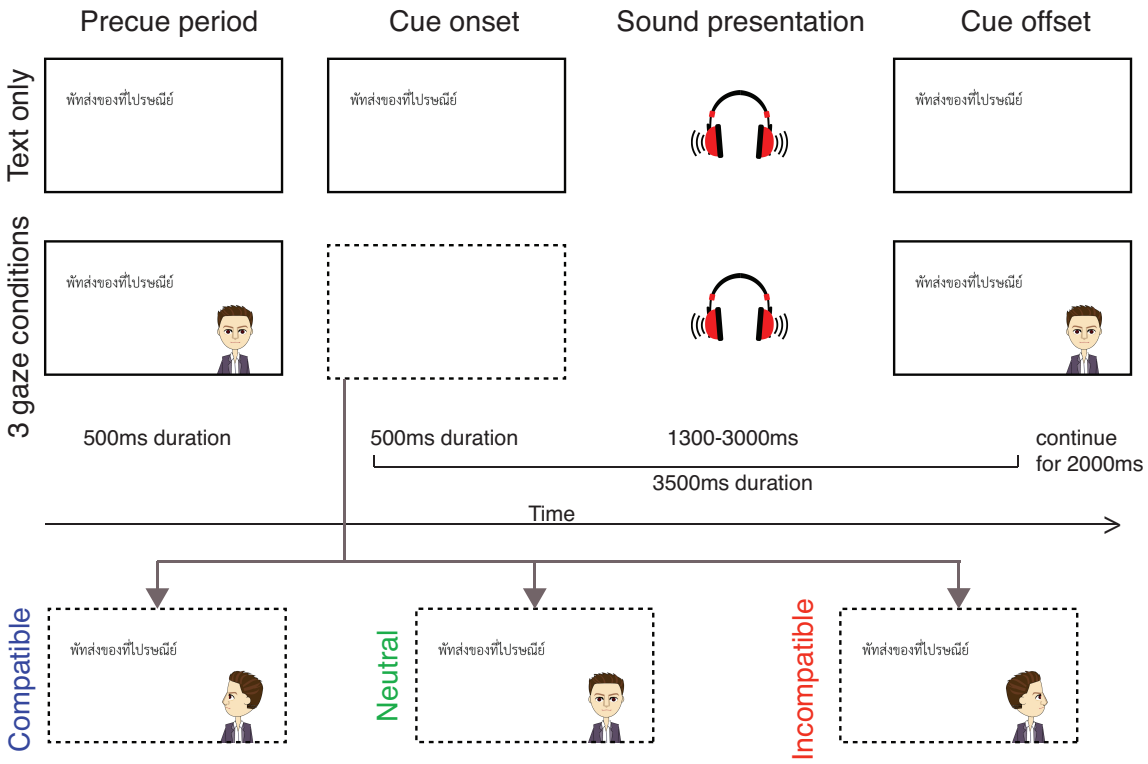

Figure 1. Task design. Each trial, participants freely viewed a sentence with or without an animated character (i.e., text-only). During trials where an animated character was presented, the character could either gaze toward the sentence (i.e., compatible gaze), gaze directly to the participant (i.e., neutral gaze), or gaze away from the sentence (i.e., incompatible gaze). Participants were allowed to move their eyes freely, while their eye movements were recorded by an infrared video camera.

2. Results

2.1 Horizontal Eye-Tracking Data

Averaged horizontal eye positions (i.e. HEP) relative to the center of the display were plotted across time as illustrated in Figure 2. In this figure, positive and negative values indicate the distance (in degree visual angle) to the left and to the right of the central position, respectively. From the start of the trial (0ms), HEP was higher than zero (~4 degrees) for all experimental conditions, indicating that participants prioritized the sentence, which located on the upper left side of the screen over the animated character with or without its presence. In the text-only condition, HEP increased until it reached ~9 degrees (near the start of the sentence) at ~500ms and gradually dropped until it reached the baseline level (~4 degrees) at ~2,000ms and maintained at this same level until the trial ended. In contrast, HPE in the other three gaze conditions only peaked at ~8 degrees and dropped to about -2 degrees, which was over the right side of the central position where the animated character was presented.

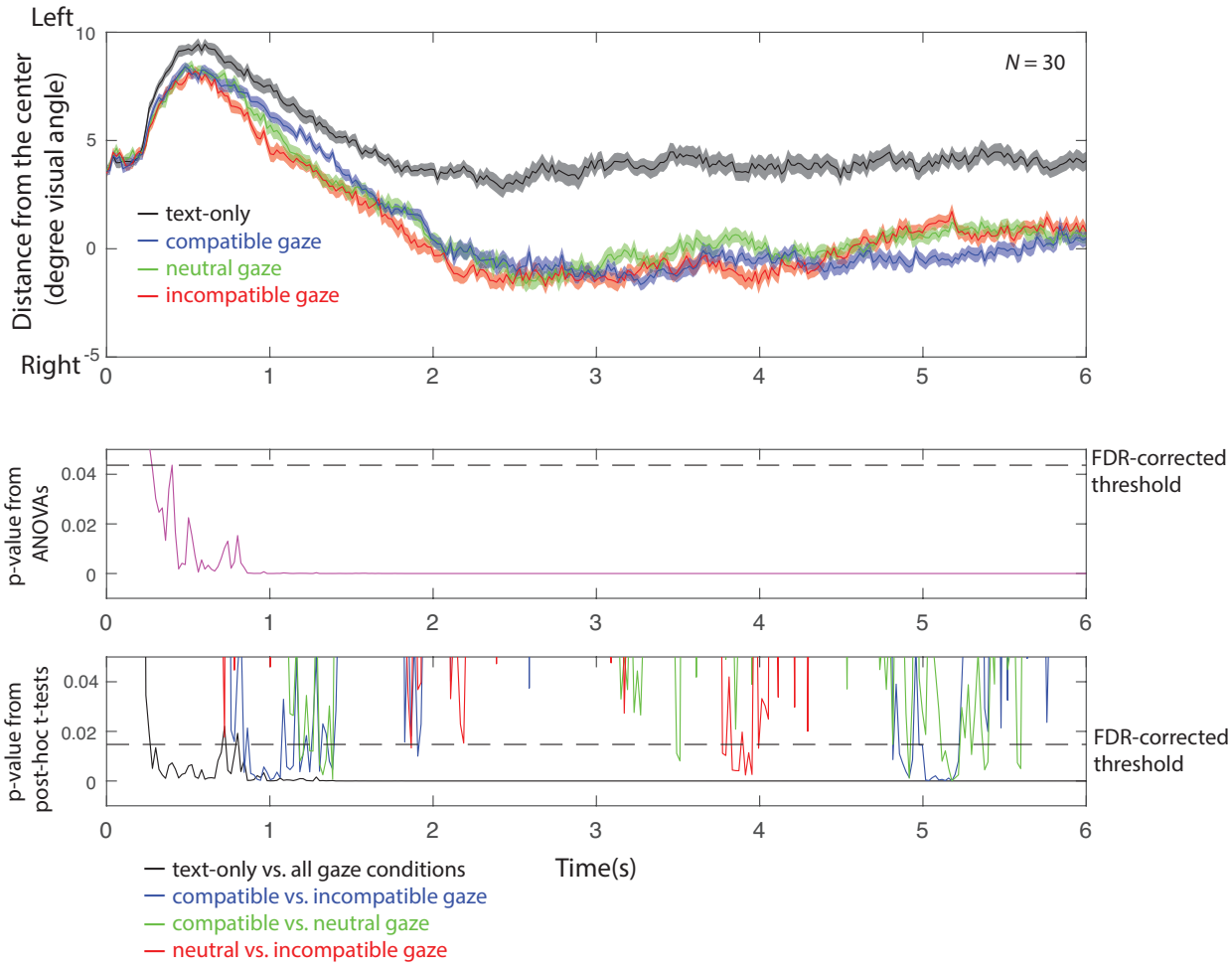

Figure 2. Averaged horizontal eye positions (i.e. HEP). Positive and negative values indicate the distance (in degree visual angle) to the left and to the right of the central position, respectively. The middle and bottom panels show p-vales from one-way repeated-measures ANOVAs and post-hoc paired t-tests, respectively. Dashed lines indicate significant thresholds corrected by the FDR method.

To officially examine HEP differences across experimental conditions, we used one-way repeated-measures ANOVAs to compare the data across all time points from 0ms to 6,000ms after trial onset. We found that HEP differed significantly across all experimental conditions from ~300ms to 6,000ms ($p \leq 0.043$, FDR-corrected across all time points). Post-hoc paired t-tests showed that this result was primarily driven by the fact that HEP in the text-only condition was significantly higher (i.e. more leftward) than HEP averaged across the other three gaze conditions ($p \leq 0.0148$, FDR-corrected). The difference in eye movement patterns between the text-only and the other gaze conditions was consistent with reports from previous studies finding that the instructor face could distract students from paying attention to relevant learning materials (e.g., texts and sentences on lecture slides) (Kizilcec, et al., 2014).

To test whether the compatible gaze could better facilitate the encoding of relevant information compared to the neutral and incompatible gaze conditions, we used post-hoc paired t-tests to compare HEP differences between the compatible and neutral gaze conditions, between the compatible and incompatible gaze conditions, and between the neutral and incompatible gaze conditions. As hypothesize, we found that HEP in the compatible gaze condition was significantly higher (more leftward) than HEP in the incompatible gaze condition from ~800ms to ~1,400ms, and than HEP in the neutral gaze condition from ~1,100ms to ~1,400ms ($p \leq 0.0148$, FDR-corrected). The result suggests that the compatible gaze automatically induced an overt shift of attention to the sentence, and this automatic attentional shift occurred rapidly (~300ms after cue onset) without instructing participants about the purpose of having different gaze directions in this video-based learning experiment. Additionally, at a much later time window, we found the reversed pattern of gaze direction modulations on HEP. Specifically, the HEP in the compatible gaze condition was significantly lower (more rightward) than HEP in the incompatible gaze condition from ~4,800ms to ~5,200ms ($p \leq 0.0148$, FDR-corrected). This was consistent with the general idea that attention could be redirected to the target location even when the cue and the target locations mismatch but this attentional reallocation process may happen too late for efficient information encoding.

2.2 Vertical Eye-Tracking Data

Averaged vertical eye positions (i.e. VEP) relative to the center of the display were plotted across time as illustrated in Figure 3. In this figure, positive and negative values indicate the distance in degree visual angle to the top and to the bottom of the central position, respectively. From the start of the trial (0ms), VEP was higher than zero (~2 degrees) for all experimental conditions, indicating that participants prioritized the sentence over the animated character with or without its presence. In the text-only condition, VEP slightly increased until ~500ms then stayed in the range between ~3-4 degrees and maintained at this same level until the trial ended. In contrast, VPE in the other three gaze conditions increased to the peak at ~3 degrees and slightly dropped to ~1 degree and stayed between 0-2 degree until the trial ended. One-way repeated-measures ANOVAs revealed that VEP differed significantly across all experimental conditions from ~1,000ms to 6,000ms ($p \leq 0.0402$, FDR-corrected across all time points). Post-hoc paired t-tests showed that this result was driven by the fact that VEP in the text-only condition was significantly higher (i.e. more upward) than VEP averaged across the three gaze conditions ($p \leq 0.0118$, FDR-corrected). Unlike the HEP data, there was no significant difference between different gaze directions at any time point before 3,000ms (n.s. with $p > 0.0118$, FDR-corrected). This could be due to the fact that the distance between the sentence and the animated character on the vertical axis was much shorter than that on the horizontal axis, thus reducing the gaze direction effect on the VEP value. However, similar to the HEP data, we observed the

reversed pattern of gaze direction modulations on VEP from ~4,800ms to ~5,200ms ($p \leq 0.0118$, FDR-corrected), consistent with the idea that attention could be redirected to the
target location even when the cue and the target locations mismatch but at later time points.

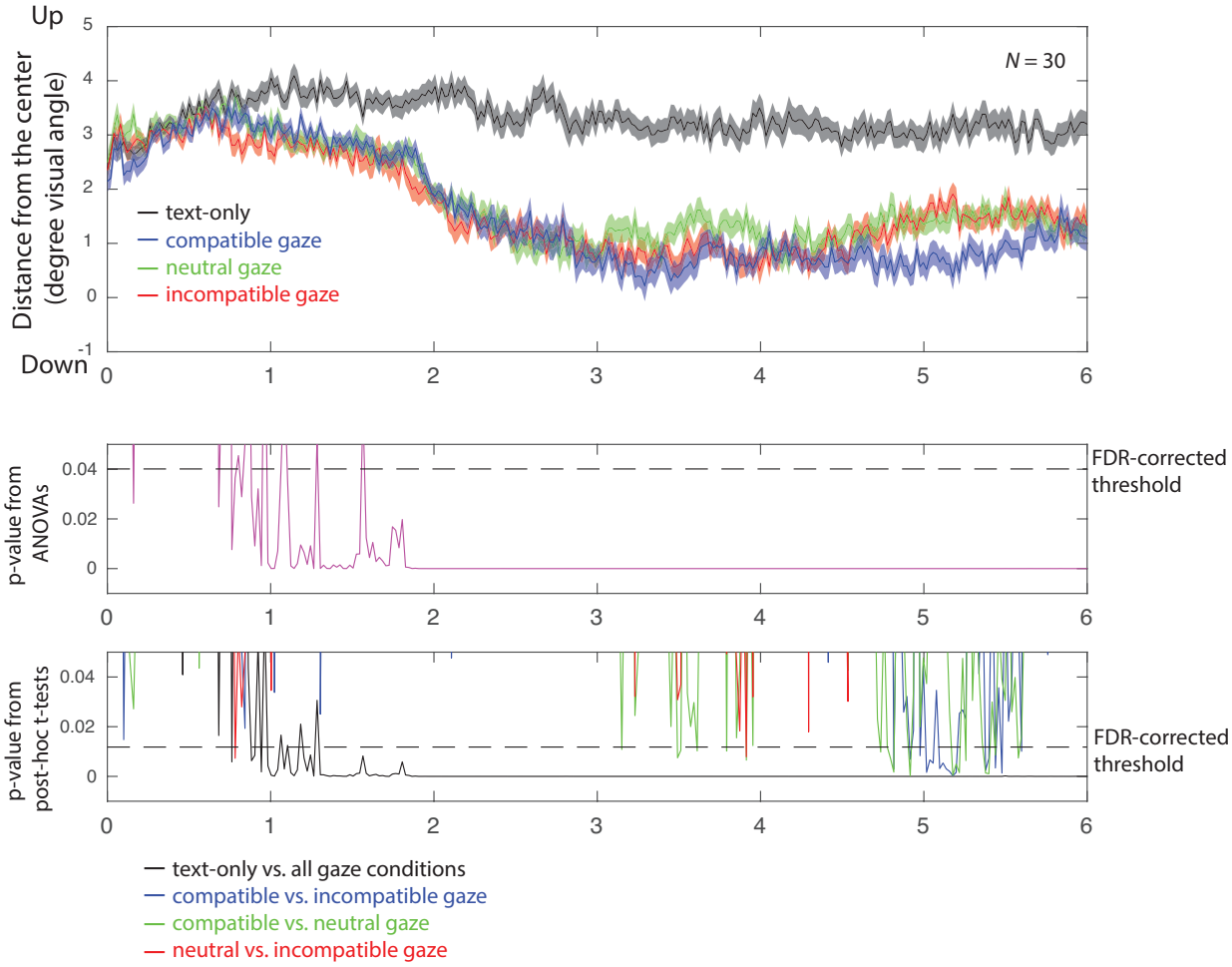

Figure 3. Averaged vertical eye positions (i.e. VEP). Positive and negative values indicate the distance (in degree visual angle) to the top and to the bottom of the central position, respectively. The middle and bottom panels show p-vales from one-way repeated-measures ANOVAs and post-hoc paired t-tests, respectively. Dashed lines indicate significant thresholds corrected by the FDR method.

3. Discussion

Recent studies have tried to improve the quality of video-based learning materials by adding the video of the instructor to increase the overall engagement level of students. However, previous studies have demonstrated that this method might not be efficient because the face of the instructor may distract students from encoding task-relevant information (e.g., reading sentences) (Kizilcec, et al., 2014; Kizilcec, et al., 2015; Cui, et al., 2012; Mayer 2005). In the present study, we examined whether gaze, which is thought to be a strong attentional cue, could help reduce this distraction by enhancing the encoding of task-relevant information during video-based learning. Consistent with the previous report, we found that when the instructor was present, the eye movement pattern shifted away from the sentence area (Kizilcec, et al., 2014). However, as predicted, the compatible gaze

induced a significant shift of the eye movement pattern toward the sentence area more than both neutral and incompatible gaze conditions. Taken together, our results suggest that gaze could alter the distribution of overt attention during video-based learning providing a new way to help improve the quality of video-based learning materials. In this study, we found significant gaze-induced attentional modulations in the eye-tracking data even if participants did not know about the purpose of different gaze conditions. This suggests that the gaze effect on attentional functions was automatic and robust. However, at this stage of the experiment, we had not investigated whether this gaze-induced attention effect would have a significant benefit on subsequent memory performance or whether the degree of attentional modulations in the eye-tracking data could predict the degree of attentional modulations in the memory performance across individuals. To examine this, in the future, memory testing has to be done after participants finishing our video-based learning task. It is highly likely that we would find a significant effect of gaze-induced attention in memory performance because a previous study has found that spatial attention could significantly enhance long-term memory encoding even when participants were not instructed about the subsequent memory test (i.e., incidental memory encoding) (Uncapher et al., 2011). Importantly, this cognitive phenomenon is supported by the increase in hippocampal activity which is regulated by the enhanced activity of the fronto-parietal attentional network in the brain (Uncapher et al., 2011). Lastly, in this study, we used animated cartoon characters instead of video recordings of human lecturers because we wanted to reduce the distractibility of the instructor's face following a previous study (Carter, et al., 2013). However, using artificial stimuli could possibly reduce the level of social engagement of students thus it may decrease the gaze effect on attention. Future studies are needed to determine the naturalness level of the facial stimulus that would produce the highest degree of attentional benefit and the lowest degree of distraction from the facial stimulus itself.

Acknowledgements

We thank Mr.Raksa Nawamalasakul and Miss.Sutatip Buatip for helping create auditory stimuli and Mr.Kodchahem Kamolwit for technical supports with an eye-tracking equipment. This study is funded by Learning Institute and School of Architecture and Design, King Mongkut's University of Technology Thonburi.

References

[1] Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, 289-300.

[2] Birmingham, E., & Kingstone, A. (2009). Human social attention. *Annals of the New York Academy of Sciences*, *1156*(1), 118-140.

[3] Borup, J., West, R. E., & Graham, C. R. (2012). Improving online social presence through asynchronous video. *The Internet and Higher Education*, *15*(3), 195-203.

[4] Carter, E. J., Mahler, M., & Hodgins, J. K. (2013, August). Unpleasantness of animated characters corresponds to increased viewer attention to faces. In *Proceedings of the ACM Symposium on Applied Perception* (pp. 35-40). ACM.

[5] Caruana, F., Cantalupo, G., Russo, G. L., Mai, R., Sartori, I., & Avanzini, P. (2014). Human cortical activity evoked by gaze shift observation: an intracranial EEG study. *Human brain mapping*, *35*(4), 1515-1528.

[6] Cui, G., Lockee, B., & Meng, C. (2013). Building modern online social presence: A review of social presence theory and its instructional design implications for future trends. *Education and information technologies*, *18*(4), 661-685.

[7] Kizilcec, R. F., Papadopoulos, K., & Sritanyaratana, L. (2014, April). Showing face in video instruction: effects on information retention, visual attention, and affect. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 2095-2102). ACM

[8] Kizilcec, R. F., Bailenson, J. N., & Gomez, C. J. (2015). The instructor's face in video instruction: Evidence from two large-scale field studies. *Journal of Educational Psychology*, *107*(3), 724.

[9] Li, J., Kizilcec, R., Bailenson, J., & Ju, W. (2016). Social robots and virtual agents as lecturers for video instruction. *Computers in Human Behavior*, *55*, 1222-1230.

[10] Mayer, R. E. (Ed.). (2005). *The Cambridge handbook of multimedia learning*. Cambridge university press.

[11] Loftus, G. R., & Masson, M. E. (1994). Using confidence intervals in within-subject designs. *Psychonomic bulletin & review*, *1*(4), 476-490.

[12] Nummenmaa, L., & Calder, A. J. (2009). Neural mechanisms of social attention. *Trends in cognitive sciences*, *13*(3), 135-143.

[13] Risko, E. F., Laidlaw, K. E., Freeth, M., Foulsham, T., & Kingstone, A. (2012). Social attention with real versus reel stimuli: toward an empirical approach to concerns about ecological validity. *Frontiers in human neuroscience*, *6*.

[14] Shepherd, S. V., Klein, J. T., Deaner, R. O., & Platt, M. L. (2009). Mirroring of attention by neurons in macaque parietal cortex. *Proceedings of the National Academy of Sciences*, *106*(23), 9489-9494.

[15] Uncapher, M. R., Hutchinson, J. B., & Wagner, A. D. (2011). Dissociable effects of top-down and bottom-up attention during episodic encoding. *Journal of Neuroscience*, *31*(35), 12613-12628.