#### **ORIGINAL ARTICLE**



# Assessing hands-free interactions for VR using eye gaze and electromyography

Yun Suen Pai<sup>1</sup> · Tilman Dingler<sup>2</sup> · Kai Kunze<sup>1</sup>

Received: 16 February 2018 / Accepted: 30 October 2018 / Published online: 20 November 2018 © Springer-Verlag London Ltd., part of Springer Nature 2018

#### Abstract

With the increasing popularity of virtual reality (VR) technologies, more efforts have been going into developing new input methods. While physical controllers are widely used, more novel techniques, such as eye tracking, are now commercially available. In our work, we investigate the use of physiological signals as input to enhance VR experiences. We present a system using gaze tracking and electromyography on a user's forearm to make selection tasks in virtual spaces more efficient. In a study with 16 participants, we compared five different input techniques using a Fitts' law task: Using gaze tracking for cursor movement in combination with forearm contractions for making selections was superior to using an HTC Vive controller, Xbox gamepad, dwelling time, and eye-gaze dwelling time. To explore application scenarios and collect qualitative feedback, we further developed and evaluated a game with our input technique. Our findings inform the design of applications that use eye-gaze tracking and forearm muscle movements for effective user input in VR.

Keywords Virtual reality · Physiological sensing · Eye gaze · Electromyography

# 1 Introduction

VR as a platform is increasingly immersive as they began with tracking the user's head direction and moved toward tracking their hands' position, to allow for various forms of gesture and button-based inputs. This method has become the conventional input system, yet also the current limitation for VR; positionally tracked physical controllers often come with a directional pad for analog input and various buttons for each finger for digital input. Though reliable, we wish to delve deeper into the next step of input modalities for VR. We propose the use of physiological signals as an additional input method to enhance the overall VR experience. We chose to leverage the use of eye-gaze tracking and muscle activity sensing as a form of point and select for several

**Electronic supplementary material** The online version of this article (https://doi.org/10.1007/s10055-018-0371-2) contains supplementary material, which is available to authorized users.

⊠ Yun Suen Pai yspai1412@gmail.com

<sup>2</sup> University of Melbourne, Melbourne, Australia

reasons: Firstly, we wish to investigate hands-free methods of input to complement conventional hand controllers that currently exist today. Secondly, both of these signals allow reliable and explicit input compared to other forms of signals like electroencephalogram (EEG) and electrocardiogram (ECG) which measure brain and heart activity, respectively.

To validate our proposed method, we conducted a Fitts' law experiment to evaluate its accuracy and speed. We compared it to five other conventional input methods; mouse, gaze direction with dwell time, eye-gaze direction with dwell time, gamepad input and using motion-tracked controllers. Motion-tracked controller input is treated as the gold standard baseline since it is the most wide-spread input method for VR now, whereas dwell time uses a timer of 1 s for activation. Figure 1 illustrates a user using our proposed method as input for this study.

We then conducted a second study, which applied our method as an input mechanic for participants to evaluate their gaming experience. The game we created allows pointing a gun using the eyes (selection) and firing it using arm muscle contraction (activation) to shoot several targets in a given time. Based on the participants' qualitative feedback, we found that eye-gaze-based selection was deemed to be overall fast and intuitive for targeting purposes, whereas arm muscle contraction simulates the sensation of a gun recoil,

Keio University Graduate School of Media Design, Yokohama, Japan



Fig. 1 Eye tracking with EMG technique

thus creating an additional sense of realism. We, therefore, showed our interaction method's potential to enhance the overall VR experience. Finally, we discuss several applications for our system in more generic tasks like pick-andplace, teleportation, user interface (UI) interaction, and gaming.

The contributions of this work are as follows:

- 1. We investigated additional input modalities using a combination of eye gaze and EMG activation.
- 2. We built a system to compare its effectiveness and accuracy.
- We validated our approach in a user study by comparing it to conventional input methods and implemented a game using our method based on physiological sensing.

## 2 Related work

Our approach is based on related research in VR interaction techniques, physiological sensing, and select-and-point task assessments. For VR interaction mechanics, we look into non-conventional input and sensing mechanics for VR like gesture-based detection and peripheral-based input devices. Physiological signals, however, have been gaining popularity as several related works in VR show. We, therefore, focus on point-and-select tasks in VR and how they are evaluated in terms of accuracy and performance.

#### 2.1 Interaction techniques

Several works of research in VR introduced novel methods for manipulation. For example, a spherical manipulation device was made using a tilt sensor, electric compass, and flex sensors that allows gesture recognition. This implementation combines a peripheral with gesture recognition, though the study was conducted only for spherical interfaces (Lee et al. 2005). This work is similar to the AcceleGlove, which presents a whole hand input device using accelerometers (Hernandez-Rebollar et al. 2002). The device was developed to recognize sign languages and can also be used as a substitute for a computer mouse for selection and activation. Another work combined gaze with gesture using a Microsoft Kinect with Mirametrix tracker for desktop use (Slambekova et al. 2012). Though no user study was conducted, it served as a proof of concept for a multimodalbased input that may increase the effectiveness of interactions in virtual space.

Other peripherals that were developed catered more toward a specific use of input, such as locomotion. For example, the omnidirectional treadmill was developed for training the army in virtual environments (Darken et al. 1997). In this regard, it was extremely effective, and the cost and overall size of the device, however, prevented it from entering the mainstream market. Another work used a Wii balance board as a low-cost input device (Williams et al. 2011). Besides locomotion, it could also be used for 3D object manipulation and application-specific task due to its discrete and continuous signals.

#### 2.2 Physiological sensing in VR

The usage of EMG may not be mainstream yet, but its feasibility as a daily interaction device is still highly promising. EMG is commonly applied in various medical applications such as muscular rehabilitation, muscular disease, and prosthetics control (Barry et al. 1990; Hernandez Arieta et al. 2006; Jacobsen and Jerard 1974; Kiguchi et al. 2004; Moseley et al. 1992). During a feasibility analysis, it was found that an accuracy of up to 95% was achievable, implying that EMG interaction can be highly promising, especially given its discrete nature (Moseley et al. 1992). EMG has the potential for gesture-based recognition such as fingers, hand, and arm motions. In terms of accuracy, however, EMG performs best simply with direct muscle contraction. Activities recognition like carrying a heavy bag or mug has a lower error percentage compared to recognition of pinching gestures (Saponas et al. 2009). Furthermore, precise calibration is required for accurate gesture recognition due to the data being extremely user dependent.

Previous work used gesture-less EMG for various inputs by determining the length of muscle contraction time; however, relying solely on EMG greatly limits any form of selection (Costanza 2005). Using the length of contraction also suffers from the same issue with dwell time. One of the solutions for this is to pair EMG with other forms of input, thus creating a multimodal system. Multimodal inputs are no stranger to human–computer interface (Zander et al. 2010) and can certainly improve an interaction mechanic if done correctly. For example, coupling EMG with a touch screen allows the system to recognize which finger is touching the surface, as well as the amount of pressure exerted through muscle sensing without the use of a pressure sensor (Benko et al. 2009).

Besides EMG, other forms of multimodal input like combining eye gaze with hand gestures have also been explored (Chatterjee et al. 2015). Regarding eye tracking, it is a sensing mechanic that is very likely to be adopted into future VR headsets. In research, eye tracking has been used for foveated rendering (Pai et al. 2016), multi-user scenarios (Saraiji et al. 2016), or other forms of input modality. Coupling eye-gaze tracking with EMG is a novel proposal and useful input, seeing as how it has previously been used for motor disabilities (Chin et al. 2008). This proves that such a multimodal input can be beneficial given the right context.

#### 2.3 Select-and-point task assessment

Evaluating the performance of a new input method, whether it is for VR or any platform, requires the correct assessment method. A simpler method would measure the time required to complete a certain task, such as walking from point A to point B using the assigned input method (Cardoso 2016). However, a more thorough evaluation can be seen in text entry studies such as work related to keyboard input (Mac-Kenzie and Soukoreff 2002). In HoVR type where the user developed a system for smartphone keyboard input in VR, they measured the time and accuracy to complete a given phrase or sentence, coupled with a general usability questionnaire (Kim and Kim 2017). One of the more common evaluation methods is the Fitts' law method as a relative measure of performance, which was also used in a study that combined gaze detection with gesture (MacKenzie and Soukoreff 2002; Chatterjee et al. 2015). Essentially, it is a selection and activation task for the user that includes a proposed solution being compared with a baseline. This was also utilized in a study to compare several gamepads (Ramcharitar and Teather 2017).

Based on the related works, it can be seen that there exist many forms of input and interaction techniques utilizing sensors, gestures, and currently available devices. Among these input methods, the use of physiological signals carries the benefit of being intuitive and more accurate depending on the context. To evaluate this, utilizing Fitts' law is appropriate as it is a well-recognized evaluation method, though a user experience study would benefit for additional qualitative feedback.

## 3 Input through physiological sensors

Our proposed system integrates eye tracking into commercially available VR headsets as a form of selection, while coupling it with a muscle sensing device for activation. This opens several forms of use cases for the user, such as aiming and firing, pick and place, navigation, and so on.

#### 3.1 Interaction mechanics

The core interaction method proposed in this multimodal solution is to have the eye gaze used as a selection mechanic while the muscle contraction used as an activation. This can be similarly dubbed as the eye gaze being part of the target acquisition phase and the muscle contraction as the target action phase (Chatterjee et al. 2015). One of the core benefits as opposed to dwelling techniques is that the user is not required to focus on an object for an extended period of time which is fatiguing (Jacob 1990). Therefore, only muscle contraction is used for the target action phase. We will perform a comparison by using eye gaze for the action phase as well in our user study.

Special considerations are required when designing an interface that relies on eye tracking. For other forms of applications that are not VR based, determining the accuracy threshold of the eye tracker is important (Chatterjee et al. 2015). However, UI design should be scaled accordingly, which is also parallel with the design consideration for VR UI (Unity 2015). Generally, if a UI is too small and requires really accurate eye tracking for a precise solution, then it is too small for proper viewing in a VR HMD. Most UIs are directly attached to the player camera; however, this does not apply for spatial UI or 3D objects that are placed in the virtual environment. This is because the user can physically move closer to the object for a better view for VR solutions that use spatial tracking.

For the target action phase, two kinds of interaction can be performed: a discrete action in the form of a single activation method like a lamp switch, whereas continuous action provides a stream of data for as long as a condition is true, such as holding down a keyboard button. Figure 2 depicts these interactions showing EMG data.

In this context, discrete action means a single muscle contraction, while continuous action means continuous muscle contraction. This allows for several interaction mechanics for the user in a VR environment. For instance, a discrete action is great for locomotion mechanic and activation of a menu. In a shooting game, this is akin to firing a semiautomatic pistol. Continuous action, on the other hand, is suitable for dragging objects around like a paintbrush, manipulating a graph bar, or firing a fully automatic rifle in VR. These two distinct methods of interaction create a taxonomy similarly Fig. 2 EMG data visualization against time for no muscle contraction (left), discrete action (middle), and continuous action (right)



proposed by Chatterjee et al. (2015). More generic actions like pick-and-place or drag-and-drop may freely utilize either interaction mechanic.

## 3.2 Hardware and software

All of the peripherals used in this study can be obtained off the shelf. Granted, the accuracy may not be as high as industrial- or medical-grade tool, but consumer electronics are the more accessible option for regular users. For the VR HMD, we use the HTC Vive VR headset. For the eye tracker, we used the IR trackers by Pupil Labs. These are a pair of 120 Hz cameras that can be bought with direct integration with the lenses of the Vive. Figure 3 shows the user's eyes being tracked.

There are several eye-tracking solutions for VR; however, they can be quite expensive, or not available for consumers at this point of writing. Finally, for EMG sensing, we opt to use the Myo Armband, which contains eight medical-grade EMG sensors placed on the user's forearm (Myo 2013). Figure 4 shows the EMG readings of each of the Myo sensors. Myo also comes equipped with an accelerometer and gyroscope for gesture input, but for the purpose of this study, we will be excluding these.

In terms of software, we interfaced with the Unity Engine due to its seamless integration with VR. C# is the primary development language. To obtain the eye-tracking data from the Pupil trackers, we use Open Sound Control (OSC) and ZeroMQ to pipe the data over to Unity, allowing us to both obtain the raw data and toggle Pupil's calibration directly from Unity (Saraiji 2016). For the Myo armband, Myo provided a Unity plug-in that directly pipes raw EMG data for each of the sensors. The machine used for this study is equipped with a Core i7-6700 processor, an Nvidia Geforce GTX 980 graphic card, and 8 GB of memory. It is important that the machine used is at least above or equal to the minimum requirement for VR suggested by Oculus (2015) to avoid any additional delays that may cause motion sickness.



**Fig. 3** Eye tracking from the Pupil software



Fig. 4 Visualization of EMG readings against iterations from the eight sensors on the Myo Armband

# 4 User study

The user study is divided into two parts: one for evaluating the performance and throughput, as well as the overall qualitative feedback in a use case scenario. Both parts of the study will be explained thoroughly in their respective subtopics. The participants first need to perform a brief calibration to map the tracking data into Unity's 3D space, though the gaze data are only in 2D. The calibration is performed by looking at nine points that appear on the FOV sequentially. No calibration for the EMG sensor is necessary.

## 4.1 Participants

A total of 16 participants were recruited for this user study, comprising of seven females and nine males aged between 22 and 38 (mean = 26.19, SD = 4.55), where their feedback and opinions were also collected at the end. All the participants had no prior knowledge of the association of EMG and eye tracking with VR, though some of them have experience with interaction in VR environments. For participants who wear glasses, it needs to be taken off in order for the eye trackers to detect the eyes accurately for calibration. Each



Fig. 5 Screen of the user study, with the red point denoting the target

participant also provided their informed consent, and no identifying information is provided in this study.

## 4.2 Method

We used a Fitts' law study shown in Fig. 5 to evaluate the throughput of our proposed interaction based on a varying index of difficulties (ID) and input methods. This particular form of study has been used to evaluate multimodal input method before and has been shown to be a viable method for the evaluation of pointing techniques (Chatterjee et al. 2015). It is worth noting that Fitts' law has been deemed suitable for gaze input as well (Miniotas 2000: Ware and Mikaelian 1986: Wilson and Cutrell 2005: Zhang and MacKenzie 2007), though none has evaluated the use of EMG in a VR environment. We directly compare our method with gamepad input, motion controllers, dwell time, and dwell time with gaze, each explained thoroughly below. Even though mouse input has been considered for comparison due to most people's comfort in using it, we deemed it unnecessary for VR as conventional VR input leans more toward a gamepad or motion controller.

- Gamepad: The gamepad was the first consumer VR input method before the introduction of motion controllers as most users are accustomed to its layout. Therefore, the gamepad's analog stick is used for selection and a button (A-button on the Xbox gamepad) is used for activation.
- Dwell time: Since one of the primary benefits of our proposed method is that it saves time, it is only fitting to compare it with a dwell time method that is currently being used for most hands-free interaction mechanics. Dwell time places a reticule in the middle of the FOV and requires the user to use head rotation to place the reticule over interactive objects for a period of time, in this case, 750 ms. Dwell time by eye gaze has shown that this value provides the highest throughput for dwell-based interactions (Zhang and MacKenzie 2007). Though this input method does not rely on using eye gaze, a similar dwell time value is used.
- Dwell time with eye gaze: Similar to conventional dwell time implementation mentioned above, this method uses eye gaze for selection instead. At the point of writing, this method has so far not been popular with consumers simply because eye tracking is not yet made mainstream to the average users. After a brief calibration, a similar dwell time of 750 ms is also assigned.
- Eye gaze with EMG: Our proposed method replaces dwell time with muscle contraction activation. The user simply needs to contract the forearm muscle to enable activation at the point of gaze, akin to clicking a mouse button for where the cursor is placed.
- Motion controllers: Motion controllers are essentially gamepads that are tracked in the virtual space. Since most current VR interfaces rely on them now, we used

the conventional "laser pointer" interaction where a raycast is produced from a motion controller to point at selected targets. Activation is achieved by pushing the trigger button on the controller. We deem this input method as the gold standard baseline for VR input and interaction.

Therefore, the independent variables are the five different input types, with six levels of IDs (2.81, 2.94, 3.07, 3.2, 3.33, and 3.46 bits). The dependent variables are the movement time (MT) as well as the effective index of difficulty (IDe) to calculate the final throughput (TP). We built the entire study in the Unity Engine, which is able to read the currently set independent variable through different modes that can be toggled with a keyboard trigger (1, 2, 3, 4, and 5)to change the input, and Q, W, E, R, T, and Y to change the ID). This means that all of the input devices are connected at the same time and can be toggled freely. The dependent variables, on the other hand, are logged into a comma-separated value (CSV) file so that they can be plotted and analyzed. The explanation for each of these variables will be further elaborated in the following subsection. We initiate the user study with a pilot study to determine the parameters for the gamepad, followed by study 1 based on Fitts' law, and finally study 2 to obtain qualitative feedback and rough accuracy estimation. Figure 6 shows the proposed framework of the



Fig. 6 Framework of the proposed user study showing the data from user input to data output

link between the variables and the implementation of the system for the user study in general.

#### 4.2.1 Pilot study

Among these five input methods, only the gamepad requires a preset cursor velocity as other methods depend on the efficiency of the user themselves. The default speed of a virtual object according to the maximum displacement of the analog stick is 120 units/s. We ran an informal pilot study with eight participants who are unrelated to the main study, with a varying amount of experience using a gamepad, to determine the most suitable speed of the cursor. Each participant is simply required to enter the VR environment and move a cursor between two points placed left and right at the opposite end of each other using the gamepad. Six levels of distances, which were equivalent to the amplitude (3, 3.2, 3.2)3.4, 3.6, 3.8, and 4 units), were set, and each participant is given the freedom to adjust the speed of the cursor until they find a comfortable speed balance between the distances. We found that users with little to no experience prefer the cursor to be slower around 96 units/s. because they have trouble positioning the cursor accurately, whereas participants with moderate experience in using a gamepad found that the lower speed is too slow, yet higher speed was very easy to overshoot. However, all participants agreed that cursor speed above 144 units/s was quite easy to overshoot, even at a maximum distance of 4 units. Therefore, we decided to use the default speed of 120 units/s without any speed multiplier.

For the main study, each participant is required to use each input method according to each preset ID, three times per ID. Therefore, each participant will run this study 90 times. To eliminate any ordering effect, the sequence of input method was counterbalanced according to the Latin Square. The sequence of ID was randomized for a more realistic range of difficulty (Ramcharitar and Teather 2017).

#### 4.2.2 Apparatus

The ISO 9241-9 Fitts' study method was used for this experiment (Chatterjee et al. 2015; ISO 1998). According to Fitts' law, the index of difficulty, ID is influenced by the distance from the center point to any of the targets and the width of the said targets.

$$ID = \log_2\left(\frac{A}{W} + 1\right) \tag{1}$$

where A is the amplitude or distance and W is the width of the target, in unity units. The targets in this experiment are modeled as green spheres in a 3D environment. When the experiment begins, a green target will be rendered red and the participant is required to activate the sphere (depending on the input method), followed by the next target which is opposite the previously selected target until each of the targets is activated. The trial ends when the last sphere is the same as the first. We also record the movement time (MT) of the whole trial for each input method. Since it refers to the time the user spends moving a pointing device, we exclude the dwell time for dwell-based interactions as it is assumed that the pointer is static. At the beginning of each trial round, the position of the cursor is reset to the middle of the scene. At the point of selection, we also measure the standard deviation (SD) of overshoots and undershoots from the center of the target spheres. We use the SD values to calculate the effective width, We, shown below:

$$We = 4.133 * SD$$
 (2)

This allows us to compute the effective index of difficulty, IDe, and the final throughput, TP.

$$IDe = \log_2\left(\frac{A}{We} + 1\right) \tag{3}$$

$$TP = \frac{IDe}{MT}$$
(4)

During the entirety of the study, we employed a think-aloud protocol where at any time, the participants are free to express their opinions and provide constructive feedback, which is recorded. Finally, after each participant experiences all the IDs of an input method, we ask them to answer the NASA Task Load questionnaire for each input to understand their perceived load.

#### 4.2.3 Qualitative study

The second study focuses on qualitative feedback from participants where we obtain informal feedback regarding the proposed input method in a simple game (Jones et al. 2013). The game places the participant in a room, equipped with a pistol that aims in the direction the participant looks, as shown in Fig. 7. The participant is simply required to look at targets that spawn in front of them and fire a bullet based on muscle contraction.

To complement the main study, we also calculated the accuracy of each bullet shot, by finding the distance between where the bullet lands and the center point of the target, and evaluated the error rate (Henze et al. 2011). However, we did not compare this with other input method and simply treat this as a minor performance evaluation.



Fig. 7 Shooting game to obtain qualitative feedback, score, and accuracy

# 5 Results

We assessed the Fitts' law index of performance (IP),

**Fig. 8** Throughput is highly dependent on the target selection method (lines), whereas level of difficulty does not yield a statistically significant effect (slope). There is no interaction effect between the two-fixed factor selection method and level of difficulty recorded subjective workload via a NASA TLX questionnaire, and collected qualitative feedback from a simple shooting game.

## 5.1 Target selection

We conducted a two-way repeated measure ANOVA with the dependent variable throughput and the two-fixed factor target selection method and level of difficulty. An interaction between condition and level of difficulty could not be demonstrated, F(20, 450) = .241, p = 1.0. There was a statistically significant effect of the target selection method on throughput, F(4, 450) = 89.211, p < .0001. With  $\eta_p^2 = .442$  this is a large effect accounting for 41.5% (adjusted  $\mathcal{R}^2$ ) of the variance in throughput. Level of difficulty did not have a statistically significant effect on throughput, F(5, 450) = 1.433, p = .211. Figure 8 visualizes the estimated marginal means of throughput and shows the dependency on the target selection method. Post hoc analysis conducted with a Tukey's range test showed that target selection by motion yielded significantly higher throughput than all other methods (p < .0001). Dwell lead to significantly higher throughput than gamepad (p = .011) and gaze



#### Estimated Marginal Means of Throughput

(p < .0001), whereas gaze and EMG resulted in significantly higher throughput than gaze alone (p = .001).

#### 5.2 Workload

To assess the mental workload perceived by participants, we calculated the raw TLX score after each condition and applied a Friedman test on the five control conditions. Here, we found a statistically significant difference between conditions,  $\chi^2(2) = 25.509, p < .0001$ ). We conducted the post hoc analysis with Wilcoxon signed-rank tests applying a Bonferroni correction, which resulted in a significance level of p < .005. Median perceived work load scores for Dwell were 40 (SD = 17.6), for Dwell and Gaze 51.7 (SD = 12.3), for Gaze and EMG 29.2 (SD = 12.7), for gamepad 25.8 (SD = 15.2), and for Motion 29.6 (SD = 10.3) (see Fig. 9). There was a statistically significant reduction in perceived work load when using Gaze and EMG (Z = -3.124, p = .002), the Gamepad (Z = -3.362, p = .001), or Motion (Z = -3.465, p = .001)compared to Dwell and Gaze. There was no statistically significant difference between the remaining conditions.

#### 6 Discussion

Overall, we demonstrated the feasibility of combining eye gaze with EMG for VR interactions. Based on the results, we found that dwell gaze performed the worst, which is to be expected because the participants were required to fixate on a point for a period of time, which can be uncomfortable or straining, especially given the length of time and number of trials they were required to do so. However, by removing fixation and replacing it with instant activation using EMG, the performance significantly improves. Using the motion controller still outperforms other input methods which can be due to several factors, namely due to how it feels like a direct pointing gesture, though it is not a handsfree solution and therefore could be more limiting regarding introducing additional input or for mobile VR solutions. Finally, the gamepad was the slowest since unlike eye, head, or arm-based selection, using the gamepad requires a preset maximum speed and thus cannot be as fast as the participant wishes it to be.

When asked to provide qualitative feedback during the Fitts' study, a total of seven participants preferred to use the combination of eye gaze with EMG, eight participants preferred using the motion controller, and one participant preferred using the conventional gamepad the most. When we interviewed the one participant that chose gamepad input, he or she mentioned that it was the most relaxing and

**Fig. 9** Perceived work load as a result of the NASA raw TLX scores: target selection with Dwell and Gaze was significantly more strenuous than using gaze and EMG, the controller, or motion



least taxing, requiring near to zero physical movements, thus reducing motion sickness because he or she did not have enough rest the previous night and was feeling slight nausea and disorientation prior to participating in our study.

For the eight participants who preferred the motion controller, three of them mentioned that it was simply more fun since the virtual laser pointer from the tip of the motion controller looked visually similar to a fictional sword in a popular science fiction movie, thus increasing the enjoyment. It was also overall fun for these three participants who had never experienced a motion-based controller for VR before. The remaining five participants who chose motion controller was simply due to its balance between speed and stability. For lower ID, a wrist movement was enough to perform the selection, though, at higher ID, arm movement was necessary, thus increasing fatigue over time. Furthermore, all eight of these participants experienced some difficulty in calibration for the eye tracking, thus reducing the overall score of both the eye-gaze dwell and eye gaze with EMG. For the participants who preferred eye gaze with EMG, five of them stated that it was extremely fast, since the selection was achieved using eye gaze which feels natural, and activation was using arm muscle without much force, thus negating any fatigue. The other three participants enjoyed the novelty of the interaction over other more conventional methods since these participants also have moderate to expert experience in VR interactions.

Among other comments, one of the participants mentioned that the loading time used for dwell and dwell gaze is frustrating. Another participant had no issue with the loading time, but dwell solution was tiring for the neck, especially for higher ID. A participant who tried the gaze solutions first followed by dwell found that he or she felt strange that activation did not start after looking at the points with the eye, possibly due to ordering effect.

## 7 Limitations

One of the main limitations at this point in time is, in fact, the accuracy of eye-gaze tracking in VR. With precise calibration, it is possible to obtain near-perfect tracking. However, in most cases, participants need to calibrate several times to achieve the desired level of accuracy. Furthermore, eye tracking is a very delicate method that uses IR cameras. The occurrence of slippage in the HMD after calibration will completely render the tracking unusable and further calibration is required. This means that the user cannot make any sudden or extreme head movement to maintain eye-tracking precision. However, it is worth noting that as this technology becomes more mainstream, the tracking will certainly improve. Another common issue was that since the position of the trackers is fixed in the HMD, it largely depends on the participant's compatibility with the trackers. From our study, we found that the eye trackers have difficulty detecting the eyes of Asian participants, whereas they seem to work relatively well for European participants. This can be due to several factors such as eye color, size, and relative position of the eyes. As of this moment, the currently used setup is suitable for our user studies; however, a separate MacBook was required to operate the eye trackers. Nevertheless, we believe eye tracking will be the next evolution for VR and AR technology once it matures. Regarding the utilization of EMG, since minor muscle contraction can be easily detected by the sensor, it is possible for accidental activation to still occur. This is a common problem for input methods that do not involve buttons. Nevertheless, software tweaks on the UI behavior or increasing the activation threshold can easily circumvent this issue. High muscle contraction only even happens if the user is lifting something heavy or performing an extreme motion like punching unless done on purpose. Comparing to interactions like arm or finger gesture, or pressure sensors on certain body parts, EMG activation is less likely to accidentally occur.

## 8 Proposed applications

Based on the obtained results, we now understand the potential of using eye gaze with EMG for VR interaction. However, the Fitts' law study merely demonstrates its interactivity with context to a point-and-click task, as opposed to real-world applications like gaming or interacting with objects. In this section, we designed several VR applications that can fully benefit from the proposed interaction. Since VR is a diverse platform, the applications are divided into their respective fields for interior and engineering design, entertainment, and gaming, as well as UI interface selection and media consumption. For each of these applications, we show how the user can easily select and activate elements that are present depending on the application that is handsfree, time-saving, and unobtrusive. These applications serve as a proof of concept on how eye gaze with EMG can provide a unique alternative to interaction, whether it is gaming on the go or have a VR business conference. Each of this applications will be explained through its categorization for both the target acquisition phase and target action phase. We explored four application use cases based on our novel input modality: (1) interior design explorations, (2) gaming, and (3) text input.

#### Fig. 10 Pick-and-place task



#### 8.1 Interior design

One of the core benefits when it comes to VR is the ability to place users in a virtual environment as though they are actually there. For professional use, one such usage would be assessing an interior design. In such an environment, minimalistic UI is best so that it will not obstruct the designer's view to allow careful observation and assessment of the environment. In our implemented prototype, navigation and pick-and-place are provided through a discrete action of the target action phase. If the user wishes to navigate to a position, they can simply look at the ground of that point and perform a muscle contraction to teleport to that location. Teleportation navigation is increasingly more popular as it negates the effect of motion sickness (Cloudhead 2015). We also implemented a pick-and-place tool shown in Fig. 10, allowing the user to look at a particular object of interest, pick it up via short muscle contraction, and place it at any designated spot by performing another muscle contraction. To summarize, this particular application is a pick-andplace scenario by moving furniture while navigating interior spaces.

#### 8.2 Gaming

This generation of VR products focus on gaming; therefore it is only natural to consider some kind of gaming function. We created a simple shooting scene shown in Fig. 11 where the user views the world in a first-person view with the gun placed at the lower right corner, similar to most first-person shooter (FPS) games. We equip the user with a pistol and allow the user to toggle between full-auto and semi-auto firing. Full-auto allows the user to continuously fire the rifle for as long as their muscle contracts, while



**Fig. 11** Firing a gun in full-auto and semi-auto



semi-auto fires a single bullet per muscle contraction. This effectively switches between both kinds of target action phase. One of the interesting benefits of using physiological signals is that in this use case, contraction of the arm muscle could simulate the recoil of a gun, whereas aiming with the eyes simply feels natural, leading to a greater level of immersion in gaming though further studies are required to test this hypothesis.

#### 8.3 Input interfaces: number pads and keyboards

We created a generalized UI system to demonstrate the feasibility of the proposed interaction mechanic. The first UI is a number pad illustrated in Fig. 12 that allows the user to simply look at a number and contract their muscle to select it. The second prototype is a series of knobs and graphs. Each of this interface depends on the user's amount of muscle contraction. Maximum contraction will maximize the knobs and graphs while relaxing the muscle reduces it back to 0. These UIs serve as a proof of concept for them to be applied in various other usages like menu selection, scrolling, and media control without the presence of physical buttons or other devices that occupy the users' hands. To evaluate this, a text entry study by reporting the words per minute (WPM) and error rate would prove useful.

## 9 Conclusion and future works

We present a novel multimodal assessment for VR interaction using both eye gaze and EMG for selection and activation. In two user studies, we have shown the applicability of these input methods for future VR solutions, as machines become more portable, powerful, and well equipped with the necessary sensors. The hands-free and time-saving nature of this interaction mechanic provides insights into its feasibility, effectiveness, and reliability for VR applications. As these input modalities become seamlessly integrated with the next generation of VR tools, more users may find themselves indulging in VR anytime, anywhere. For future works, we plan to further evaluate the proposed application scenarios by performing an immersion study for gaming, text entry study for input interfaces, and user feedback in pick-and-place and navigation scenarios.

Acknowledgements This work was supported by the JSPS KAKENHI Grant Number 18H03278.

## References

- Barry DT, Gordon KE, Hinton GG (1990) Acoustic and surface EMG diagnosis of pediatric muscle disease. Muscle Nerve 13(4):286–290
- Benko H, Saponas TS, Morris D, Tan D (2009) Enhancing input on and above the interactive surface with muscle sensing. In: Proceedings of the ACM international conference on interactive tabletops and surfaces—ITS '09, p 93. https://doi.org/10.1145/1731903.17319 24. http://portal.acm.org/citation.cfm?doid=1731903.1731924
- Cardoso J (2016) Comparison of gesture, gamepad, and gaze-based locomotion for VR worlds. In: Proceedings of the 22nd ACM conference on virtual reality software and technology. ACM, pp 319–320
- Chatterjee I, Xiao R, Harrison C (2015) Gaze + gesture : expressive , precise and targeted free-space interactions. In: Proceedings of the 2015 ACM on international conference on multimodal interaction (C), pp 131–138. https://doi.org/10.1145/28183 46.2820752
- Chin CA, Barreto A, Cremades JG, Adjouadi M (2008) Integrated electromyogram and eye-gaze tracking cursor control system for computer users with motor disabilities. J Rehabil Res Dev 45(1):161
- Cloudhead G (2015) VR Navigation. http://cloudheadgames.com/ cloudhead/vr-navigation/

- Costanza E, Inverso SA, Allen R (2005) Toward subtle intimate interfaces for mobile devices using an EMG controller. In: Proceedings of the SIGCHI conference on Human factors in computing systems CHI 05:481. https://doi.org/10.1145/10549 72.1055039. http://eprints.soton.ac.uk/270956/
- Darken RP, Cockayne WR, Carmein D (1997) The omni-directional treadmill: a locomotion device for virtual worlds. In: Proceedings of the 10th annual ACM symposium on User interface software and technology—UIST '97, pp 213–221. https:// doi.org/10.1145/263407.263550. http://dl.acm.org/citation. cfm?id=263550
- Henze N, Rukzio E, Boll S (2011) 100,000,000 taps: analysis and improvement of touch performance in the large. In: Proceedings of the 13th international conference on human computer interaction with mobile devices and services. ACM, pp 133–142
- Hernandez Arieta A, Katoh R, Yokoi H, Wenwei Y (2006) Development of a multi-DOF electromyography prosthetic system using the adaptive joint mechanism. Appl Bionics Biomech 3(2):101–111
- Hernandez-Rebollar JL, Kyriakopoulos N, Lindeman RW (2002) The AcceleGlove: a whole-hand input device for virtual reality. In: ACM SIGGRAPH 2002 conference abstracts and applications. ACM, p 259
- ISO (1998) Ergonomic requirements for office work with visual display terminals (VDTs)—part 11: guidance on usability. ISO 9241-11:1998, p 22
- Jacob RJK (1990) What you look at is what you get: eye movementbased interaction techniques. In: Proceedings of the SIGCHI conference on Human factors in computing systems. ACM, pp 11–18
- Jacobsen SC, Jerard RB (1974) Computational requirements for control of the Utah arm. In: Proceedings of the 1974 annual conference, vol 1. ACM, pp 149–155
- Jones BR, Benko H, Ofek E, Wilson AD (2013) IllumiRoom: peripheral projected illusions for interactive experiences. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM, pp 869–878
- Kiguchi K, Tanaka T, Fukuda T (2004) Neuro-fuzzy control of a robotic exoskeleton with EMG signals. IEEE Trans Fuzzy Syst 12(4):481–490
- Kim YR, Kim GJ (2017) HoVR-type: smartphone as a typing interface in VR using hovering. In: 2017 IEEE international conference on consumer electronics (ICCE). IEEE, pp 200–203
- Lee JS, Lee SE, Jang SY, Park KS (2005) A simplified hand gesture interface for spherical manipulation in virtual environments. In: Proceedings of the 2005 international conference on Augmented tele-existence. ACM, p 284
- MacKenzie IS, Soukoreff RW (2002) Text entry for mobile computing: models and methods, theory and practice. Hum Comput Interact 17(2–3):147–198
- Miniotas D (2000) Application of Fitts' law to eye gaze interaction. In: CHI '00 extended abstracts on Human factors in computer systems—CHI '00, pp 339–340. https://doi.org/10.1145/63348 2.633496. http://portal.acm.org/citation.cfm?doid=633292.63349
- Moseley JRJB, Jobe FW, Pink M, Perry J, Tibone J (1992) EMG analysis of the scapular muscles during a shoulder rehabilitation program. Am J Sports Med 20(2):128–134

- Myo (2013) Myo Gesture Control Armband. http://www.myo.com/ techspecs
- Oculus (2015) The Rift's Recommended Spec, PC SDK 0.6 Released, and Mobile VR Jam Voting. https://www3.oculus.com/en-us/blog/ the-rifts-recommended-spec-pc-sdk-0-6-released-and-mobile-vrjam-voting/
- Pai YS, Tag B, Outram B, Vontin N, Sugiura K, Kunze K (2016) GazeSim: simulating foveated rendering using depth in eye gaze for VR. In: ACM SIGGRAPH 2016 Posters. ACM, p 75
- Ramcharitar A, Teather RJ (2017) A Fitts' law evaluation of video game controllers: thumbstick, touchpad and gyrosensor. In: Proceedings of the 2016 CHI conference extended abstracts on human factors in computing systems. ACM, New York, NY, CHI EA '17, pp 2860–2866. https://doi.org/10.1145/3027063.3053213. http:// doi.acm.org/10.1145/3027063.3053213
- Saponas TS, Tan DS, Morris D, Balakrishnan R, Turner J, Landay JA (2009) Enabling always-available input with muscle–computer interfaces. In: Proceedings of the 22nd annual ACM symposium on User interface software and technology, pp 167–176. https:// doi.org/10.1145/1622176.1622208
- Saraiji Y (2016) PupilHMDCalibration. https://github.com/mrayy/ PupilHMDCalibration
- Saraiji Y, Sugimoto S, Fernando CL, Minamizawa K, Tachi S (2016) Layered telepresence: simultaneous multi presence experience using eye gaze based perceptual awareness blending. In: ACM SIGGRAPH 2016 posters. ACM, p 20
- Slambekova D, Bailey R, Geigel J (2012) Gaze and gesture based object manipulation in virtual worlds. In: Proceedings of the 18th ACM symposium on virtual reality software and technology. ACM, pp 203–204
- Unity (2015) User interfaces for VR. https://unity3d.com/learn/tutor ials/topics/virtual-reality/user-interfaces-vr
- Ware C, Mikaelian HH (1986) An evaluation of an eye tracker as a device for computer input. ACM SIGCHI Bull 17(SI):183–188. https://doi.org/10.1145/30851.275627
- Williams B, Bailey S, Narasimham G, Li M, Bodenheimer B (2011) Evaluation of walking in place on a Wii balance board to explore a virtual environment. ACM Trans Appl Percept 8(3):1–14. https ://doi.org/10.1145/2010325.2010329
- Wilson AD, Cutrell E (2005) Flowmouse: a computer vision-based pointing and gesture input device. In: Human-computer interaction-INTERACT 2005, vol 3585, pp 565–578. https://doi. org/10.1007/11555261. http://www.springerlink.com/index/ DJ6AENMGBTYBC0EE.pdf
- Zander TO, Gaertner M, Kothe C, Vilimek R (2010) Combining eye gaze input with a brain? Computer interface for touchless human? Computer interaction. Int J Hum Comput Interact 27(1):38–51. https://doi.org/10.1080/10447318.2011.535752
- Zhang X, MacKenzie IS (2007) Evaluating eye tracking with ISO 9241—part 9. In: Proceedings of the 12th international conference on human-computer interaction: intelligent multimodal interaction environments, HCI'07. Springer, Berlin, pp 779–788. http:// dl.acm.org/citation.cfm?id=1769590.1769678