Identifying Language-induced Mental Load from Eye Behaviors in Virtual Reality

Johannes Schirm[1] , Monica Perusquía-Hernández[1] ,
Naoya Isoyama[1] , Hideaki Uchiyama[1] and Kiyoshi Kiyokawa[1]

[1]Nara Institute of Science and Technology

Abstract

We compared content-independent eye tracking metrics under different levels of language-induced mental load in virtual reality (VR). We designed a virtual environment to balance consistent recording of eye data with user experience and freedom of action. We also took steps towards quantifying the phenomenon of not focusing exactly on surfaces by proposing "focus offset" as a VR-compatible eye metric. Responses to conditions with higher mental load included larger and more variable pupil sizes and less fixations. We also observed less voluntary gazing at distraction content, and a tendency of looking through surfaces.

CCS Concepts
• Human-centered computing → Virtual reality; HCI theory, concepts and models;

1. Introduction

The relation between cognitive states and eye behaviors has been identified in many studies [Bea82]. In this work, we analyze quantitative eye data to identify mental load in users of virtual reality (VR) during a language task. We considered five of the most commonly used eye metrics [WCP*21]. **Pupil size** is known to increase with mental load [Bea82]. **Blinks** were found to play a role in blocking out visual input to reduce distraction while thinking. **Fixations** are the times at which gaze velocity and acceleration stay below a predefined threshold, while **saccades** happen when the gaze wanders around and breaks this threshold. **Eye vergence** refers to the angle between the right and left gaze directions. It tends to fall back into a resting pose during internal thought [HLN*19].

In VR, eye behaviors are commonly analyzed at the level of virtual objects [CKK19]. This can be done by adding colliders to relevant scene objects and detecting when the user's gaze hits their surface. Only few works explore the above metrics in a VR context, e.g., fixation detection from 3D gaze data. More general, content-independent metrics [WCP*21] might improve the detection of cognitive states from eye behaviors in VR. However, it is not self-evident whether screen-based approaches also apply to VR, where the conditions for visual perception are vastly different from looking at a screen in the real world [HGAB08].

We induced mental load in users passively viewing a VR scene only by aural cues, and then statistically confirmed the generated mental load from their eye behaviors. We hypothesized that users listening to speech samples they rated as more difficult to understand would experience high mental load. High mental load would

Figure 1: *The VR scene with the phone booth to walk in.*

lead to increased pupil size, longer and more frequent blinks, longer and less frequent fixations, and more intense and less frequent saccades. We also expected users with high mental load to more frequently have their gaze fall back into a resting pose, as if looking through/behind the virtual 3D surface they are facing. To measure this effect in VR, we propose a new metric "focus offset," which we define as the mismatch between **the virtual 3D surface the user is gazing at** and **the actual intersection of the left and right gaze rays**. Finally, we expected high mental load to inhibit users from voluntarily looking around.

We chose to elicit cognitive load using a listening task primarily because of two relevant advantages. First, the necessary listening skills might already be acquired [Kra81], so mental load decreases the more fluent a person gets. This further moves the extremes of our mental load scale apart from each other and makes them easier to distinguish. Second, since there is no visual element required for listening, eye tracking data can be assumed to represent the common use case of natural, passive viewing of a VR environment. We expected this to reduce biases towards task-related behavior.

2. Environment

Figure 1 shows the virtual phone booth used as the environment for listening to the audio stimuli. Participants were only instructed to understand the contents of the speech samples and could look around freely if they desired. The green glowing sphere in a glass bulb on top of the phone was pulsing slightly as the amplitude of the currently played audio in- or decreased. Opposing this default fixation point, we added distraction posters left and right to the phone, allowing the detection of voluntary exploration. The VR experience lasted on average 4.15 minutes ($SD = 0.67$).

3. Experiment

14 males and 1 female university students with a mean age of 22.87 years ($SD = 0.92$) participated in an experiment. Following the declaration of Helsinki, they gave informed consent after an explanation of the experiment. They were all Japanese native speakers, and with English being a regular part of Japanese education, all participants were expected to be learners of English. Four natural samples of speech were chosen to increase difficulty step by step: *JP-Familiar* was familiar speech by a close Japanese person, *JP-Unfamiliar* was a literary and more dense sample from a Japanese science fiction movie, *EN-Familiar* was an intentionally easy sample for learners of English, and *EN-Unfamiliar* was a high-level lecture in English. Each sample lasted about 40 seconds. After the VR experience, participants answered the 5-point Likert item "How well do you feel that you understood speaker number *x*?" for each speech sample on a 5-point Likert scale. As our main dependent variable, we recorded eye data using the inbuilt eye tracker of the HTC VIVE Pro Eye headset. This included head transforms, gaze collision points on scene surfaces, the closest point between the left and right gaze rays, and names of the gazed-at virtual objects.

4. Results

After performing blink detection using the R package GazeR and linearly interpolating eye data during blinks, pupil size values were normalized using the data between 500 ms and 2000 ms of each audio sample trial as a baseline. Due to strong noise, the focus offset feature was smoothed using a Savitzky-Golay filter (window size 201, order 0). In the absence of an easy-to-use toolkit for 3D gaze analysis, the 3D eye gaze data was mapped onto a virtual screen in participants' head space and fixations were detected using the R package gazepath.

Participants rated their perceived comprehension to be 4.91 ($SD = 0.30$) for *JP-Familiar*, 3.09 ($SD = 1.51$) for *JP-Unfamiliar*, 3.55 ($SD = 0.93$) for *EN-Familiar*, and 1.64 ($SD = 0.67$) for *EN-Unfamiliar*. This means that *JP-Familiar* could be followed by all participants, while most of them had problems following *EN-Unfamiliar*. The randomization of audio samples was afterwards found to be biased so that easier samples are presented first. The data of each of the 40-second long listening phases was summarized in chunks of 10 s each by taking mean, SD, and event count.

The most significant differences occurred in pupil size means, which were largest for *EN-Familiar* compared to *JP-Familiar* ($p < .05$, $d = 0.38$, 0.072 mm larger), *JP-Unfamiliar* ($p < .01$, $d = 0.66$, 0.095 mm larger) and *EN-Unfamiliar* ($p < .05$, $d = 0.6$, 0.103 mm larger), with d representing Cohen's effect size of each test. This indicates that the highest mental load was induced during *EN-Familiar*—we think that many participants might have given up on following *EN-Unfamiliar*. Focus offset did not follow the same pattern. Participants focused more behind surfaces during *EN-Familiar* ($p < .01$, $d = 0.88$, 7.0% higher) and *EN-Unfamiliar* ($p < .05$, $d = 0.72$, 6.5% higher) compared to *JP-Unfamiliar*. In addition to known stability issues of eye vergence, focus offset might have suffered from the aforementioned randomization issue. Fixation frequency dropped during *EN-Familiar* compared to *JP-Unfamiliar* ($p < .05$, $d = 0.53$). Blinks and saccades did not significantly differ between audio samples—this might have been due to the comparatively low sample rate. Finally, participants who looked at the distraction posters near the phone had a tendency of doing so during the easier audio samples.

5. Discussion and future work

We used general techniques for detection of mental load from eye data in VR, and think their application to VR should be further explored. Whenever feasible, experimenters can take advantage of colliders outside the main area of interest to identify gaze behavior in VR indicating distraction. As the focus offset metric did not follow the pattern of pupil size, it might not be a direct indicator of mental load. However, with increasingly accurate eye tracking hardware, it might feature useful connections, e.g., to attention and alertness. These two states (just as an example) have in common that they can be controlled more intentionally by users compared to mental load. In the current study, we learned how many prerequisites there are for true mental load to occur—often depending on the specific individual tested. We recommend that future studies focusing on mental load and its quantitative evaluation should first verify responses to the specific stimuli in use separately.

References

[Bea82] BEATTY J.: Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin 91*, 2 (1982), 276–292. doi:10.1037/0033-2909.91.2.276. 1

[CKK19] CLAY V., KÖNIG P., KÖNIG S. U.: Eye tracking in virtual reality. *Journal of Eye Movement Research 12*, 1 (Apr. 2019). doi:10.16910/jemr.12.1.3. 1

[HGAB08] HOFFMAN D. M., GIRSHICK A. R., AKELEY K., BANKS M. S.: Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision 8*, 3 (03 2008), 33–33. doi:10.1167/8.3.33. 1

[HLN*19] HUANG M. X., LI J., NGAI G., LEONG H. V., BULLING A.: Moment-to-moment detection of internal thought during video viewing from eye vergence behavior. In *Proceedings of the 27th ACM International Conference on Multimedia* (New York, NY, USA, 2019), MM '19, Association for Computing Machinery, p. 2254–2262. doi:10.1145/3343031.3350573. 1

[Kra81] KRASHEN S. D.: *Second language acquisition and Second language learning*. Pergamon Press Inc., 1981. 1

[WCP*21] WALCHER S., CEH S. M., PUTZE F., KAMPEN M., KÖRNER C., BENEDEK M.: How reliably do eye parameters indicate internal versus external attentional focus? *Cognitive Science 45*, 4 (2021), e12977. doi:https://doi.org/10.1111/cogs.12977. 1