

# Rabbit Run: Gaze and Voice Based Game Interaction

J. O'Donovan<sup>†1</sup>, J. Ward<sup>‡2</sup>, S. Hodgins<sup>2</sup> and V. Sundstedt<sup>3</sup>

<sup>1</sup>MSc Interactive Entertainment Technology, Trinity College Dublin, Ireland

<sup>2</sup>Acuity ETS Ltd., Reading, United Kingdom

<sup>3</sup>Graphics Vision and Visualisation Group, Trinity College Dublin, Ireland

---

## Abstract

*Modern eye tracking technology allows an observer's gaze to be determined in realtime by measuring their eye movements. Recent studies have examined the viability of using gaze data as a means of controlling computer games. This paper investigates the combination of gaze and voice recognition as a means of hands-free interaction in 3D virtual environments. A novel game evaluation framework is implemented controllable by input from gaze and voice as well as mouse and keyboard. This framework is evaluated both using quantitative measures and subjective responses from participant user trials. This is the first evaluation study performed comparing gaze and voice against mouse and keyboard input. The main result indicates that, although game performance was significantly worse, participants reported a higher level of immersion when playing using gaze and voice.*

Categories and Subject Descriptors (according to ACM CCS): I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Virtual Reality H.5.2 [Information Interfaces and Presentation]: User Interfaces—Input Devices and Strategies

---

## 1. Introduction

Recent innovations in the video game industry include alternative input modalities for games to provide an enhanced, more immersive user experience. For example, the Nintendo Wii sold over twice as many units as the PlayStation 3 from when they were released to the middle of 2007 [San07]. Eye tracking is a process that allow us to determine where the observer's focus lies on the computer screen at any given time. As eye trackers become cheaper and less intrusive to the user the technology could well be integrated into the next generation of games. It is important therefore to ascertain its viability as an input modality and see if it can be used to enhance the gamer experience.

Gaze based interaction is not without its issues. It tends to suffer from the “Midas touch” problem. This is where everywhere you look, another command is activated; you cannot look anywhere without issuing a command [Jac90]. To combat this problem gaze is often used in conjunction with another input mechanism such as a mouse click. The intention

of this work is to show that the “Midas touch” problem can be overcome by combining voice recognition with gaze to achieve a completely hands-free method of game interaction. Alternative means of interaction in games are especially important for disabled users for whom traditional techniques, such as mouse and keyboard, may not be feasible. Given that gaze and voice are entirely hands-free it presents a real opportunity for disabled users to interact fully with computer games.

This paper explores if gaze and voice controlled game interaction is a viable alternative to mouse and keyboard. A multi modal game framework is developed which can take input from mouse, keyboard, gaze, and voice. This allows comparisons to be drawn between different input types. The remainder of the paper is organised as follows:

Section 2 summarises the relevant background information and reviews the state of the art with regard to gaze and voice in gaming. The design and implementation of the game evaluation framework are described in Section 3. Section 4 describes the experimental design of the user evaluation and Section 5 presents the results obtained. Finally, in Section 6 conclusions are drawn and related future work is discussed.

---

<sup>†</sup> {jodonov|sundstev}@tcd.ie

<sup>‡</sup> {jon|scott}@acuity-ets.com

## 2. Background and Related Work

### 2.1. Eye Tracking

Eye-tracking [Duc07] is a process that records eye movements allowing us to determine where an observer's gaze is fixed at a given time. The direction of gaze indicates where humans focus their attention. Eye trackers measure the physical rotations of the eyes. The most common system for measuring the orientation of the eyes in space is the video-based corneal reflection eye tracker [Duc07].

In video-based eye-trackers, the light source reflection on the cornea (caused by infra-red light) is measured relative to the location of the pupil's centre. These two points are used to compensate for head movements. Video-based eye trackers compute the *point-of-regard* (POR) in real-time using cameras and image processing hardware. The Tobii T60 eye tracking system used in this paper, as shown in Figure 1, is of this type.

Eye tracking systems are normally classified into diagnostic and interactive systems [Duc07]. This paper is concerned with interactive eye tracking systems. There are two main types of interactive applications using eye tracking technology: *selective* and *gaze-contingent*. Selective uses an eye tracker as an input device, in a similar way to a mouse. Gaze-contingent applications manipulate the presented information to match the processing of the human visual system. This often matches foveal/peripheral perception in real-time.

### 2.2. Gaze in Gaming

Several computer games have exploited the concept of attention in gaming. For example, in *The Legend of Zelda: The Wind Waker* the avatar of the player can direct the attention to important objects or approaching enemies by looking at them. This can aid the player in puzzle solving.

Jacob [Jac90] presented one of the first papers studying the feasibility of gaze based selective interaction and identified the "Midas touch" problem. Jacob suggested dwell time as a selection mechanism to overcome the problem. Starker and Bolt [SB90] introduced one of the first documented prototypes with computerised real-time eye tracking and intentionally constructed storytelling. It featured a gaze responsive self disclosing display.

Leyba and Malcolm's work [LM04] examined the performance variations of eye tracking versus mouse as an aiming device in a computer gaming environment. A simple 3D game was created in which the task was to remove balls on the screen by aiming at them using a form of "Manual and Gaze Input Cascaded (MAGIC)" pointing. The performance was measured using target accuracy and completion time. The Midas touch problem was overcome by allowing users to point with their eyes, but make selections using the mouse. They found that mouse input was more accurate, but completion time also proved longer using only mouse.



Figure 1: The Tobii T60 eye tracker used in the user trials.

Kenny *et al.* [KKD\*05] used eye tracking for diagnostic purposes. A FPS game was created using mouse and keyboard as input. Using a head-mounted eye tracker the players' gaze was recorded while they played. It was found that the cross hair in a FPS game acts as a natural fixation point. 88% of fixations and 82% of game time took place in the near centre of the screen (the inner 400 x 300 rectangle from the 800 x 600 resolution screen). Jönsson [Jön05] experimented with the FPS *Half Life* and the Shoot-em-up *Sacrifice*. These games were adapted to accept gaze input from an eye tracker. Jönsson showed that in *Sacrific*, in which aim was controlled by mouse or gaze, a higher score could be obtained with eye control. Smith and Graham [SG06] also examined different game genres and proposed to overcome the Midas touch problem using gaze based interaction in conjunction with mouse and keyboard.

Isokoski and Martin [IM06] and Isokoski *et al.* [IHKM07] developed a FPS style game which decoupled aiming from viewing. The game used gaze for aim, mouse to control the camera and the keyboard to move the player around the scene. Dorr *et al.* [DBMB07] adapted the game *Breakout*, to allow input from mouse or gaze. The game objective was to dissolve bricks using a paddle which could be moved horizontally to control a ball that hit them. *Breakout* proved to be well suited to gaze control and participants in the study performed better using gaze input. While gaze data cannot compete with the accuracy of a mouse, eye movements are faster.

Gaze-contingent displays attempt to balance the amount of information displayed against the visual information processing capacity of the observer [DCM04]. Hillaire *et al.* [HLCC08] developed an algorithm to simulate depth-of-field blur for first-person navigation in virtual environments. Using an eye-tracking system, they analysed users' focus point during navigation in order to set the parameters of the algorithm. The results achieved suggest that the blur effects could improve the sense of realism experienced by players. O'Sullivan and Dingliana [OD01] and O'Sullivan *et al.* [ODH02] take another approach, instead of resolution degradation of peripheral objects, collision handling outside

the foveal region of interest was degraded. An extensive overview of gaze controlled games is given by Isokoski *et al.* [JSM09]. Isokoski *et al.* also discuss game genre implications and challenges for gaze input. For example, eye trackers are not competitive with the sub-pixel positioning accuracy of modern mice.

## 2.3. Voice Recognition in Gaming

There are two different categories of speech recognition technologies: *speaker dependent* and *speaker independent* [MZG04]. Speaker-dependent requires each user to go through a process of training the engine to recognise his/her voice. Speech independent recognition avoids this by training with a collection of speakers in the development phase. The game used in this paper uses a speaker independent technology. The process of speech recognition can be divided into the following steps [Lar02]:

**Recognition grammar:** specifies and defines the speech input and its pattern to the speech recognition engine.

**Phoneme identification:** incoming audio signals are analysed and compared to language phonemes.

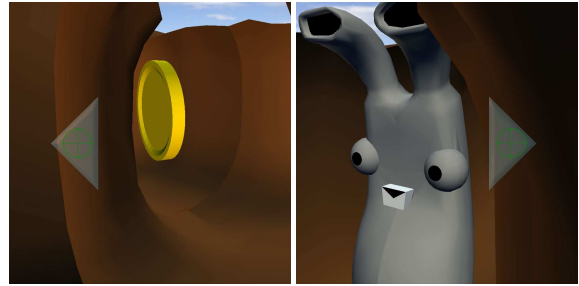
**Word identification** the resulting output of phonemes are compared to words in the recognition grammar.

**Output:** the output produced is the best guess the engine can construct from the user's input.

Mehdi *et al.* [MZG04] used natural language speech recognition to instruct virtual characters in a 3D environment. Speech synthesis was also used to respond to the user. Härmäläinen *et al.* [HMPPA04] developed a musical edutainment game in which the game character was controlled using pitch. The player had to sing in tune to move a character up and down a path to its destination. Sometimes the pitch detection reacted to the background music when the user was not singing. Acoustic echo cancellation methods were suggested as a possible solution to remove the music from the input signal. This could be used in other games where voice recognition is used as an input mechanism. This would allow game music to be played creating a more normal game environment. However, background noise would still be an issue. An extensive overview of video games which incorporate speech recognition is given in [Spe09].

## 2.4. Gaze and Voice in Gaming

Wilcox *et al.* [WEP\*08] created the first and only other game which used gaze based interaction and voice control. The game, a 3rd person adventure puzzle, could be controlled by both gaze and voice and by gaze alone. In the second modality blinks and winking were also used to activate commands. The work included some interesting features which utilised the characteristics of both gaze and voice input. For example a time lag in selecting items was used, which allowed time for voice commands to be recognised and processed. Unfortunately the work did not involve a user evaluation so it is difficult to judge the benefits or shortfalls of their approach.



**Figure 2:** Example images from the Rabbit Run game: (left) coins to be collected and (right) rabbits to be shot.

## 3. Game Design

This paper explores if gaze and voice controlled game interaction is a viable alternative to mouse and keyboard. To study this a game evaluation framework was needed that was controllable by input from gaze and voice, as well as mouse and keyboard. This framework is evaluated both using quantitative measures and subjective responses from participant user trials described further in Section 4.

The game needed to be relatively simple so users could understand it quickly and finish within a reasonable time frame. It also needed to include common gaming tasks, such as navigation and object selection. The premise decided upon was *Rabbit Run*. The player is trapped in a rabbit warren, inhabited by evil rabbits, from which he/she must escape. Example images from the game are shown in Figure 2. The main objective is to navigate through the warren maze and find the exit in the shortest time possible. To earn extra points coins distributed throughout the maze could be collected. In order to pass by the evil rabbits they needed to be shot. Once the exit was reached the game ended. A map would also be provided (upon a key press or voice command depending on the input) in order to assist players finding their way through the maze. The game was developed in the first-person perspective since this is how we view the world.

### 3.1. Game Evaluation Framework

The game evaluation framework was developed using Microsoft XNA. The framework allowed the game to be controlled by mouse/keyboard (MK), gaze/voice (GV), mouse/voice (MV), and gaze/keyboard (GK). Ultimately only MK and GV were tested in the user evaluation. A menu system was also provided controllable by MK and GV. All relevant game data such as shots fired and coins collected were stored in XML format once a game trial was completed.

A map was provided in order to assist the players navigate the warren. The map was generated on the fly showing the places where the player had been in the warren. There was a subtle difference between its implementation using MK

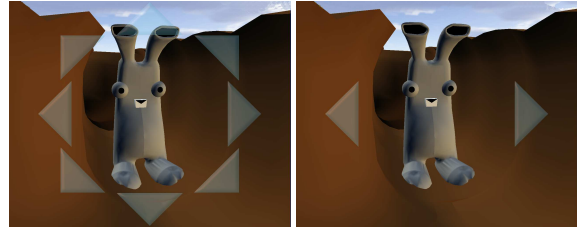
versus GV. For MK the map was only updated with the coordinates of where the player currently was. So if the player travelled to a new location that location would be revealed on the map. A novel game feature was used for GV input. In this case locations were revealed based on where the player had looked. So if the player looked at a particular location that area would be shown on the map without requiring the player to physically move to those positions. This novel game feature could be useful in puzzle based games where the player needs to memorise parts of the virtual environment or objects seen.

Calibration and processing of gaze data from the Tobii T60 eye tracker was accomplished using the Tobii SDK. Upon start of the application all previous calibration data was cleared and a new calibration was initiated. Once calibrated real-time gaze data was relayed to the application. The gaze data includes the area being looked at on screen by both the left and right eyes and the distance of both eyes from the screen. This information was averaged to give the most likely POR.

Given that the game was implemented in the first person perspective, the cameras in the game played a dual role of showing the game play and acting as the player's avatar. The position of the camera was checked for collisions with the surrounding game objects. The camera implementation differed between the GV and MK input types. The MK camera works like most FPS cameras where the target is at the centre of the screen and the mouse movement is used to shift the camera in a given direction. For targeting rabbits to shoot a cross hairs was rendered at the current position of the mouse, which is almost always at the centre of the screen.

The GV camera builds upon the idea of Castellina and Corno [CC08]. They used semitransparent buttons to rotate the camera and move the avatar. As shown by Kenny *et al.* [KKD\*05] the vast majority of game time in a FPS games is spent looking in the inner 400 x 300 rectangle of a 800 x 600 resolution screen. Our implementation is based on this idea by using the outer rectangle of the screen to place the semitransparent gaze activated buttons. The inner rectangle could be left alone to allow normal game interaction. By placing the buttons in this outer area it was hoped that they would not interfere with game play. The buttons would not be displayed unless activated, by looking in that area of the screen, to avoid distracting the player. The purpose of the buttons was to rotate the camera in a given direction. By looking at the left and right part of the screen the camera would shift left and right respectively. The original idea had been to place the buttons in such a way as to form an eight-sided star, as shown in Figure 3 (left).

An early user test showed that the star gaze camera buttons were difficult to use. The camera rotated when the user did not want it to, causing it to spin in a disorientating manner. This could be due to the narrow screen of the integrated eye tracker monitor used. Perhaps if a wider screen was used



**Figure 3:** Original eight-sided star gaze camera is shown on the left. This was abandoned in favour of a simpler gaze camera with only left and right arrows.

this issue may not have occurred. After feedback this was simplified to only use the left and right arrows, as shown in Figure 3 (right). The buttons act as a visual aid to the player indicating the direction the camera was shifting. For targeting rabbits a cross hairs was displayed using the current POR as screen coordinates. This separated the targeting from the camera view much in the same way as Jönsson [Jön05] did in her *Half Life* demo.

Navigation for MK in the game environment was implemented using the arrow keys. Players could move forwards, backwards, left, or right relative to the direction the camera was pointing. This updated the position at a “walking” pace. If players wanted to increase their speed to a “run” pace they needed to hold down the shift key. Navigation with GV was achieved using three commands “Walk”, “Run” and “Stop”. When the “Walk” command was issued the camera proceeded to move, at a walking pace, in the direction the camera was facing until it encountered an obstacle, such as a wall, a rabbit, or the “Stop” command was issued. The “Run” command worked in a similar way except at a faster pace. Microsoft Speech SDK 5.1 was used to implement voice recognition. A few problems were encountered with the voice recognition in an early user test. More intuitive voice commands were not always recognised so instead more distinct voice commands were chosen. For example, instead of saying “Map” to bring up the game map, “Maze” had to be used instead. When selecting menu items “Select” also proved to be inconsistent so the command “Option” was used instead.

#### 4. User Evaluation

A user evaluation was performed to evaluate how suitable GV is as a means of video game control compared to MK. This meant that each participants would need to play the game using both modes of interaction. One of the main objectives of the user study was to gather both quantitative measures and subjective comments. Because the game could be controllable by MK as well as GV it facilitated direct comparisons between the two methods of interaction. In addition to saved game data questionnaires were given to the participants to



ascertain subjective data, such as how effective GV was perceived to be and how immersive or entertaining the experience was.

#### 4.1. Stimulus

The game was designed to bear relevance to a real game, while being controlled enough to allow for analysis. Each game needed to be played twice in order to yield comparable results. It was decided that the exact same layout would be used in each trial by only swapping the start and exit points. The different start and end points would eliminate the possibility of skewed results from learning. The exact same number of coins and rabbits were used, distributed in the same positions in each setup. Figure 4 shows the two different game layouts that were used in an alternated order. The game had three main objectives: (1) to find the exit in the fastest time possible, (2) to collect coins, and (3) to shoot rabbits. These objectives were chosen to encourage the player to perform various tasks.

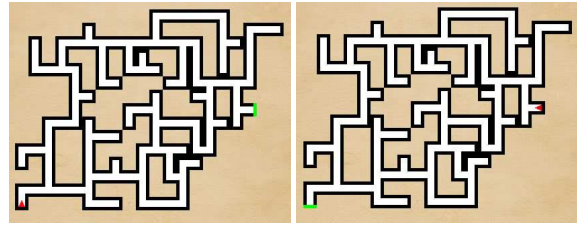
#### 4.2. Equipment and Setup

The Tobii binocular T60 eye tracker, as shown in Figure 1, was placed at a distance of 60 cm away from the user. The T60 is a portable stand-alone unit, which puts no restraints on the user. The freedom of head movement is 44 x 22 x 30 cm. The eye-tracker is built into a 17 inch TFT, 1280 x 1024 pixels monitor. The eye-tracker has an accuracy of 0.5 degrees and has a data rate of 60 Hz. This degree of accuracy corresponds to an error of about 0.5 cm between the measured and actual gaze point (at a 60 cm distance between the user and the screen). An adjustable chair was provided to allow participants to make themselves comfortable and place themselves at the correct distance from the monitor.

The user trial took place in a sound proof video conferencing room. This was to avoid any interference background noise might have on voice recognition. The hardware setup consisted of a laptop running the application while connected to the T60 eye tracker via an Ethernet connection. A keyboard, a mouse, and a microphone headset were also connected to the host laptop to allow for keyboard, mouse and voice input. The lighting was dimmed throughout the experiment. A calibration was carried out for each participant, prior to each trial, to ensure that the collected data would be reliable.

#### 4.3. Participants

Participants were sought by recruiting volunteering post-graduates and staff in the college. Ten participants (8 men and 2 women, age range 22-33) with normal or corrected to normal vision were recruited for the user evaluation. As the eye tracker occasionally loses calibration during a trial, especially for participants wearing glasses, trials which clearly



**Figure 4:** Different warren setups used in each user trial. The red arrow indicates the player position and direction while the green bar shows where the exit is located. In these examples the entire map has been revealed.

produced unreliable data were removed. In total eight users successfully completed their trials. The participants had a variety of experience with computer games. Each participant played the game using both GV and MK. Half the participants played the GV game first, while the other half played the MK first.

#### 4.4. Procedure

Upon arrival each participant were asked to fill in a consent form and answer a short background questionnaire about age, profession, and their gaming habits. After this the participants read a sheet of instructions on the procedure of the particular game they would play. It was decided that an instruction pamphlet would be used to inform all users in a consistent way. The participants were first asked to play a demo version of the game using the selected interaction method. They were allowed to play the demo version of the game for as long as they wanted. They were also encouraged to ask any questions they might have at this stage. Once satisfied with the controls participants were asked to complete the full user trial version of the game. Immediately after playing each trial, participants were asked to answer a second questionnaire. This questionnaire aimed to get the participants subjective opinions on that particular input. After the second and final game was played (and the post-trial questionnaire was completed) a third and final questionnaire was given to the participants. This final questionnaire aimed to compare the two different games played by the participant to gauge which one they preferred. An open question at the end invited comments from the participants.

#### 5. Results

This section examines the results gathered in the user trial both from the saved game data and the user questionnaires. It is not completely straightforward to quantify how well a player did in the game. Different quantitative and subjective measures based on the game and its three main objectives, discussed in Section 4.1, are described below. Related *t*-tests

Measure	MK (mean)	GV (mean)	Related t-test (two tailed)
Distance travelled (squares)	122.5	73.25	(T(df = 7) = 2.556, p < 0.05)
Time (seconds)	228.47	418.86	(T(df = 7) = 4.683, p < 0.01)
Speed (distance/time)	0.583	0.176	(T(df = 7) = 7.4961, p < 0.001)
Coins collected (#)	9.125	5.625	(T(df = 7) = 3.004, p < 0.02)
Rabbits shot (#)	10	6.625	(T(df = 7) = 2.791, p < 0.05)
Shooting accuracy (%)	72.6	64	<b>(T(df = 7) = 1.034, p &gt; 0.2)</b>
Map references (#)	10.1	9.8	<b>(T(df = 7) = 0.08, p &gt; 0.1)</b>

**Table 1:** Statistics for each quantitative measure. Values in bold indicate no significant differences.

Measure	MK (mean)	GV (mean)	Wilcoxon Signed-Ranks test (two tailed)
How fast did you play the game? (1 slow - 7 fast)	6	3	(T = 1, N = 8, 0.01 < p < 0.02)
How well do you think you played? (1 poor - 7 good)	5.5	3	(T = 0, N = 8, 0.01 < p < 0.01)
How difficult was it to shoot rabbits? (1 difficult - 7 easy)	6	6	<b>(T = 11, N = 7, p &gt; 0.2)</b>
How much control did you feel you had? (1 little - 7 lots)	6	4	(T = 0, N = 6, p < 0.001)
How precise did the game react to your controls? (1 imprecise - 7 precise)	6	3.5	(T = 2, N = 8, 0.02 < p < 0.05)
How did you find moving through the game? (1 difficult - 7 easy)	6	4	(T = 0, N = 7, p < 0.001)
How difficult was it to collect coins? (1 difficult - 7 easy)	6.5	4.5	(T = 0, N = 8, p < 0.001)
How did you find menu navigation? (1 difficult - 7 easy)	7	6	(T = 0, N = 6, p < 0.001)
How difficult did you find the game? (1 difficult - 7 easy)	6	4	(T = 1.5, N = 7, 0.02 < p < 0.05)
How much effort was it to play? (1 no effort - 7 lots of effort)	3	5	(T = 0, N = 7, p < 0.001)
How natural did the controls feel? (1 unnatural - 7 natural)	6	4.5	(T = 0, N = 6, p < 0.001)
How immersive was the game? (1 not immersive - 7 immersive)	4.5	6	(T = 2.5, N = 8, p < 0.05)
How much did you enjoy the game? (1 not at all - 7 very enjoyable)	4.5	4	<b>(T = 3.5, N = 6, p &gt; 0.2)</b>
How useful was the map in completing the game? (1 useless - 7 useful)	6	6	<b>(T = 9, N = 7, p &gt; 0.2)</b>

**Table 2:** Statistics for each subjective measure. Values in bold indicate no significant differences. Note that the mean score is higher for GV than MK in the immersive ratings.

were used to see if there were statistically significant differences between the quantitative performance measures in the two games. The results of this statistical analysis can be seen in Table 1. Wilcoxon Signed-Ranks tests were used to determine any significant difference in the subjective ratings. The statistical outcome of these tests are shown in Table 2.

### 5.1. Performance

There was a significant difference in the distance travelled using MK versus GV. Each square in the rabbit warren represents one distance unit. The shortest correct path through the warren corresponds to 33 squares. The obtained statistical result indicates that participants travelled further on average when playing with MK. The difference in time taken to finish the game was also significant indicating that participants finished quicker using MK. Since a time limit of eight minutes was imposed the speed (distance covered divided by time taken) was also measured to see if players covered less distance using GV. The difference was statistically significant and the speed of the participants was on average 3.3 times faster playing using MK.

The difference between rabbits shot and coins collected was also shown to be statistically significant. In both cases participants performed better when playing with MK. The Wilcoxon Signed-Ranks test showed that the participants perceived that they played faster using MK as opposed to when playing with GV. They also reported that they played better overall using MK. The statistical results indicates that the participants performed worse using GV interaction across all measures.

### 5.2. Accuracy and Control

To gauge shooting accuracy the number of shots taken was measured against the number of rabbits killed. There was no statistically significant difference between shooting accuracy using MK versus using GV. Again this is backed up by a two-tailed Wilcoxon Signed-Ranks test which showed there was no statistically significant difference between how difficult it was perceived to shoot rabbits using MK or with GV. This indicates that GV can be an alternative to MK when it comes to shooting enemies. However, further tests are needed to identify the effect of target object size on shooting accuracy. The Wilcoxon Signed-Ranks tests also showed that participants felt MK offered significantly more control and that the game reacted more precisely to the controls than when using GV.

### 5.3. Game Navigation and Difficulty

Navigation was also perceived to be significantly easier when using MK than with GV. Coin collection can also be thought of as an indirect measure of how easy it was to navigate through the game. Coins were all placed in small corner areas of the maze where the navigation would be most testing. Coin collection was found to be easier using MK than when using GV. This result could be explained by the collision detection issues discussed further in Section 5.6. Menu navigation using MK was also statistically ranked easier than menu navigation with GV. Participants also found the game to be less effort to play using MK as opposed to with GV. Overall the MK was also ranked easier as opposed to playing with GV.

## 5.4. Game Naturalness and Enjoyment

While participants thought using MK was more natural than using GV, GV was ranked as being more immersive than MK, as shown in Table 2. This result is interesting showing that playing with GV can increase how immersive the player feels. The game was also evaluated based on how much the player enjoyed the game. A two-tailed Wilcoxon Signed-Ranks test showed no statistically significant difference between MK and GV in terms of how much they enjoyed the game. However, when asked which interaction type was more enjoyable to use, 75% of participants selected GV as the more enjoyable.

## 5.5. Game Map Usefulness

One of the novel features in the GV game interaction was the map generator. This map displayed all areas that the player had seen as opposed to have visited. The participants were not informed about this difference between the two games to avoid influencing them on which one they found more useful for game completion. To ascertain the worth of this feature participants were asked to rank how useful they found the map. The number of the times the map was referenced throughout the game was also recorded. As shown in Table 1, the difference in the number map references using MK versus using GV showed that the difference was not statistically significant. Participants' perception of how useful they found the map was also tested for significance. It showed that participants did not perceive the map to be any more useful in completing the game using GV than using MK.

## 5.6. Discussion

From the results it is clear that MK performed better than GV. Apart from the fact that people are more used to MK the primary problem seems to have been difficulty in navigation. The collision response appears to be the main problem in this regard. The camera was stopped moving if any collisions occurred (between the camera and the maze walls or rabbits) to prevent it from interpenetrating walls. For MK this worked well since users quickly tried the other direction keys if they felt they could not move out of a corner. However, when using GV users issue commands and moved until a collision occurred (or until the "Stop" command was issued). In order to navigate out of a corner the player needed to shift the gaze camera in the opposite direction and re-issue the "Walk" command. This led some players to continually hit off walls and get a sense that they were trapped. One participant said "I felt I got trapped sometimes near the walls" and another "I got stuck a lot which was easier to get out of using the keyboard rather than the gaze". A more intuitive response could have been to make players smoothly travel along the wall rather than stop entirely. Future work involves making such an improvement in the collision response. Other problems noted by users included the delay

between issuing a voice command and the game reacting to this command. Comments obtained from the participants included: "it took a while to get used to since the responsiveness from the voice input was a bit slower than mouse and keyboard" and "using the voice commands felt slower than pressing a key". This is one of the facets of voice recognition and it is difficult to see how this could have been avoided.

The participants found the gaze aiming helpful and one commented that "the gaze worked well at targeting the rabbits to be shot". However, one participant reported that the cross hair slightly jumped around the screen. A smoothing filter could possibly have been applied to the gaze data to make it more stable. Another participant commented favourably on the gaze camera saying "I really liked the rotation movement (left/right) of the gaze camera. It felt natural to look left when I wanted to go left, was nice, could be useful in games". Despite the negative effect these problems had on players ability to navigate around the game easily, 75% of the participants reported GV to be the more enjoyable form of interaction. Table 2 also shows that participants felt more immersed playing with GV as opposed to MK.

## 6. Conclusions and Future Work

This paper presents a game which can be controlled using gaze, voice, mouse and keyboard. A game evaluation framework was built in order to evaluate these input modalities. This is the first evaluation study performed evaluating MK versus GV. The framework was implemented as a scalable tool and having been designed in an object orientated way it could easily be used with other games and input types. The complete game was implemented, but not without some issues. The greatest problem with the implementation was in the collision response. It is difficult to say whether or not GV could compete with MK as an input modality with the results achieved. Although previous research [LM04] has reported similarly performance results when comparing gaze against common input mechanisms, GV would have assuredly done better had these collision issues been resolved in a more satisfactory manner.

However, some promising results were achieved. First of all, although the game performance was significantly worse, the game could be played entirely hands-free. This does present an opportunity for disabled users for whom traditional interaction techniques are not sufficient. The GV option was also selected by most participants as the more immersive form of interaction in the user evaluation. This is a trend of studies involving gaze based interaction in video games. Participants in previous trials have reported similar positive feelings towards gaze interaction [Jön05, SG06]. Perhaps it is only a novelty factor but it does show that people are interested in alternative forms of computer interaction. Most previous studies looking at gaze interaction have adapted open source games to work with gaze data [Jön05, SG06, DBMB07]. There are a few exceptions

who created game scenarios to compare different input types [IM06,CC08]. Although adapting open source games reduces the implementation time required, these games were originally created with mouse and keyboard in mind. So the adapted game is restricted to using gaze as a replacement for the mouse rather than an input device in its own right. The gaze input acting only as a *mouse emulator*. When a game is developed from scratch it should be free of this restriction. Unfortunately this was not the case in this study. Perhaps the goal of comparing gaze and voice directly against mouse and keyboard unwittingly manifested this idea on the project. A further study could exploit the way we use gaze and voice in the real world to create novel input modalities for video games. Rather than comparing directly against mouse and keyboard the game experience itself could be measured using a game experience questionnaire [IdKP08]. There were no sounds used in the game and the user evaluation took place in a sound proof room. This was to counteract any ill effects background noise may have had by creating false positive recognitions. This is not an ideal gaming scenario. Further work could look at the effects of background noise on voice recognition in games. Another area worth investigating is how the animation and AI of game characters could be adapted to react to the player's gaze and voice. More socially realistic scenarios could be created if game characters responded in ways appropriate to the tone of the player's voice and/or the focus of their gaze. It would also be interesting to evaluate how the user's head direction, obtained using Natural Point's TrackIR system, could be used with gaze to reduce issues related to the camera motion.

## 7. Acknowledgements

The authors would like to thank Acuity Ets Limited. for providing the loan of an eye tracker. We would also like to thank Paul Masterson for his help in all matters hardware related and all participants that took part in the experiments.

## References

- [CC08] CASTELLINA E., CORNO F.: Multimodal Gaze Interaction in 3D Virtual Environments. In *Proceedings of the 4th COGAIN Annual Conference on Communication by Gaze Interaction, Environment and Mobility Control by Gaze* (2008).
- [DBMB07] DORR M., BÖHME M., MARTINETZ T., BRATH E.: Gaze beats mouse: a case study. In *Proceedings of COGAIN* (2007), pp. 16–19.
- [DCM04] DUCHOWSKI A., COURNIA N., MURPHY H.: Gaze-Contingent Displays: A Review. In *CyberPsychology & Behavior* (2004), pp. 621–634.
- [Duc07] DUCHOWSKI A.: *Eye Tracking Methodology, Theory and Practice*, second ed. Springer, 2007.
- [HLCC08] HILLAIRE S., LÉCUYER A., COZOT R., CASIEZ G.: Depth-of-Field Blur Effects for First-Person Navigation in Virtual Environments. *IEEE Computer Graphics and Applications* 28, 6 (2008), 47–55.
- [HMPPA04] HÄMÄLÄINEN P., MÄKI-PATOLA T., PULKKI V., AIRAS M.: Musical computer games played by singing. In *Proc. 7th Int. Conf. on Digital Audio Effects (DAFx'04), Naples* (2004).
- [IdKP08] IJSELSTEIJN W., DE KORT Y., POELS K.: The Game Experience Questionnaire: Development of a self-report measure to assess the psychological impact of digital games. *Manuscript in preparation* (2008).
- [IHKM07] ISOKOSKI P., HYRSKYKARI A., KOTKALUOTO S., MARTIN B.: Gamepad and Eye Tracker Input in FPS Games: data for the first 50 min. In *Proceedings of COGAIN* (2007), pp. 10–15.
- [IJSM09] ISOKOSKI P., JOOS M., SPAKOV O., MARTIN B.: Gaze controlled games. *Universal Access in the Information Society* (2009).
- [IM06] ISOKOSKI P., MARTIN B.: Eye Tracker Input in First Person Shooter Games. In *Proceedings of the 2nd COGAIN Annual Conference on Communication by Gaze Interaction: Gazing into the Future* (2006), pp. 78–81.
- [Jac90] JACOB R.: What You Look At Is What You Get: Eye Movement-Based Interaction Techniques. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (1990).
- [Jön05] JÖNSSON E.: *If Looks Could Kill - An Evaluation of Eye Tracking in Computer Games*. Master's thesis, KTH Royal Institute of Technology, Sweden, 2005.
- [KKD\*05] KENNY A., KOESLING H., DELANEY D., MCLOONE S., WARD T.: A Preliminary Investigation into Eye Gaze Data in a First Person Shooter Game. In *Proceedings of the 19th European Conference on Modelling and Simulation* (2005).
- [Lar02] LARSON J.: *VoiceXML: Introduction to Developing Speech Applications*. Prentice Hall PTR Upper Saddle River, NJ, USA, 2002.
- [LM04] LEYBA J., MALCOLM J.: Eye Tracking as an Aiming Device in a Computer Game. In *Course work (CPSC 412/612 Eye Tracking Methodology and Applications by A.Duchowski)* (2004), Clemson University.
- [MZG04] MEHDI Q., ZENG X., GOUGH N.: An interactive speech interface for virtual characters in dynamic environments.
- [OD01] O'SULLIVAN C., DINGLIANA J.: Collisions and Perception. *ACM Trans. Graph.* 20, 3 (2001), 151–168.
- [ODH02] O'SULLIVAN C., DINGLIANA J., HOWLETT S.: Eye-movements and Interactive Graphics. *The Mind's Eyes: Cognitive and Applied Aspects of Eye Movement Research* (2002).
- [San07] SANCHANTA M.: Nintendo's wii takes console lead. *Financial Times* (2007).
- [SB90] STARKER I., BOLT R.: A Gaze-Responsive Self-Disclosing Display. In *CHI '90: Proceedings of the SIGCHI conference on Human factors in computing systems* (New York, NY, USA, 1990), ACM, pp. 3–10.
- [SG06] SMITH J. D., GRAHAM T. C. N.: Use of Eye Movements for Video Game Control. In *Proceedings of the 2006 ACM SIGCHI International Conference on Advancements in Computer Entertainment Technology* (2006).
- [Spe09] SPEECHTECHMAG: Speech Technologies Make Video Games Complete. Accessed 28 August 2009 at <http://www.speechtechmag.com/>, 2009.
- [WEP\*08] WILCOX T., EVANS M., PEARCE C., POLLARD N., SUNDSTEDT V.: Gaze and Voice Based Game Interaction: The Revenge of the Killer Penguins. In *Proceedings of International Conference on Computer Graphics and Interactive Techniques* (2008).