

**2017 Master Thesis**

**VPresentationHelper – A presentation practice  
application in VR**

Date: July 27 2017  
Supervisor: Hayato Yamana

Department of Computer Science and Communications Engineering,  
Graduate School of Fundamental Science and Engineering,  
Waseda University  
Student ID: 5115FG17-1

**Yixuan He**

# Abstract

In this paper we propose a presentation rehearsal application in Virtual Reality (VR) that could be used to help users practice, improve their presentation skills, and overcome stage fright. Good presentations can be defined on three criteria including “Verbal,” “Vocal,” and “Visual”: Verbal represents words presenters choose; Vocal includes tone of voice, pitch range, and loudness; Visual includes postures, gestures, and eye contacts. Here in this paper, our goal is to practice “Visual” part in Virtual Reality. By using our proposed application, users can import their own materials and thus they can see the slides during their speech in virtual environment. When they finish the rehearsal, a score will be given as a feedback. The whole presentation is evaluated by the user’s stage usage and interaction with audience, such as eye contacts. Additionally, an eye tracking heat map will also be generated. It will allow users to see where and how they interact within the virtual environment. Recently, in Virtual Reality Applications, there are several domains show great potentials: Video Games, Tourism, Education, etc. Especially, Interactive Learning in VR gradually becomes more useful and approachable compared to traditional education ways. It is proved to be more effective to enable users to get more involved in the learning process. In this research, our goal is not only to provide the user a place to practice presentation despite the limitations of time and space, but also to help them overcome the performance anxiety and improve the quality. To the best of our knowledge, our paper is the first to research on the “Visual” impact of presentation in Virtual Reality.

Three types of experiments and evaluations are conducted and we have verified that: 1) PhD students present better than bachelor students, 2) the familiarity of the slides influences “Visual” performance, and 3) many users do not pay much attention to “Visual” aspects when they present.



# Contents

1. Introduction
  2. Background
    - 2.1 Virtual Displays
      - 2.1.1 Head-Mounted Displays (HMD)
      - 2.1.2 World-Fixed Displays
    - 2.2 HTC Vive
      - 2.2.1 Lighthouse
      - 2.2.2 Chaperone
  3. Related Work
  4. Objectives
  5. Implementation
    - 5.1 Overview
    - 5.2 Presentation Slides
    - 5.3 Audience Design
    - 5.4 Data Collection
      - 5.4.1 Position Data, Head Rotation data, Time data
      - 5.4.2 Gaze Data
      - 5.4.3 Interaction Data
    - 5.5 Heat Map
      - 5.5.1 Heat Map of Position Data
      - 5.5.2 Heat Map of Gaze Data
    - 5.6 Evaluation of the presentation
  6. Results and Discussion
  7. Evaluation
  8. Conclusion and Future Work
- Acknowledgements
- References
- Appendix
- Publications

# 1. Introduction

Presentations gradually become common and important in our everyday life. Besides the context and layout of the materials, nonverbal communication, such as volume and speed of voice, eye contacts, stage usage, are key elements of a successful presentation. However, millions of people suffer from performance anxiety, commonly called “stage fright”<sup>1</sup>, which causes trembling hands and voice, lacking eye contacts and interactions. One of the best ways to deal with stage fright is to practice more. However, practice is limited to the space and time. Because of lacking of proper places and time to practice, people are seeking to find a solution using technology. Unable to simulate nervousness and presence, traditional platforms, such as computers and phones, failed to give a satisfying solution. With the ability to provide an immersion, virtual reality is therefore being expected to solve this kind of problems<sup>3</sup>. Virtual reality gives a solution, which works well to provide users a presence. And at the same time, unlike other<sup>1</sup> ways, practicing in VR is not dependent on people and scheduling, place availability and costs.

According to the research, there are mainly three aspects when defining a good presentation: “Verbal,” “Vocal,” and “Visual”<sup>4</sup>. Verbal represents words people choose, vocabulary, grammar, etc. Vocal includes tone of voice, pitch range, loudness, etc. Visual includes postures, gestures, eye contacts, etc. Researchers also identify communication as 55% Visual, 38% Vocal, and 7% Verbal.<sup>5</sup> It indicates that when we make speaking, the words we choose only have a small impact on the presentation quality, the way we convey the message matters much more. Especially the visual effect, for example how we look when we say it, is a dominant element for a good presentation.

In previous research, there existed many research focusing on “Verbal”, using natural language processing to analyze the context of the speech, using image processing to analyze the layout, color, and quantity of the contents on the presentation materials [1]. Some of them are focusing on “Vocal”. Researchers do audio analysis to categorize utterance duration, pitch, filled pause, and so on by using the input from headset microphone. They also build a real-time alarm system as a feedback to the user [2]. Due to the difficulty of collecting data, such as stage usage data, eye tracking data, there is a few research focusing on “Visual.” In [3], they integrate signals obtained from variety of sources, such as the gestures and voice. Facial expressions are collected as videos by using cameras. The data are analyzed using image processing.

In our research, HTC Vive head-mounted headset is used. Lighthouse, which is HTC Vive’s tracking system, is used to collect position data. For eye tracking data, there existed several ways: 1) using eye-tracker glasses, 2) using eye-tracker for VR, and 3) estimation. However, glasses cannot be fit into VR, while eye-tracker (SMI) for VR is very expensive. There is also one specialized VR, FOVE2, which is said to be the world’s first eye tracking VR headset. However, it is still under developing. Therefore, we will collect eye-tracking data by using an estimation method, which we will introduce in

---

<sup>1</sup><http://www.webmd.com/anxiety-panic/guide/stage-fright-performance-anxiety#2>

<sup>3</sup> <http://www.vrphobia.com/therapy.htm>

<sup>4</sup> <http://successfully-speaking.com/blog/the-three-vs-of-communication>

### Chapter 3.

In our research, we are mainly focusing on Visual. We evaluate the presentation based on the usage of stage, and eye contacts, and how well the presenter interacts with the audience. To our best knowledge, our proposal is the state-of-the-art approach to build presentation rehearsal application in VR.

In the reminder of this paper, Section 2 describes the background of the research, and in Section 3 we introduce prior works related to virtual reality and evaluation of good presentation. We show a list of the objectives of this research in Section 4. Our implementation is explained in Section 5. Then, we present our results and make some discussions in Section 6. The evaluation results of the whole system are discussed in Section 7. Finally we conclude the paper and talk about our future work in Section 8.

## 2. Background

### 2.1 Virtual Displays

There are two ways that today's virtual reality systems are mainly implemented in: Head-Mounted Displays and World-Fixed Displays. [5]

#### 2.1.1 Head-Mounted Displays (HMD)

The Google Cardboard is a low-cost platform that enables users to adapt any phone to become a Virtual Reality (VR) device. It was created to attract people's interests of developing virtual reality applications. Developers can build their own Cardboard using simple components provided by Google. For normal users, they can also purchase a pre-assembled one. There are also various applications for Cardboard available at Google Play store and iTunes store. Users can then begin their virtual reality experience by simply inserting their own phone into the box.

Samsung Gear VR is another popular smartphone-based VR headset. In collaboration with Samsung, Samsung Gear VR is a lighter-weight mobile version of Oculus. The latest version contains one controller, with which users could use to interact in Virtual environment using their one hand.

Oculus was the first successful attempt to market consumer virtual reality headsets. It is a typical VR Display with built-in head motion tracking sensors, which allows developers to collect and analyze head position and rotation data. In 2017, Oculus Rift has launched their Touch controllers, which enables a movements tracking within a desired play area thanks to the twin sensors built in Touch controllers.

HTC Vive was the first HMD that provides whole-room tracking features, prior to which only sitting and standing in one spot movement is possible to be captured. Being able to physically move within a space can let users better replicate real-world actions, and therefore provides a better immersion experience. Besides, HTC Vive controllers were the first to show consumers a solution to interact in virtual environment. We will make a detailed explanation in section 2.2.



Fig.1 HTC Vive (2016)

The following table gives a general comparison of the above HMDs.

Table.1 Comparison of HMDs

	Resolution per eye	Environment	Whole-Room scale	Sensors
<b>Google Cardboard VR Viewer</b>	Low	Android, iOS, Unity	NA	Accelerometer, Gyro Sensor, Proximity Sensor
<b>Samsung Gear VR</b>	1440 x 1280	Samsung's flagship smartphones	NA	Accelerometer, Gyro Sensor, Proximity Sensor
<b>Oculus Rift</b>	1220 x 1080	PC	5 x 11 feet	Accelerometer, gyroscope, magnetometer, Constellation tracking camera.
<b>HTC Vive</b>	1220 x 1080	PC	15 x 15 feet	Accelerometer, gyroscope, Lighthouse laser tracking system, front-facing camera

### 2.1.2 World-Fixed Displays

Unlike HMD, which is more or less attached to user's head, World-Fixed Displays usually render graphics onto monitors' surfaces. Display surfaces can be planar or curved or any other shapes if well defined. CAVEs and CAVE-like displays shown in Fig.5 are typical examples belonging to this category.

A typical CAVE usually includes: projection walls (monitors), speakers, tracking sensors, etc. We can also obtain the position and rotation for head tracking, however the requirements of accuracy and latency is not critical as HMD, since the stimuli are independent on users' head motion. In many fields, especially in scientific and engineering, the CAVE provides a better solution than HMDs from many perspectives. For example, it allows multiple users to share the virtual environment at the same time easily.

## 2.2 HTC Vive

As the most valuable HMD in today's consumer virtual reality market, HTC Vive has made many innovations and provides many novel operations and functions compared with other HMDs.

### 2.2.1 Lighthouse

Lighthouse is the key technology of whole-room experience in Virtual Reality. It can

track the position and rotation of user's HMD and controllers in real time. It is achieved by flooding non-visible lights between two base stations, and sensors placed on HTC Vive's HMD and controllers. At the setting-up step, two base stations are required to be mounted on shelves or any higher places. This is to ensure the stations can "see" the HMD and controllers. Together with SteamVR, which we will introduce in section 2.2.3, user's position in 3D could be tracked by determining where the lights emitted by base stations are intercepted by receptors of HMD or controllers.

### **2.2.2 Chaperone**

Chaperone system is a safeguard technology. It aims to protect the user from collision with other real objects, such as walls, furniture, etc. by using the front camera of HMD. Once the user gets close to any obstacle, it will recognize and represent the real objects in blue grids in virtual environment. There is also another mode of this operation, which can enable users see an instant silhouette of the surroundings. This mode can be activated by double tapping the menu button on the controllers.

## 3. Related Work

In this section, we refer to some researchers on virtual reality and discuss previous approaches on tracking gaze, input method, and achieving room-scale interactive area. Also, as we have explained before, there are mainly three pillars when determining whether a presentation is good or not: Vocal, Verbal, and Visual. We investigated the research related to creating presentation rehearsal system based on one or several pillars to help user practice and improve public speech skills. We found that although there are some studies evaluating based on “Vocal” and “Verbal”, there are a few works done using “Visual” data.

### 3.2 Virtual Reality

In Tarid’s work [7], they discussed several technologies in a forensic psychiatry context. They present a work in the development of a multimodal brain-computer interface (BCI) combining neuroimaging and eye tracking with VR devices. In their work, SensoMotoric Instrument (SMI) Eye-Tracking Glasses V2.0 is used to track the gaze in the 3D environment. The Volfoni Glasses are synchronized with the VR System’s display using a Volfoni ActiveHub IR100. They do not use any VR device, instead, a modified version of Volfoni Active Eyes Glasses is set on top of the SMI ETG glasses to enable stereoscopic vision.

In [8], Evgeniy showed us an approach aiming to predict gaze data of omnidirectional visual content in virtual reality. Mainly the head angular velocity data is used to complete the analysis. However, in their research, only panoramic pictures (2D) were used. The problem of estimating gaze data in 3D virtual reality is reminded to solve.

In Nikolaos’s work [9], they present a gaze-controlled Multimedia User Interface and developed six applications including mail composing, multimedia viewing for VR headsets. They mentioned that now various companies, such as SensoMotoric Instruments, provide eye-tracking add-on to the Oculus Rift Development Kit 2 (DK2) HMD. Within this, immersive User Interface paradigms embedded in a VR setup controlled via eye tracking can be designed, implemented and evaluated. In their approach, they used twin-CCD binocular eye-tracker by Arrington Research to record the eye-tracking data. Then, to model good 3D Characters that meet their needs, they developed ARViPL Character Designer, which is an in-house Unity 3D plugin. Together with Singular Inversion FaceGen SDK, it can generate full bodies.

In [10], Wanhong designed an input interface in virtual reality by users’ gestures. With the leap motion attached on the Oculus Rift headset, users are able to interact with surrounded objects. They tested the usability and throughput of gesture input using leap motion and provided the optimized features for building such input system in virtual reality.

### 3.3 Presentation

In [11], Kazutaka proposed an automated basic quantitative evaluation of presentation materials. They evaluated by analyzing the quantity of the contents, font size, and how well the font color matches with the background color. The score of font size is given as:

$$\frac{H_m f_p l_{min}}{H_p l f_{min}} \geq 1 \quad (1)$$

where  $H_m$  is the Height/Length of the display,  $H_p$  is the Vertical resolution of the display,  $f_p$  represents the number of pixel of the font,  $l$  represents the distance between audience and display,  $l_{min}$  equals to 0.4m, and  $F_{min}$  equals to 0.003m.

They provide two ways to evaluate how well the font color and background color matches. The first is to compare the Contrast, and the formula is given as follows:

$$0.299 \times R_d + 0.587 \times G_d + 0.114 \times B_d \geq 125 \quad (2)$$

The other is to calculate the color-difference:

$$\begin{aligned} & \max(R_1, R_2) - \min(R_1, R_2) \\ & + \max(G_1, G_2) - \min(G_1, G_2) \\ & + \max(B_1, B_2) - \min(B_1, B_2) \geq 500 \end{aligned} \quad (3)$$

Mathieu [3] compared three feedback strategies to improve public speaking: a non-interactive virtual audience, direct visual feedback, and nonverbal feedback from an interactive virtual audience. They take self-assessment questionnaires, expert assessments, eye contacts, and avoidance of pause fillers into account when evaluate the speaking. And they found that among these three conditions, interactive virtual audience works the best in: (1) increasing engagement and challenges, (2) improving public speaking.

In Anderson work [12], they have implemented a virtual job interview simulation platform for NEETs, representing young people not in employment, education or training. Their work is based on TARDIS project. They collect lots of user's voice and facial expressions during virtual interviews and then design a tool for researchers to design interview and provide coaching for users.

Esin [13] built a comprehensive Virtual Environment of treatment for public speaking anxiety. They have made great contribution on the physical and vocal cues that virtual audiences make. Virtual audiences, with automated animations to the speaker's arrival, delivery, pauses and exit, are created in their system to lower the public speaking anxiety levels of users. However, the correlations between these cues and anxiety, and the interactions between presenter and audience could be further analyzed.

In [14], Trinh present a robotic coach for oral presentation. The coach offers the user feedback from three aspects: speech quality, content coverage, and audience orientation. Speech quality, including pitch, speaking rate, and filler rate, and content coverage of the slides are measured basically by using natural language processing. While for audience orientation, eye contact, Microsoft Kinect is used to track user's head orientation. One of the limitations of this system is that there only exists one audience, which is the robotic coach, during the rehearsal.

As described above, there exist several research working on helping users improve



their speaking based on “Verbal” and “Volume”, but there is only a few research on evaluating “Visual”, eye contacts, interactions, and usage of the stage due to the difficulties of collecting such data. On the other hand, the innovation of VR, which has the ability to provide an immersion, has a great potential.

Table.2 gives a summary of methods used in previous research on evaluating presentation quality.

Table.2 Aspects used in previous work on evaluating presentation

	<b>Verbal</b>	<b>Vocal</b>	<b>Visual</b>
<b>Kazutaka [11]</b>	Quantity of the contents, font Size, color match		
<b>Mathieu [3]</b>		pause fillers	eye contacts
<b>Anderson [12]</b>		voice	facial expressions
<b>Esin [13]</b>			physical and vocal cues of audiences
<b>Trinh [14]</b>	content coverage	speech quality	eye contacts with only one audience

In our paper, we will present a novel approach and an application that could help users practice, evaluate, and improve their presentation based on their “Visual” performance. We will also evaluate “Visual” performance from three perspectives: 1. User’s stage usage, 2. Eye contacts with audience, 3. Interactions with audiences.

## 4. Objectives

The objective of this research is to build a virtual reality application that can help users practice and improve their presentation by giving them practical advice based on their “Visual” performance. According to [15] [16], we can conclude the key elements for “Visual” behavior includes:

1. Stage usage: the position of user, whether the user makes the full use of the whole stage.
2. Interaction: whether the user is able to notice the actions of audience. Actions include sleeping, shaking head, and raising hand.
3. Eyes contacts: whether the user gives enough eyes contacts with the audiences, or is he/she keep looking at the wall, etc.
4. Time: whether the presenter is able to finish their presentation within an expected period of time.

When finish the presentation practice, in real time the system presents the following outputs:

1. Score of the practice: A comprehensive score is given based on the four aspects of “Visual” behavior we mentioned above. The score for each aspect is also calculated using user’s position data, times of interactions with virtual audience, and rehearsal time duration respectively.
2. Advice to improve the presentation: Advice is given according to the obtained four individual scores.

The tasks of this research are as following:

1. Construct different scenes for the application, including menu scene, presentation scene, scores scene.
2. Import user’s own presentation slides into virtual reality. Allow the user see and interact with their slides in virtual environment.
3. Create virtual audience with customized animations. It enables us to control the interactions between presenter and audience.
4. Collect user’s position data and head rotation data.
5. Estimate and collect user’s gaze data.
6. Count the times of interaction between presenter and audience.
7. Generate heat map of position data and gaze data.
8. Evaluate the presentation based on position, rotation, and gaze data.
9. Evaluate the system.

Implementation of task 1 is presented in Section 5.1. Approaches for task 2 and task 3 is discussed in Section 5.2 and Section 5.3. The data collection steps, task 4, 5, and 6, are shown in Section 5.4. Then, task 7, which related to the method we use to evaluate the presentation rehearsal, is explained in Section 5.5. Finally, task 8 about the evaluation of our system will be discussed in Section 5.6.

# 5. Implementation

In this research, our experiments are implemented on HTC Vive. Resolution of the HMD is 2160×1200. Eyesight range is 110°. Game Engine Unity3D 5.5 is used when develop the application.

## 5.1 Overview

We use unity3d to build our system. Fig. 4 shows the structure of virtual classroom that we construct. HTC Vive is used as our VR platform.

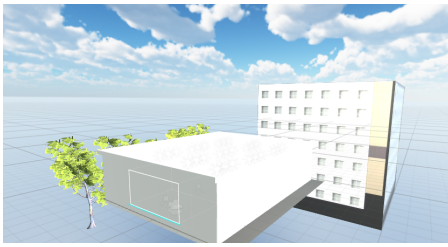


Fig.4 Structure designed in unity3d

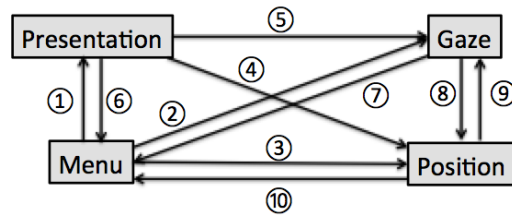


Fig.5 Relations between scenes

Our system consists of four scenes in total: “Menu”, “Presentation”, “Position”, and “Gaze”. “Presentation” scene is the scene where user do the presentation practice. The heat map of user’s position data during practice could be found in “Position” scene while the heat map of gaze data could be found in “Gaze” scene. Fig.5 shows the relationship among these three scenes. In the figure, arrays represent possible sequences of scene-change, while ①~⑩ represent 10 triggers set in four scenes, with which user can easily jump from one scene to another by gaze input. A red cursor is placed at the center of the screen in VR to represent user’s gaze. As described above, gaze-based input is used in our application. We will introduce these triggers more in detail in the following paragraphs.

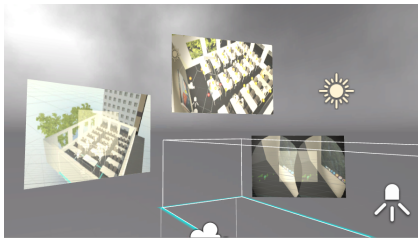


Fig. 6 “Menu” scene captured in unity

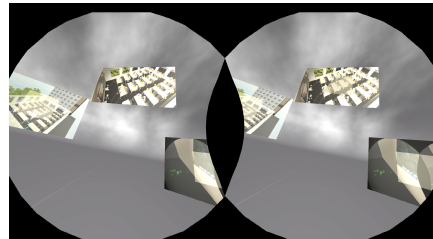


Fig.7 “Menu” scene captured in VR

At the beginning when a user starts the practice, “Menu” scene will be presented, as shown in Fig. 6 and Fig.7. The three planes represent “Presentation”, “gaze”, and “position” respectively. User can select which scene to go by gaze. “Menu” scene contains triggers ①~③, representation starting a presentation practice, checking previous heat map of gaze, checking previous heat map of position respectively.



Fig.8 “Presentation” scene captured in unity3d Fig.9 “Presentation” scene captured in VR

When the “Presentation” scene is chosen, a virtual classroom, as shown in Fig.8 and Fig.9, will be presented. Then when the user presses down the “grip button” on the controller, as shown in Fig.10, the practice will begin and the system will start collecting data for later calculation.



Fig.10 Controllers in Virtual Reality

If the user presses down the “grip button” again, three optional screens will be presented in front of the user, as shown in Fig.11. These screens are triggers ④~⑥, which can lead to “Position”, “Gaze”, and “Menu” scene respectively. At the same time, our system will start calculating a comprehensive score based on user’s duration of speech (time), stage usage (position data), eye contact (gaze data), and interaction with audience.

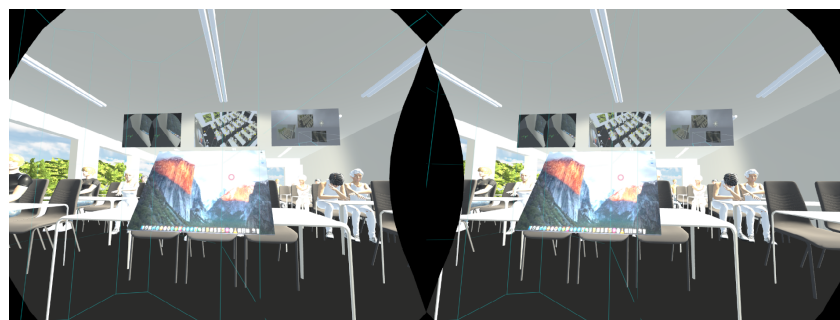


Fig.11 Three optional screens pop up

Fig.12 and Fig.13 are screenshots of “Gaze” and “Position” scene. It is easy to find that different from “Presentation” scene, there are two extra screens above the classroom in these two scenes. These are triggers ⑦~⑩. Right-hand side screen enables the user to switch between “Position” and “Gaze” scene, while left-hand side screen is to quit the practice and go back to “Menu” screen.

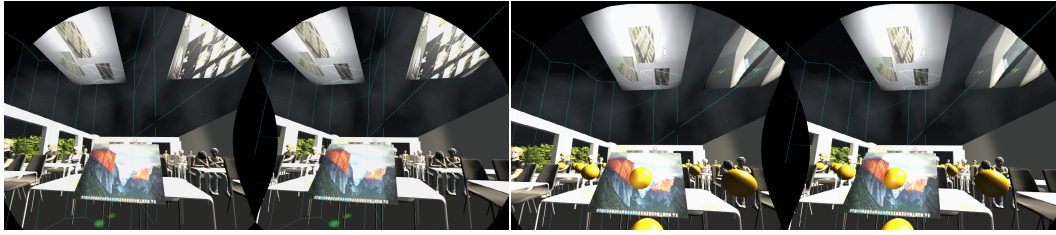


Fig.12 “Position” scene captured in VR

Fig.13 “Gaze” scene captured in VR

## 5.2 Presentation Slides

As shown in Fig.14, there is one small screen and one large screen in virtual environment, which represents PC and projector respectively. Our system enables users to import and use their own slides when practice their presentation in virtual reality. Additionally, to imitate the functions of laser pointer, it is also possible for user to use controllers to switch slides during their presentation practice. User can change slides in virtual environment by simply pushing down the triggers on controllers, as shown in Fig.10.

Steps how we import users’ presentation materials and how we make it possible to be interacted using controllers in VR environment are as following:

1. Create two empty objects, one for projector, the other one for PC.
2. Convert and save all the slides as image files.
3. Attach image files as materials to projector and PC.
4. Use controllers to change the “material” of the objects simultaneously.

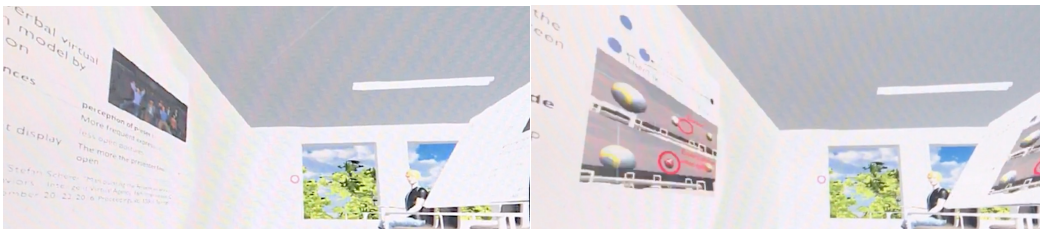


Fig. 14 Presentation slides in Virtual Reality

## 5.3 Audience Design

We have created in total 22 virtual audience as shown in Fig.15. Free assets [17] [18] used when building. In order to collect users’ gaze data, which we will explain more in detail in Section 5.4.2, we add a “sphere collider” to each audience. Therefore, once the presenter is looking at certain audience, the position of that audience can be obtained and collected. In this way user’s gaze can be tracked. In Fig. 16, the glowing green sphere shows an example of the collider.



Fig.15 Virtual Audience

As concluded in [4], the virtual audiences' attentiveness is the most dominating audience characteristic that presenter perceive. Also in [3], the authors have confirmed that virtual audiences with interactions bring better engagement experience for presenter. Additionally, they found that interactive audiences can be helpful for improving speaking skills. Therefore, series of animations that could indicate audiences' attentiveness are also added to the virtual audience. Actions are designed beforehand, with certain timing and duration. We keep only one audience move at a time. In total we have added 10 animations to 10 different virtual audiences.

Three types of actions that we use, and their durations that we set are as following:

1. Sleep, 30s, as shown in Fig.17
2. Raise hand, 15s
3. Shake head, 15s

We give each action different time duration, meaning that the score of "Interaction" will be deduced if the user fails to make a notice within a certain time. A good presenter is supposed to notice all these obvious actions that audiences make within this certain time period. This kind of interactions between the presenter and audience is also one of the standards that we use when we evaluate the presentation practice.

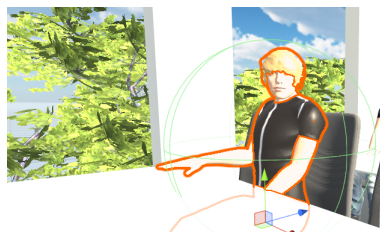


Fig.16 Collider for Virtual Audience



Fig.17 "Sleeping" event

## 5.4 Data Collection

In our system, four kinds of data are collected:

1. User's position data and head Rotation Data
2. Eyes-tracking Data
3. Time duration of the whole presentation
4. Times of interaction

All the data are collected using Unity Scripting API. Unity provides diverse categories of API. In this research, namespace “UnityEngine.VR” is mostly used. The following table shows the main functions that we use to obtain data.

Table.3 Main functions that we use in this section

Function	Description
<b>GetLocalPosition</b>	Gets the position of a specific node.
<b>GetLocalRotation</b>	Gets the rotation of a specific node.
<b>Transform.eulerAngles</b>	The rotation as Euler angles in degrees.
<b>timeSinceLevelLoad</b>	The time this frame has started.
<b>gameObject.name</b>	The name that the GameObject

All the data are saved into csv file for further calculation and analysis.

In the following subsections, we will introduce how we collect these data.

#### 5.4.1 Position Data, Head Rotation data, Time data

Lighthouse is used when we collect user’s position, rotation data in VR. Lighthouse is a position tracking system developed for HTC Vive and SteamVR. It tracks the position, head rotation, and controllers accurately.

For time, we collect two types of time data:

1. Duration of presentation
2. Time stamp for each event

Types of events are listed in Table.4. User’s position data is all the locations that presenter stands at, and Head Rotation Data is users’ head movements during the practice. Gaze data represents eyes tracking data. And “interaction with audience” is to see whether the user could make enough notice to the audience.

Table.4 Events that we collect the time stamp data

<b>User’s Position Data</b>
<b>Head Rotation Data</b>
<b>Gaze Data</b>
<b>Interaction with audience</b>

UnityEngine.Time is used for collecting time data. Unity returns data every 0.04 second. Due to user’s rotation and position usually does not change too much in a short time, we change to collect the data every one second. As for the whole duration of presentation practice, we calculate using the following formula:

$$D = T_{1stPress} - T_{2ndPress}$$

,where D,  $T_{1stPress}$  and  $T_{2ndPress}$  represents time duration, the first time, and the second



time when the user press down the “grip button” of controllers, respectively. And Table.5 shows an example of retrieved data.

Table.5 Sample data structure of position & rotation data (part of)

Time (s)	Position (m)			Rotation (°)		
	x	y	z	x	y	z
0.03	0.68	1.26	0.74	356.98	226.61	3.54
1.05	0.68	1.26	0.74	356.22	228.79	3.23
2.09	0.66	1.28	0.76	354.65	212.37	3.59
3.12	0.68	1.27	0.78	353.2	140.52	0.54

All the data is rounded to the second decimal place.

### 5.4.2 Gaze Data

The way we collect gaze data is similar to the way we implement for gaze-based input. A cursor, which represents user’s gaze, is placed at the center of the screen. Then we add collider to each of the object in “Presentation” scene. We collect all the positions where the cursor and objects collide. Since we build the whole virtual environment using unity3d, positions are all already determined and therefore position data of objects can all be obtained by simply using Scripting API of Unity.

Fig. 18 shows an example. Here we use the color-change of ball to show the movement of gaze. The red circle at the center represents user’s gaze. We collect all the data once the color-changing of ball has been noticed. For instance, in Fig.10, we could find that the color of the ball laved “8” has changed. It indicates that the user’s gaze is collided with the “8” ball, thus the position of the “8” ball will be collected as gaze data.

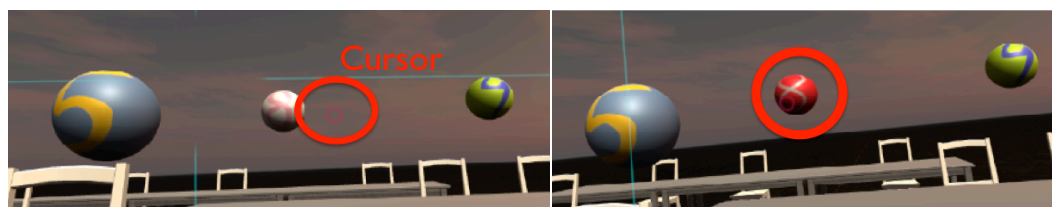


Fig. 18 The detection of user’s gaze

Table.5 gives a sample structure of the gaze data.

Table.5 Sample data structure of gaze data (part of)

Time (s)	Position (m)			Object name
	x	y	z	
0.09	-0.39	1.6	-1.93	ppt
8.82	-2.18	-0.56	2.57	Female1
8.9	-2.84	-0.55	3.82	Male1
8.99	-2.18	-0.56	2.57	Female1
9.08	-4.12	-2.02	0.82	Chair



### 5.4.3 Interaction Data

The purpose of collecting Interaction data is to examine whether the presenter could give necessary notice to each audience's reaction towards their presentation. For example if there is an audience raising hand, a good presenter should notice. We apply very similar method as the way we obtain gaze data to interaction data since gaze is also used as the method for users to interact with virtual audience.

Table.6 gives a sample structure of the interaction data.

Table.6 Sample data structure of interaction data (part of)

Name of objects	Action	Count of total notice
Male1	Sleep	1
VAFemale1	Raise hand	2
Male2	Shake head	3

## 5.5 Heat Map

When a user finishes their presentation, two kinds of heat map will be present:

1. Heat map of position data
2. Heat map of gaze data

Here we introduce how we generate these two kinds of data.

### 5.5.1 Heat Map of Position Data

The way to present heat map in virtual reality is to attach a transparent plane with a material, in which the color contribution for each pixel is calculated by using our collected position data. The process of attaching a material is shown in Fig.19.

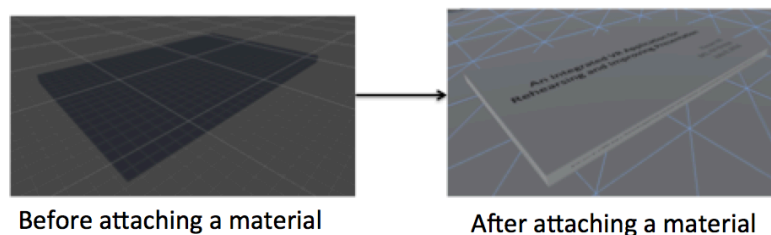


Fig.19 Process of attaching a material to a plane

Therefore, there are three types of information need to be collected and calculated in order to generate the heat map: position data, radius, and intensity. For position data, we read and utilize the position data we gathered as described in Section 5.4.1. All the y-axis data are discarded, since y-axis only represents the height of the user, which does not change much during the speech. From here we can see that the heat map of position data

is in 2 dimensions since it mainly depends on user's x-axis and z-axis positions. In our system, radius of each point is set to 0.1 meter. The color gradient bar we use to represent different intensity is as shown in Fig. 19. The intensity of each point is calculated by counting the times the user stands at each point. All above are calculated and applied to be part of parameters of the material by using Shaders. Shader is a kind of code that defines a lot of properties to affect what will be displayed when the object is rendered. This shader program is written in HLSL (Cg) language.



Fig. 20 Gradient bar for heat map of position data

### 5.5.2 Heat Map of Gaze Data

Different from heat map of position data, which is in 2d, what we create for gaze data is in 3d. While there is various ways to represent gaze in virtual reality, in this research, spheres are used to present a rough image of the eyes movements. The steps we adapted to create for gaze data is described below:

1. Read and load the gaze data we collect, as described in Section 5.4.2.
2. For each object, count the times of being collided/looked.
3. Create and instantiate spheres at the same positions as the objects that appear in gaze data.
4. According to the color gradient bar shown in Fig. 20, assign different color to each ball based on counted times (intensity).



Fig.21 Gradient bar for heat map of gaze data

## 6. Results and Discussion

In order to find more effective solution for users with different levels of speech skills to improve their presentation using our application, the following research questions are accessed:

1. PhD students present better then bachelor students.
2. The familiarity of the slides influences “Visual” performance
3. People do not pay much attention to “Visual” aspects when they present

To validate these research questions, three groups of experiments are conducted:

1. Six users, including three bachelor students and three PhD students, who practice their presentations using their own slides. Evaluation methods informed beforehand. (EG1)
2. Four users practice presentations using the same slides. Evaluation methods informed beforehand. (EG2)
3. Four users practice presentations, without knowing the evaluation method. (EG3)

Table.7 shows an overview of hypotheses and experiments.

Table.7 Overview of hypotheses and experiments

	Hypothesis 1	Hypothesis 2	Hypothesis 3
Experiment	EG1	EG1 & EG2	EG1 & EG3

Fig.21 shows the layout of the virtual classroom.



Fig.22 Layout of the virtual classroom

### 6.1 Group 1 – own slides, evaluation method told

In this group, three students from bachelor course and three students from doctor

course were invited to practice their presentation with their own slides imported into virtual reality. Before the practice, we explained to them the aspects that we focus on when evaluating their presentations, which includes position, gaze, and interaction.

Table.7 shows a brief summary of the slides each user uses in Group 1. Table.8 provides the results of objects that have been looked at for the most in total for each user. Table.9 gives the results related to the object that has been looked by the user for the longest time continuously.

Table.7 Summary of slides of each user

User	Slides number	Total time (s)	Time per slide (s)
PhD 1	29	296	10.21
PhD 2	19	401	21.11
PhD 3	20	453	22.65
Bachelor 1	20	399	19.95
Bachelor 2	33	942	28.55
Bachelor 3	16	302	18.88

Table.8 Summary of collected data #1

User	Percentage of Projector	Percentage of PC	Percentage of audience
PhD 1	13.36%	5.30%	6.91%
PhD 2	1.25%	0.84%	13.89%
PhD 3	2.10%	0.26%	16.21%
Bachelor 1	6.52%	2.41%	21.81%
Bachelor 2	0.28%	2.53%	19.13%
Bachelor 3	0.16%	2.38%	14.37%

Table.9 Summary of collected data #2

User	Longest time continuously and percentage over whole time (s, %)	Object name
PhD 1	18.22, 6.16%	Projector
PhD 2	34.09, 8.50%	Outside
PhD 3	9.80, 2.16%	Projector
Bachelor 1	19.50, 4.89%	Projector
Bachelor 2	9.15, 0.01%	Male2-6
Bachelor 3	13.76, 4.56%	PC

The results of users' gaze data and position data are shown below.

From all the tables and figures we can find that: 1. PhD students have better eye contacts with all of the audience than bachelor students. A good eye contacts means having an average rate of eye contacts with each audience, not too much, not too little. Besides, we can find that bachelor students are more often reading the contents on projector and PC than PhD students. Research question 1 is verified.

### 6.1.1 Gaze Data

The heat maps of gaze data for three PhD students are shown below.

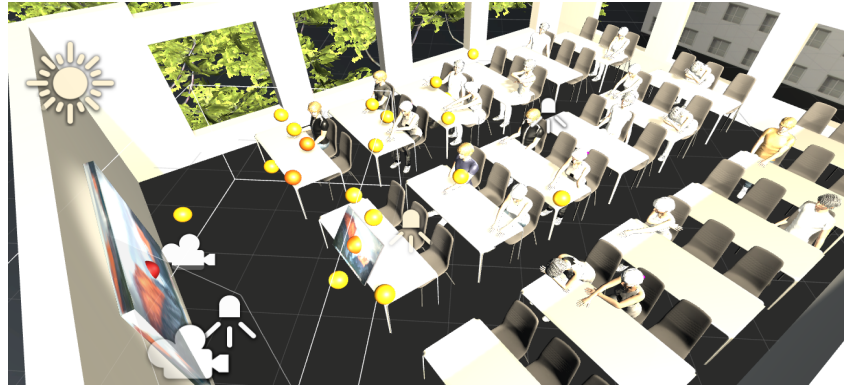


Fig. 23 Heat map of PhD student's gaze data captured in unity3d #1

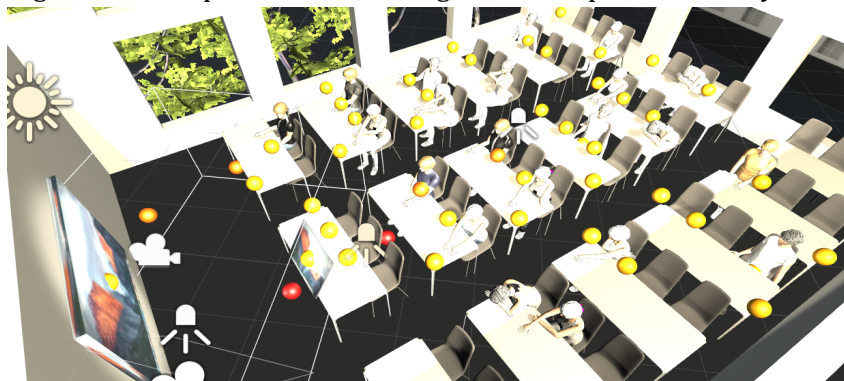


Fig. 24 Heat map of PhD student's gaze data captured in unity3d #2

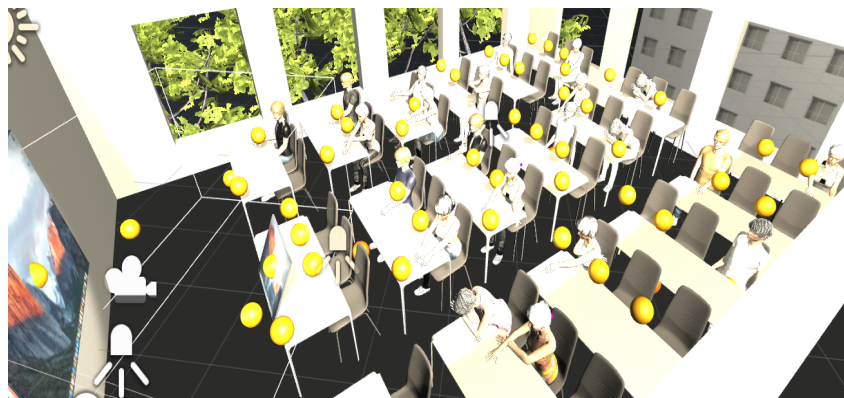


Fig.25 Heat map of PhD student's gaze data captured in unity3d #3

And the following are screenshots from Virtual Reality.

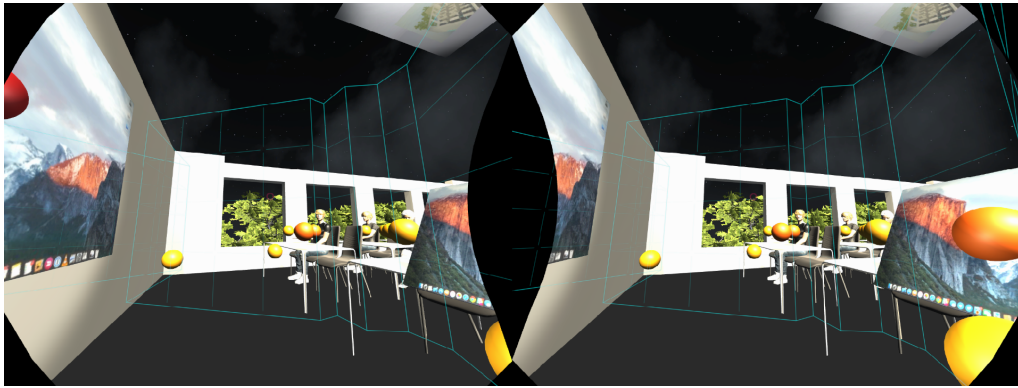


Fig. 26 Heat map of PhD student's gaze data captured in VR #1

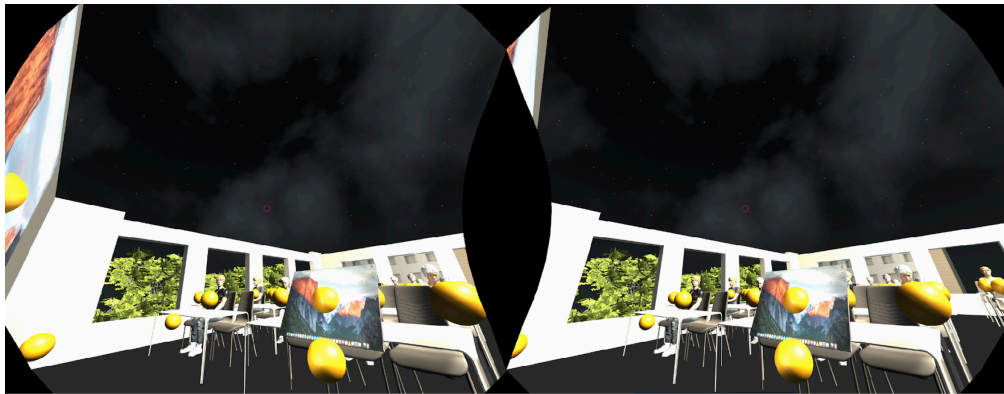


Fig.27 Heat map of PhD student's gaze data captured in VR #2

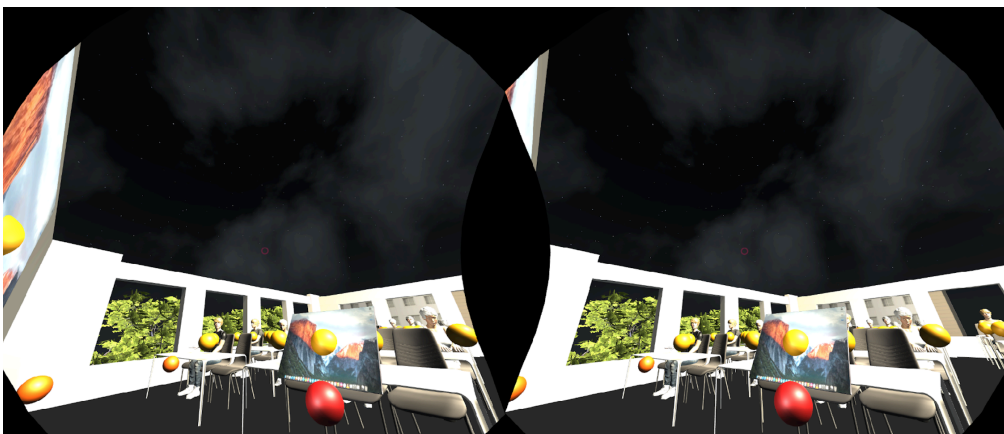


Fig. 28 Heat map of PhD student's gaze data captured in VR #3



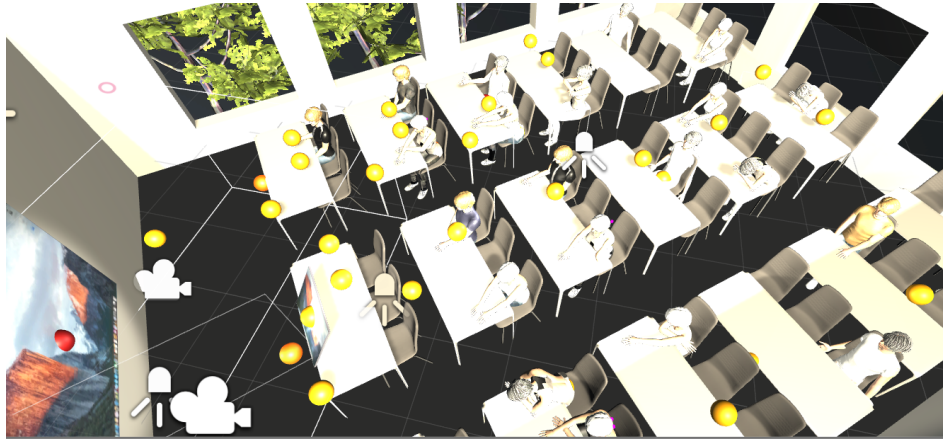


Fig.29 Heat map of Bachelor student's gaze data captured in unity3d #1

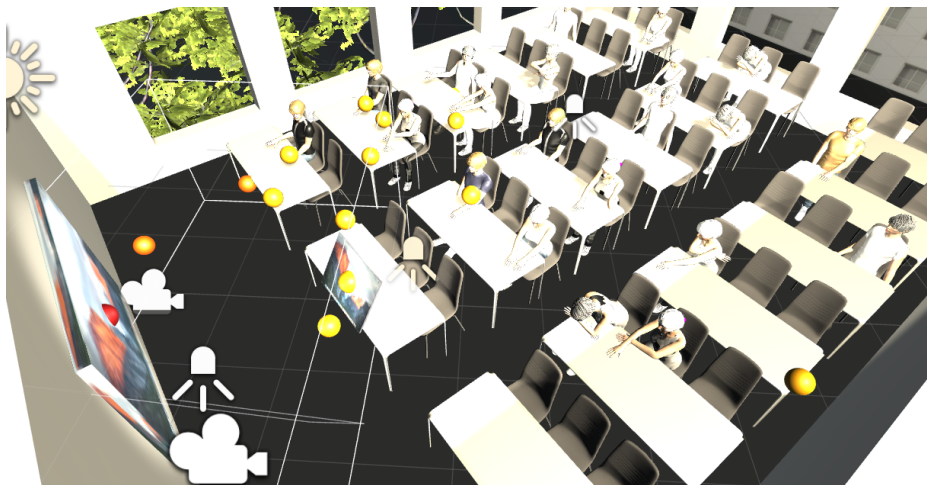


Fig.30 Heat map of Bachelor student's gaze data captured in unity3d #2

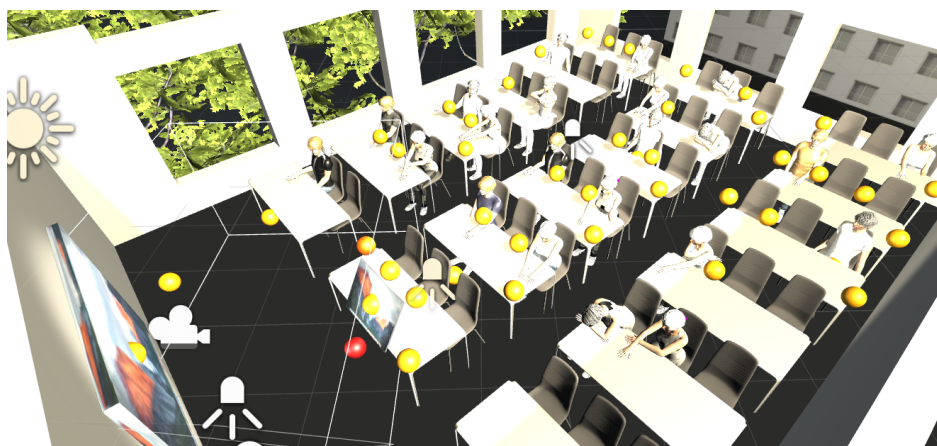


Fig.31 Heat map of Bachelor student's gaze data captured in unity3d #3

The following are screenshots from Virtual Reality.

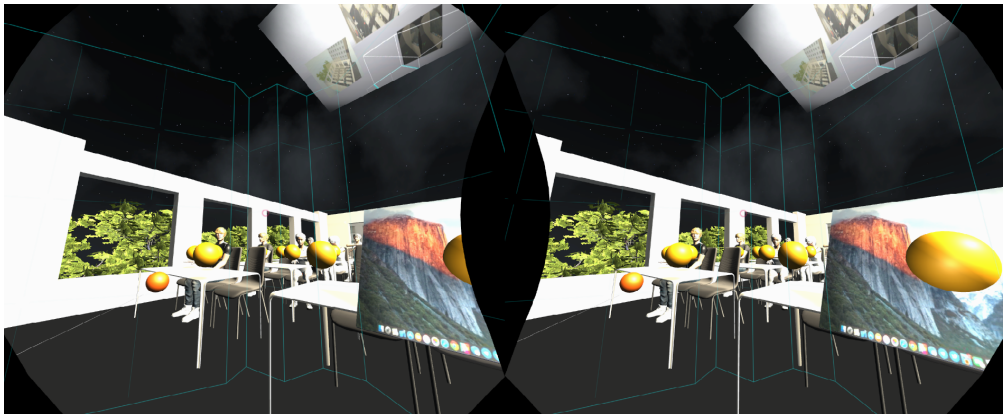


Fig.32 Heat map of Bachelor student's gaze data captured in VR #1

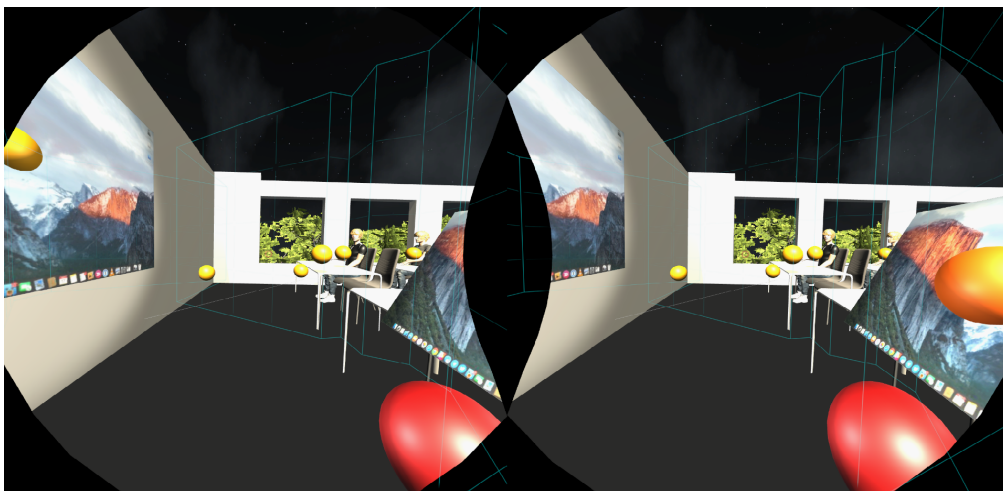


Fig.33 Heat map of Bachelor student's gaze data captured in VR #2

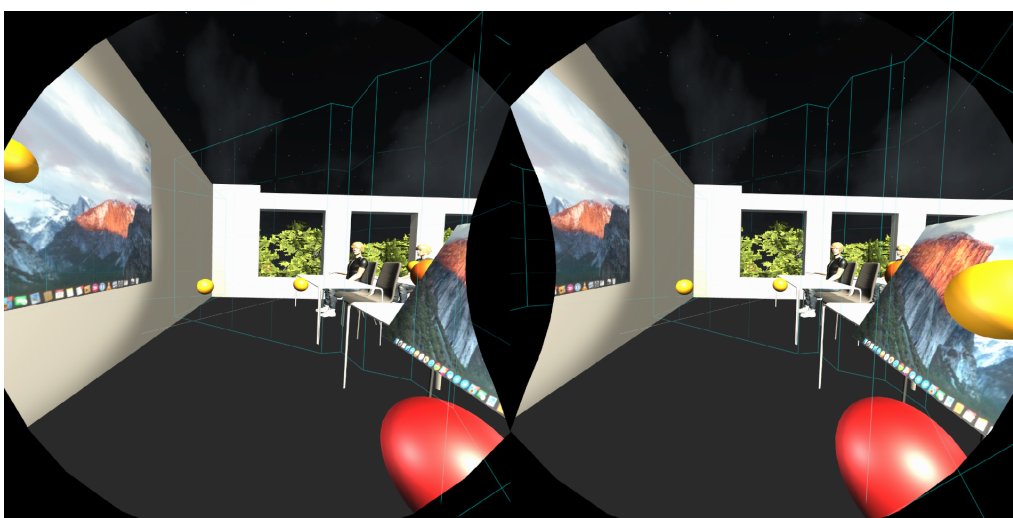


Fig.34 Heat map of Bachelor student's gaze data captured in VR #3



### 6.1.2 Position Data



Fig.35 Heat map of PhD student's gaze data captured in unity3d #1



Fig.36 Heat map of PhD student's gaze data captured in unity3d #2

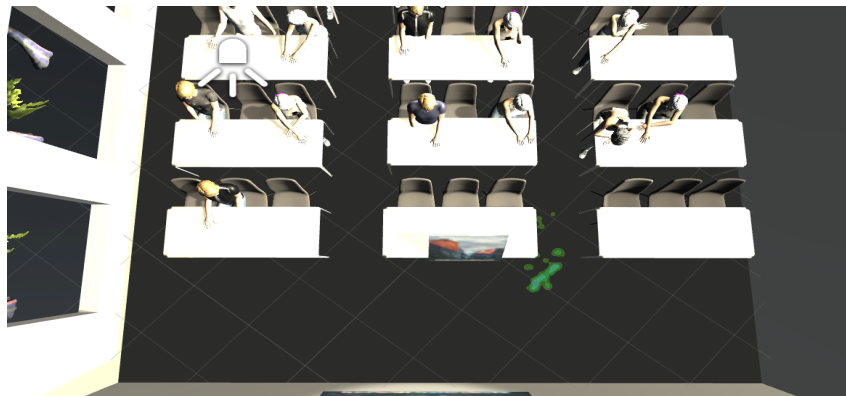


Fig.37 Heat map of PhD student's gaze data captured in unity3d #3

The following are screenshots from Virtual Reality.

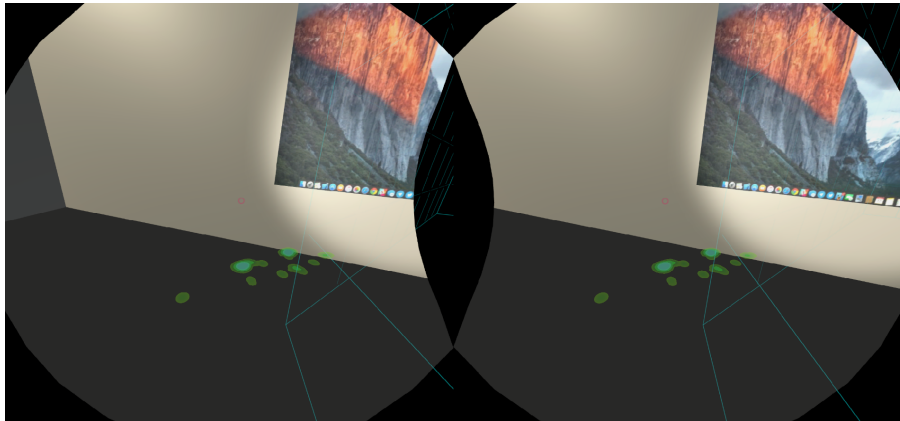


Fig. 38 Heat map of PhD student's position data captured in VR #1

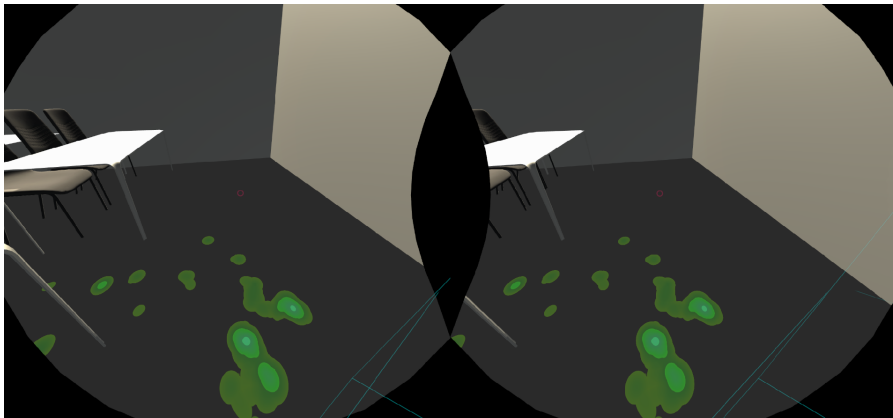


Fig. 39 Heat map of PhD student's position data captured in VR #2

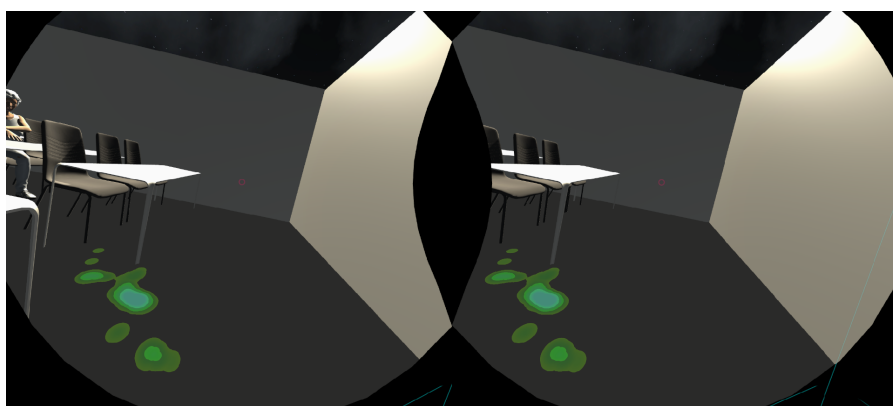


Fig. 40 Heat map of PhD student's position data captured in VR #3

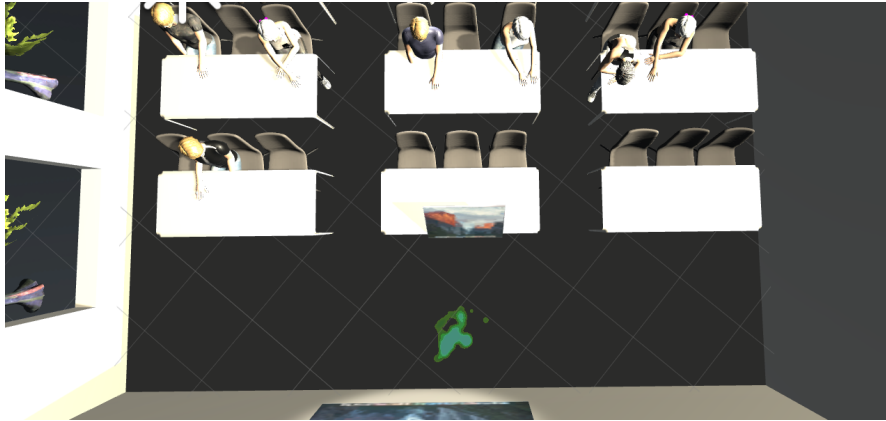


Fig.41 Heat map of Bachelor student's position data captured in unity3d #1



Fig.42 Heat map of Bachelor student's position data captured in unity3d #2



Fig.43 Heat map of Bachelor student's position data captured in unity3d #3

The following are screenshots from Virtual Reality.

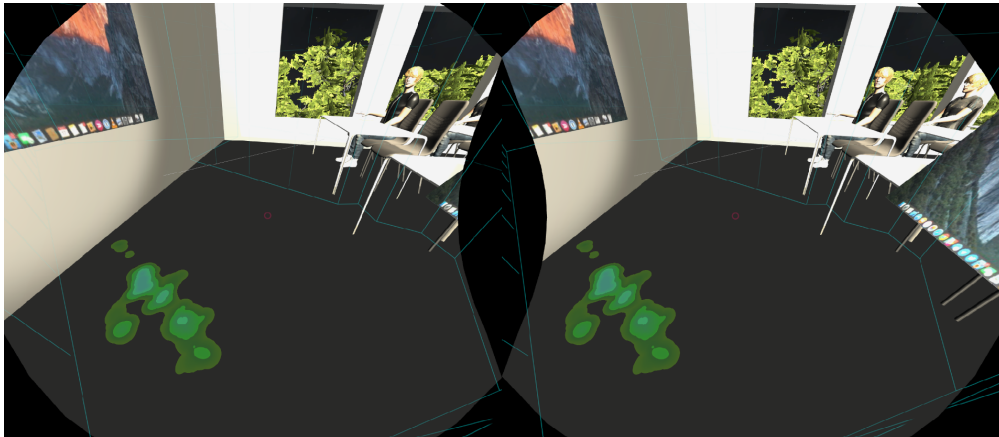


Fig. 44 Heat map of PhD student's position data captured in VR #1

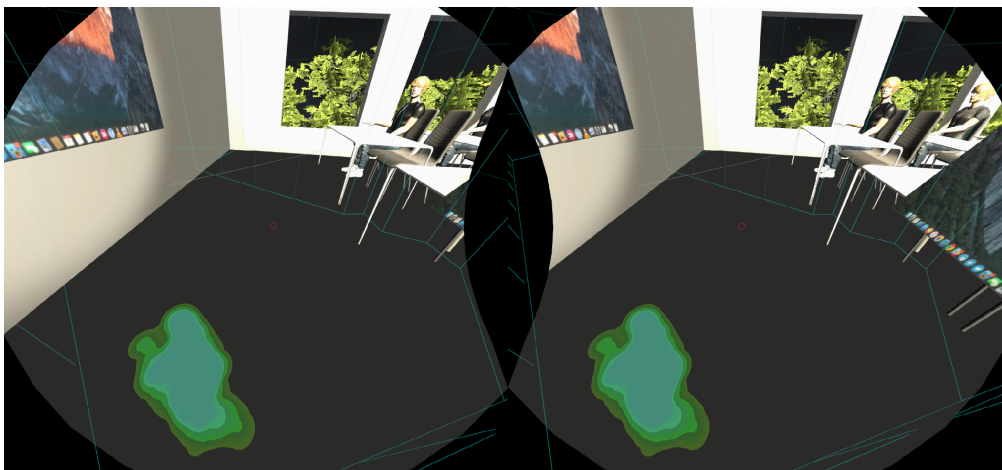


Fig. 45 Heat map of PhD student's position data captured in VR #2

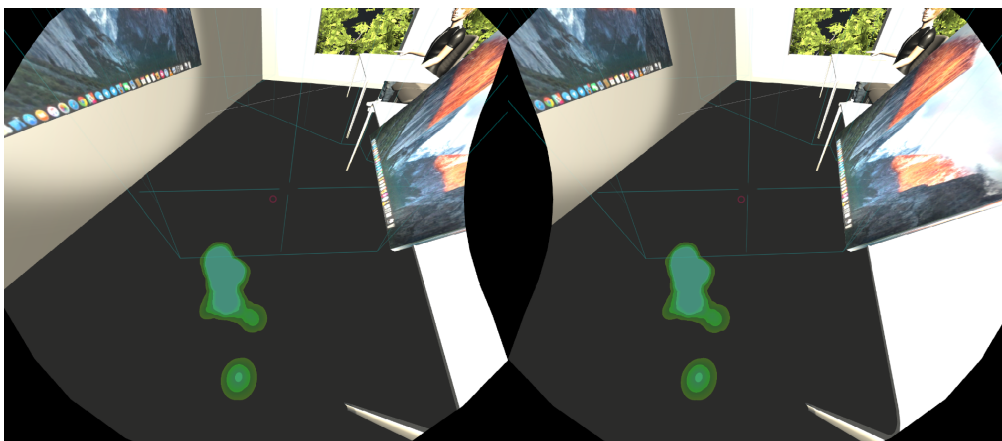


Fig. 46 Heat map of PhD student's position data captured in VR #3

## 6.2 Group 2 – Same slide, evaluation method told

In this group, we invited four students practicing presentation using the same slides. Slides that we use in this group are well considered that it should be familiar to all the students tested. In this case, we use the slides of “Introduction to our laboratory”. Unlike in Group 1, which does not consider the time duration, there exists a preferred time duration for Group 2 since the slides were present strictly follow a fixed time period. Users are told about the aspects that would be evaluated in advance.

Table.10 presents a brief summary of the usage of slides of each user uses in Group 2. Table.11 shows the results of objects that have been looked for the most in total for each user while Table.12 shows the results related to the object that has been looked by the user for the longest time continuously.

Table.10 Summary of slides of each user

User	Slides number (fixed)	Preferred duration (fixed)	Total time (s)	Time per slide (s)
1	25	240	520	20.80
2	25	240	247	9.88
3	25	240	502	20.08
4	25	240	832	33.28

Table.11 Summary of collected data #1

User	Percentage of Projector	Percentage of PC	Percentage of audience
1	3.02%	0.72%	14.39%
2	4.97%	3.07%	15.64%
3	6.17%	2.94%	14.21%
4	24.31%	2.91%	13.76%

Table.12 Summary of collected data #2

User	Longest time continuously and percentage over whole time (s, %)	Object name
1	10.50, 2.02%	Outside
2	6.66, 2.70%	Projector
3	11.35, 2.26%	Projector
4	25.26, 3.04%	Projector

Comparing from the results we obtained from Group 1 and Group 2, it is obvious to see that all the users spent more than necessary time to make the presentation. Additionally, from the heat map of gaze we generated, as shown in Section 6.2.2, we can also find that users in Group 2 are more relying on the contents on projector and PC than users in Group 1. Therefore we could conclude that presenter’s “Visual” performance is somehow depending on the familiarity of the slides being used. Research question 2 is verified.

## 6.2.2 Gaze Data

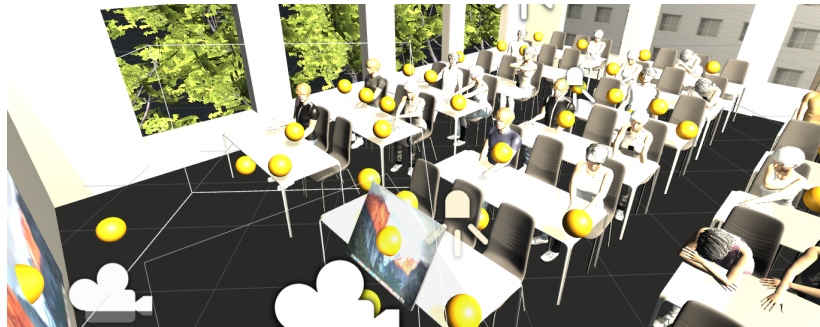


Fig.47 Heat map of User 1's gaze data captured in unity3d

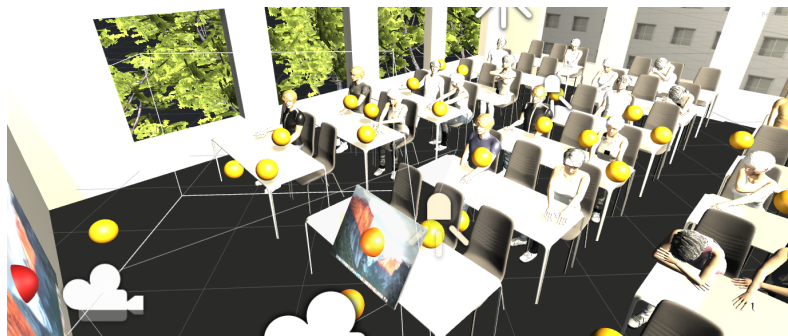


Fig.48 Heat map of User 2's gaze data captured in unity3d

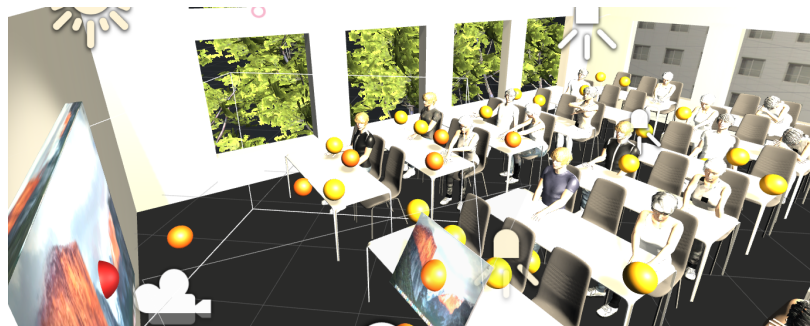


Fig.49 Heat map of User 3's gaze data captured in unity3d

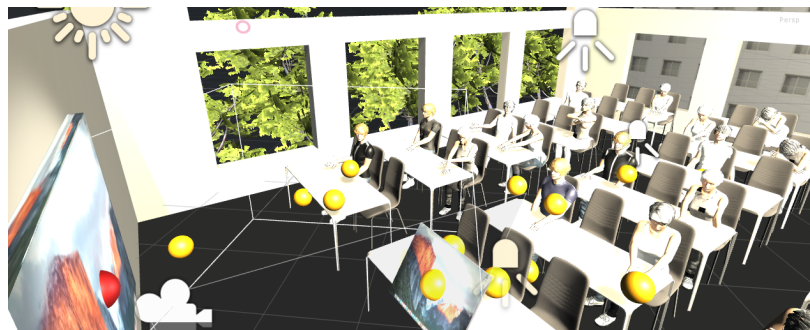


Fig.50 Heat map of User 4's gaze data captured in unity3d



The following are screenshots from Virtual Reality.

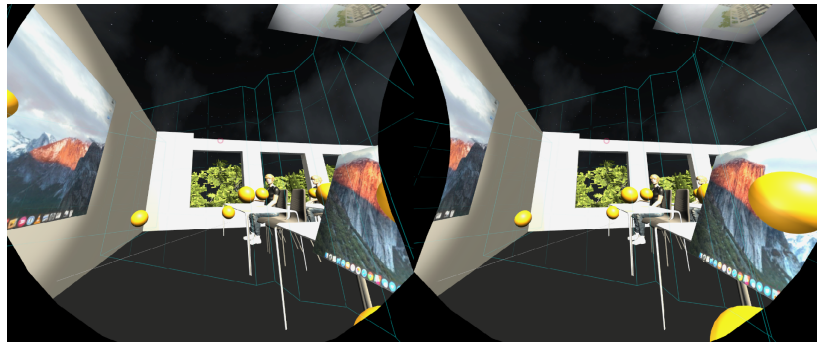


Fig.51 Heat map of User 1's gaze data captured in VR

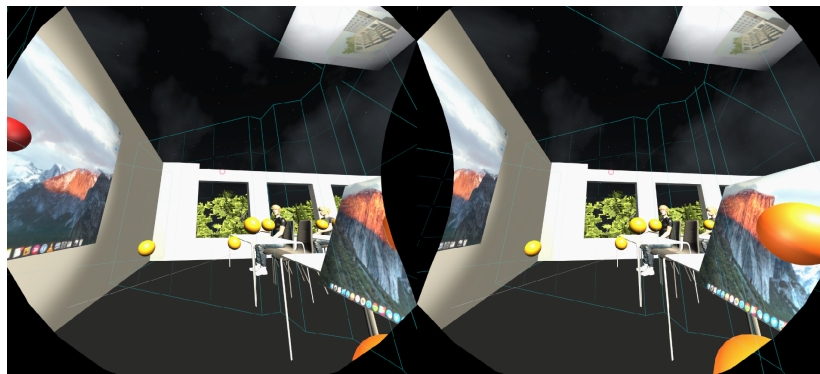


Fig.52 Heat map of User 2's gaze data captured in VR

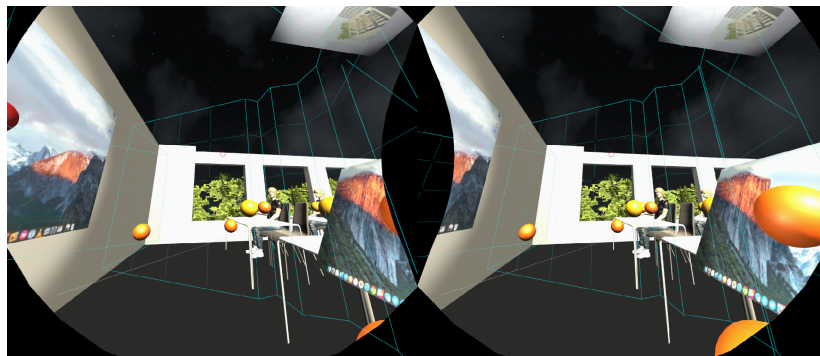


Fig.53 Heat map of User 3's gaze data captured in VR

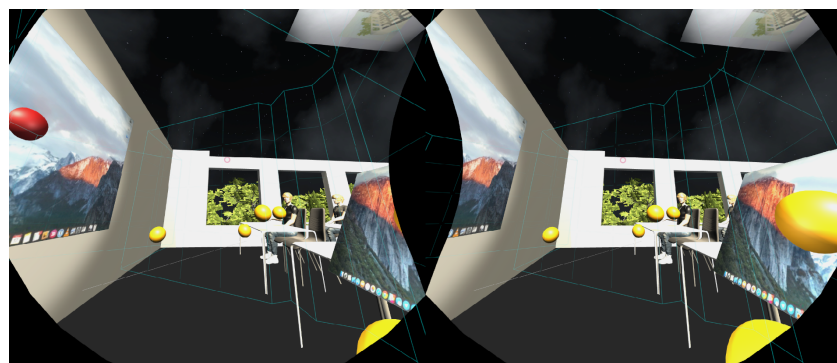


Fig.54 Heat map of User 4's gaze data captured in VR

## 6.2.2 Position Data

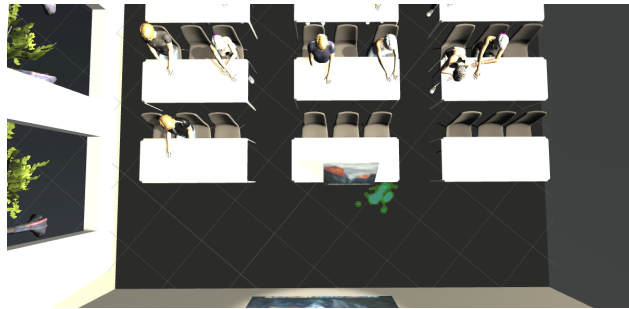


Fig.55 Heat map of User 1's position data captured in unity3d

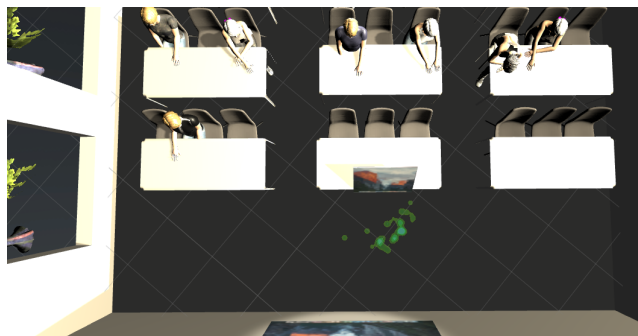


Fig.56 Heat map of User 2's position data captured in unity3d

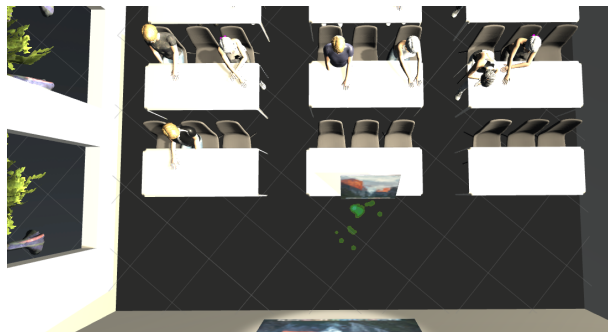


Fig.57 Heat map of User 3's position data captured in unity3d



Fig.58 Heat map of User 4's position data captured in unity3d



The following are screenshots from Virtual Reality.

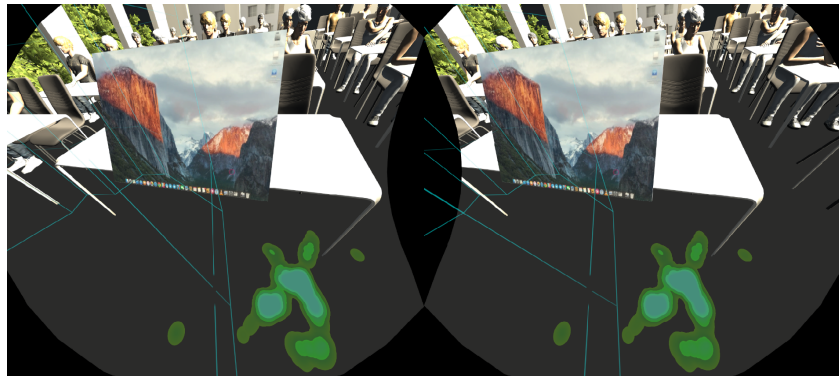


Fig. 59 Heat map of User 1's position data captured in VR

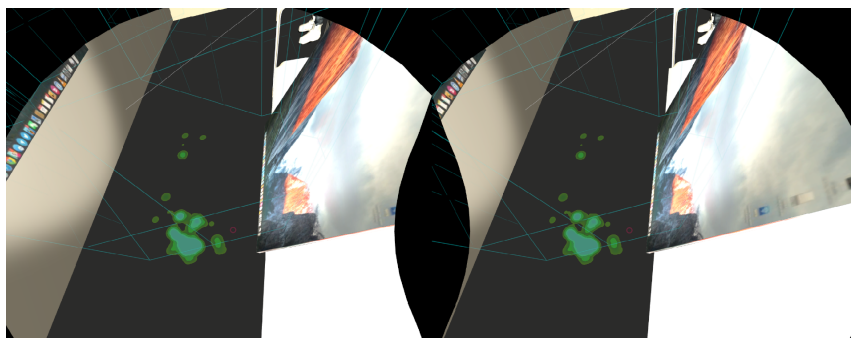


Fig. 60 Heat map of User 2's position data captured in VR

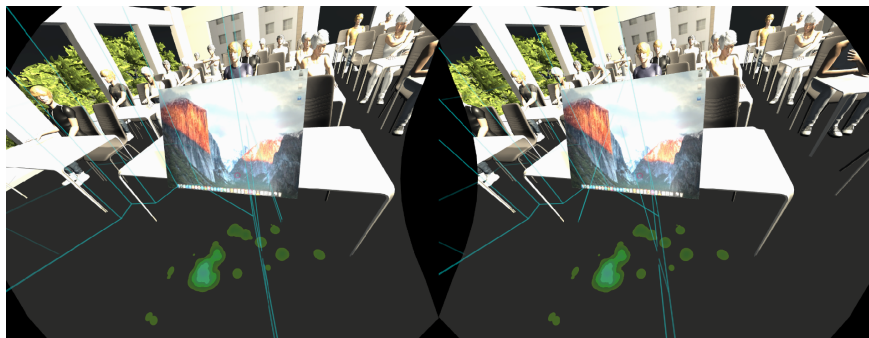


Fig. 61 Heat map of User 3's position data captured in VR

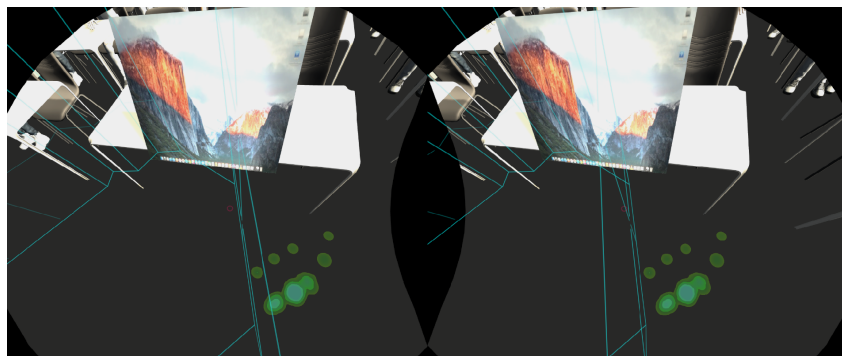


Fig. 62 Heat map of User 4's position data captured in VR

### 6.3 Group 3 – Own slide, evaluation method not told

For Group 3, we invited four students without telling them what aspects will be evaluated during presentation. Similar to Group 1, we allowed each user to import and use their slides.

Table.13, Table.14, and Table.15 give some summery of the data. Similar to Section 6.1 and Section 6.2, Table.13 shows a summery of the slides each user uses. Table.14 gives a summery of objects that have been looked for the most for each user. And Table.15 presents the results related to the object that has been looked by the user for the longest time continuously.

Table.13 Summary of slides of each user

User	Slides number	Total time (s)	Time per slide (s)
1	9	431	47.89
2	16	685	42.81
3	29	895	30.86
4	17	699	41.18

Table.11 Summary of collected gaze hit data #1

User	Percentage of Projector	Percentage of PC	Percentage of audience
1	12.96%	0.93%	13.89%
2	28.79%	3.03%	11.36%
3	2.46%	8.04%	9.38%
4	0.91%	6.36%	14.32%

Table.12 Summary of collected data #2

User	Longest continuously time and percentage over whole time (s, %)	Object name
1	99.57, 23.10%	Projector
2	160.01, 23.36%	Projector
3	210.35, 23.50%	Projector
4	39.94, 5.71%	Projector

### 6.3.1 Gaze Data



Fig. 63 Heat map of User 1's gaze data captured in Unity3d



Fig. 63 Heat map of User 2's gaze data captured in Unity3d



Fig. 63 Heat map of User 3's gaze data captured in Unity3d



Fig. 63 Heat map of User 4's gaze data captured in Unity3d

The following are screenshots from Virtual Reality.

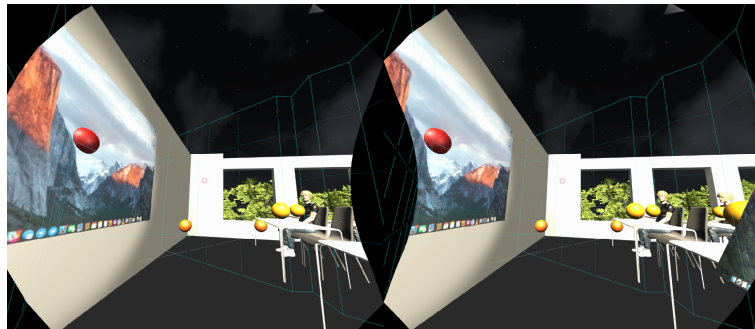


Fig. 63 Heat map of User 1's gaze data captured in VR

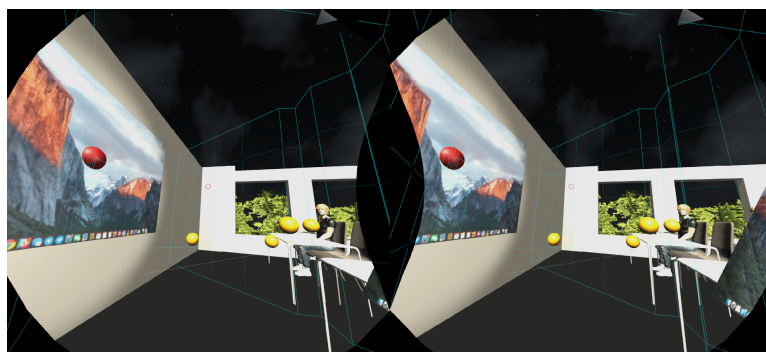


Fig. 63 Heat map of User 2's gaze data captured in VR

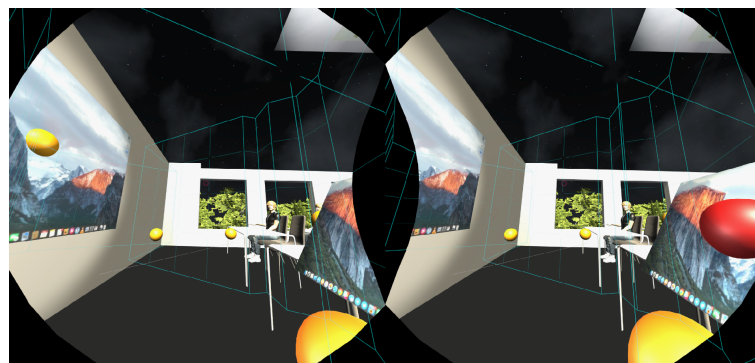


Fig. 63 Heat map of User 3's gaze data captured in VR

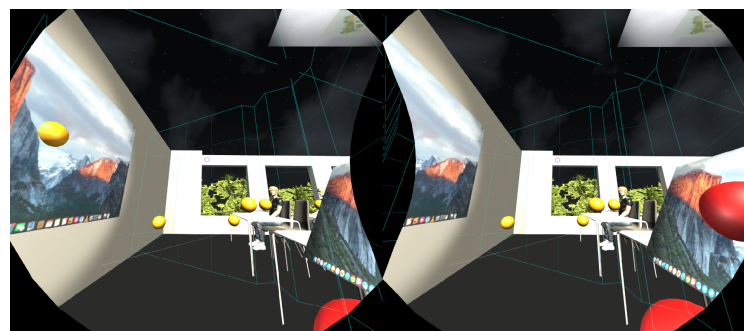


Fig. 63 Heat map of User 4's gaze data captured in VR

### 6.3.2 Position Data

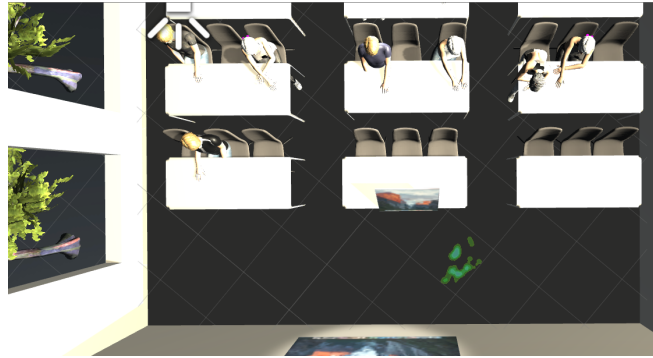


Fig.64 Heat map of User 1's position data captured in unity3d



Fig.65 Heat map of User 1's position data captured in unity3d

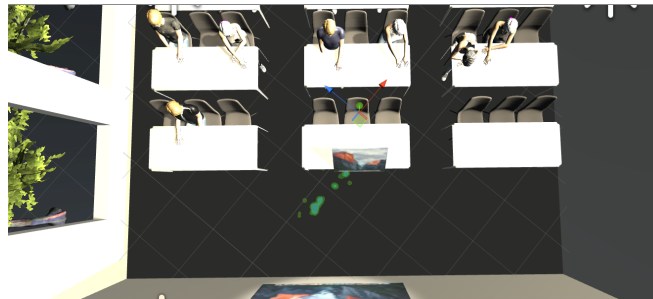


Fig.66 Heat map of User 1's position data captured in unity3d

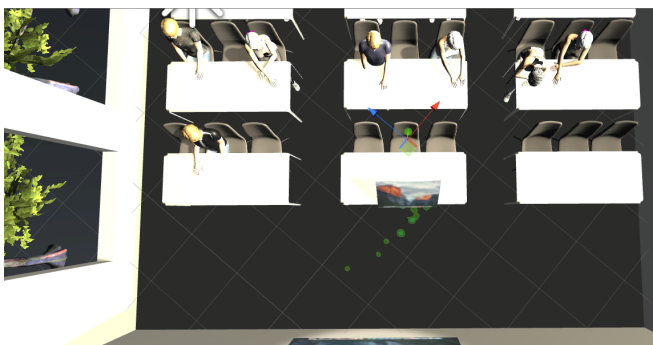


Fig.67 Heat map of User 1's position data captured in unity3d



The following are screenshots from Virtual Reality.

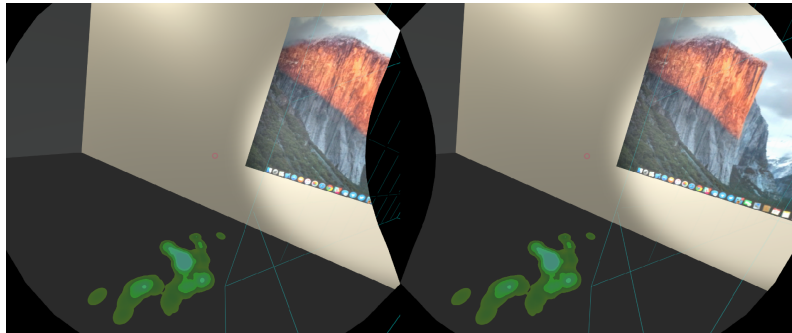


Fig.68 Heat map of User 1's position data captured in VR

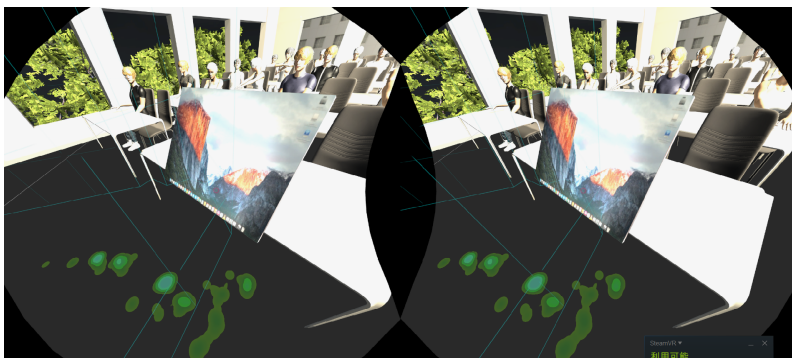


Fig.69 Heat map of User 1's position data captured in VR

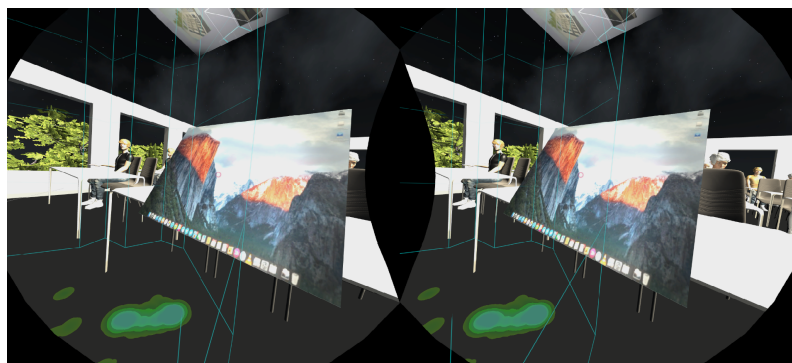


Fig.70 Heat map of User 1's position data captured in VR



Fig.71 Heat map of User 1's position data captured in VR

## 7. Evaluation

A questionnaire with 10 questions is used for the evaluation of the system. In total 12 people have answered. 7 are males and 5 females.

The questions and answers are shown as below.

1. How easier did you feel when switch between scenes?

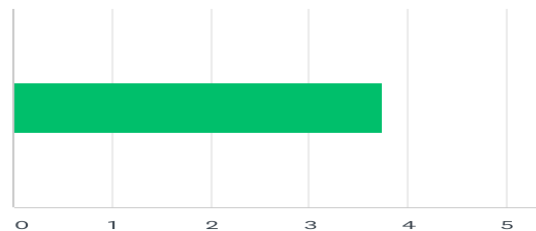


Fig.72 Evaluation results of Q1

Table.13 Evaluation results of Q1

Score	1	2	3	4	5	Average Score
Votes #	0	0	5	5	2	3.75

And we have also received some comments such as “it is easy to press the button by mistake”.

2. How much did you feel comfortable with controllers?

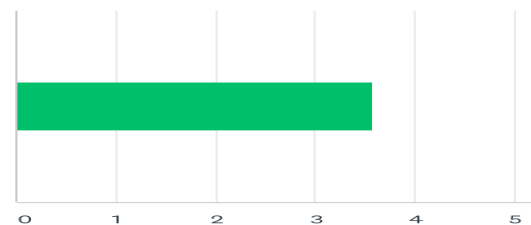


Fig.73 Evaluation results of Q2

Table.14 Evaluation results of Q2

Score	1	2	3	4	5	Average Score
Votes #	0	1	4	6	1	3.58

Most of users feel satisfied with the controllers, however some of them also mentioned that it would be better if a “tutorial of the application” scene could be added since it is hard at the beginning to know the all the functions for a new user.

3. How well did you feel when interacting with the presentation slides?

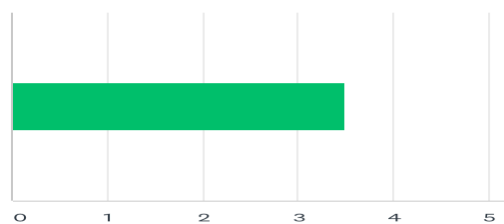


Fig.74 Evaluation results of Q3

Table.15 Evaluation results of Q3

Score	1	2	3	4	5	Average Score
Votes #	0	1	5	5	1	3.50

Most users felt satisfied with the slides in VR. But we have also received some advices. For example, the “go next” button, which enables users to forward slide, is convenient for use, but a “go back” button may also be very helpful.

4. How comfortable were you when moving in virtual environment?

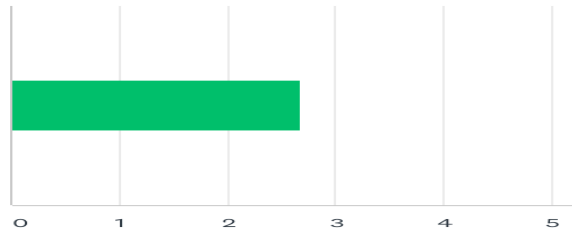


Fig.75 Evaluation results of Q4

Table.16 Evaluation results of Q4

Score	1	2	3	4	5	Average Score
Votes #	0	7	3	1	1	2.67

We have received a low score for moving. Some users mentioned that the images shake when they move therefore they tend to stand at the same point. It also explains the results that we obtained in Section 6. The dizzy caused by moving in our system is also considered as a limitation of this research.

5. How comfortable were you when interacting, for example having eye contacts, with the virtual audience?

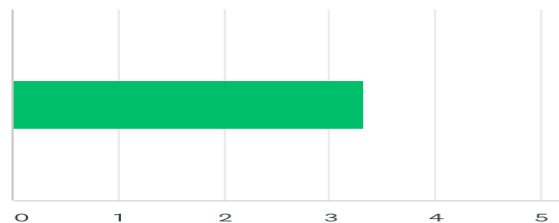


Fig.76 Evaluation results of Q5

Table.17 Evaluation results of Q5

Score	1	2	3	4	5	Average Score
Votes #	0	3	4	3	2	3.33

Most users felt comfortable to make eye contacts with the virtual audience. But some users also mentioned that they felt a little bit dizzy when they move their head.



6. How much did your presentation in virtual environment seem consistent with your real presentation?

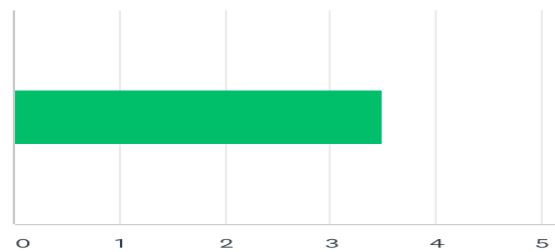


Fig.77 Evaluation results of Q6

Table.18 Evaluation results of Q6

Score	1	2	3	4	5	Average Score
Votes #	0	0	8	2	2	3.50

Most users gave an average score for this question. Some users said there is no difference with real presentation. While some users felt unreal when interact with the virtual audiences due to the lack of animations of virtual audiences.

7. How much did you feel dizzy when doing the presentation? (1: now at all, 5: very much)

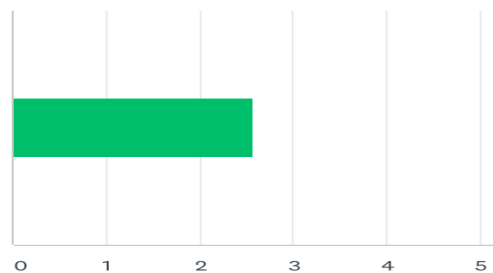


Fig.78 Evaluation results of Q7

Table.19 Evaluation results of Q7

Score	1	2	3	4	5	Average Score
Votes #	3	3	3	2	1	2.58

Users hold different feelings towards this question. For some user they do not feel any dizzy during the whole practice, while for some users they do.

8. If you felt dizzy, what do you think might be the biggest problem?

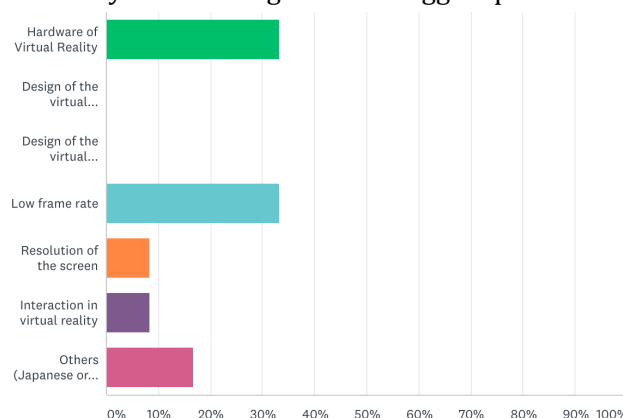


Fig.79 Evaluation results of Q7

From the users, we could find most users considered “Hardware of Virtual Reality” and “Low frame rate” as the main reasons of dizzy.

#### 9. How much did you feel involved? (Immersion)

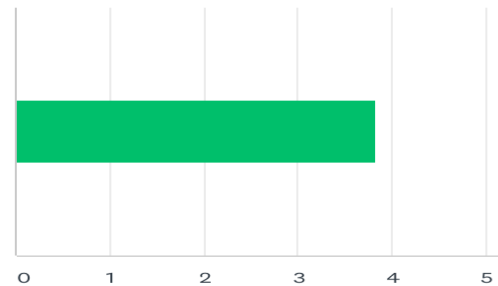


Fig.80 Evaluation results of Q9

Table.20 Evaluation results of Q9

Score	1	2	3	4	5	Average Score
Votes #	0	2	2	4	4	3.83

Most users felt involved during the practice. And reasons for not feeling so involved are discussed in the Q10.

#### 10. If you do not feel so much involved, what do you think might be the biggest cause?

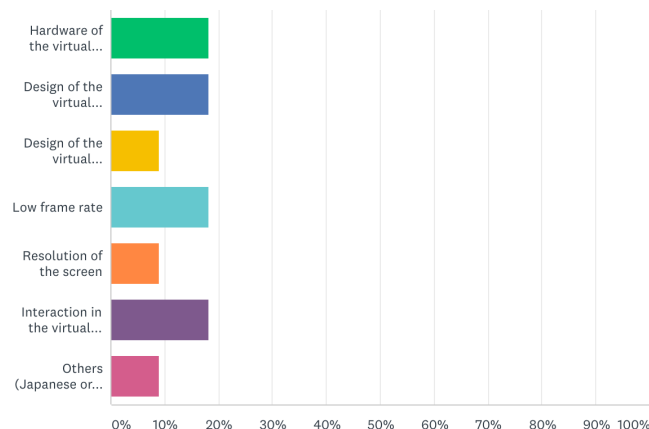


Fig.81 Evaluation results of Q10

Users hold different opinions towards this question. We could conclude that the immersion in VR can be improved from all the aspects that we listed above.

## 8. Conclusion and Future Work

In this paper, we present a VR application for presentation practice. Currently we construct four stages: “Menu”, “Presentation”, where the user could practice presentation, “Position” and “Gaze”, where the user could check the heat map based on their position data and gaze data. In our system, users are able to import, use, and interact with their presentation materials in virtual environment. We evaluate each practice mainly based on presenter’s Visual behaviors. Visual behaviors include: stage usage, eye contacts, and interactions. In order to evaluate, three types of data are collected: position data, gaze data, and interaction times. As feedbacks, two kinds of heat maps are generated in real time: heat map of position and gaze data.

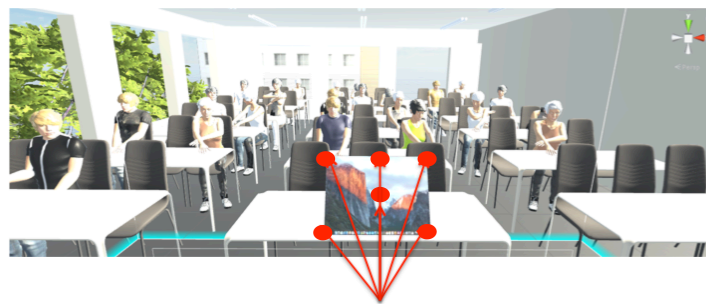


Fig.82 Combining head rotation with object collider position data

There are several aspects that could be considered for further improvement: 1) Accuracy of gaze. And the accuracy could be improved from the following perspectives: 1 Since currently the collisions with object colliders are utilized as the gaze data, the accuracy could be increased if put more colliders in the virtual environment. 2 Combine with head rotation data. As shown in Fig.82, the angle and a more precise position for gaze could be calculated. 2) Frame rate. From the evaluation results of the system we could conclude that the more works could be done to increase the frame rate. Especially because low frame rate when moving causes dizzy, the system makes the user feel more comfortable if standing at the same point, which is opposite of what we expect from a good presenter. Currently the following two aspects are considered to be the main reasons for low frame rate: 1) CPU 2) The application is heavy due to too many objects and colliders. Obviously we could find that there exists a trade-off between gaze data, and frame rate. More researches are needed to solve these problems.

# Acknowledgements

I would like to give my thanks to Professor Yamana, who gives me many advice, ideas and support for my bachelor thesis. Also I would like to thank people who helped me do the experiments. And finally I would like to express my thanks to all the lab members in Yamana Lab. Thanks to all of you who have given me a perfect school life.

# Publications

Yixuan, He, Hayato Yamana. *"A VR Application for Rehearsing and Improving Presentation."* In DEIM 2017.

# References

- [1] 栗原一貴, 加藤公一, and 大浦弘樹. "SlideChecker: プレゼンテーション資料の基礎的な定量的自動評価手法." WISS 第 17 回インタラクティブシステムとソフトウェアに関するワークショップ論文集 (2009): 89-94. Kazutaka Kurihara, Masataka Goto, Jun Ogata, Yosuke Matsusaka, Takeo Igarashi, *Presentation Sensei: A Presentation Training System using Speech and Image Processing*, ICMI'07
- [2] Mathieu Chollet, Torsten Wortwein, Louis-Philippe Morency, Ari Shapiro, Stefan Scherer, *Exploring Feedback Strategies to Improve Public Speaking: An Interactive Virtual Audience Framework*, UbiComp'15, 2015
- [3] Ni Kang, Willem-Paul Brinkman, M. Birna van Riemsdijk, Mark Neerincs, *The design of virtual audiences: Noticeable and recognizable behavioral styles*, Computers in Human Behavior, Volume 55, Part B, Pages 680-694, 2016
- [4] Jason Jerald, ph.D. *The VR Book, Human-Centered Design for Virtual Reality*, ACM Books
- [5] 久山 貴大 *VOYAGE GROUP VR 室ブログ*. Retrieved July 18, 2017, from <http://vr-lab.voyagegroup.com/>
- [6] Tarid Boukhalfi, Christian Joyal, Stephane Bouchard, Sarah Michelle Neveu, Patrice Renaud, *Tools and Techniques for Real-time Data Acquisition and Analysis in Brain Computer Interface studies using qEEG and Eye Tracking in Virtual Reality Environment*, International Federation of Automatic Control Hosting by Elsevier Ltd. 2015
- [7] Evgeniy Upenik, Touradj Ebrahimi, *A simple Method to Obtain Visual Attention Data in Head Mounted Virtual Reality*, IEEE International Conference on Multimedia and Expo, 2017
- [8] Nikolaos Sidorakis, George Alex Koulieris, Katerina Mania, *Binocular Eye-Tracking for the Control of a 3D Immersive Multimedia User Interface*, IEEE, 2015
- [9] Wanhong Lin, Lear Du, Carisa Harris-Adamson, Alan Barr, David Rempel, *Design of Hand Gestures for Manipulating Objects in Virtual Reality*, HCI User Interface Design, Development and Multimodality pp 584-592, 2017
- [10] 栗原一貴, 加藤公一, 大浦弘樹, *SlideChecker: プレゼンテーション資料の基礎的な定量的自動評価手法*, WISS 第 17 回インタラクティブシステムとソフトウェアに関するワークショップ論文集, pp.89-94, 2009
- [11] ] Keith Anderson Elisabeth André T. Baur Sara Bernardini M. Chollet E. Chrysafidou I. Damian C. Ennis A. Egges P. Gebhard H. Jones M. Ochs C. Pelachaud Kaška Porayska-Pomsta P. Rizzo Nicolas Sabouret, *The TARDIS framework: Intelligent virtual agents for social coaching in job interviews*, Advances in Computer Entertainment pp. 476-491, 2013
- [12] Esin Söyler, Chathika Gunaratne, Mustafa İlhan Akbaş, *Towards a Comprehensive Simulator for Public Speaking Anxiety Treatment*, Advances in Applied Digital Human Modeling and Simulation pp 195-205, 2016
- [13] H. TRINH, R. Asadi, D. Edge, T. Bickmore, *RoboCOP: A Robotic Coach for Oral Presentations*, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies archive Volume 1 Issue 2, June 2017 Article No. 27, 2017
- [14] S. Delivering an effective presentation. Retrieved July 18, 2017, from <http://www2.le.ac.uk/offices/ld/resources/presentations/delivering-presentation>
- [15] *10 Tips for a Good Presentation*. Retrieved July 18, 2017, from <http://blog.ut.ee/10-tips-for-a-good-presentation/>
- [16] *MCS Male*. Retrieved July 23, 2017, from <https://www.assetstore.unity3d.com/jp/#!/content/45805>
- [17] *MCS Female*. Retrieved July 23, 2017, from <https://www.assetstore.unity3d.com/jp/#!/content/45807>