Examining User Engagement via Facial Expressions in Augmented Reality with Dynamic Time Warping

by

Kushal Reddy Papakannu

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved April 2021 by the
Graduate Supervisory Committee:

Ihan Hsiao, Chair
Chris Bryan
Mina Johnson-Glenberg

ARIZONA STATE UNIVERSITY

May 2021

ABSTRACT

Augmented Reality (AR) has progressively demonstrated its helpfulness for novices to learn highly complex and abstract concepts by visualizing details in an immersive environment. However, some studies show that similar results could also be obtained in environments that do not involve AR. To explore the potential of AR in advancing transformative engagement in education, I propose modeling facial expressions as implicit feedback when one is being immersed in the environment. I developed a Unity application to record and log the users' application operations and facial images. A neural network-based model, Visual Geometry Group 19 (VGG19, Simonyan and Zisserman (2014)), is adopted to recognize emotions from the captured facial images. A within-subject user study was designed and conducted to assess the sentiment and user engagement differences in AR and non-AR tasks. To analyze the collected data, Dynamic Time Warping (DTW) was applied to identify the emotional similarities between AR and non-AR environments. The results indicate that users showed an increase in emotion patterns and application operations throughout the AR tasks in comparison to non-AR tasks. The emotion patterns observed in the analysis show that non-AR provides less implicit feedback compared to AR tasks. The DTW analysis reveals that users' emotion change patterns appear to be more distant from neutral emotions in AR than non-AR tasks. Succinctly put, the users in the AR task demonstrated more active use of the application and yielded ranges of emotions while operating it.

DEDICATION

Dedicated to

My Professors, My Family and My Friends

ACKNOWLEDGMENTS

I would like to express my sincere gratitude and indebtedness to my thesis mentor Dr. Sharon Hsiao for her valuable suggestions, constant supervision, timely guidance, keen interest and encouragement throughout my thesis. This work has only been possible because of the knowledge I have gained under her guidance. The lessons learnt would help me immensely in my future endeavors. I am thankful to Dr. Mina Johnson-Glenberg and Dr. Chris Bryan for being supportive of my research and for willing to serve on my thesis committee.

I would also like to thank Cheng-Yu Chung for mentoring me throughout my thesis. I would also like to thank Setu Shah for her continuous support in all aspects of my life. Finally, I would like to thank Arizona State University for providing the amenities required in the successful completion of my thesis and the department of CIDSE for the constant guidance and assistance throughout my master's program and thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

Chapter 1

INTRODUCTION

Augmented Reality (AR) is a promising technology that can change learning and teaching processes (FitzGerald *et al.* (2013)). In addition to the opportunities for using AR in education, diverse populations, such as gamers, also develop/use AR in day-to-day life (Merchant *et al.* (2014)). The unique characteristics of AR make it different than other technologies. These characteristics are 1. Contextual visualization: overlaying digital information on physical objects; 2. Intuitive interaction: easing the complexity in using computers and eliminating the need for intermediate input devices such as mouse and keyboard; and 3. Continuity of virtual and physical objects: supporting real and virtual learning by adopting tangible user interfaces and promoting embedded learning.

Nincarean *et al.* (2013); Joan (2015) identify that Mobile Augmented Reality (MAR) learning environments, especially when combined with game-based features, could positively affect students' achievements and improve their problem-solving skills. Also, it was discovered that using MAR could motivate students and engage them in accomplishing tasks (Lai *et al.* (2019)). The paper also states that MAR-based learning environments have been witnessed/watched as more immersive for students than non-AR environments.

Jerry and Aaron (2010) show promises that having AR content increases the study performance in the Augmented Reality Learning Experience(ARLE). Despite these studies, there is also research demonstrating that having non-AR content is just as helpful and suggests that engaging students with AR and non-AR contents could yield quite similar learning outcomes (Radu and Schneider (2019)).

However, relatively few studies have focused on investigating the students' emotions during the execution of the intervention itself. More research is needed to investigate the types of emotions the students experience dynamically and understand how these emotions are related to the MAR characteristics and the task properties. This thesis aims to examine the users' learning in a MAR learning environment. The users are engaged in gaming-based problem-solving tasks. The current research focused on highlighting users' emotions and performances in AR tasks compared to non-AR tasks. Less prior work is addressed on using mobile applications enacting front-end camera to capture users' facial expressions simultaneously detecting their emotions. Hence, it motivates me to conduct a user study with a mobile augmented reality application. This research could help draw empirical evidence-based findings by examining the facial emotions of users to determine their engagement.

Dynamic Time warping introduced in 60s (Bellman and Kalaba (1959)) is a very popular tool that finds optimum alignment between two time series given one time series may be warped by stretching or shrinking it non-linearly along its time axis. It was extensively used to compare speech recognition in the 70's (Myers *et al.* (1980)). Dynamic time warping has also been useful in many other disciplines (Keogh and Pazzani (2001)). To analyze the emotion changes patterns from the user study, I applied Fast Dynamic Time Warping (FastDTW) (Salvador and Chan (2007)) to model the collected data according to the temporal sequences. Thus, emotion data and operational task data will be aligned and permitted to examine the emotion pattern changes. FastDTW clusters/warps parts of the temporal sequence data before calculating the distance. Transforming the emotion patterns by one-hot encoding the emotions to '0' or '1' depending on the detection of the emotion allows the use of these sequences in DTW analysis.

1.1 Research Questions

In light of the previous literature review, the questions that guide the research are:

1. Can users' emotions be measured to gauge their engagements in AR/non-AR by using Dynamic Time Warping?

2. What are the emotional pattern differences observed in AR versus non-AR environments?

Chapter 2

LITERATURE REVIEW

2.1 AR in the Education Field

AR has emerged as a mainstream technology that affects education and other areas of our life, and researchers have defined it differently. Klopfer and Squire (2008) state that AR technology combines real and virtual information in a meaningful way. Milgram *et al.* (1995) defines it as "augmenting natural feedback to the operator with simulated cues." Azuma *et al.* (2001) outline three main properties that AR systems should have: (1) combining real and virtual objects in a real environment; (2) running real-time interaction; and (3) the 3D registration of virtual and real objects. Bujak *et al.* (2013); Santos *et al.* (2013) show that AR could help students by promoting authentic and contextual learning, assisting in demonstration and visualization of abstract concepts, and reducing their cognitive load in learning processes. However, Ibáñez *et al.* (2014); Bellucci *et al.* (2018) still claim that AR is a mere new technology that could not bring a substantial change. They even assert that other technologies can achieve such a visualization provided by AR. Recently, with the significant advancement in mobile technology, AR has become more promising. The popularization of mobile technology has contributed to AR adoption (Nadolny (2017)). The popularization enabled researchers and teachers to design rich AR learning environments, which positively affected the students' conceptual understanding, their learning experiences, and motivation (Cheng and Tsai (2013)). In addition, the adoption of gaming principles in these learning environments could enhance students' engagement and foster their motivation (Smiderle *et al.* (2020)). Here is, for example, how Yuen *et al.* (2011) describe AR advantages:

1. AR is used in object modeling, allowing learners to envision how a given item would look in different settings.
2. AR gaming can be utilized to assist students in quickly grasping class concepts. AR games offer a unique opportunity for making learning highly visual and highly interactive.
3. AR can support discovery-based learning. For example, historic sites may use AR to convey virtual and audio information when visitors are walking. Field trips can be turned into "scavenger hunts" with specific details provided in AR systems
4. AR can support collaborative peer learning, allowing individuals to work with other learners in object-modeling, gaming, and discovery learning.

To summarize, researchers reported an increase in the positive effects of ARLE, such as improving content understanding, facilitating long-term memory retention, and fostering students' motivation. On the other hand, they point out an increase in adverse effects of ARLE, such as attention tunneling, ineffective classroom integration, and learning differences (Dede (2009)). Although most of the advantages and limitations were discussed, more research is still needed to investigate the preliminary results and inconsistent findings around AR. Furthermore, less research focuses on examining the emotional aspects related to using AR.

2.2 Emotions in AR

The study of emotions is critical in education, especially in AR, which is known as a technology for developing learning mediums with immersive, motivating, and positively effective engaging qualities. This definition goes in line with Pekrun and

Perry (2014) who highlight the significant relationship between emotions and learning; emotions are expected to predict learning outcomes through their influence on motivational, cognitive, and meta-cognitive processes (Harley *et al.* (2016)). To note, since emotions are general, it is accepted to use the term achievement emotions to describe emotions that a person typically experiences in relation to achievement activities and outcomes within a given situation (Poitras *et al.* (2019)). In this context, the control-value theory of achievement emotions Pekrun (2006) asserts that emotional states fluctuate based on subjective appraisals of control and value; learners who feel in control of the activity and value it are more likely to feel positive emotions, such as enjoyment, and be engaged in adaptive learning outcomes. However, learners who do not value activity are more likely to experience boredom and experience maladaptive learning outcomes. Also, emotion awareness is essential in designing rich multimedia-based learning environments for arousing students' positive emotions and assisting the learning process (Heidig *et al.* (2015)). According to Mayer (2014) emotions derived from immersiveness could relate to motivational features, and in turn, help students engage in learning processes and improve their learning. Although research about AR already exists, there is still a lack of empirical studies that examine emotions related to AR and the role they might play in learning. To address the lack of research, I built a mobile augmented reality application engaging users in two AR and non-AR tasks. The users' facial expressions were tracked dynamically. These emotions were later used to detect their emotional state, such as anger, disgust, fear, happy, sad, surprise, and neutral.

Chapter 3

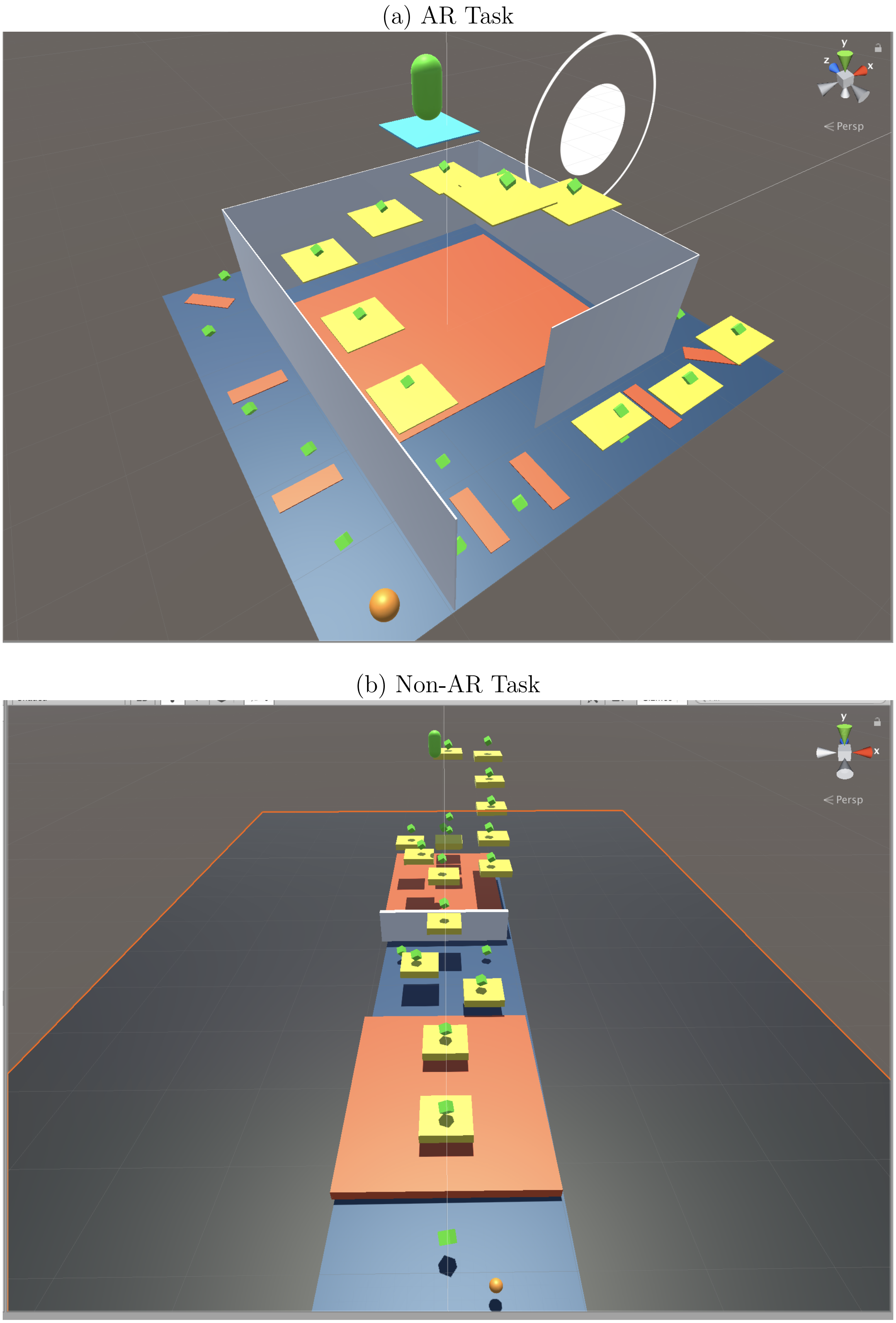
RESEARCH DESIGN

In this chapter, I will discuss creating an application that enables the users to engage in a learning environment and allows the collection of implicit feedback from users. A neural network will also be created which takes the implicit feedback as the input. This neural network will be able to detect the emotions recorded from the implicit feedback. The objective is also to develop a time series analysis to predict the differences of emotions in both AR and non-AR scenarios.

3.1 Research Platform

A Unity application using Vuforia Software Development Kit (SDK) was developed for the user study. As seen in figure 3.1, the within-subject user study consists of two tasks: AR task and non-AR task. In the application, the users are given two virtual controls: jump and joystick, through which they can move the sphere across the map. The objective of the task is to traverse through a map using a sphere and reach the destination (green capsule). Each task consists of four sections of increasing difficulty. The idea is to help users understand and master controls in each of the first three levels and use their knowledge to collect the capsule in the final section of the task. The design of the sections are as follows:

- **Section 0:** Simple joystick movement is introduced to the users. This movement helps them to traverse through the map and collect objects as shown in 3.2.



(a) AR Task

(b) Non-AR Task

Figure 3.1: A Snapshot of the Task Designed for the User Study. The Objective Is to Collect All the Rewarding(Green) Objects While Avoiding Non-Rewarding(Red) Objects

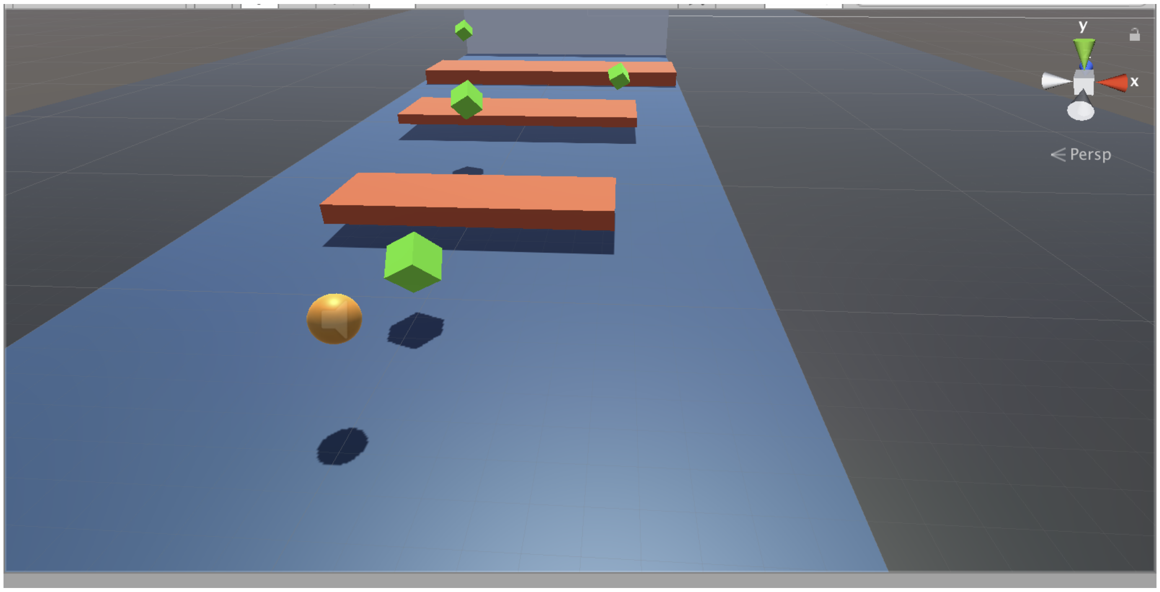


Figure 3.2: A Snapshot of Section 0

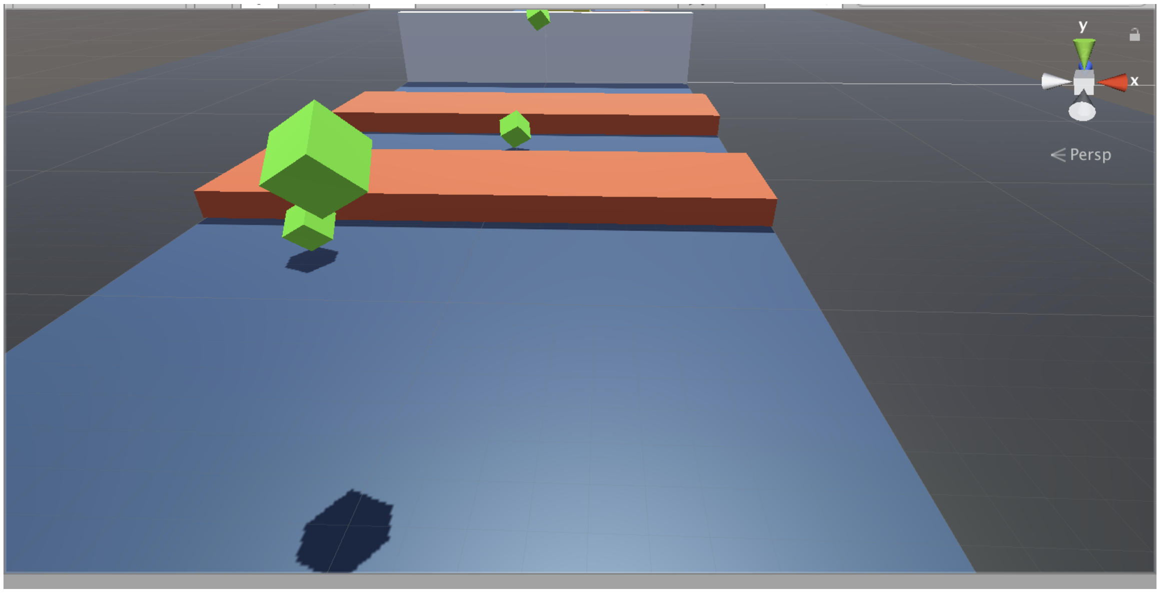


Figure 3.3: A Snapshot of Section 1

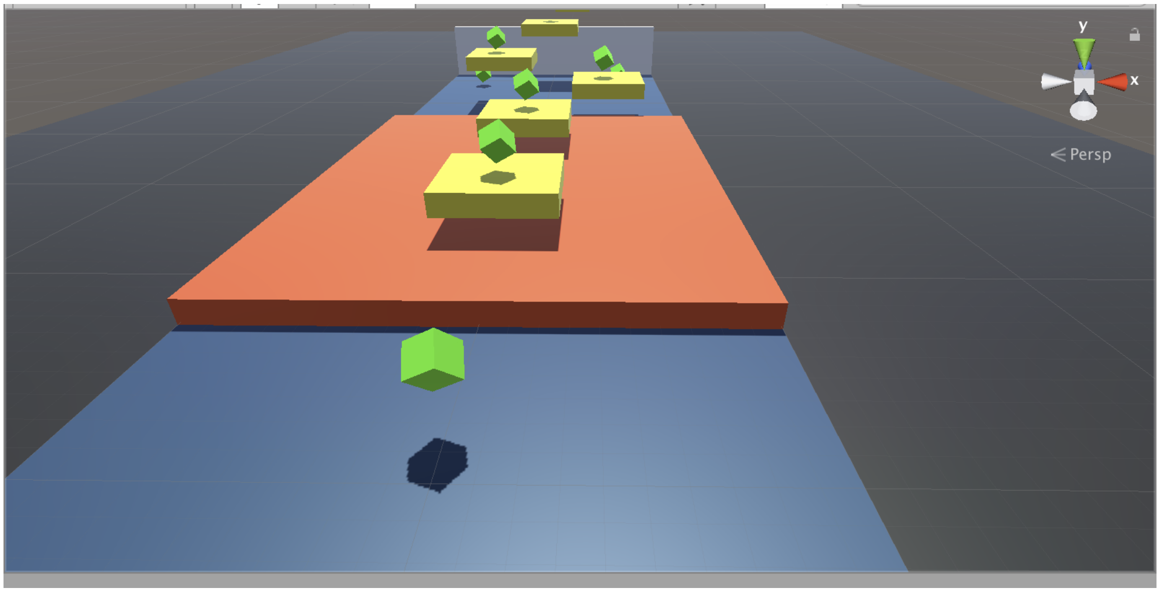


Figure 3.4: A Snapshot of Section 2

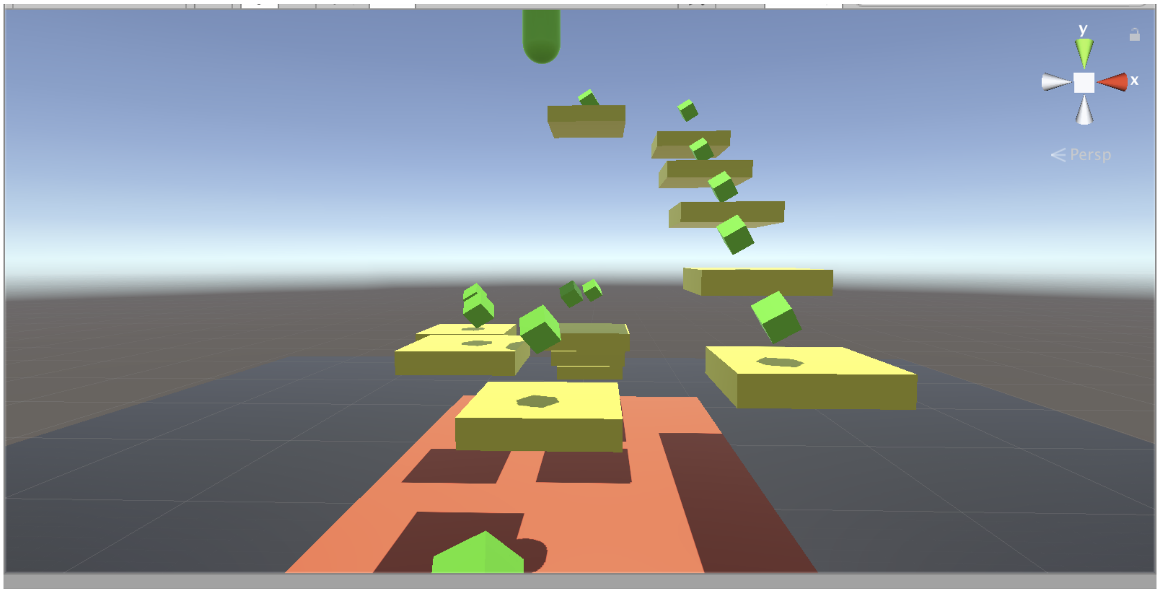


Figure 3.5: A Snapshot of Section 3

- **Section 1:** While some objects are on the ground, some objects are a little higher from the ground as shown in 3.3. In order to collect these objects, users would have to jump. Hence, the users are acquainted with jump controls.

- **Section 2:** As seen in 3.4, bridges(yellow objects) are used to help user understand the combination of jumps and movement.

- **Section 3:** Finally, in Section 3, users use the learning from previous sections to jump over multiple bridges to reach the capsule as shown in 3.5.

During the task, users are encouraged to collect as many rewarding objects as possible, increasing the score while avoiding the non-rewarding objects that decrease the score. Both the AR and non-AR tasks have an equal number of rewarding elements and non-rewarding objects. There is a capsule at the end of the task, which upon collecting will end the task.

The non-AR version needs the rear camera to work. However, to play the AR version, the user is required to scan the QR code. The AR task uses an image target(QR code) placed on the desk to identify and portray the augmented objects using Vuforia. While in AR, both the front and back cameras are used simultaneously to capture the facial image and the image target. Out of all the devices, I ran this application on, only HTC One M8 works fine without crashing. The movement in the non-AR task is applied by adding a small force to the sphere. Contrastingly, the movement in AR is applied by adding force to the coordinates of the sphere with respect to the real-world coordinates. The restrictions/implications of Vuforia SDK cause a movement in AR that is slightly different from non-AR tasks.

3.2 Study Design

The user study consists of four phases:

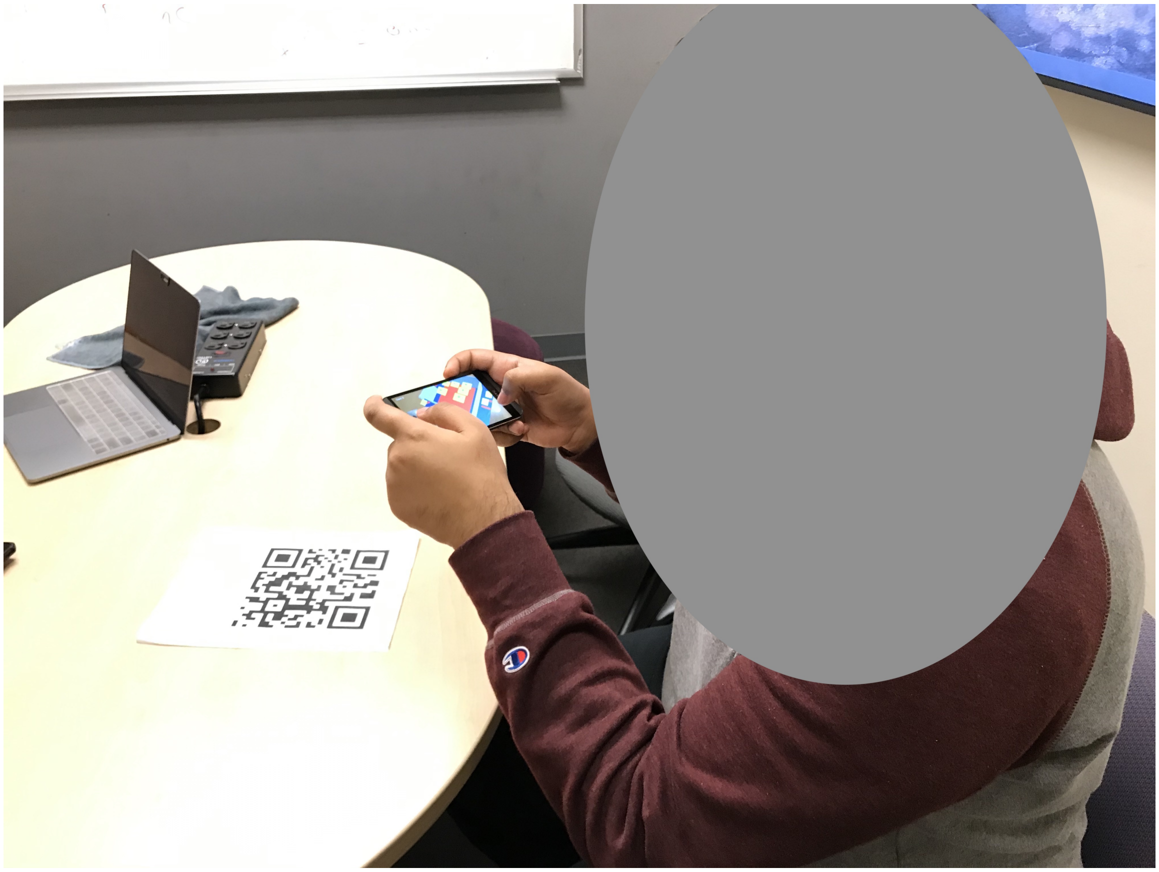


Figure 3.6: A Snapshot of User Study Being Conducted. The User Uses the Mobile Device Back Camera to Scan the Qr Code to Produce AR Content

- Introduction- 5 minutes
- Pre-Task - 5 minutes
- Task - 25 minutes
- Post-Task - 5 minutes

Introduction

The introduction phase begins by explaining to users about the user study and AR technology. A brief introduction about the research along with the distribution of tools such as operating devices was performed. I explicitly notified each user that the entire data used for the user study would be anonymous.

Pre-Task

In the pre-task phase, I explained the instructions for operating the application. Other controls, such as movement and jumping, were explained. As shown in figure 3.6, the users learn to focus the smartphone's back camera on the QR code, enabling the AR content to appear on the screen. Users were given an ample amount of time to get familiar with holding the device and operating it. Users were also explained not to cover the camera with fingers resulting in obstructing the facial data logging.

Task

In the task phase, the users were given two tasks: an AR task and a non-AR task. Half of the users were prompted to start the AR version first, while I prompted the other half to start the non-AR version first. Users were encouraged to speak about what they were doing during this stage. After completing a task, a new task is started not to lose the focus from the previous task. The program ends if the user collects the capsule or if 25 minutes have elapsed.

Post- Task

The device is collected from the user and plugged into the desktop to transfer the logs. I indulged in a few question-and-answer sessions with the user. In the Post-Task phase, I asked a few questions about the user experience and tasks. Users were also given a chance to ask questions about the user study to learn more about the technology.

3.2.1 Logging the Data

As soon as the user starts a task, a log file is created. A logger was also built into the application to capture data from the study. The role of the logger is to capture all operational data such as jumps, the objects collected by the user, and the facial images captured by the camera. The application uses the front camera to capture the users' facial images at 60 Hz/minute or one frame per second.

Data Collection

A logger collected various data from the users such as:

- Jumps
- Objects collected
- Section changes
- Scores
- Facial images

The data has been saved in the following way:

- I saved the scores in the log file with the format of [Section-Action-Timestamp], giving me the exact location of the action that has been made with the timestamp.
- I saved the images with .jpg file format with [Section-YYYY-MM-DD-hh-mm-ss.jpg] as the file name in the same folder. This format helps us identify the sections in which the picture is taken.
- I created an individual folder for the individual user and a sub-folder to determine AR task or non-AR task.

3.2.2 Pre-Processing

The collected data is then sent for pre-processing. The log text file is converted into users' action dataset and the image emotion is annotated by VGG19 and converted to users' emotions dataset.

The image file is first checked for the existence of a face with the use of the HaarCascade front face classifier. If the face is not detected, then a record with null emotion is recorded in the dataset. If a face is detected, then it is sent to the VGG19 neural network for emotion detection. The detected emotion is then recorded into the dataset with other helpful information.

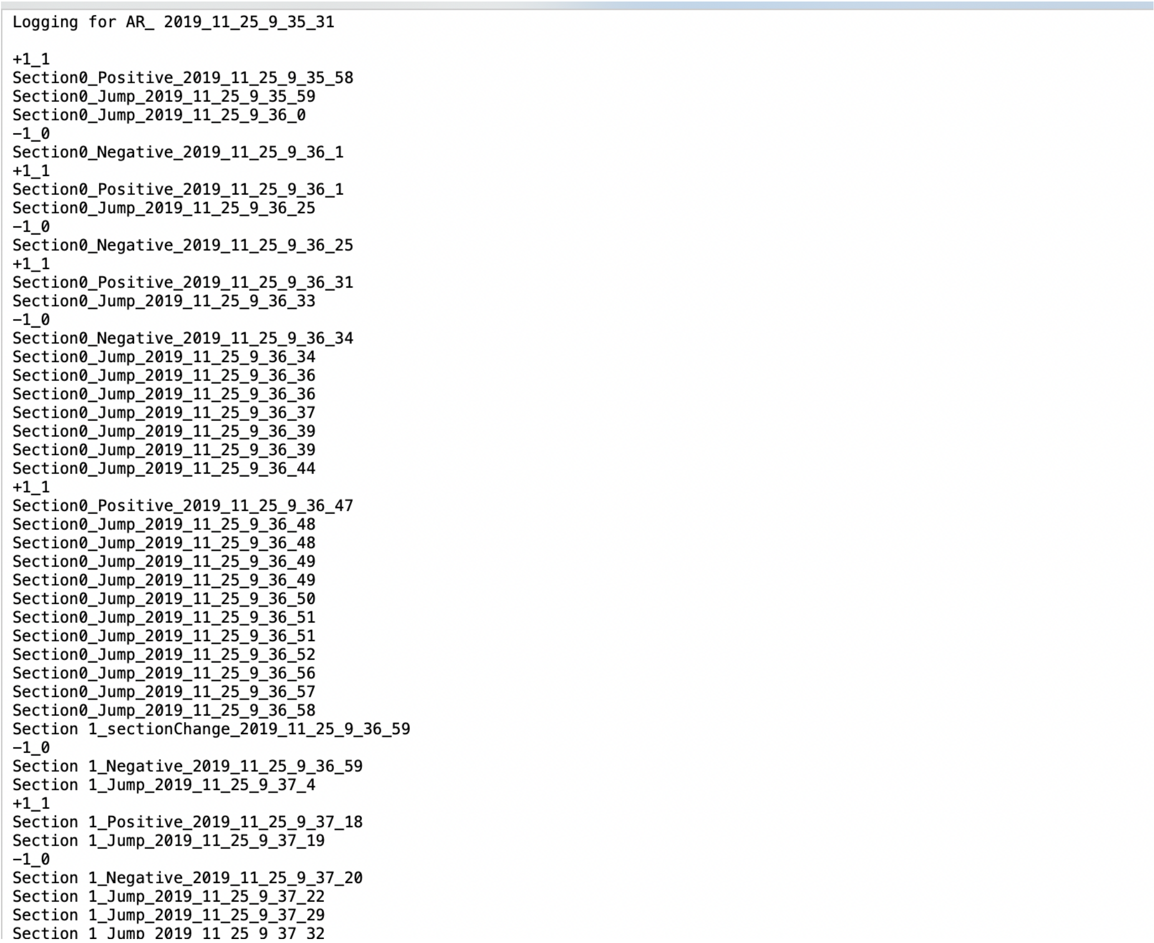
Finally, based on our preliminary round of experimental UI/UX study with four users, we discovered that the time required to complete the AR tasks was significantly longer. It restricted us from designing the user study with the uneven time for conducting non-AR tasks and AR tasks. Therefore, the varying lengths resulted in different amount of images being captured. That's why we normalize the data in the analysis for comparable comparisons.

3.3 Methodology

Users were given two similar tasks. The data from the task is logged to determine the results later.

3.3.1 Scene Settings

I recruited twelve undergraduate students (seven male and five female) from computer science backgrounds. All users have below two years of programming experience. Each user was given a unique identity to keep track of the logs. Simultaneously, the application also records the facial images of the users every second. The opera-



```
Logging for AR_ 2019_11_25_9_35_31

+1_1
Section0_Positive_2019_11_25_9_35_58
Section0_Jump_2019_11_25_9_35_59
Section0_Jump_2019_11_25_9_36_0
-1_0
Section0_Negative_2019_11_25_9_36_1
+1_1
Section0_Positive_2019_11_25_9_36_1
Section0_Jump_2019_11_25_9_36_25
-1_0
Section0_Negative_2019_11_25_9_36_25
+1_1
Section0_Positive_2019_11_25_9_36_31
Section0_Jump_2019_11_25_9_36_33
-1_0
Section0_Negative_2019_11_25_9_36_34
Section0_Jump_2019_11_25_9_36_34
Section0_Jump_2019_11_25_9_36_36
Section0_Jump_2019_11_25_9_36_36
Section0_Jump_2019_11_25_9_36_37
Section0_Jump_2019_11_25_9_36_39
Section0_Jump_2019_11_25_9_36_39
Section0_Jump_2019_11_25_9_36_44
+1_1
Section0_Positive_2019_11_25_9_36_47
Section0_Jump_2019_11_25_9_36_48
Section0_Jump_2019_11_25_9_36_48
Section0_Jump_2019_11_25_9_36_49
Section0_Jump_2019_11_25_9_36_49
Section0_Jump_2019_11_25_9_36_50
Section0_Jump_2019_11_25_9_36_51
Section0_Jump_2019_11_25_9_36_51
Section0_Jump_2019_11_25_9_36_52
Section0_Jump_2019_11_25_9_36_56
Section0_Jump_2019_11_25_9_36_57
Section0_Jump_2019_11_25_9_36_58
Section 1_sectionChange_2019_11_25_9_36_59
-1_0
Section 1_Negative_2019_11_25_9_36_59
Section 1_Jump_2019_11_25_9_37_4
+1_1
Section 1_Positive_2019_11_25_9_37_18
Section 1_Jump_2019_11_25_9_37_19
-1_0
Section 1_Negative_2019_11_25_9_37_20
Section 1_Jump_2019_11_25_9_37_22
Section 1_Jump_2019_11_25_9_37_29
Section 1_Jump_2019_11_25_9_37_32
```

Figure 3.7: Operations Captured by Logger

tions data captured was in the following format.

As shown in 3.7 the log text file is created with the following rules:

1. Application start should trigger the creation of a new log file with a timestamp as the first line.

2. The application should record every action in the user study with the section as the predecessor and timestamp as the successor.

3. Objects collected, Jumps recorded, and section changes should all follow rule 2.

4. Upon collecting an object, the application should also log the respective score.

5. Images stored should have the timestamp as the filename.

3.3.2 Emotion Modeling

Convolutional neural networks are very successful in large-scale image and video recognition (Russakovsky *et al.* (2015); Zeiler and Fergus (2014)). Many attempts were made to improve the original architecture of Krizhevsky *et al.* (2017) (Zeiler and Fergus (2014); Sermanet *et al.* (2014)). Simonyan and Zisserman (2014) address another important factor in the architecture: the depth. As mentioned in Simonyan and Zisserman (2014), VGG19 is a 19 layer variant of the VGG model, a deep convolution neural network model used for image classification. This model has been developed by Visual Geometry Group at Oxford's. VGG19 consists of sixteen convolution layers, three fully connected layers, five Maxpool layers, and one Softmax layer. The layers are defined as follows:

1. Convolution layer: This calculates a convolution operation to the input, passing the result to the next layer.

2. Maximum Pooling (or Max Pooling): Calculates the maximum value for the patch of the feature map. This layer applies a max-pooling operation on its inputs.

3. SoftMax: SoftMax layer produces output by applying the Softmax function as activation to the net input from the previous layer. It transforms the input to be between 0 and 1.

The convolutional layers compute the dot product between their weights and the small regions to which they are linked. Relu and Softmax are used as activation functions that activate the neural network nodes for computation when activation conditions are satisfied. Pooling Layers will down-sample the operation along the

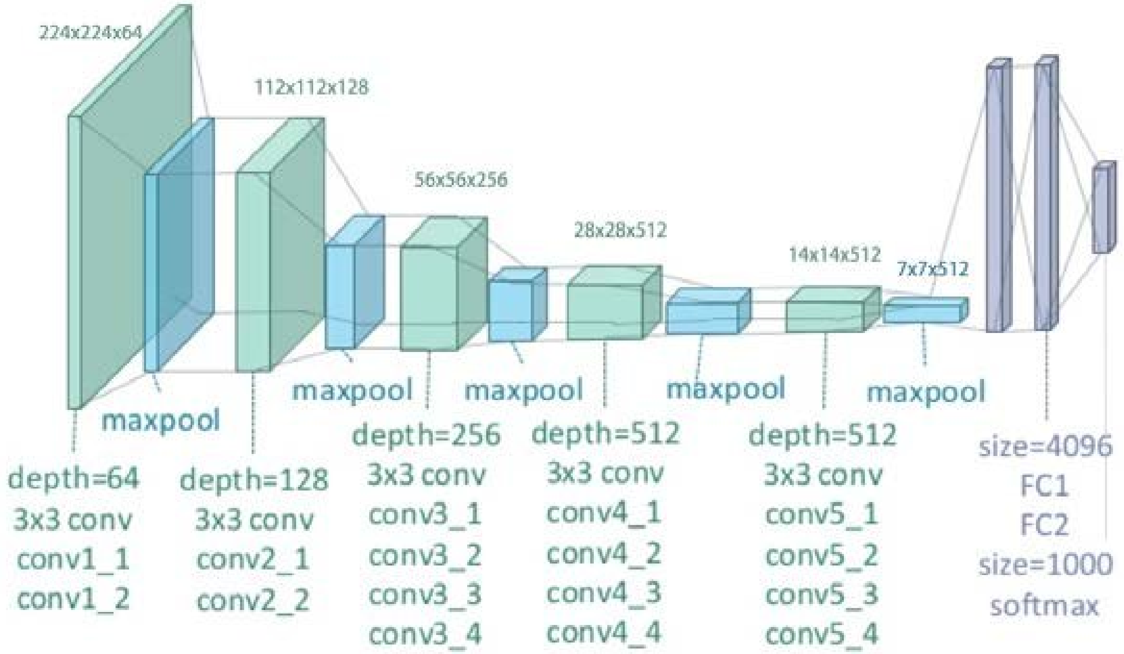


Figure 3.8: Illustration of the Network Architecture of VGG-19 Model: Conv means Convolution, FC means Fully Connected

dimensions. Thus, it reduces the number of parameters to learn and computation performed in the network. The pooling layer summarises the features present in a region of the feature map generated by a convolution layer. This summarization, in turn, helps in minimizing processing power. The dense layers do the decision-making. To prevent overfitting, dropout layers randomly turn off a few neurons in the network. Sometimes, while training the neural network, the training accuracy remains constant without a change. This is known as a plateau. If a plateau is detected in validation loss, the learning rate is altered by a specified factor. The design of the layers is as shown in figure 3.8.

Training and Testing

The objective is to train the model on human faces to detect facial expressions.

Dataset

The paper Goodfellow *et al.* (2013) provides one of the most significant datasets that consist of databases of human faces, the FER2013 dataset. The FER2013 dataset is used to train a deep convolution neural network (in this case, VGG19). The dataset consists of approximately 30,000 images with RGB values of facial images converted to a 48x48 2-dimensional array. The images in this dataset are pre-classified into the seven emotions, namely 0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral. I used the HaarCascade front face classifier to remove the background from the image and only focus on the face. The processed facial images data are then fed into the neural network for training.

Results

The trained classifier can detect seven types of emotions: angry, disgust, fear, happy, sad, surprise, and neutral. The model achieved a private test classification accuracy of 73%.

3.3.3 Emotion Change Pattern

The main objective of this subsection is to identify the users' emotional changes before and after the activity (action/object collection). I identified the moments of activities actions such as object collections and jumps. Once the moment is identified, I cross-referenced the occurrence time-frame to identify the associated facial images. Meanwhile, I used VGG19 emotion detected model to label the images with the emotions. So that allows me to identify the before, current, and after action's emotions. In addition, I classified the emotion change patterns into *Positive*, *Neutral*, and *Negative* emotion valence. Lastly, the trends were observed in both AR and non-AR conditions.

The rules for identifying the before and after emotions are:

1. The emotion recorded before the current activity record should be considered as the before emotion.
    - If the before emotion is similar to the current emotion, repeat step 1 to find a different emotion.
    - If a different activity occurs before an emotion change, then consider the change final and return the current emotion

The same rules apply for the after emotion too. Instead of looking for prior emotion, the program looks for emotions following the current activity. Once the before and after emotions are identified, the emotions are classified into one of the three categories:

- **Positive:** Happy, Fear, Surprise
- **Neutral:** Neutral
- **Negative:** Angry, Sad, Disgust

According to Bantinaki (2012), fear emotion can be regarded as positive valence emotion when it satisfies two conditions, one is that there should be a benefit, and two, there should be a reward. These conditions match perfectly for this research. Hence, I classified fear as a positive valence emotion. Based on the previous classification of emotions, I annotated the sequence of emotions with one of the three following classifications:

- Upward Trend
- Neutral Trend

- Downward Trend

The rules for classifying trends are:

1. **Upward Trend:** If emotion changes from negative to positive valence or neutral to positive valence.
2. **Neutral Trend:** If emotion changes from negative to negative valence or neutral to neutral valence or positive to positive valence.
3. **Downward Trend:** If emotion changes from positive to negative valence or neutral to negative valence.

Table 3.1 shows sequence of emotions and their trends.

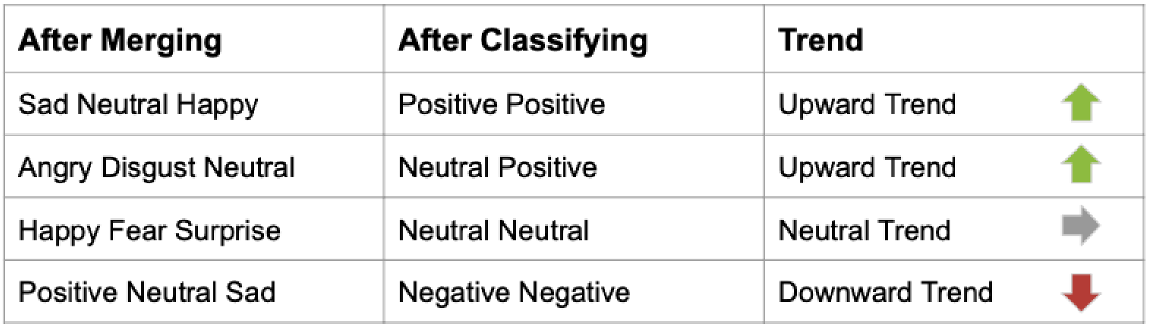
Table 3.1: Pattern Mapping. $< - >$ Indicates Empty Emotion Record

Pattern Mapping	
Emotion Pattern	Mapping
$< - >< - >$	Neutral Trend
$< - >$ disgust	Downward Trend
$< - >$ angry	Downward Trend
$< - >$ sad	Downward Trend
$< - >$ neutral	Neutral Trend
$< - >$ fear	Upward Trend
$< - >$ happy	Upward Trend
$< - >$ surprise	Upward Trend
disgust $< - >$	Neutral Trend
disgust disgust	Neutral Trend

Continuation of Table 3.1

Emotion Pattern	Mapping
disgust angry	Neutral Trend
disgust sad	Neutral Trend
disgust neutral	Upward Trend
disgust fear	Upward Trend
disgust happy	Upward Trend
disgust surprise	Upward Trend
angry $< - >$	Neutral Trend
angry disgust	Neutral Trend
angry angry	Neutral Trend
angry sad	Neutral Trend
angry neutral	Upward Trend
angry fear	Upward Trend
angry happy	Upward Trend
angry surprise	Upward Trend
sad $< - >$	Neutral Trend
sad disgust	Neutral Trend
sad angry	Neutral Trend
sad sad	Neutral Trend
sad neutral	Upward Trend
sad fear	Upward Trend
sad happy	Upward Trend
sad surprise	Upward Trend
neutral $< - >$	Neutral Trend

Continuation of Table 3.1	
Emotion Pattern	Mapping
neutral disgust	Downward Trend
neutral angry	Downward Trend
neutral sad	Downward Trend
neutral neutral	Neutral Trend
neutral fear	Upward Trend
neutral happy	Upward Trend
neutral surprise	Upward Trend
fear $< - >$	Neutral Trend
fear disgust	Downward Trend
fear angry	Downward Trend
fear sad	Downward Trend
fear neutral	Downward Trend
fear fear	Neutral Trend
fear happy	Neutral Trend
fear surprise	Neutral Trend
happy $< - >$	Neutral Trend
happy disgust	Downward Trend
happy angry	Downward Trend
happy sad	Downward Trend
happy neutral	Downward Trend
happy fear	Neutral Trend
happy happy	Neutral Trend
happy surprise	Neutral Trend



After Merging	After Classifying	Trend
Sad Neutral Happy	Positive Positive	Upward Trend ⬆
Angry Disgust Neutral	Neutral Positive	Upward Trend ⬆
Happy Fear Surprise	Neutral Neutral	Neutral Trend ➡
Positive Neutral Sad	Negative Negative	Downward Trend ⬇

Figure 3.9: Example of How Trends are Observed From the Emotion Change Sequence.

Continuation of Table 3.1

Emotion Pattern	Mapping
surprise < − >	Neutral Trend
surprise disgust	Downward Trend
surprise angry	Downward Trend
surprise sad	Downward Trend
surprise neutral	Downward Trend
surprise fear	Neutral Trend
surprise happy	Neutral Trend
surprise surprise	Neutral Trend

End of Table

Figure 3.9 shows an example of emotions after merge, emotion valence mapping, and the trends observed.

3.3.4 Time Series Analysis

Dynamic Time Warping (DTW) is used to measure the similarities or calculate the distance between two temporal sequences, varying in length. It can be used in many scenarios, for example, stock market analysis, sound pattern recognition, etc.

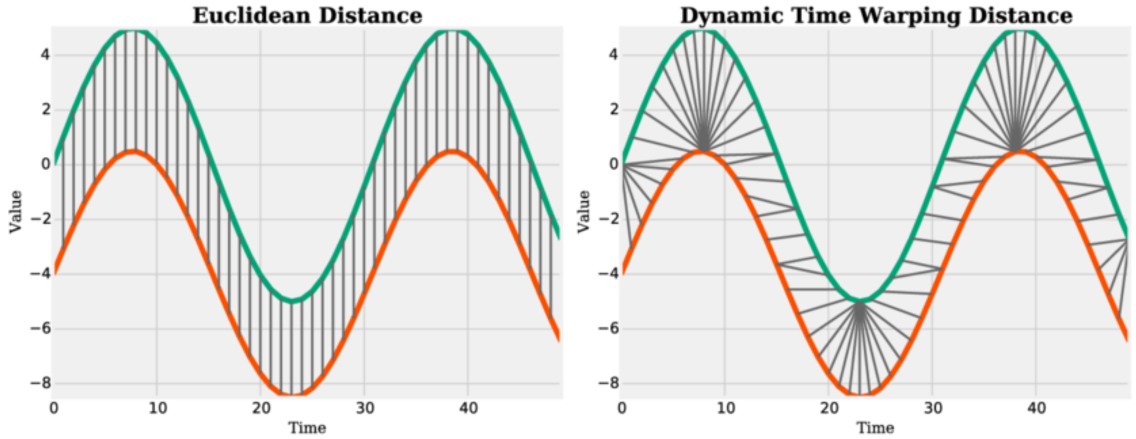


Figure 3.10: Source: Schäfer (2015): One-to-One Mapping(Euclidean Distance) Vs One-to-Many Mapping(Dynamic Time Warping)

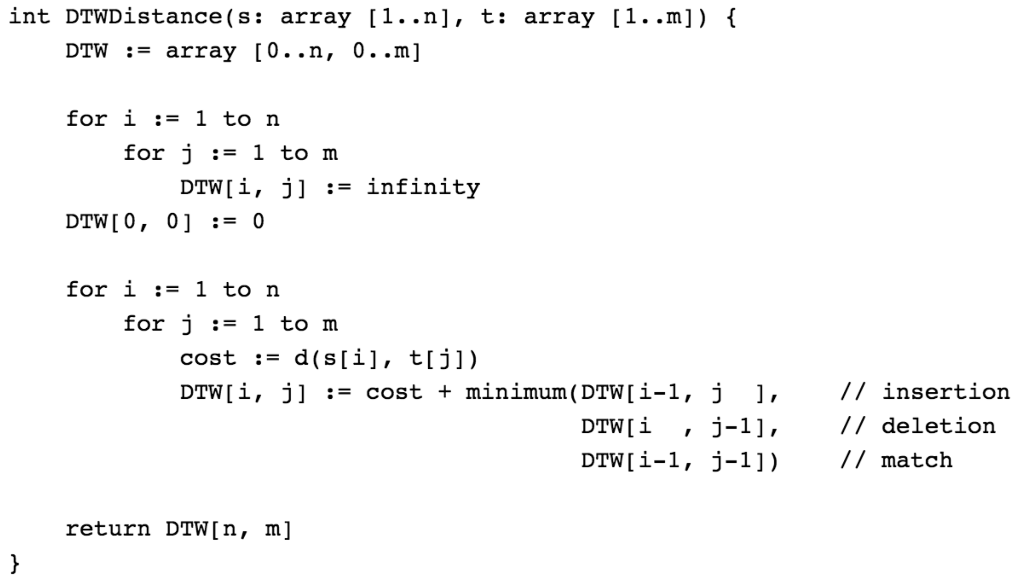
The idea to measure similarity distance is to build one-to-many and many-to-one matches to minimize the total distance between the two temporal sequences.

Consider the two different time-series data, Red and Blue, as seen in 3.10. Although they have similar patterns, the blue curve is larger than the red. Applying one-to-one mapping leaves out the tail of the blue curve. Applying one-to-many mapping by utilizing DTW can overcome this issue.

Rules

Concise rules from Senin (2008) suggest:

- Every index from the primary sequence should be mapped with one or more indices from the other sequence and vice versa.
- The first index from the first sequence must be mapped with the first index from the other sequence (but it does not have to be its only match).
- The last index from the first sequence must be mapped with the last index from the other sequence (but it does not have to be its only match).



```
int DTWDistance(s: array [1..n], t: array [1..m]) {
    DTW := array [0..n, 0..m]

    for i := 1 to n
        for j := 1 to m
            DTW[i, j] := infinity
    DTW[0, 0] := 0

    for i := 1 to n
        for j := 1 to m
            cost := d(s[i], t[j])
            DTW[i, j] := cost + minimum(DTW[i-1, j  ],    // insertion
                                        DTW[i  , j-1],    // deletion
                                        DTW[i-1, j-1])    // match

    return DTW[n, m]
}
```

Figure 3.11: DTW Algorithm

- The mapping of the indices from the first sequence to the other sequence must be monotonically increasing, and vice versa. In other words, if j > i are indices from the first sequence, then there must not be two indices l > k in the other sequence, such that index i is matched with index l and index j is matched with index k, and vice versa.

Many inbuilt modules in python provide an accurate calculation of DTW. For ease of comparison, I used python implementation of DTW based on Salvador and Chan (2007) known as FastDTW. The advantage of FastDTW is that it warps the data into clusters before computing the distance. This clustering enables us to use euclidean distance on the data and reduce the complexity of the distance calculation. Hence, in this implementation, I proceeded with Euclidean distance instead of Manhattan distance.

Dynamic Time Warping

The emotion annotated dataset is used to create seven different emotion sequences for both AR and non-AR. For example, to create a neutral emotion sequence, all the records with 'neutral' as emotion are termed '1', and all other records with an emotion other than neutral were termed '0'. This emotion sequence can be used later in DTW analysis.

Although the VGG19 model exhibits high accuracy while detecting user emotions, there is always room for improving accuracy and minimizing errors. Hence to minimize the error, I used Dynamic Time Warping (DTW). Detecting user engagement with neutral emotions can be unreliable than detecting with non-neutral emotions. Hence I compared the sequences observed from neutral emotion in the user study task to the other non-neutral emotions. The reason for choosing neutral emotion is because it is the highest observed emotion in the dataset. If a non-neutral emotion has a similar sequence to a neutral emotion, it can be considered ineffective to detect users' engagement.

The objective here is to understand the difference in distance between emotions in AR and non-AR tasks. Hence, neutral emotion is compared with every other emotion in the task. Each users' neutral pattern in a task is compared with the other emotions in the task for that user. FastDTW module in python is used to calculate the distance between the patterns. The mean value of twelve users' emotion distance is recorded.

Chapter 4

FINDINGS

Twelve users participated in the user study, performing both the tasks. The users' activity and emotion data are recorded continuously throughout the user study. The recorded data includes

- The number of jumps performed
- The number of non-rewarding objects collected
- The number of rewarding objects collected
- The images in the emotion dataset per user per section

This level of data helps us understand the emotions/stress the user might experience in each section. The activity data and the emotion data are analyzed. The findings help us answer the research questions.

4.1 Findings from Emotion Outcome

The emotions detected with the help of VGG19net are analyzed to provide insights into the user experience. Facial images captured from the user study are converted into 256*256 bit data of RGB values. The trained VGG19 model then annotates the emotion. Seven types of emotions are detected: angry, disgust, fear, happy, sad, surprise, and neutral. In non-AR, the total images collected are $\approx$1900, while in AR, the total images collected are $\approx$3600. Due to camera blur or users covering the front camera, some of these images($\approx$45% in each task) could not be used. To overcome the dissimilarity, the mean value of combined emotions data for all the users per minute are observed. This normalization ensures that the data will always be comparable for analysis.

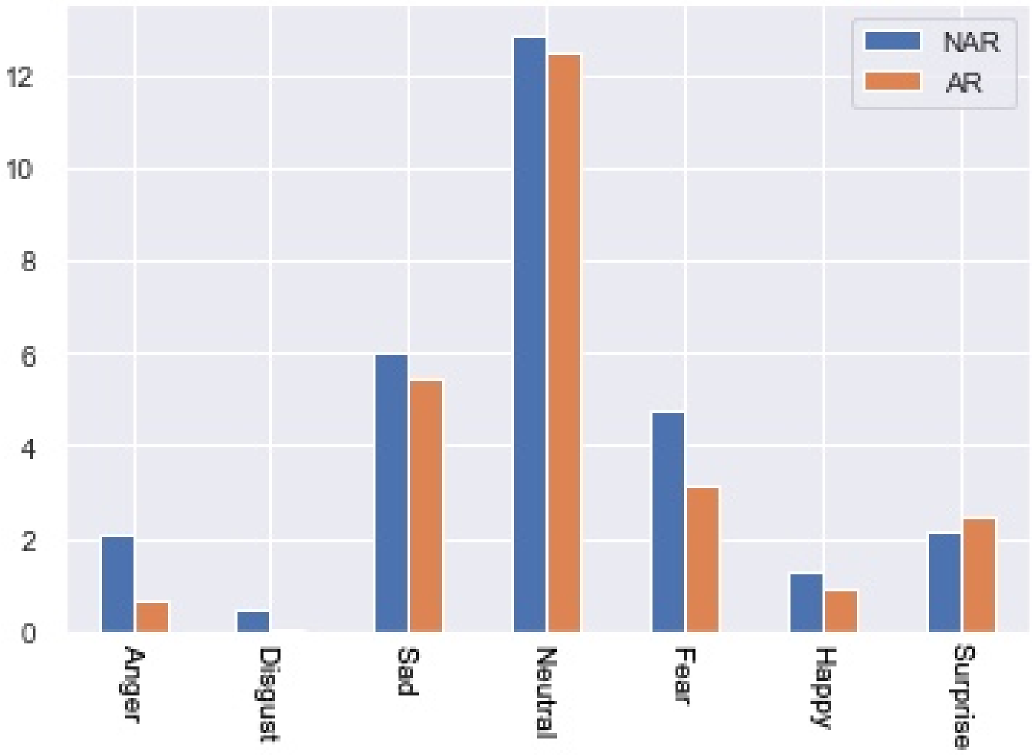


Figure 4.1: Emotion Counts Per Minute in AR and Non-AR Tasks

Figure 4.1 represents the average emotion counts observed per minute from all the users in a task. Six emotions out of seven had higher values in non-AR than AR. Emotions that are angry, disgust, sad, neutral, fear, and happy are higher in non-AR tasks than in AR tasks. The values of neutral emotions mask the entire dataset. In the AR task, forty-nine percent of the total emotions recorded were of the type neutral. While in the non-AR task, neutral emotions were detected forty-four percent of the time. Out of the remaining emotions, users displayed more fear and sad emotions in both AR and non-AR. Disgust emotion was not recorded or negligible in the AR task. Surprise emotion was the only emotion that had higher emotion counts per minute in AR than non-AR.

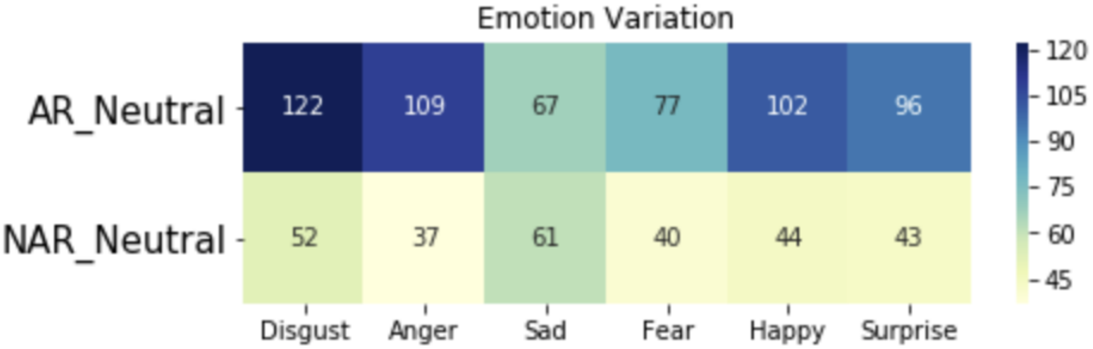


Figure 4.2: Non-Neutral Emotion Distances From Neutral Emotion

Dynamic Time Warping is an important tool used to compare the similarity distance between two signals/patterns of varying time/lengths. DTW is used to understand how effective it is to assess their learning experience based on users' emotion patterns. If the emotions pattern from neutral emotion is similar to any other emotion's pattern, then the similar emotion could be considered ineffective in analyzing users' engagement. If emotion remains constant over time without a change, then relying on VGG19 data for assessing users' engagement might not reveal true values since the VGG19 neural network model only showed an accuracy of 73%. In order to understand the likelihood of assessing users' engagement accurately, DTW is used. Using this methodology, a neutral emotion pattern is compared with every other emotion pattern. As seen in the figure 4.2, emotions patterns recorded in AR are more distant from the neutral emotion than non-AR, meaning the detection of users' experience is higher in AR than non-AR. On the other hand, non-AR emotions show lower distance values than AR, meaning that in comparison, they are more similar to neutral than AR emotions. The emotions disgust, anger, happy, and surprise have high distances in AR compared to non-AR. These values do not indicate that users

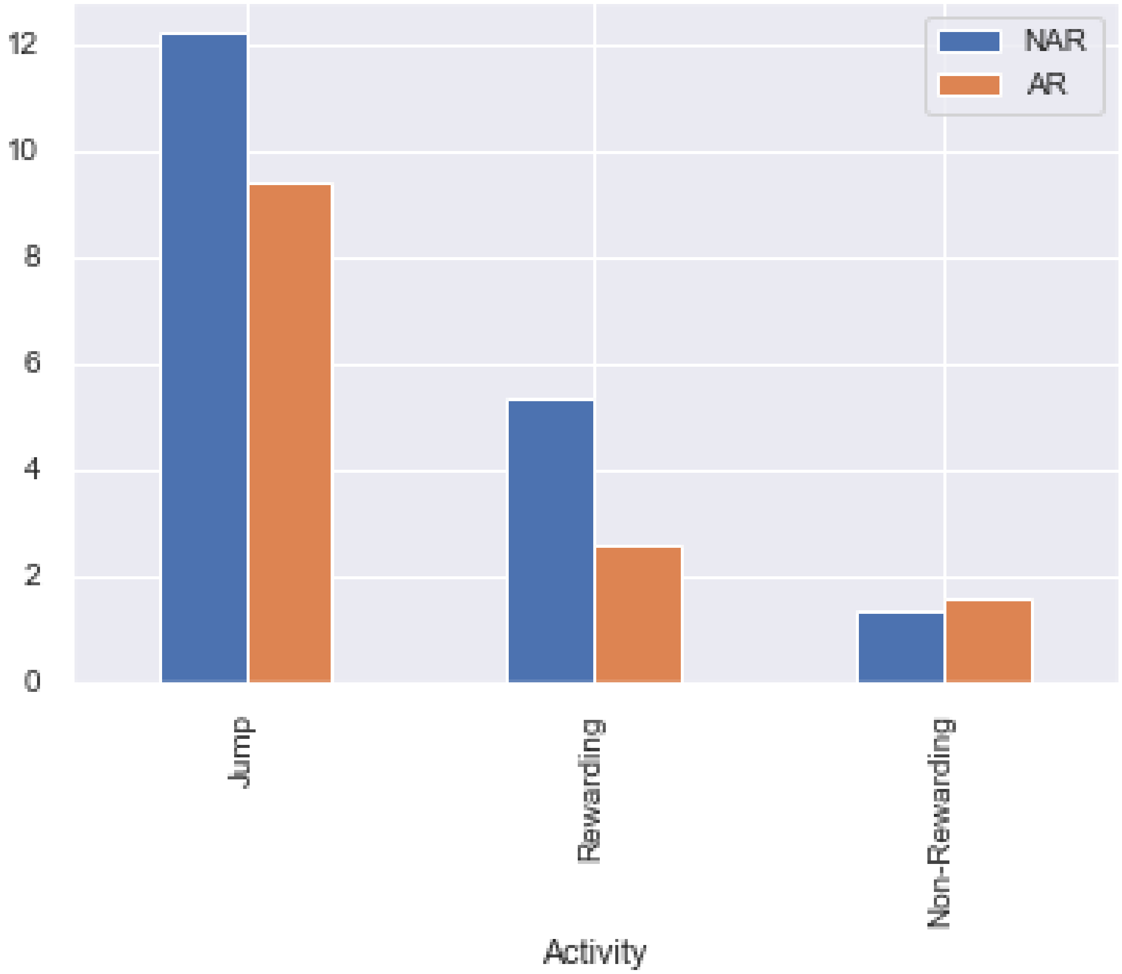
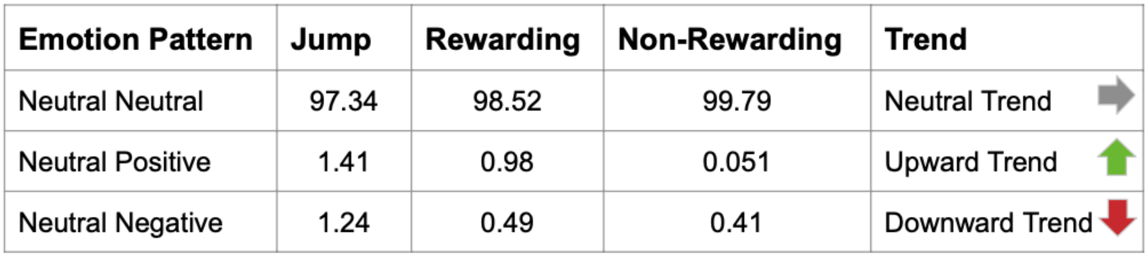


Figure 4.3: Activity Counts Per Minute Observed in AR and Non-AR Tasks

are happier in AR than non-AR. This data only indicates that users' engagement can be analyzed more reliably in AR tasks than non-AR tasks.

4.2 Findings from Activity Outcome

Activity data is classified into three types: jump data, rewarding data, and non-rewarding data. The number of rewarding objects and non-rewarding objects in both AR and non-AR are equal. The mean value of combined activities data for all the users per minute is observed. Figure 4.3 explains the average emotion counts observed per minute from all the users in a task. The jump activity is higher in non-AR tasks



Emotion Pattern	Jump	Rewarding	Non-Rewarding	Trend
Neutral Neutral	97.34	98.52	99.79	Neutral Trend →
Neutral Positive	1.41	0.98	0.051	Upward Trend ↑
Neutral Negative	1.24	0.49	0.41	Downward Trend ↓

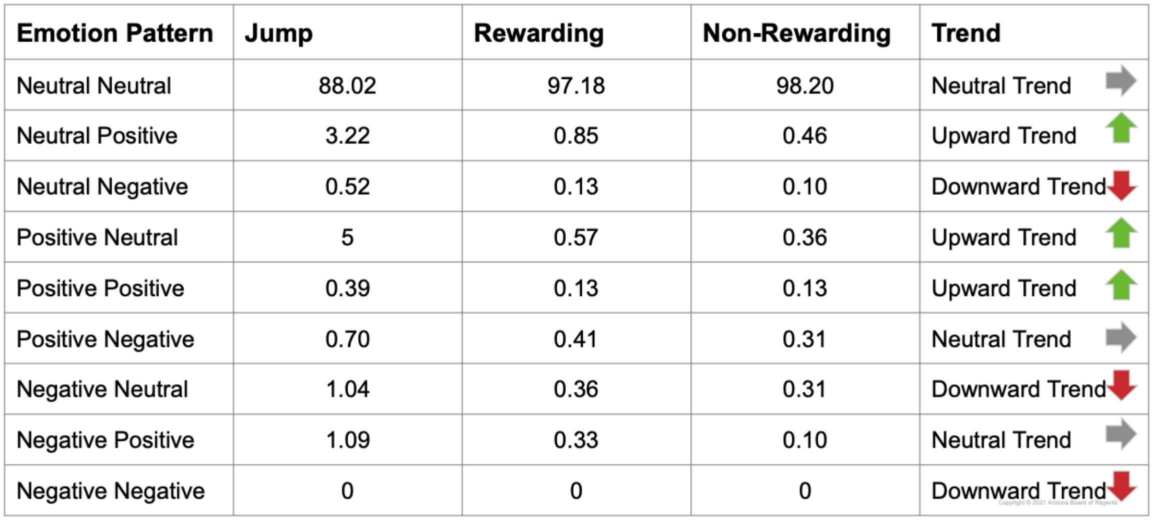
Figure 4.4: Percentage of Number of Emotion Patterns and Trends Observed in Non-AR Task

than AR tasks. The users collected a higher amount of rewarding objects in non-AR tasks than the AR tasks. The collection of non-rewarding objects are lower in non-AR tasks and higher in AR tasks.

4.3 Findings from combination of emotion and activity records

The activity records are cross-referenced with emotion records to determine the emotion valence users are experiencing before, during, and after collecting an object. Figure 4.4 shows the number of such patterns observed in the non-AR task and their percentage. *Neutral* to *Neutral* emotion change can be regarded as no change in emotion valence. Most of the patterns recorded are *Neutral* to *Neutral*. After collecting rewarding objects, the users have a positive valence emotion. Thirty-eight counts of *Neutral* to *Positive* emotion change patterns are observed when the user collected a rewarding object. When collecting a non-rewarding object, users seem to have very few changes from *Neutral* to *Negative*.

In table 4.5, most of the patterns recorded are *Neutral* to *Neutral*. Overall, users' emotions changed from *Neutral* to *Positive* after a jump or collecting rewarding objects, or collecting non-rewarding objects. Many changes in emotion show a change from *Negative* to *Positive* or *Neutral* to *Positive* emotions. Emotion change from *Negative* to *Negative* pattern never happened in the AR task. It has been observed AR task provided more emotion change patterns than non-AR.



Emotion Pattern	Jump	Rewarding	Non-Rewarding	Trend
Neutral Neutral	88.02	97.18	98.20	Neutral Trend →
Neutral Positive	3.22	0.85	0.46	Upward Trend ↑
Neutral Negative	0.52	0.13	0.10	Downward Trend ↓
Positive Neutral	5	0.57	0.36	Upward Trend ↑
Positive Positive	0.39	0.13	0.13	Upward Trend ↑
Positive Negative	0.70	0.41	0.31	Neutral Trend →
Negative Neutral	1.04	0.36	0.31	Downward Trend ↓
Negative Positive	1.09	0.33	0.10	Neutral Trend →
Negative Negative	0	0	0	Downward Trend ↓

Figure 4.5: Percentage of Number of Emotion Patterns and Trends Observed in AR Task

On average, users took five minutes to complete the task in a non-AR setting and twenty minutes in an AR setting. The users produced more patterns in the AR task than in the non-AR task. To substantiate the claim that the AR task is not comparable to the non-AR task, I use the data from figure 4.4 and figure 4.5 and compare the most prominent trends in emotion changes using a chi-square test. The emotion patterns as shown in figure 4.4 and figure 4.5 are grouped into the neutral trend, upward trend, and downward trend. I have grouped the activity in the experiment into jump, rewarding, and non-rewarding. A Chi-square test for each activity has been performed.

The null hypothesis H_0 is that the AR episodes are comparable to the non-AR episodes. The chi-square test is performed with a 5% significance level ($\tilde{\chi_E}^2 = 5.99$). The formula to calculate chi-square critical is as follows:

$$\tilde{\chi_c}^2 = \sum_{k=1}^{n} \frac{(O_k - E_k)^2}{E_k}$$

The subscript “c” is the degrees of freedom. “O” represents the observed value, and E represents the expected value.

Jump Task	Upward Trend	Downward Trend	Neutral Trend
AR (Observed values)	206	239	3395
non-AR (Expected values)	49	43	3375

Table 4.1: Number of Jump Activity Trends Detected

Rewarding Task	Upward Trend	Downward Trend	Neutral Trend
AR (Observed values)	60	43	3737
non-AR (Expected values)	38	19	3810

Table 4.2: Number of Rewarding Activity Trends Detected

Tables 4.1, 4.2, and 4.3 represent the number of trends detected in jump, rewarding and non-rewarding activity respectively. The chi-square critical analysis is performed on the observed and expected values from the above tables. The obtained chi-square critical value for the activity are:

1. **Jump activity**: $\tilde{\chi_c}^2 = 1396.548$
2. **Rewarding activity**: $\tilde{\chi_c}^2 = 44.44$
3. **Non-Rewarding activity**: $\tilde{\chi_c}^2 = 609.77$

In the above three scenarios, chi-square critical is significantly greater than chi-square expected that is $\tilde{\chi_c}^2 >> \tilde{\chi_E}^2$. Thus, the Null Hypothesis was rejected. AR showed significantly different emotion patterns compared to Non-AR.

Non-Rewarding Task	Upward Trend	Downward Trend	Neutral Trend
AR (Observed values)	34	30	3776
non-AR (Expected values)	2	6	3859

Table 4.3: Number of Non-Rewarding Activity Trends Detected

Chapter 5

CONCLUSION

5.1 Summary

In this work, an augmented reality application is developed in Unity. The application can capture both operations and facial images. A deep neural network, VGG19, has been trained using the FER2013 dataset. This model can annotate the user emotions with 72% accuracy. Also, a within-subject user study was designed to compare users' engagement in AR and non-AR tasks. Twelve undergraduate students participated in the study. Each user engaged in both AR and non-AR tasks. The data was gathered and pre-processed to feed into the emotion recognition neural network, and DTW models the activity data for emotion pattern mining. The results help us to answer the research questions.

5.2 Research Questions

In the following section, we sought to address the research questions:

R1) Can users' emotions be measured to gauge their engagements in AR/non-AR by using Dynamic Time Warping?

Figure 4.2 show that users produce more distinct non-neutral emotions in AR tasks when compared with the non-AR task. The more distant the values are from neutral emotion, the more engaged the user is regardless of the type of emotion. Emotions like happy, angry, disgust, and surprise have more distinct values in AR when compared to non-AR. As more non-neutral emotions in AR have higher distance values than neutral emotions, the users' engagement can be detected/understood with higher accuracy. The non-neutral emotions in non-AR are not as distant from the neutral emotion as AR emotions are. Although, because of this reason, the detection of users' engagement can be harder, it is not impossible. Low distance means DTW can still predict users' engagement in non-AR but with lower accuracy.

R2) What are the emotional pattern differences observed in AR versus non-AR environments?

The patterns observed in AR are much higher than the patterns observed in non-AR. In the post-task phase, the users' majority of the users (nine out of twelve users) mentioned that the non-AR task seems mundane and the AR task seemed more interesting. The patterns show a similar reality. In the non-AR task, users did not put much effort into clearing the task. As the controls were easy for users to follow in the non-AR task, they completed the task quickly with minimum changes in patterns. On the other hand, as the AR task was new, users took an ample amount of time to get used to the controls and displayed higher upward emotion trends.

The number of upward trends, downward trends, and neutral trends observed in AR tasks are higher compared to non-AR tasks. Having more knowledge of users' emotion changes helps in better understanding/ preparing for future outcomes. Having to deal with minimal emotion change patterns is more challenging and can affect analyzing users' engagement. If a user collects a rewarding object and the emotion pattern changes from neutral to positive valence emotion after collecting the object, it can be considered a helpful pattern. However, if the exact pattern change is observed when the user collects a non-rewarding object, then the pattern can be regarded as ineffective in analyzing the user's engagement. In such scenarios, it is helpful to have more patterns of emotion change to identify the engagement the user is experiencing accurately.

This result proves that AR-task brings out realistic emotion values when collecting rewarding or non-rewarding objects. Users are likely to experience positive feelings when doing something right and negative when doing something wrong in the AR task than the non-AR task.

5.3 Limitation

One of the major limitations in the current work is the quality of the collected facial images quality. Some of the data was excluded due to camera blurriness or covered by hand from the front camera. Additionally, the capacity of the hardware was limited, specifically (a) the ram on the device only permitted to capture one frame per second and, (b) the three-dimensional content in AR can only be projected by using an image target.

When a user moves the camera more in the AR setting, the images can appear very blurry. As the movement in the AR task could be more complex than the non-AR task, for instance, the users have to move their device to focus on the sphere/map, such a scenario could cause multiple blurry images. These blurry images might be why the detected emotion records in the AR tasks were fewer than the non-AR tasks. Also, the room's lighting produced a glow in some parts of the users' facial images. This change of lighting often resulted in facial emotions not being detected. Therefore, the higher computing power of mobile devices is desired to capture the quality resolution of facial images.

Using an image target limits the amount of AR content that can be displayed on the screen. While using Vuforia SDK may seem like the better option for AR application, each user input/ movement needs to be converted from the device level to the software level and processed by the SDK. Once Vuforia is done processing, it will relay the commands to the device, decoding them and displaying the projections on the screen. The mobile device could not handle Vuforia XR technology to support extended reality options and crashed all the time. Hence, image target is used for more straightforward computation.

Moreover, the application only focused on few types of operations. Having more diverse operations can help in better understanding users' experience or thought processes in the tasks.

5.4 Recommendations and Future Work

Using a tripod and controlled lighting in the user study will improve the quality of the dataset. The paper Wei *et al.* (2018) talks about how the lighting conditions in user study can be controlled to utilize transdermal optical imaging to identify the user's stress level. Using this method could help collect additional implicit feedback. User study can use the feedback to determine user experience. Vuforia XR technology enables the device to be moved outside of a defined area and still save AR content. Using better devices like smartphones with lidar sensors or laser sensors can natively calculate the positions of the AR objects on the device level and display them on the screen without any further need of processing. This performance upgrade makes the application faster and improves the smoothness of controls users felt was very lagging/tough in AR tasks. Sequential analysis models or prediction models such as Bayesian neural networks can help identify the missing emotions in the dataset based on previous emotion sequences. The data set that is used to train the model is the FER2013 dataset. Although it is referred to as one of the best data sets available online to train an emotion recognition model, better datasets such as FER2013+ or CK+ can be used to increase the efficiency of the neural network.

VGG19 model can annotate seven different emotions: anger, disgust, sad, neutral, fear, happy, and surprise. As mentioned in Chapter 4, around forty-five percent of the emotions recorded in the dataset are neutral emotions. Furthermore, the reason for this could be because the user is just concentrating on the task. Although emotions can describe the user's emotions, it will be better to develop/ train a model that can identify emotions like boredom or concentration in the learning environment. These emotions can demonstrate better when a user is excelling or falling behind in the learning environment.

During the user study, a third-party audio recording device is also used to record the audio of the entire user study. Although users were encouraged to talk throughout the user study to analyze their thought patterns, they did not speak. They were pretty involved in the task to express their thought process. Introducing a reward/ bonus for talking can make the users talk during the study when the task becomes interesting.

REFERENCES

Azuma, R., Y. Baillot, R. Behringer, S. Feiner, S. Julier and B. MacIntyre, "Recent advances in augmented reality", IEEE computer graphics and applications **21**, 6, 34–47 (2001).

Bantinaki, K., "The paradox of horror: Fear as a positive emotion", The Journal of Aesthetics and Art Criticism **70**, 4, 383–392 (2012).

Bellman, R. and R. Kalaba, "On adaptive control processes", IRE Transactions on Automatic Control **4**, 2, 1–9 (1959).

Bellucci, A., A. Ruiz, P. Díaz and I. Aedo, "Investigating augmented reality support for novice users in circuit prototyping", in "Proceedings of the 2018 International Conference on Advanced Visual Interfaces", pp. 1–5 (2018).

Bujak, K. R., I. Radu, R. Catrambone, B. MacIntyre, R. Zheng and G. Golubski, "A psychological perspective on augmented reality in the mathematics classroom", Computers & Education **68**, 536–544 (2013).

Cheng, K.-H. and C.-C. Tsai, "Affordances of augmented reality in science learning: Suggestions for future research", Journal of science education and technology **22**, 4, 449–462 (2013).

Dede, C., "Immersive interfaces for engagement and learning", science **323**, 5910, 66–69 (2009).

FitzGerald, E., R. Ferguson, A. Adams, M. Gaved, Y. Mor and R. Thomas, "Augmented reality and mobile learning: the state of the art", International Journal of Mobile and Blended Learning (IJMBL) **5**, 4, 43–58 (2013).

Goodfellow, I. J., D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee *et al.*, "Challenges in representation learning: A report on three machine learning contests", in "International conference on neural information processing", pp. 117–124 (Springer, 2013).

Harley, J. M., E. G. Poitras, A. Jarrell, M. C. Duffy and S. P. Lajoie, "Comparing virtual and location-based augmented reality mobile learning: emotions and learning outcomes", Educational Technology Research and Development **64**, 3, 359–388 (2016).

Heidig, S., J. Müller and M. Reichelt, "Emotional design in multimedia learning: Differentiation on relevant design features and their effects on emotions and learning", Computers in Human Behavior **44**, 81–95 (2015).

Ibáñez, M. B., Á. Di Serio, D. Villarán and C. D. Kloos, "Experimenting with electromagnetism using augmented reality: Impact on flow student experience and educational effectiveness", Computers & Education **71**, 1–13 (2014).

Jerry, T. F. L. and C. C. E. Aaron, “The impact of augmented reality software with inquiry-based learning on students’ learning of kinematics graph”, in “2010 2nd international conference on education technology and computer”, vol. 2, pp. V2–1 (IEEE, 2010).

Joan, D., “Enhancing education through mobile augmented reality.”, Journal of Educational Technology **11**, 4, 8–14 (2015).

Keogh, E. J. and M. J. Pazzani, “Derivative dynamic time warping”, in “Proceedings of the 2001 SIAM international conference on data mining”, pp. 1–11 (SIAM, 2001).

Klopfer, E. and K. Squire, “Environmental detectives—the development of an augmented reality platform for environmental simulations”, Educational technology research and development **56**, 2, 203–228 (2008).

Krizhevsky, A., I. Sutskever and G. E. Hinton, “Imagenet classification with deep convolutional neural networks”, Communications of the ACM **60**, 6, 84–90 (2017).

Lai, A.-F., C.-H. Chen and G.-Y. Lee, “An augmented reality-based learning approach to enhancing students’ science reading performances from the perspective of the cognitive load theory”, British Journal of Educational Technology **50**, 1, 232–247 (2019).

Mayer, R. E., “Incorporating motivation into multimedia learning”, Learning and Instruction **29**, 171–173 (2014).

Merchant, Z., E. T. Goetz, L. Cifuentes, W. Keeney-Kennicutt and T. J. Davis, “Effectiveness of virtual reality-based instruction on students’ learning outcomes in k-12 and higher education: A meta-analysis”, Computers & Education **70**, 29–40 (2014).

Milgram, P., H. Takemura, A. Utsumi and F. Kishino, “Augmented reality: A class of displays on the reality-virtuality continuum”, in “Telemanipulator and telepresence technologies”, vol. 2351, pp. 282–292 (International Society for Optics and Photonics, 1995).

Myers, C., L. Rabiner and A. Rosenberg, “Performance tradeoffs in dynamic time warping algorithms for isolated word recognition”, IEEE Transactions on Acoustics, Speech, and Signal Processing **28**, 6, 623–635 (1980).

Nadolny, L., “Interactive print: The design of cognitive tasks in blended augmented reality and print documents”, British Journal of Educational Technology **48**, 3, 814–823 (2017).

Nincarean, D., M. B. Alia, N. D. A. Halim and M. H. A. Rahman, “Mobile augmented reality: The potential for education”, Procedia-social and behavioral sciences **103**, 657–664 (2013).

Pekrun, R., “The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice”, Educational psychology review **18**, 4, 315–341 (2006).

Pekrun, R. and R. P. Perry, “Control-value theory of achievement emotions”, in “International handbook of emotions in education”, pp. 130–151 (Routledge, 2014).

Poitras, E. G., J. M. Harley and Y. S. Liu, “Achievement emotions with location-based mobile augmented reality: An examination of discourse processes in simulated guided walking tours”, British Journal of Educational Technology **50**, 6, 3345–3360 (2019).

Radu, I. and B. Schneider, “What can we learn from augmented reality (ar)?”, in “Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems”, pp. 1–12 (2019).

Russakovsky, O., J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein *et al.*, “Imagenet large scale visual recognition challenge”, International journal of computer vision **115**, 3, 211–252 (2015).

Salvador, S. and P. Chan, “Toward accurate dynamic time warping in linear time and space”, Intelligent Data Analysis **11**, 5, 561–580 (2007).

Santos, M. E. C., A. Chen, T. Taketomi, G. Yamamoto, J. Miyazaki and H. Kato, “Augmented reality learning experiences: Survey of prototype design and evaluation”, IEEE Transactions on learning technologies **7**, 1, 38–56 (2013).

Schäfer, P., “Scalable time series similarity search for data analytics”, (2015).

Senin, P., “Dynamic time warping algorithm review”, Information and Computer Science Department University of Hawaii at Manoa Honolulu, USA **855**, 1-23, 40 (2008).

Sermanet, P., D. Eigen, X. Zhang, M. Mathieu, R. Fergus and Y. LeCun, “Overfeat: Integrated recognition, localization and detection using convolutional networks. 2nd international conference on learning representations, iclr 2014”, in “2nd International Conference on Learning Representations, ICLR 2014”, (2014).

Simonyan, K. and A. Zisserman, “Very deep convolutional networks for large-scale image recognition”, arXiv preprint arXiv:1409.1556 (2014).

Smiderle, R., S. J. Rigo, L. B. Marques, J. A. P. de Miranda Coelho and P. A. Jaques, “The impact of gamification on students’ learning, engagement and behavior based on their personality traits”, Smart Learning Environments **7**, 1, 1–11 (2020).

Wei, J., H. Luo, S. J. Wu, P. P. Zheng, G. Fu and K. Lee, “Transdermal optical imaging reveal basal stress via heart rate variability analysis: a novel methodology comparable to electrocardiography”, Frontiers in psychology **9**, 98 (2018).

Yuen, S. C.-Y., G. Yaoyuneyong and E. Johnson, “Augmented reality: An overview and five directions for ar in education”, Journal of Educational Technology Development and Exchange (JETDE) **4**, 1, 11 (2011).

Zeiler, M. D. and R. Fergus, “Visualizing and understanding convolutional networks”, in “European conference on computer vision”, pp. 818–833 (Springer, 2014).