User interaction in hands-free gaming: a comparative study of gaze-voice and touchscreen interface control

Muhtar Çağkan ULUDAĞLI1,∗, Cengiz ACARTÜRK2
1Department of Game Technologies, Graduate School of Informatics, Middle East Technical University, Ankara, Turkey
2Department of Cognitive Science, Graduate School of Informatics, Middle East Technical University, Ankara, Turkey

Received: 12.10.2017 • Accepted/Published Online: 03.05.2018 • Final Version: 27.07.2018

Abstract: The goal of the present study is to contribute to the research on the use of eye tracking in hands-free gaming. We report the results of an experimental study by using a platform game designed by the authors. The game allowed gaze-voice command control, as well touchscreen control of the interface by the player. We employed a set of performance measures and we measured player engagement as an indicator of user experience. The findings revealed that the players did not improve in game performance when they used the gaze-voice control compared to the touchscreen control. On the other hand, they exhibited stronger engagement, which indicates the potential of gaze-voice combination as an acceptable interaction method compared to touchscreen in mobile games.

Key words: Eye movement, eye tracking, hands-free gaming, user engagement, touchscreen control, gaze-voice control, human computer interaction

1. Introduction

The use of gaze as an input modality in games is a relatively recent technological advance compared to the use of haptic or touch-based devices, such as keyboard, mouse, and joy-stick. Given that the revenues of the mobile gaming industry exhibited an annual increase (21% from 2015 to 2016), the use of gaze in mobile gaming is likely to attract popular demand from society in the incoming years. On the other hand, our knowledge is limited on the user experience of the players in mobile gaming by gaze and complementary modalities to gaze. The main objective of our study is to investigate the capability of gaze and voice control as complementary modalities in a game setting. We hypothesize that the combination of gaze and voice control has potential as an efficient and successful game control method in hands-free platforms in various ways. From a theoretical perspective, the allocation of attentional resources through different sensory channels, in this case vision and audition, provide an appropriate method of multimodal interaction in contrast to allocation through the same sensory channel [1–3]. Moreover, from the perspective of user interaction, games have high potential for user engagement. For this, we developed a multimodal (gaze-voice) control method for a hands-free, fast-paced game platform to investigate the potential of gaze and voice control as complementary and efficient modalities. In particular, we aim at studying how users perceive gaze-voice command control in hands-free gaming by comparing game performance and player engagement with touchscreen command control.

∗Correspondence: e199582@metu.edu.tr
2. Eye tracking in computer games

2.1. Control methods for interaction in computer games

Gaze has been employed as an interaction modality in a variety of game genres, including platform games such as chess [4], first-person shooter games [5], breakout games [6], massively multiplayer online games [7], and flight simulators [8]. The control methods for interaction depend on the game genre. For instance, in chess, the players may pick the piece by gaze by a blink, a specific pattern of gaze gesture, or an elongated dwell time on the piece. In games that involve character navigation, the player may control a tool by gaze, or the view of sight in the dynamic game scene. The common aspect of gaze control in games is that it usually mimics the input from the player, which can also be provided by alternative modalities, such as a mouse click, keyboard input, or joystick direction control for navigation. However, it is a limited control method alone for gaming since using solely gaze control is infeasible to perform tasks for which any two of other modalities must be used at the same time.

2.2. Player engagement as a measure of user experience in computer games

Player engagement is described as deep play [9]. It may be target-related or empathetic [10]. Several concepts have been employed to measure user engagement in games, including traditional measures, such as presence, flow, and enjoyment [11]; measurements that are based on emotion recognition [12]; and process-oriented frameworks for measurement, which focus on specific aspects of engagement such as objectives, accomplishment, activity, and affect [13,14]. In the present study, we employ the Game Engagement Questionnaire (GEQ), which has been proposed as a reliable measure for evaluating player engagement in computer games [15]. The GEQ specifies four modules, namely immersion, presence, flow, and psychological absorption. Immersion is the experience of becoming engaged in gameplay experience while keeping some awareness of the surrounding environment [16]. Presence is the experience of being inside a virtual environment [17]. Flow is the feeling of enjoyment that happens when stability between skill and challenge is achieved while performing an intrinsically rewarding activity [18]. Finally, psychological absorption is the total engagement in the present condition [19].

2.3. User experience in gaze-controlled games

The previous research on user experience in gaze-controlled games has revealed divergent findings. For instance, [20] reported that the participants found gaze input easy to learn and more enjoyable than mouse control in first-person shooter games. For the same game genre, in a similar study, [6] reported no improvement in players' game performance when they used gaze as a game interface control method. On the other hand, in a breakout game clone, participants exhibited higher game performance when they controlled the game interface by gaze compared to mouse control. The players also found the gaze control method enjoyable [7]. The majority of the findings in previous research reveal either similar or lower game performance when players employ gaze control in a game compared to traditional control methods such as mouse control [7,8,21,22]. There are a few exceptional studies though, which reported that players exhibited higher performance by gaze control than traditional control methods [23]. On the other hand, in all aforementioned studies, players reported higher game engagement when they controlled the interface by gaze. Gaze control in gaming was proposed as a promising method of interaction, given a set of positive indicators of eye tracker use (e.g., one-player mode, turn-based gameplay) and a set of negative indicators (e.g., continuous position control, dissociation of focus of attention and control, and large number of commands) as a function of game genre [24]. In conclusion, the previous
research emphasizes various aspects of multimodal interaction in games and relevant contexts. Recently and more specifically, more research is needed to understand the role of gaze as a game control interface, both as a single modality and as a modality in multimodal contexts.

2.4. Gaze control in multimodal contexts

The relationship between gaze and potential complementary modalities to gaze in game playing has attracted research interest since the last decade. For instance, [25] investigated two commonly used tasks in computer game control, namely target acquisition and target tracking. The authors compared target acquisition performance in pointing by synchronous clicking between gaze-mouse control, mouse-electromyography (EMG) control, and mouse-only control. They found that gaze-EMG control was faster than mouse-only control. In [26], Petrini and Forslin employed a brain-computer interface as a complementary modality to gaze input. They used gaze input or mouse input for camera rotation and aiming and used brain-computer interface or mouse input to shoot a target, and all of the different combinations of these were evaluated. Their findings revealed that gaze and brain-computer interface inputs resulted in a decrease in player performance. In [27], Sundstedt et al. discussed possibilities and challenges in video games and virtual reality applications. They reported that gaze interaction is novel in games and virtual reality applications and further research is needed to evaluate the possibilities. In [28], the researchers investigated gaze-contingent audio enhancement methods driven by audiovisual attention, which might be used to simulate rich interaction in virtual environments. They reported that users enjoyed controlling the sound level at the attention point of gaze; therefore, gaze-contingent audio modulation methods might improve the user experience in virtual reality applications.

A synchronous use of gaze and voice in gaming is recently rare, but [29] reported a gaze- and voice-controlled game called “The Revenge of the Killer Penguins”. It is a third-person adventure puzzle game and they used a focus group to measure the effects of the interface control method. They did not conduct a user study, though; therefore, their findings are limited to the evaluation of the importance of good game design that uses eye tracking technology. Another gaze-voice controlled game is “Rabbit Run”, which can be controlled by keyboard-mouse combination, as well [30]. The participants reported a higher level of immersion for gaze-voice control compared to keyboard-mouse control, despite their worse game performance in gaze-voice control.

Finally, [31] investigated gaze input as a cursor control method and voice input as a command control method for a computer drawing software. Their participants reported the feeling of less control, speed, and precision compared to control by mouse and keyboard. The participants also reported that it could be easier to use the new control method if they had more practice with it.

These findings reveal that our knowledge of gaze control as an alternative interaction methodology in multimodal contexts has not yet reached a mature state of convergent and reliable findings. The current literature is limited to a few studies about gaze-voice control of gaming interfaces. Moreover, to our knowledge, there is no study that focuses on gaze-voice control in hands-free gaming. To address this gap in the relevant research literature, we designed an experiment that included a platform runner game, namely “Neon Glider”, in which the users were able to control the interface either by a traditional method (tapping on the touchscreen) or by gaze-voice control. Below we introduce the game infrastructure and design.

3. Neon Glider: A hands-free game interface for gaze-voice control

3.1. Game infrastructure

A gaze-voice–controlled game infrastructure involves various aspects of game development, including the programming language for the implementation, the game engine, eye tracking equipment, and voice recognition.
setup. Neon Glider is a platform runner game that was designed by the authors. It was implemented in C# for compatibility with the eye tracker (EyeTribe™), the voice recognition system (Microsoft SAPI), and the game engine (Unity). Unity was used as a game engine in the present study. It is a cross-platform game engine developed by Unity Technologies and it has been used for developing video games for most of the available gaming platforms including desktop personal computers and mobile devices. A dynamic-link library (DLL) by EyeTribe was used for integrating eye tracking capability into the game. For voice recognition, a console application and a DLL that provided intercommunication in the local network were developed by the first author. The users uttered appropriate commands while they were playing the game and the application recognized only three voice commands from the users that were necessary for controlling the game. Voice recognition was independent from the voice, tone, or volume of the user and the experimental environment was silent to eliminate interference from the real world.

We used an EyeTribe eye tracker with a reported sampling rate of 30 Hz, 0.5–1 degrees of accuracy, 45–75 cm of working range, and 40 × 30 cm trackbox at 65 cm distance from the screen. For voice recognition, we used the Microsoft Speech Application Programming Interface (SAPI version 3.0 and .NET Framework version 3.5), which supplied a high-level interface between the voice recognition engine and the application. The touchscreen was a Lenovo™ 13.3” FHD (1920 × 1080) multitouch display with an i7 2.00 GHz 1600 MHz 4 MB processor and Intel 4400 integrated graphic card. In the following section, we introduce the design of the game interface and game dynamics.

3.2. Game design

The game was a 2D platform-runner game. Figure 1 presents a screenshot from the game interface. The player controls the square character (shown on the left of the top half of the screen in Figure 1) to prevent collision with the horizontal zigzag obstacles and switching to the other platform when needed. During the gaze-voice–controlled gameplay, the player is able to shift the square to the other half (either the top half or the bottom half) of the screen by gazing to the intended half-screen area. The voice recognition engine recognizes two voice commands, which were selected to have phonetically distinct labels for better voice recognition performance.

- The “jump” command makes the square jump over an obstacle.
- The “twice” command makes the square go to the opposite edge of the platform (i.e. from top to middle and from middle to top in the top half of the screen, and from middle to bottom and from bottom to middle in the bottom half of the screen).

Figure 1. A snapshot from the game interface.
The control of the game interface by touchscreen, as an alternative to the gaze-voice control of the game, involves tapping the screen to control the actions of the square. There are three actions that can be used: up, down, and shift area. If the player wants to go up, he/she touches the bottom right side of the screen. To go down, the player touches the bottom left side of the screen. If the player wants to shift to the other half of the screen when he/she is on one of the middle platforms, he/she touches the appropriate side of the screen (i.e. bottom left or bottom right). Figure 2 shows the setup in the gaze-voice control and the touchscreen control conditions.

![Figure 2. Gaze-voice control of the game (left) and touchscreen control of the game (right).](image)

The platforms continuously move to the left. The square advances forward (i.e. to the right) in the case of no command input, thus colliding with an obstacle. For the purpose of providing a continuous flow to the player during the experiment, the game did not stop when the square collided with an obstacle.

We conducted an experimental investigation to make a comparative analysis of the use of the gaze-voice control and the touchscreen control using the game platform. The next sections present the experiment and the results.

4. The experiment

4.1. Participants

Eighteen male and six female participants participated in the experiment, with a mean age of $M = 32.35$ years (standard deviation [SD] = 7.07). The participants had normal vision, so none of them used glasses in the experiment. The demographic data about their game-playing habits revealed that 25% of them played computer games daily and 38% of them played games at least once a week. The rest played computer games not very often. They were from different departments and working areas. The game genres that were played by the participants most were action games and casual games. Sports, action/adventure, adventure, and strategy games were the other frequently played genres in that order. The majority (75%) of the participants played games on mobile platforms, and nearly 70% played games on a desktop computer. Console players were in the minority (20.8%).

4.2. Design and procedure

The experiment setup was designed to acquire game-play data and GEQ data from the participants. The participants provided informed consent before the experiment. The experiment was conducted in single sessions. The participants first filled in a demographic questionnaire. After the calibration of the eye tracker, they were presented a tutorial to make them familiar with the game environment and the interaction controls.
After the tutorial session, the participants started to play the game with one of the control methods (either by gaze-voice control or by touchscreen control). The starting conditions were randomized. The game session took approximately 2 min per participant and per control condition. After each game session, the participants filled in the GEQ form. The questionnaire form was filled in for each control method separately. The participants used a Likert scale for the scoring, i.e. a 1 to 7 scale where the ordered values were strongly disagree, disagree, slightly disagree, neutral, slightly agree, agree, and strongly agree. The whole experimental session took approximately 30 min. After the experiment, the results were analyzed with IBM SPSS v23.

4.3. Results
The results and the analyses are presented in two parts below, for game performance and for questionnaire results separately.

4.3.1. Game Performance
We specified two measures for evaluating game performance, namely success rate and action counts. The results for each measure are presented below.

4.3.1.1. Success rate
We defined the success rate of the players as the inverse ratio of the number of failures (i.e. the number of obstacle collisions) to the total number of obstacles that were displayed during the game, given by the formula below.

\[\text{successRate} = \left(1 - \frac{\text{failureCount}}{\text{totalObstacle}}\right) \times 100.\]

The results of a pairwise comparison between the two experimental conditions (i.e. the gaze-voice control and the touchscreen control) revealed a significant difference in the success rate of the participants between the touchscreen control condition (M = 96.30%, SD = 2.50%) and the gaze-voice control condition (M = 87.86%, SD = 4.05%), \(t(23) = 9.55, P < 0.01\); the players exhibited higher success rates when they played the game by touchscreen control than by gaze-voice control.

In order to understand if the game-playing habits of the participants and/or the control method that they used affected the success rates of their gameplay, a two-way ANOVA was conducted that examined the effect of control method and playing habit on success rates. No statistically significant interaction was obtained for playing habits on success rates with an effect size \(F(4, 38) = 0.66, P = 0.63\), and neither was a main effect of game-playing habit on success rates obtained, \(F(4, 38) = 1.56, P = 0.21\).

4.3.1.2. Action counts
We measured action counts as indicators of the density of the actions generated by the participants during the game. There were three actions that a participant was able to take in both the gaze-voice control condition and the touchscreen control condition: jumping, changing the platform (e.g., from the middle of the screen to the top within the same half-screen area), and changing the half-screen area (e.g., from the bottom-half platform to the top-half platform). For instance, a participant was able to jump with the “jump” voice command in the gaze-voice controlled game and he/she was able to jump by touching the rightmost bottom corner or the leftmost bottom corner of the screen in the touchscreen controlled game. Similarly, the participant was able to
use the other actions in both conditions through different modalities (i.e. gaze-voice vs. touchscreen tapping). Table 1 presents the mean number of actions generated during the game for both control conditions.

Table 1. Action count results.

<table>
<thead>
<tr>
<th>Action type</th>
<th>Game control condition</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Touchscreen</td>
<td>Gaze-voice</td>
</tr>
<tr>
<td>Jumping</td>
<td>58.33 (11.88)</td>
<td>10.04 (5.77)</td>
</tr>
<tr>
<td>Changing the platform</td>
<td>41.0 (17.35)</td>
<td>13.92 (8.42)</td>
</tr>
<tr>
<td>Changing the half-screen area</td>
<td>20.08 (9.09)</td>
<td>15.92 (9.57)</td>
</tr>
</tbody>
</table>

A two-way ANOVA revealed a main effect of action type (jumping, changing the platform, and the half-screen area), $F(2,138) = 26.42, P < 0.01$; a main effect of the type of game interface control (gaze-voice control and touchscreen control), $F(1,138) = 210.67, P < 0.01$; and an interaction between the two, $F(2,138) = 48.65, P < 0.01$.

Further pairwise comparisons revealed a significant difference between the count of the jumping actions with the touchscreen control and the one with the gaze-voice control, $t(23) = 20.25, P < 0.01$. This finding suggests that the participants used the jump command more frequently when they played the game with the touchscreen compared to gaze-voice control. A similar result was obtained for changing the platform action, $t(23) = 6.38, P < 0.05$, which shows that the participants used this action more frequently in the touchscreen condition than the gaze-voice condition.

The third game action was changing the half-screen area action. A comparison of the use of this action between the two game control conditions revealed no significant difference, $t(23) = 1.31, P = 0.203$. It is evident that the participants used the two control methods differently during game play, as indicated by significant differences for the number of actions between the two conditions. Changing the half-screen area action revealed no significant difference, thus suggesting mixed results for the game actions. Therefore, at this stage, our interpretation of the results requires further investigation of the use of alternative interface controls in mobile gaming. We present the results of the GEQ in the following section.

4.4. Game evaluation questionnaire

There are 19 questions in the GEQ, and these questions comprise four groups or modules: presence, immersion, flow, and psychological absorption [16]. Table 2 shows the mean scores given by the players for each module and for the two game interface control conditions.

Table 2. Game Evaluation Questionnaire (GEQ) results.

<table>
<thead>
<tr>
<th>Module</th>
<th>Game control condition</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Touchscreen</td>
<td>Gaze-voice</td>
</tr>
<tr>
<td>Presence</td>
<td>4.93 (1.01)</td>
<td>4.82 (1.06)</td>
</tr>
<tr>
<td>Immersion</td>
<td>5.63 (1.09)</td>
<td>5.42 (1.44)</td>
</tr>
<tr>
<td>Flow</td>
<td>4.50 (1.01)</td>
<td>4.25 (0.98)</td>
</tr>
<tr>
<td>Psychological absorption</td>
<td>3.42 (1.25)</td>
<td>4.02 (1.32)</td>
</tr>
</tbody>
</table>

The immersion module has only one question: “I really get into the game”. Four questions, “Things seem to happen automatically, My thoughts go fast, I play longer than I mean to, I lose track of time”, constitute the presence module. The questions “I don’t answer when someone talks to me, I can’t tell I’m getting tired,
If someone talks to me I don’t hear them, I feel like I can’t stop playing. The game feels real, I get wound up, Playing seems automatic, I play without thinking how to play, Playing makes me feel calm” are in the flow module. Finally, the questions “I feel different, I feel scared, Time seems to kind of stand still or stop, I feel spaced out, I lose track of where I am” are in the psychological absorption module of the GEQ.

For comparing the module scores, GEQ answers of the participants were analyzed by a mixed ANOVA test for two different game control methods. Since the assumption of sphericity was not met, as indicated by the Mauchly test, we used the Greenhouse–Geiser corrected degrees of freedom to assess the significance. The ANOVA revealed a main effect for the modules, $F(2.21, 101.54) = 50.19$, $P < 0.01$, as well as an interaction between the modules and the game control condition, $F(2.21, 101.54) = 3.42$, $P < 0.05$. Follow-up pairwise comparisons revealed a significant difference between the gaze-voice interface control and touchscreen interface control in the psychological absorption module, $t(23) = 2.23$, $P < 0.05$. This finding is of importance since it indicates the potential of gaze-voice control as a usable method of interface control.

As for the influence of game-playing habits on the questionnaire scores, analysis of the interaction between game-playing habits, the control methods, and the questionnaire modules revealed no significant interaction effect on the GEQ scores.

5. Conclusions and future directions

The present study revealed two major findings about the use of gaze-voice interface control and the use of touchscreen interface control in hands-free gaming. The first is that the participants exhibited higher game performance, or in other words they were more successful in the game, when they used the touchscreen control, as indicated by the comparison of the success rates between the two game control conditions. Secondly, the results of the questionnaire revealed a significant interaction of the GEQ modules and the interface control method in that the participants who used the gaze-voice control method to play the game were more psychologically absorbed than the ones who used the touchscreen control method to play it. These findings confirm the expected potential of gaze-voice command reported in the literature [32,33].

Recently, systematic research has been scarce on game experience evaluation of multimodal interface control methods in mobile devices. The findings of the present study suggest gaze-voice control as an acceptable interface control method in mobile devices, which has the potential to lead to higher gameplay engagement on the player’s side.

There are several limitations that should be addressed for generalizing the findings. First, the present study is limited to a single game for practical reasons, which is assumed to be representative for a similar game genre. For the generalization of the findings, the gaze-voice control method should be investigated under various game environments, as well as in nongame contexts. Moreover, the higher engagement of the players in the gaze-voice game interface control compared to touchscreen control might be due to the novelty effect [34]. The participants may have found the eye tracking interface exciting since that is a novel technology for hands-free gaming. Therefore, further research is necessary to understand if players’ engagements change as they get familiar with the eye tracking technology. In particular, future research should address the concept of “gaze load”, a sense of tiredness that some of the participants reported after the experiment. Further research should also address a comparative analysis of a richer set of conditions by employing alternative game genres, such as slow-paced platform games.

In the present design, the game did not stop for the purpose of providing a continuous flow to the player. This design decision has the potential to influence the player’s engagement, a topic that we leave for further
Conducting the experiment with people who have different game-playing habits or computer usage frequency should be considered to generalize the research findings. In the future, the complementary input method to gaze might be changed to another modality than voice input, such as body movements, brain-computer interface, EMG, or EOG to measure blinking and winking. Alternative user groups, including individuals with motor disabilities, should also be addressed [35]. Mobile game control methods will evolve towards more usable and efficient ones as the efficiency of the underlying sensor technologies improve. We propose that employing gaze as an interaction method, possibly in the multimodal context of voice and other complementary modalities in hands-free game control, has the potential to be an alternative interface control method to traditional ones such as tapping on a touchscreen. The potential of the use of gaze and voice as a game control method in hands-free gaming will go beyond the common players of mobile games by addressing players with bodily motor disabilities and elderly citizens with motor disabilities of the hands and fingers.

References


