Spatial Note System using Virtual-Agent to Enhance Family Connection

Puxuan Qu
44161530-1

Master of Engineering

Supervisor: Prof. Jiro TANAKA
The Graduate School of Information, Production and Systems
Waseda University

July 2018

Abstract

With the pace of life becoming faster, some people are too busy to communicate with their families. In this research, we proposed an augmented reality system in smart-phone to leave emotional notes in the real environment using Virtual-Agent to enhance family connection. The crucial part of the system focuses on the emotional short voice messages exchange with the Virtual-Agent. Spatial note system is based on two parts, one is the virtual agent services and the other is the AR (augmented reality) system. The virtual agent service allows users to make voice messages by recording users' voice and is able to detect the users' voice after recording it. Then generates a Virtual-Agent with the appropriate facial expression to express users' emotion. Users can also change the facial expresses as they like to express feeling. There are 4 emotions of the virtual agents: happy, sorrow, angry, calm. Each emotion has 4 levels to express. The AR (augmented reality) system can detect the nearest plane in the real environment through the phone's camera. Users can put the Virtual-Agent anywhere they want in the reality according to the AR system.

Through this system, family members can exchange information, share feeling, discuss topics, and what's more, bond their relationships. For the note-making user, he can make any note he wants in the real world and conveys his emotion to the other users easily. For the note-receiving user, he can freely choose whether to hear it according to the agent's facial expression. He can assume the others' feelings today by seeing agents' faces intuitively. We conducted a experiment to evaluate the usability of the system and positive feedback has been received.

Keywords: Augmented Reality, Virtual Agents, Emotional Notes

Acknowledgements

Thanks to Professor Jiro TANAKA who gives me great advice and points me the right direction of my research. Thanks to him for opening my mind, finding my mistakes and helping me fix them. Without his patient guidance, I won't complete this. And also I want to thank my laboratory members for inspiring me, taking my experiments and encouraging me when I felt hard.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

With the pace of life becoming faster, some people are too busy to communicate with their families. When parents are busy with their work, the children are trying to preform better at school. Even at home, children are hard to chat with parents because parents stay out late. Without sharing life with each other, their family connection probably becomes weaker as time flies.

The best way to enhance the family connection is to communicate as much as possible. Based on that, we assume to find an ideal to share each other's life without taking up their work time and disturbing the others.Messages exchange is an appropriate method in sharing information. The remaining ways to exchange messages are:

1. Leaving a sticky note at home (Fig 1.1 (a)). This way is able to let the others easy to understand what you want to express because the notes are environment supportive. It increases empathy. On the contrary, these notes are easy to be lost.

2. Using your Smart phones to send messages in LINE (Fig 1.1 (b)). By this approach, users can read messages momentarily, despite time and location. It also can transfer the real voice of users' families. But on the other hand, it is non-environment supportive.

So we proposed an augmented reality system in smart-phone to leave emotional notes in the real environment using Virtual-Agent to enhance families connection. The crucial part of the system focuses on the emotional short voice message exchange with the Virtual-

(a) Sticky notes on fridge (b) LINE

Fig. 1.1 Two ways to communicate by messages exchange

Agent. Spatial note system allows users to make voice message by recording users' voice. Additionally, the system is able to detect the users' voice after recording it and then generate a Virtual-Agent with the appropriate facial expression to express users' emotion. Users can put the Virtual-Agent anywhere they want in the reality by using their smart phones. The virtual agents can detect four main emotions from the user's voice, they are calm, happy, angry and sorrow that are showing below. These environment-supportive messages can help families to start small conversation, talking about daily life without face to face. With the help of this system, family members can exchange information, share feeling, discuss topics, and what's more, enhance their connection.

Chapter 2

Research Background

2.1 Augmented Reality

2.1.1 Definition

In one of the best definitions, a researcher named Ron Azuma says that Augmented Reality is technology that has three key requirements[1]:

- It combines real and virtual content,
- It is interactive in real time, and
- It is registered in 3D.

In the bigger picture, Augmented Reality is a new method or a new technology to make computer interfaces invisible and enhance user interaction with the real world[2]. AR allows the user to see the real world, with virtual objects superimposed upon or composited with the real world. Therefore, AR supplements reality, rather than completely replacing it. Ideally, it would appear to the user that the virtual and real objects coexisted in the same space[1]. Augmented reality is a direct or indirect live view of a physical, real-world environment whose elements are "augmented" by computer-generated perceptual information, ideally across multiple sensory modalities, including visual, auditory, haptic, somatosensory, and olfactory [3].

The main value of AR (augmented reality) is that it brings digital parts of the virtual world into the real world form people's perception, and it's not just a display of data, but also perceived as natural parts of an environment through the combination of immersive sensations. In the early 1990s, the first functional AR systems were invented. It provided immersive mixed reality experiences for users, starting with the Virtual Fixtures system developed at the U.S. Air Force's Armstrong Laboratory in 1992[4].

Augmented reality is used to augment the natural environments and provides perceptually enhanced experiences. With the advanced AR technologies' development, the information of the user's surroundings becomes interactive and digitally manipulable[2][5].Augmented Reality is related to two largely terms:mixed reality and computer-mediated reality.

2.1.2 Taxonomy

Milgram and Kishino[6] introduced the concept of "mixed reality" that combines the real world and the virtual world, and the concept of mixed reality continuum (see in Fig 2.1). Mixed reality continuum is a taxonomy of various methods of merging "virtual" and "real" elements together. In the virtual environment (VE), the user's worldview is completely replaced by computer-generated virtual content. And in a real environment (RE), any user's point of view is not replaced by virtual content[7]. At the end of the virtual environment, extended virtualization replaces most of the user's views with computer graphics, but there is still a perception of the real world. Finally, the augmented reality approaches the end of the RE, and the virtual hint improves the user's real world view[8]. As more or less virtual content is added to the AR scenario, the interface is closer to or farther away from the VE or RE endpoint[9]. The main lesson of this taxonomy is that the AR interface does not exist as a discrete point between real and virtual experiences but can appear anywhere in a mixed reality continuum.

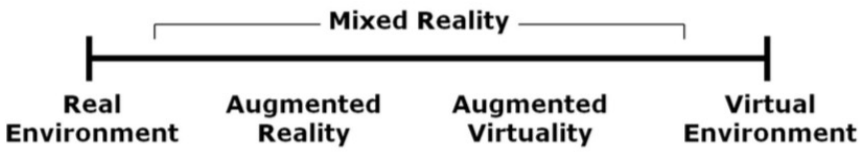


Fig. 2.1 Mixed Reality continuum of Milgram. [Milgram and Kishino, 1994]

2.1.3 Techniques

At present, augmented reality hardware includes processors, displays, sensors, and input devices. The latest mobile computing devices, such as smartphones and tablets, containing cameras and MEMS sensors that includes accelerometers, GPS, solid state compasses and others, are the appropriate augmented reality platforms[10]. A head-mounted display (HMD) is a display device connected to the forehead, such as a harness or a helmet. HMD places images of the physical world and virtual objects in the user's view. Modern HMDs typically use sensors for six degrees of freedom monitoring, allowing the system to align the virtual world with the physical world and adjust based on the movement of the user's head[11]. In the Fig.2.2, there is a head-mounted display and its brand is epson moverio bt-300.

Fig. 2.2 HMD: Epson moverio bt-300

In the aspect of the software of the augmented reality field, an important indicator of the AR system is how to truly integrate enhancement with the real world. Software must obtain real-world coordinates that are independent of the camera from the camera image. This process is called image registration and uses various computer vision methods related to video vision[12]. Many computer vision methods of augmented reality are inherited from visual odometry. Usually, these methods consist of two parts. The first step is to detect interest points, fiducial marks, or optical streams in the camera image. This step can use feature detection methods such as corner detection, spot detection, edge detection or thresholding, and other image processing methods[13]. To enable rapid development of augmented reality applications, some software development kits (SDKs) have emerged. AR SDKs are offered by Vuforia, ARToolKit, Catchoom CraftAR Mobinett AR,Wikitude, Blippar Layar, Meta and ARLab.

2.1.4 AR Applications in Annotation Visualization

Azuma[1] summarized at least six classes of potential AR applications have been explored: medical visualization, maintenance and repair, annotation, robot path planning, entertainment, and military aircraft navigation and targeting. The next describes work that has been done in annotation visualization.

In the past more than 10 years, AR technology has been verified in many different annotation applications.

Billinghurst[14] proposed a system called MagicBook. This is especially true when you concentrate on books by reading by using AR to stack interactive 3D digital content on book pages. AR can be used to transition from reading real books to exploring immersive virtual reality spaces. Since the first MagicBook prototype was developed, a wide range of other AR enhancement books have been developed by academics and companies, some of which have been commercialized since 2008.

The HIT Lab NZ developed a mobile AR application CCDU AR [15]. CCDU AR is an application designed to show building information, which is superimposed on the print pages of books made by the City Council to explain the city's reconstruction and display the conceptual image of the future landmark. Users can use their tablets or smartphones, and when they run the application and put the camera point to the city's print map, they will see the 3D virtual label appear, showing the key buildings they are rebuilding in the city.

2.2 Virtual Agent

2.2.1 Definition

In computer science, a virtual agent is a computer program that acts for a user or other program in the agent context. Agents are colloquially known as bots, from robot. They can be embodied as software such as chat bots, for example when performing a robot body pairing, or on a phone (eg Siri) or other computing device[16]. Virtual agents can be autonomous and can also work with other agents and personnel. Virtual agents interacting with humans (eg, chatbots, human-computer interaction environments) have human-like qualities such as natural language understanding, discourse, personality, or human form[17]. Due to advances in artificial intelligence and cognitive computing programs, virtual agent has improved over the past five years. Current interactive technologies allow virtual agents to go far beyond Interactive Voice Response (IVR) systems. Virtual agents can understand customer intent and provide personalized answers to customer questions, such as people. Virtual agents typically communicate with customers on corporate websites via email or chat in real time[18]. In the latter case, avatars are often used to provide a visual representation of a virtual agent.

2.2.2 Taxonomy and Techniques

Haag proposed four essential types of intelligent virtual agents[19]:

1. Buyer agents or shopping bots,
2. User or personal agents,
3. Monitoring-and-surveillance agents, and
4. Data-mining agents.

The buyer agent moves around a network (eg, the Internet) searching for goods and service information. These agents are also known as "shopping robots" and are compatible with CDs, books, electronic components and all other products. The buyer's agent is usually optimized to make it possible to use digital payment services for e-commerce and traditional services[20].User agents, or personal agents, are intelligent agents that take action on your behalf[21].Monitoring and monitoring agents are used to monitor and report devices, usually computer systems. Agents can track the company's inventory levels, monitor the prices of competitors and return them to the company, and monitor inventory operations through insider trading and rumors.[22] Data-mining agents use information technology to retrieve rich information trends and patterns from a variety of sources[23]. Users can sort this information and find the information they are looking for.

2.2.3 Applications

In the medical health area, Bickmore[24] describe the design and evaluation of a virtual agent that explains health documents to patients. The motivation foe the agents is low health literacy. These agents provides face-to-face interaction with patients, which is like communicate with the health providers. They analyze the form and distribution of pointing gestures used by experts in explaining health documents, and use this data to develop a agent to do document explanation. Preliminary results indicate that patients are more satisfied with health document explanation by a virtual agent compared to a human.

In the personal assistant area, wiend[25] presents the virtual anatomy assistant Ritchie that monitors the user's actions in a physical space and dynamically responds to them. Augmented reality opens up new ways for humans to interact with Embodied Conversational Agents (ECAs) since they allow users to meet ECAs in the physical space. But attempts to integrate ECAs as digital overlays in a physical space have been rare. High demands such an integration puts to the animation of ECAs as virtual augmentations of the physical space, their capabilities to perceive not only the virtual, but also the physical world as well as reactive behavior control. So in the research, they describe their technical contributions towards solving these challenges.

Chapter 3

Goal and Approach

3.1 Goal

The goal of my research is to create a useful and novice system to help busy families to share, to communicate, to exchange information and eventually, to enhance family connection, bond relationships.

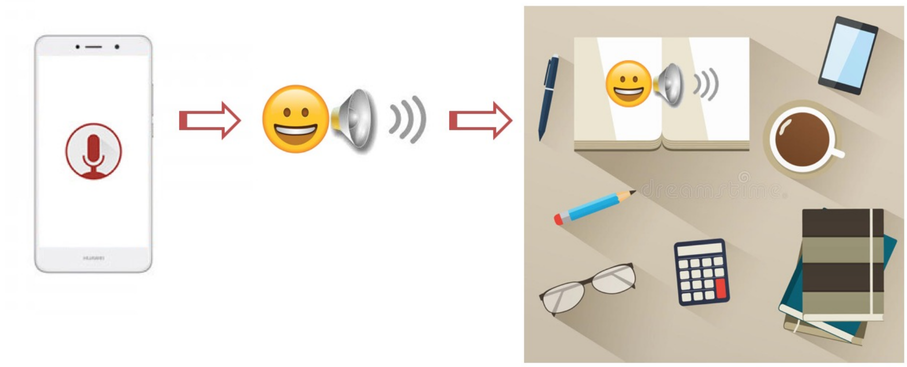
And through this system, we are able to:

1. Put messages into real environment to make environment-supportive messages (Like sticky notes),
2. Make messages be eye-catching so none are ignored (Like sticky notes),
3. Make messages be based on the real voice (Like LINE), and
4. Make messages convey emotions to improve empathy.

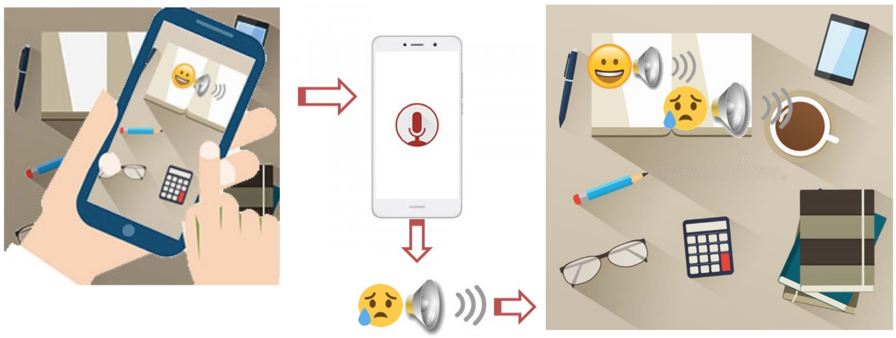
3.2 Approach

So we proposed an android AR(Augmented Reality) system using Virtual-Agent to leave emotional notes in space to enhance family connection and convey emotions. The crucial part of the system focuses on the emotional short voice message exchange with the Virtual-Agent.

Spatial note system allows users to make voice message by recording users' voice. Here is an example of how the system works. In figure 3.1 (a), when one family member A gets home, he can leave a voice message using phone and put the emotional voice message in the real environment. So in figure 3.1 (b), When the other family member B gets home, he can see that message.And after listening A's message, B can figure out what's A's meaning and what A felt at that time. And then leave back a new one. Additionally, the system is



(a) User A leaves a message



(b) User B responds

Fig. 3.1 Examples of messages exchange

able to detect the users' voice after recording it and then generate a Virtual-Agent with the appropriate facial expression. And these facial expressions express users' emotions. The virtual agents can detect four main emotions from the user's voice, they are calm, happy, angry and sorrow. Users can put the Virtual-Agent anywhere they want in the reality by using their smart phone.

Chapter 4

System Design

4.1 System Overview

Spatial note system is based on two parts, one is the virtual agent services and the other is the AR (augmented reality) system. The virtual agent service allows users to make voice messages by recording users' voice and is able to detect the users' voice after recording it. Then generate a Virtual-Agent with the appropriate facial expression to express users' emotion. Users can also change the facial expresses as they like to express feeling. There are 4 emotions of the virtual agents: happy, sorrow, angry, calm. Each emotion has 4 levels to express. The AR (augmented reality) system can detect the nearest plane in the real environment through the phone's camera. Users can put the Virtual-Agent anywhere they want in the reality according to the AR system.

Through this system, family members can exchange information, share feeling, discuss topics, and what's more, bond their relationships. For the note-making user, he can make any note he wants in the real world and convey his emotion to the other users easily. For the note-receiving user, he can freely choose whether to hear it according to the agent's facial expression. He can assume the others' feelings today by seeing agents' faces intuitively. We conducted a experiment to evaluate the usability of the system and positive feedback has been received.

4.2 Use Case Diagram

The usage case diagram of prototype system is shown as Fig 4.1. As shown in the diagram, there are 4 main use cases (choose a family role, receive new notes, respond notes, make new notes). And there are 2 sub use voice cases of the use case of receive-new-notes (new notes alert, listen new notes), 4 sub use cases of respond-notes and make-new-notes (record voice notes, change emotion detection results, choose notes' receiver, set virtual agents in reality).

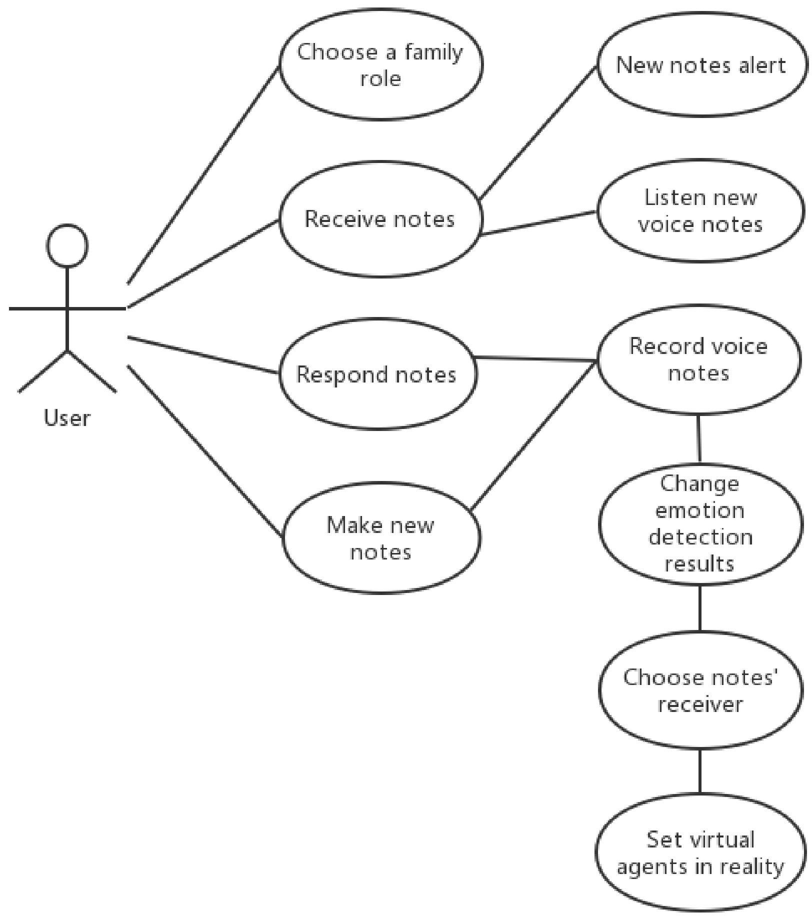


Fig. 4.1 Use case diagram

4.3 System Structure

Figure 4.2 shows the structure diagram of the proposed system. There are three layers, an interface display layer, a business logic layer, and a data access layer.

The first layer is the client. Its simple input and output function handle the information transaction of the system. This layer is mainly displayed in the user's interface, receives a series of instructions from the user, and presents the message to the user. The second layer is the server that plays the role of information transfer. If the user wants to access the database, it first needs to send the request to the server. Next, the server sends a database access request to the database. This request is usually implemented in a SQL statement. The third level is a database that stores large amounts of data. After receiving a request from the server, the database processes the SQL statement and returns the result to the server. The server converts the data to the client.

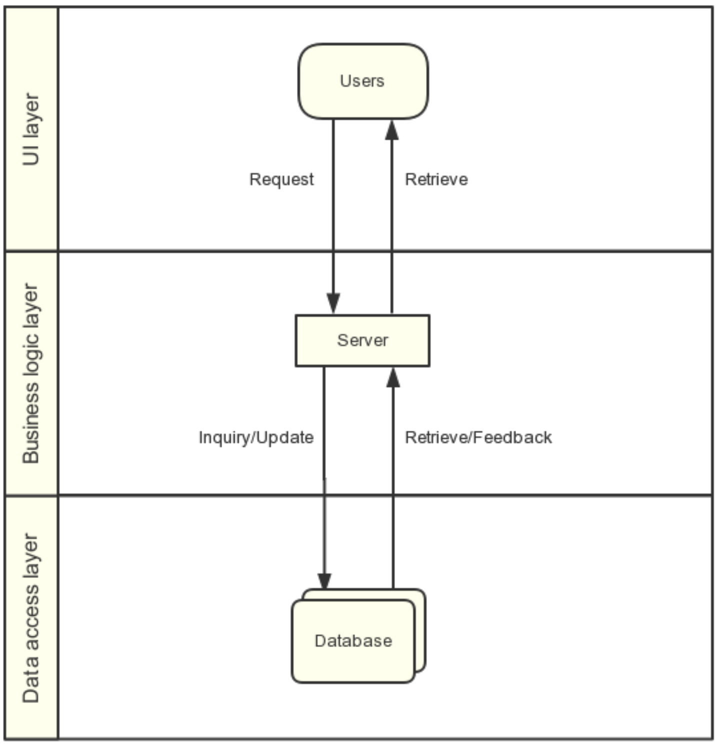


Fig. 4.2 System structure diagram

4.4 Usage Scenario

4.4.1 Persona and Frustration

- Kate (main persona): Mom of the family, 40 years-old, an office lady, leaves home early in the morning and goes back home late at night, has a rest in weekends.
- Kevin: Dad of the family, 42 years-old, a business man, leaves home early in the morning and goes back home late at night, sometimes goes to another city for business, has less rest than mom in weekends.
- Alice: Daughter of the family, 14 years-old, a high school student, leaves home to school at 8:00 in the morning and goes back home at 17:00, has a rest at home in weekends.
- Frustration: They can't communicate often, and know too little things about the others' daily life, feeling, thoughts, etc. They want to enhance relationships.

4.4.2 Notes-leaving Scenario

While getting up in the morning, Kate (mom) makes a great breakfast for the family and eats it with her husband. Before she leaves, she puts the breakfast of her daughter on the table, who is still in dreams. And then she uses the spatial note system in the smart phone to leave a voice note to her daughter. She said:"I bought your most favorite bread, please remember eat them all, mom loves you!" and then she uses the phone to put the note on the table near the breakfast. That note is represented by a smiling virtual agent and can be seen through smart phone's camera in the reality. The usage scenario diagram is showing as Fig.4.3. After that, the parents leave home to work.

4.4.3 Responding Scenario

After Alice (daughter) gets up, she knows that her parents are out. Alice sees the breakfast on the table. When she walks near the table, there is a new note alert on her phone, so Alice opens the spatial note system in the smart-phone to see that note. While she is opening the phone's camera and holding the phone towards the breakfast, a smiling virtual agent appears on the table in her phone's image. Alice finds out that mom left her a note so she clicks the agent to listen it. After hearing what's mom said, Alice records a new note and says "Thanks mom, the bread is delicious, but they are too much", then she puts this new note next to the mom's note to respond her mom, showing as Fig.4.4. At last, she goes to school.

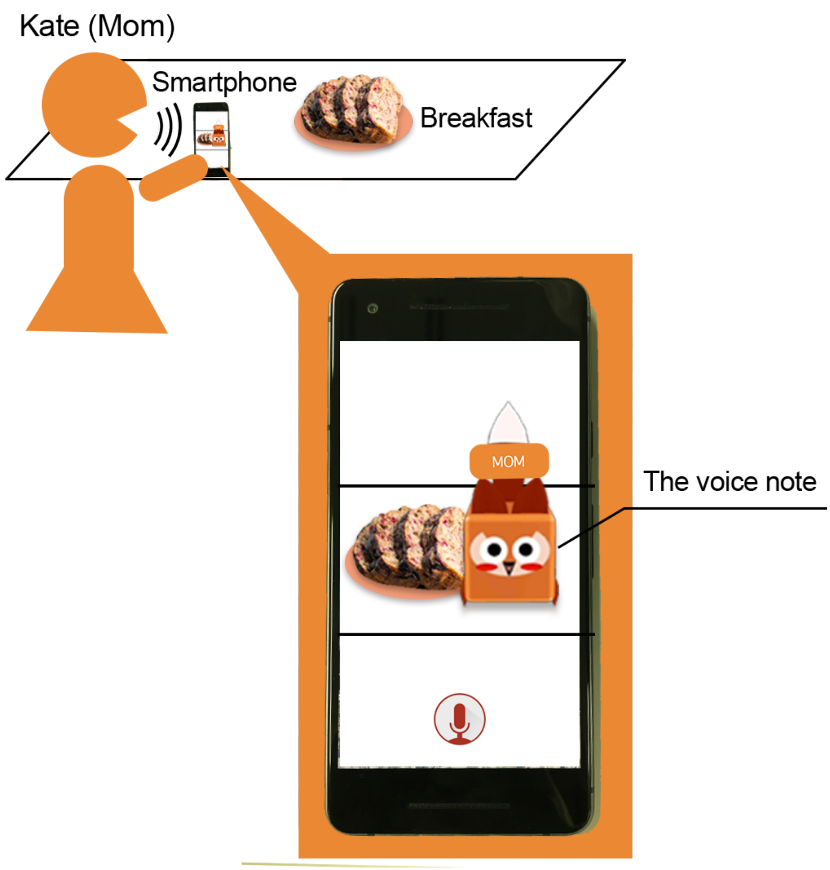


Fig. 4.3 Usage scenario diagram of notes-leaving

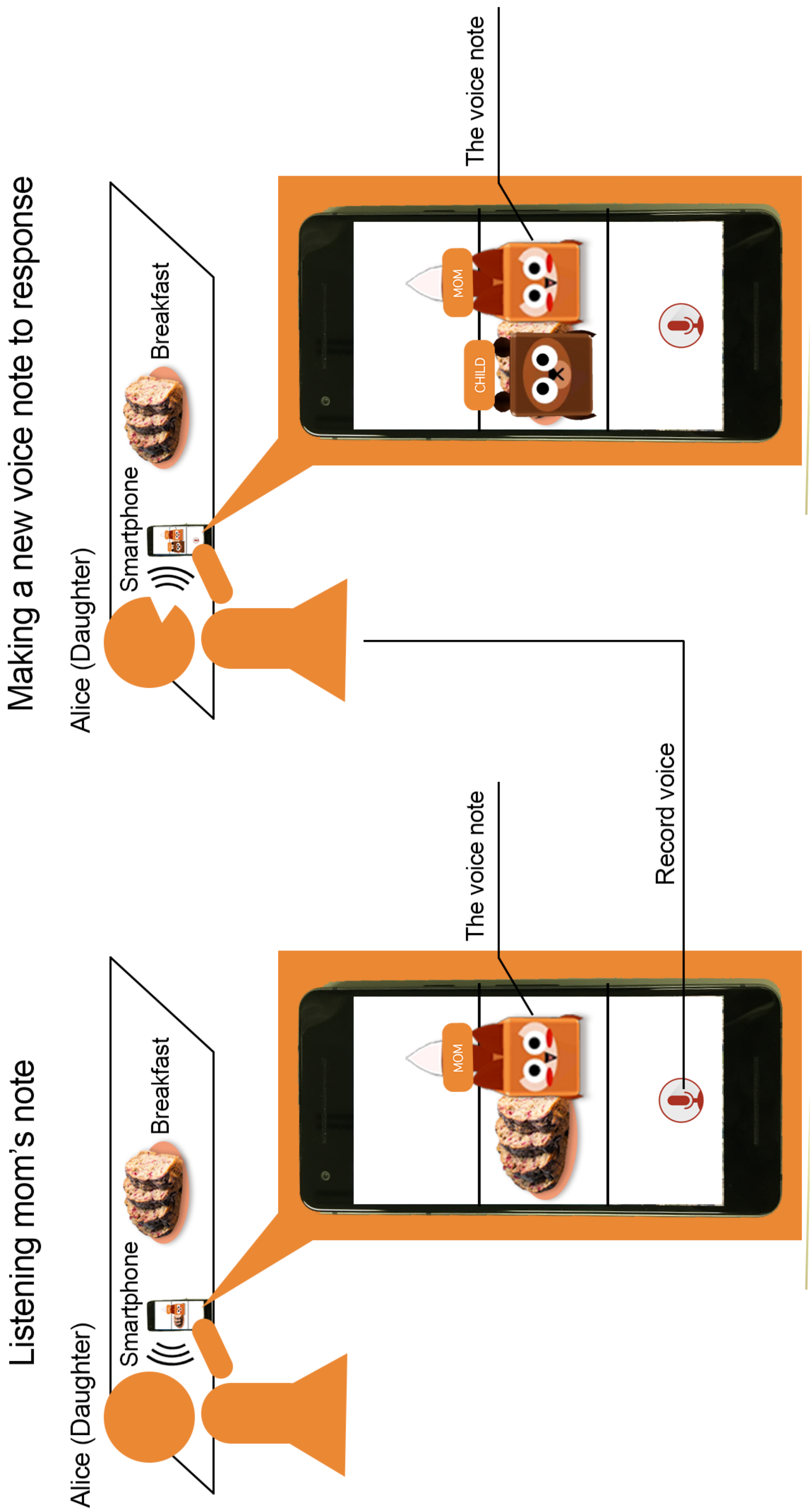


Fig. 4.4 Usage scenario diagram of responding

4.5 Role classification

Generally, there are three kind of roles existing in the family as Mom, Dad and Children, showing in the Fig.4.5. And communication sometimes happens between two of them. It is important to classify different roles when using the system, because it would be hard to identify who is the note sender from all the notes and send notes to individual member without this function. Different roles can communicate with each other, the content is only visible to each other. And three persons can also communicate, the content is visible to all members. Once the user chooses one of the roles, she/he will turn into the role's interface.

(a) Mom (b) Dad (c) Child

Fig. 4.5 Role Classification Icon

4.6 New Notes Alert

Once a family member has left a note to another family member, that family member will receive a new-note alert. Specifically, in the role's interface, a figure will appear in the lower right corner of the role's icon, indicating that several notes left for him have not been heard. When the user finds a new note in the real world and listens, the number in the lower right corner of the icon minus 1. Until the user hears all the new notes, the figure in the lower right corner of the roles' icon disappears.

This is to remind users not to forget any note left to him and also a reminder to remind user to respond notes. And how this function works is showing in the Fig.4.6.

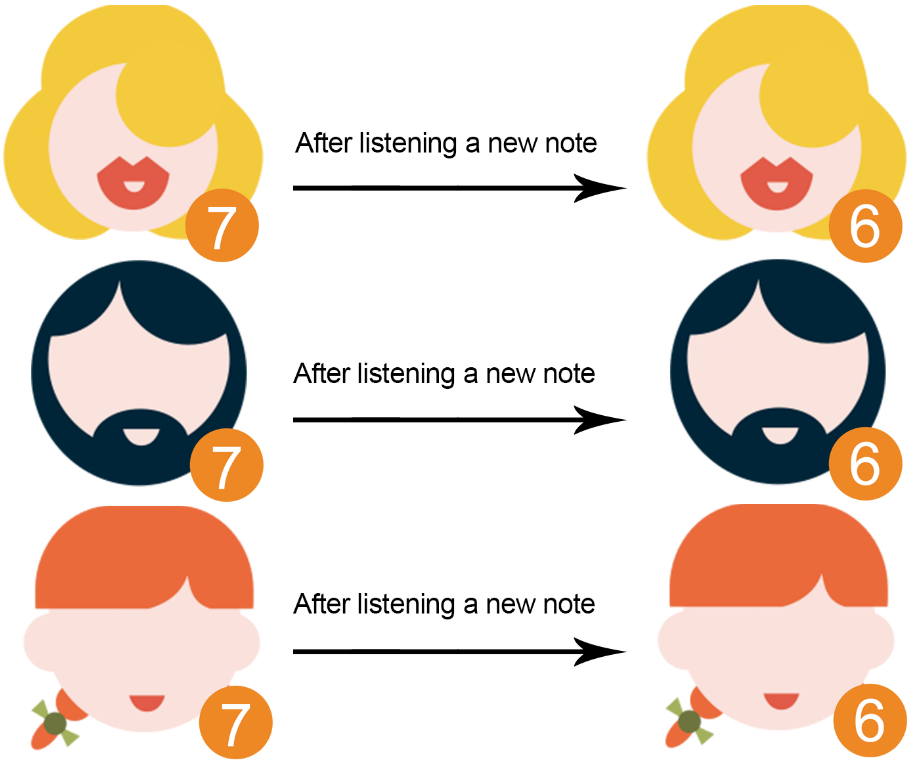


Fig. 4.6 New notes alert

4.7 Making Voice Notes

4.7.1 Environment Scanner

The system is using ARCore. In addition to identifying key points, ARCore can detect flat surfaces, like a table or the floor, and can also estimate the average lighting in the area around it. These capabilities combine to enable ARCore to build its own understanding of the world around it. ARCore's understanding of the real world lets users place objects, annotations, or other information in a way that integrates seamlessly with the real world. Users can place a napping kitten on the corner of the coffee table, or annotate a painting with biographical information about the artist. This is the fundamental part for putting virtual agent into the reality. The Fig.4.7 shows how the ARCore scans the environment.

4.7.2 Emotion detection from Voice

The system can perform voice emotion recognition when recording voice notes. When users are recording their voice to leave notes, the system will call the Empath API and get the emotion detection results. The Empath API can identify the emotion by analyzing not what you say but physical properties of users' voice such as pitch, tone, speed and power. It can detect emotion from every language.

The emotion results include 5 values as calm, angry, joy, sorrow, energy, 4 emotion results as calm, angry, happy and sorrow and each emotion has 4 emotion levels as a little, much, very much and extremely and they are represented by 4 different emojis.

The emotion detection results are calculated by the emotion value, showing in the Table 4.1. The Table 4.2 shows the emotion levels calculated from the energy value. The Fig.4.8 shows the images of emotion levels.

Fig. 4.7 Scanning the environment

Table 4.1 Correspondence between detected values and emotion detection results

Elements	Value	Results
Calm	25-50	Calm
Anger	25-50	Angry
Joy	25-50	Happy
Sorrow	25-50	Sorrow

Table 4.2 Emotion levels according to energy values and emotion detection results

Emotion Result	Emotion levels (Energy value)			
	0-10	11-20	21-30	31-40
Calm	A little	Much	Very much	Extremely
Angry	A little	Much	Very much	Extremely
Happy	A little	Much	Very much	Extremely
Sorrow	A little	Much	Very much	Extremely

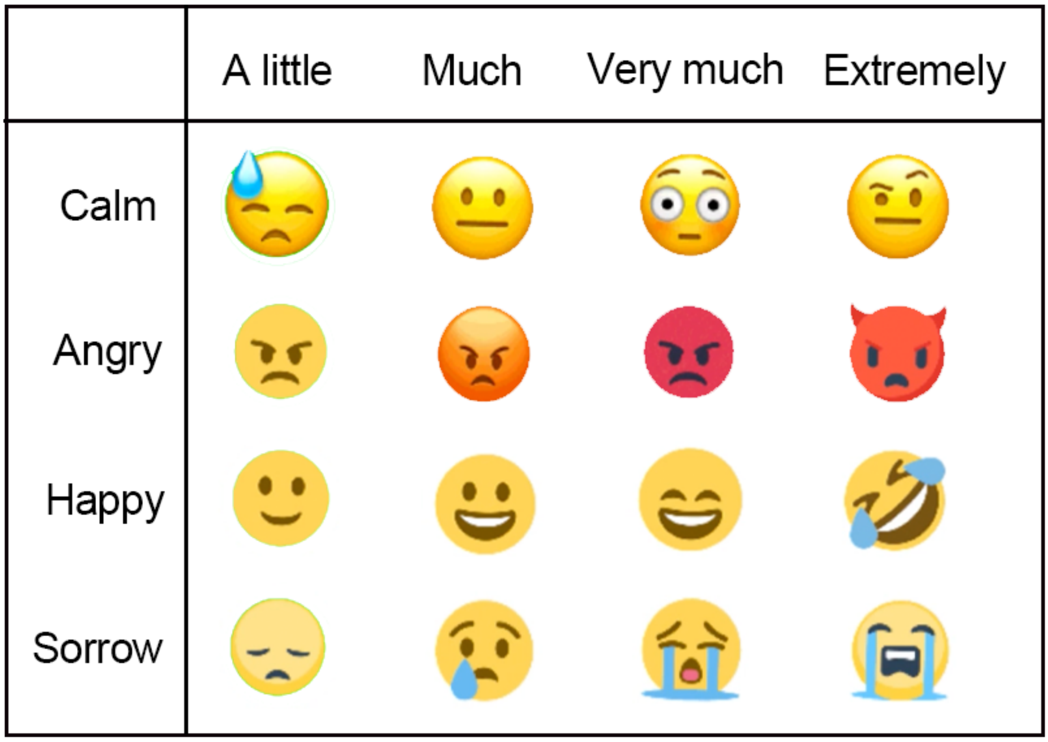


Fig. 4.8 Emotion levels

4.7.3 Emotion Result Change

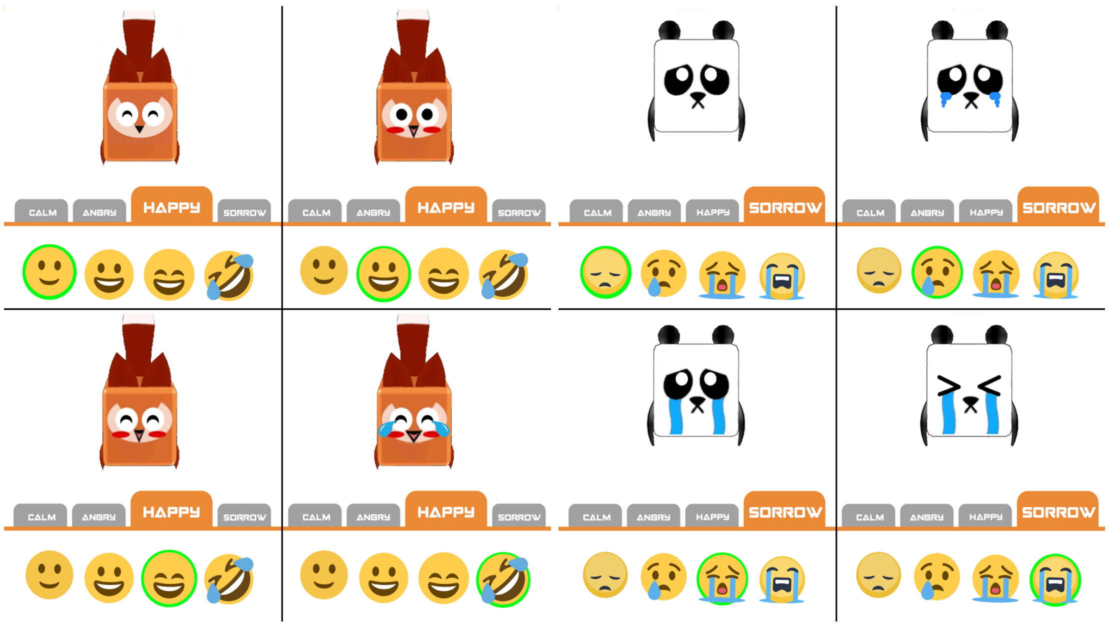
Users can change emotion results as they like, they can change the emotion results and also change the emotion levels. Choosing different emotion levels will change the facial expression of the virtual agents. The Fig.4.9 shows the 4 levels of calm emotion and its names. Its levels' definition is slightly different from the others. The Fig 4.10 shows all the emotion levels.

(a) calm and speechless (b) calm

(c) calm and curious (d) calm and confused

Fig. 4.9 Calm Emotion

(a) calm (b) angry



(c) happy (d) sorrow

Fig. 4.10 Emotion levels

4.7.4 Visible Time Setting

After the user decides the proper emotion and the system generates a virtual agent, the user can decide how long will this agent exist after been listened. The options are 1 hour, 4 hours, 8 hours and 24 hours, showing in Fig 4.11.

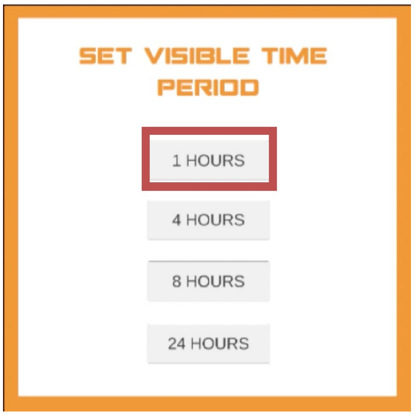


Fig. 4.11 Visible time setting

4.7.5 Notes Receiver Setting

Users can decide who can receive the notes, an individual or the whole family. If the user only sends notes to an individual, the notes can be seen only by the receiver, not anyone. But if the user chooses the whole family, then all family members can see that note. The Fig 4.12 shows the mom's choice board.

Fig. 4.12 Mom's note-receiver choice board

Chapter 5

System Implementation

5.1 Development Environment

We develop this system using Unity 2017 and development language is C sharp. We use ARCore Unity SDK to realized marker-less augmented reality, the emotion detection API is Empath API. MySQL is used as database. PHP is used to implement the communication between the application and the server.

5.2 System Hardware

To realize this system, the Android platform needs to be Android 7.0 or higher. So we use Google Pixel2 smartphone as the device. Google Pixel2 is an Android smartphone with great camera, which is good in developing the system, showing in Fig 5.1.

5.3 System Diagram

Spatial note system is based on two parts, one is the virtual agent services and the other is the AR (augmented reality) system (see Fig 5.2). The virtual agent services allow users to make voice messages by recording users' voice and are able to detect the users' voice after recording it. Then generate a Virtual-Agent with the appropriate facial expression to express

Fig. 5.1 Google Pixel 2

users' emotion. Users can also change the facial expresses as they like to express feeling. There are 4 emotions of the virtual agents: happy, sorrow, angry, calm. Each emotion has 4 levels to express. The AR (augmented reality) system can detect the nearest plane in the real environment through the phone's camera. Users can put the Virtual-Agent anywhere they want in the reality according to the AR system.

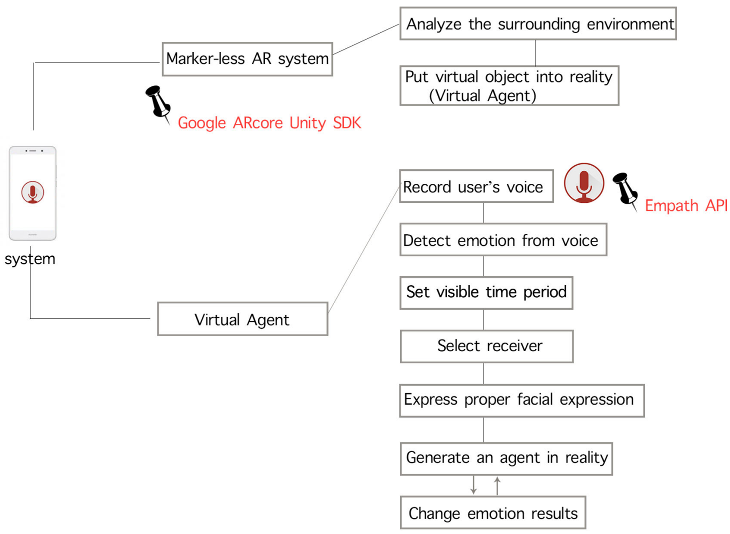


Fig. 5.2 System Design Diagram

5.4 Users Interfaces

5.4.1 Start Interface

When user opens the spatial note system in their phone, this interface will show up, displaying all the agents we have in this system. User can click left and right arrows button to see every different agents the system has. And also click the icon in the bottom to turn to the next interface (see Fig 5.3).

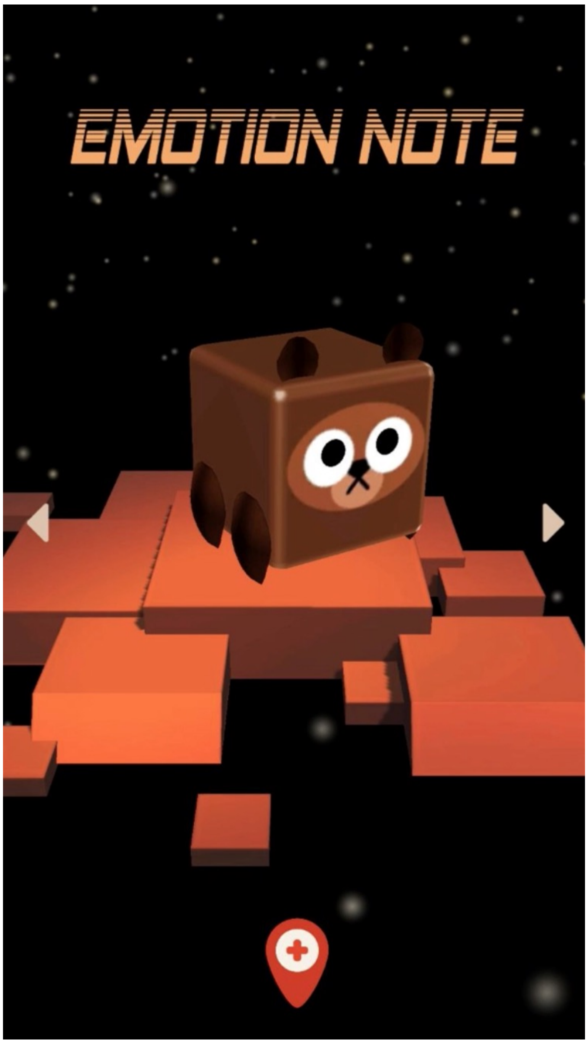


Fig. 5.3 Start Interface

5.4.2 Family Role Selection Interface

There are three kind of roles existing in the family as Mom, Dad and Children, showing in Fig 5.4. Different roles can communicate with each other, the content is only visible to each other, and three persons can also communicate. The content is visible to all members. Once the user chooses one of the roles, she/he will turn into the role's interface.

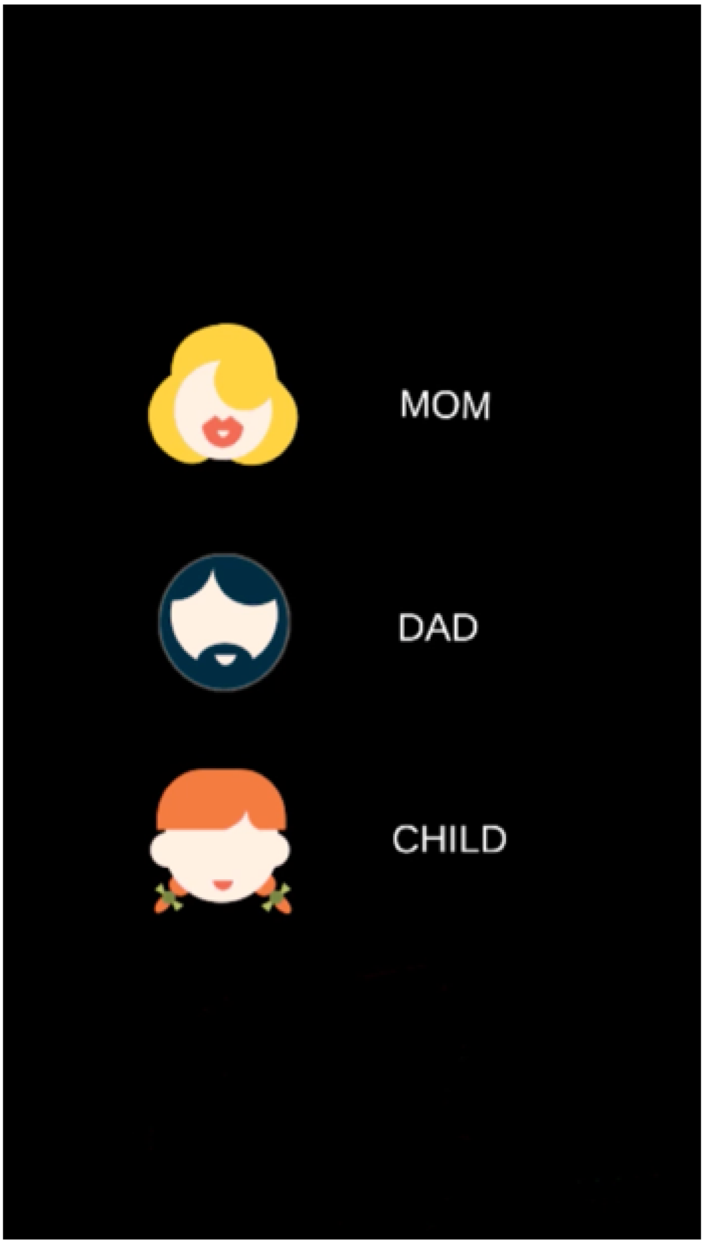
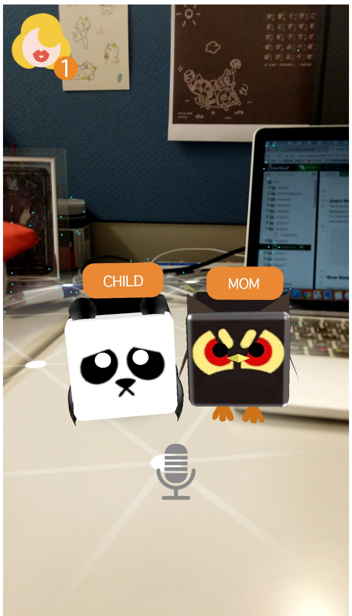


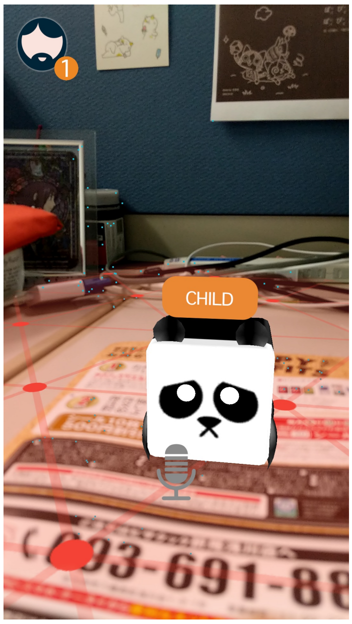
Fig. 5.4 Role Selection Interface

5.4.3 Users Main Interface

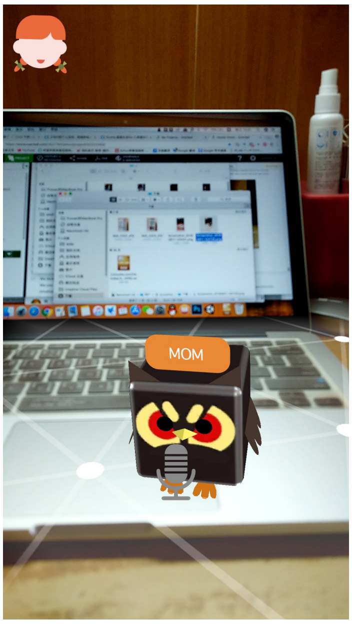
The three kind roles' interface is almost the same, except the role's icon displaying in the left up corner. The example of all the users' interface is showing in Fig 5.5.



(a) Mom's Interface



(b) Dad's Interface

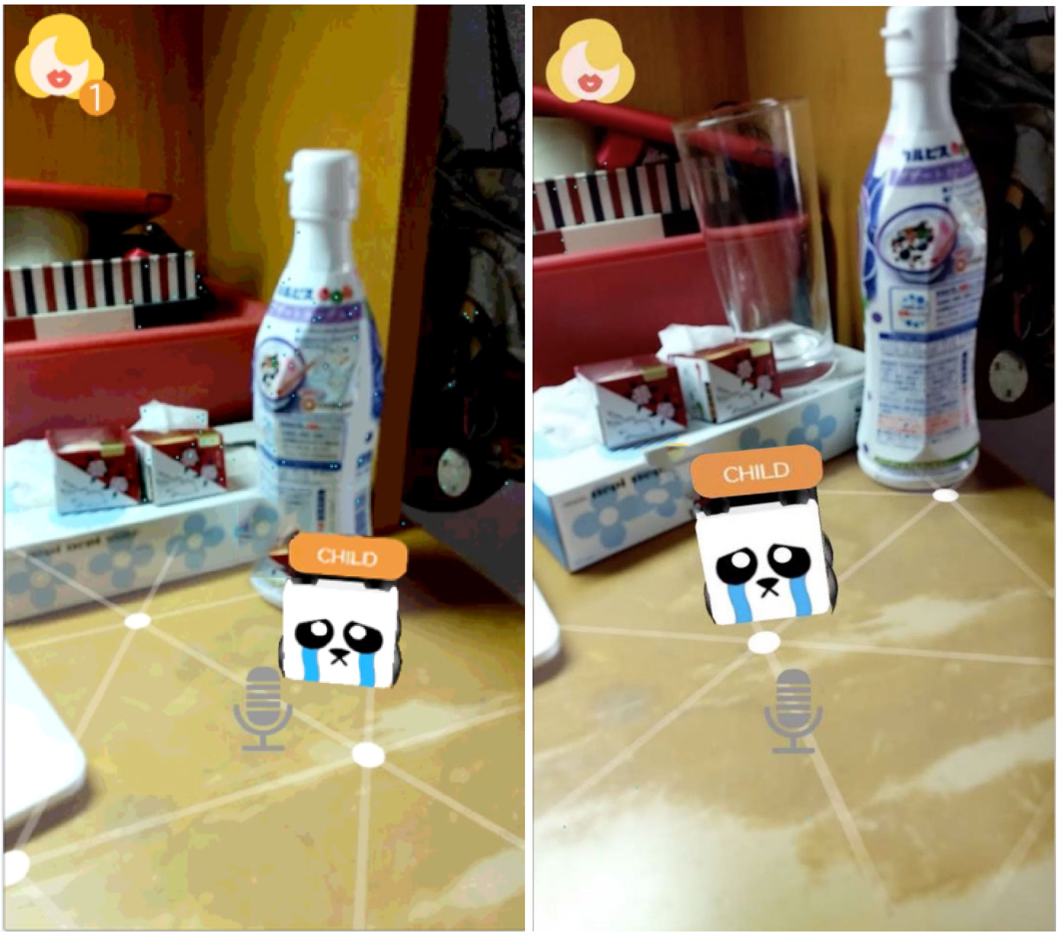


(c) Child's Interface

Fig. 5.5 Users' Interface

5.4.4 New Notes Alert Change Interface

Once a family member has left a note to another family member, that family member will receive a new-note alert. When the user finds a new note in the real world and listens, the number at the lower right corner of the role's icon minus 1, showing in Fig 5.6. After mom has listened to the child's note, the number at the lower right corner of the role's icon minus 1 so the new note alert disappears. This function aims to alert users how many notes remain right now and need to be listened. Users can also check this part to confirm receiving all the notes and listening to all of them.

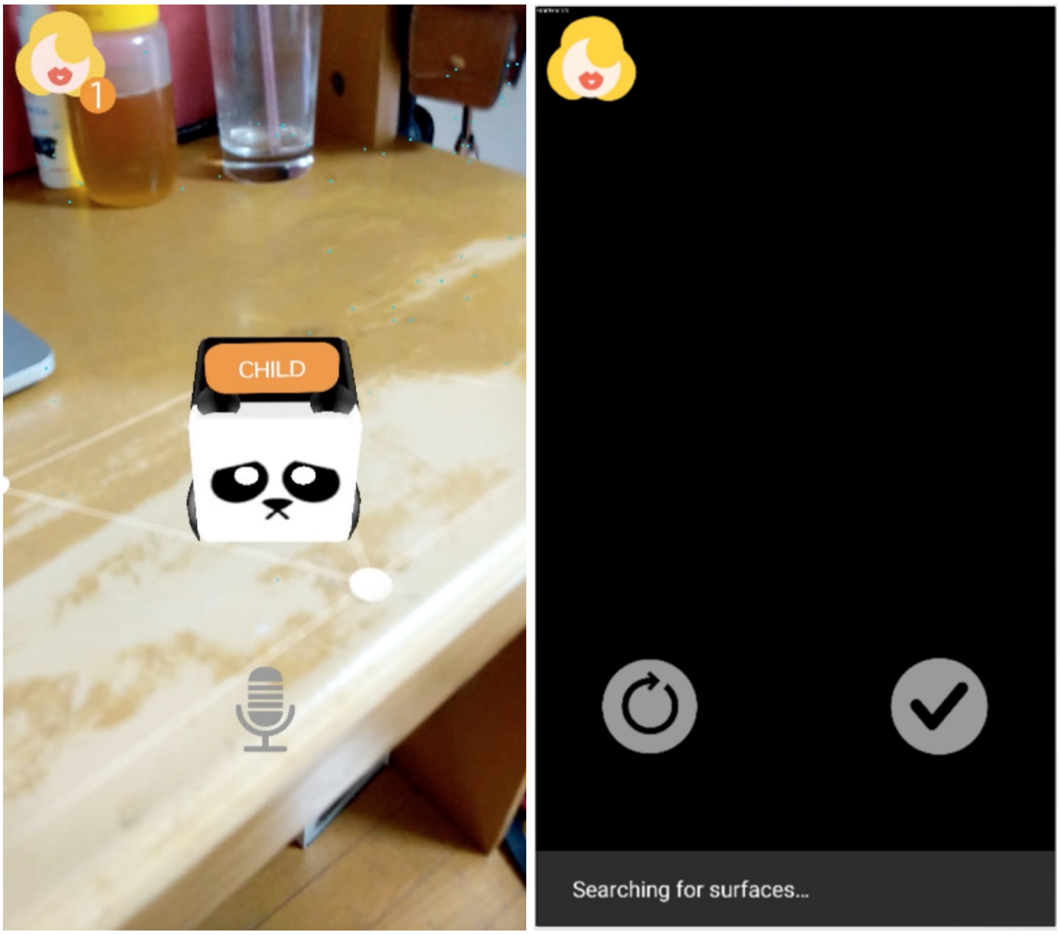


(a) Before listening to the new note (b) After listening to the new note

Fig. 5.6 New Notes Alert in Mom Interface

5.4.5 Voice Recording Interface

Users can click record button to start recording the voice. And press stop button when user finish his sentences. The record button are always showing in the main page, so the users can leave notes by recording voice any time and any place. The interface is showing in Fig 5.7.

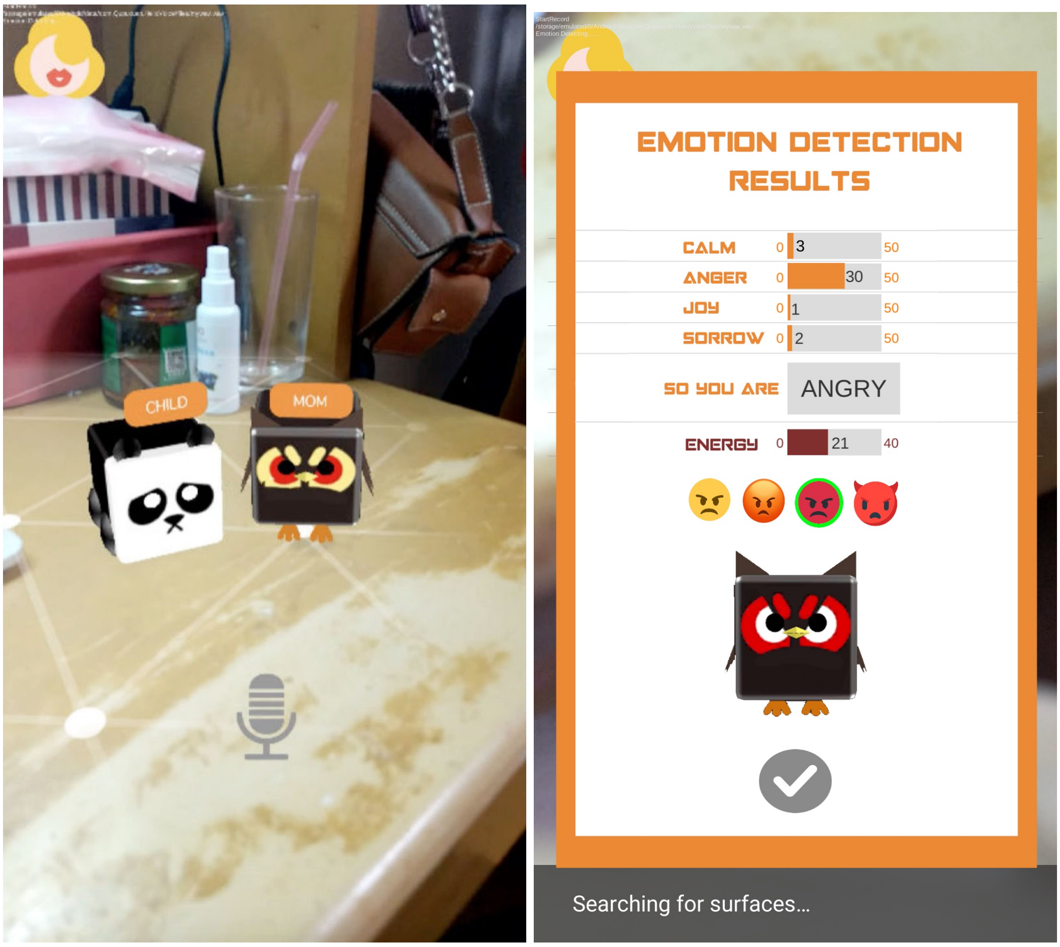


(a) Before recording the new note (b) Recording the new note

Fig. 5.7 Voice note recording

5.4.6 Emotion Detection Interface

After mom has listened to the child's note and recorded voice to respond, the system detects the mom's voice and generates a virtual agent with proper facial expression. Mom chooses a place to let the virtual agent display, showing as Fig 5.8 (a). But mom thinks the facial expression is not what she wants now so she gives the virtual agent a long press and the emotion detection result interface shows, as Fig 5.8 (b). In the interface, there are 4 emotion elements and their values which detected from user's voice, user's emotion. The emotion levels are decided by the energy value. For the emotion elements, each element is independently measured by the system from the voice and the detected values are all varied between 0 to 50. And there is only one value will be bigger than 25 detected by the Empath API, and this element whose value is bigger than 25 decides the final emotion results. The energy is detected from the sound intensity. It decides the emotion levels.



(a) Default virtual agent without change (b) Emotion detection results

Fig. 5.8 Emotion detection

5.4.7 Emotion Change Interface

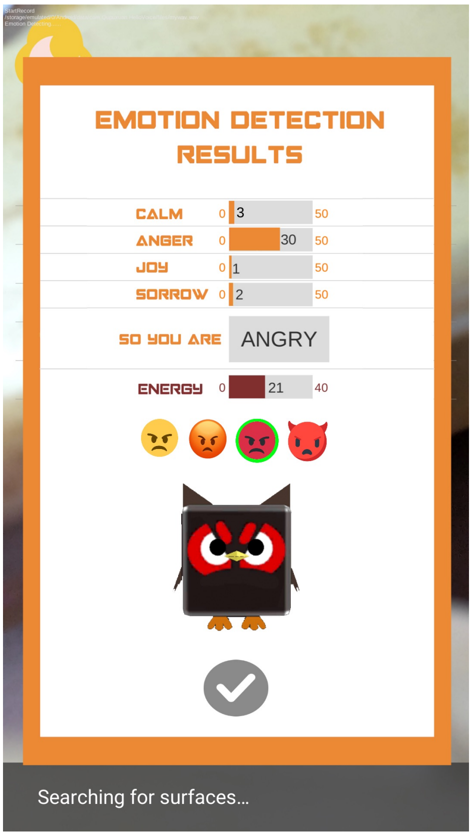
Users can freely change their notes' emotion by themselves, they can only change the emotion levels from the Emotion-Detection-Results interface, showing as Fig 5.9 or they can change the whole emotion and levels by clicking the virtual agent showing in the Emotion-Detection-Results interface. Then it will turn to emotion change interfaces. After deciding which emotion to choose, users can press OK button at the bottom of the page. Calm emotion interfaces are showing in Fig 5.10. Angry emotion interfaces are showing in Fig 5.11. Happy emotion interfaces are showing in Fig 5.12. Sorrow emotion interfaces are showing in Fig 5.13.

5.4.8 Visible Time Setting Interface

Fig 5.14 shows the interface of setting time period about how long the user's note exists after being listened. There are 4 choice for users to choose, 1 hour, 4 hours, 8 hours and 24 hours. If user choose 1 hour, it means that the note he sends can only exist 1 hour after all the receivers have checked and listened. Similar situation to the other choices.

5.4.9 Receiver Selection Interface

The fig 5.15 shows the interfaces of choosing receiver by mom. There are 3 choices for each role in family. For mom, there are dad, child and family. If mom choose dad, the note she leaves right now can be only seen by dad, child can't see that note. And if mom choose child, dad can't see that note but child can. If mom choose family, then dad and child will receive the note and can listen to it. And it is similar for dad and child. In dad's interface, dad has 3 choices as mom, child and family. In child's interface, child has 3 choices as mom, dad, family.



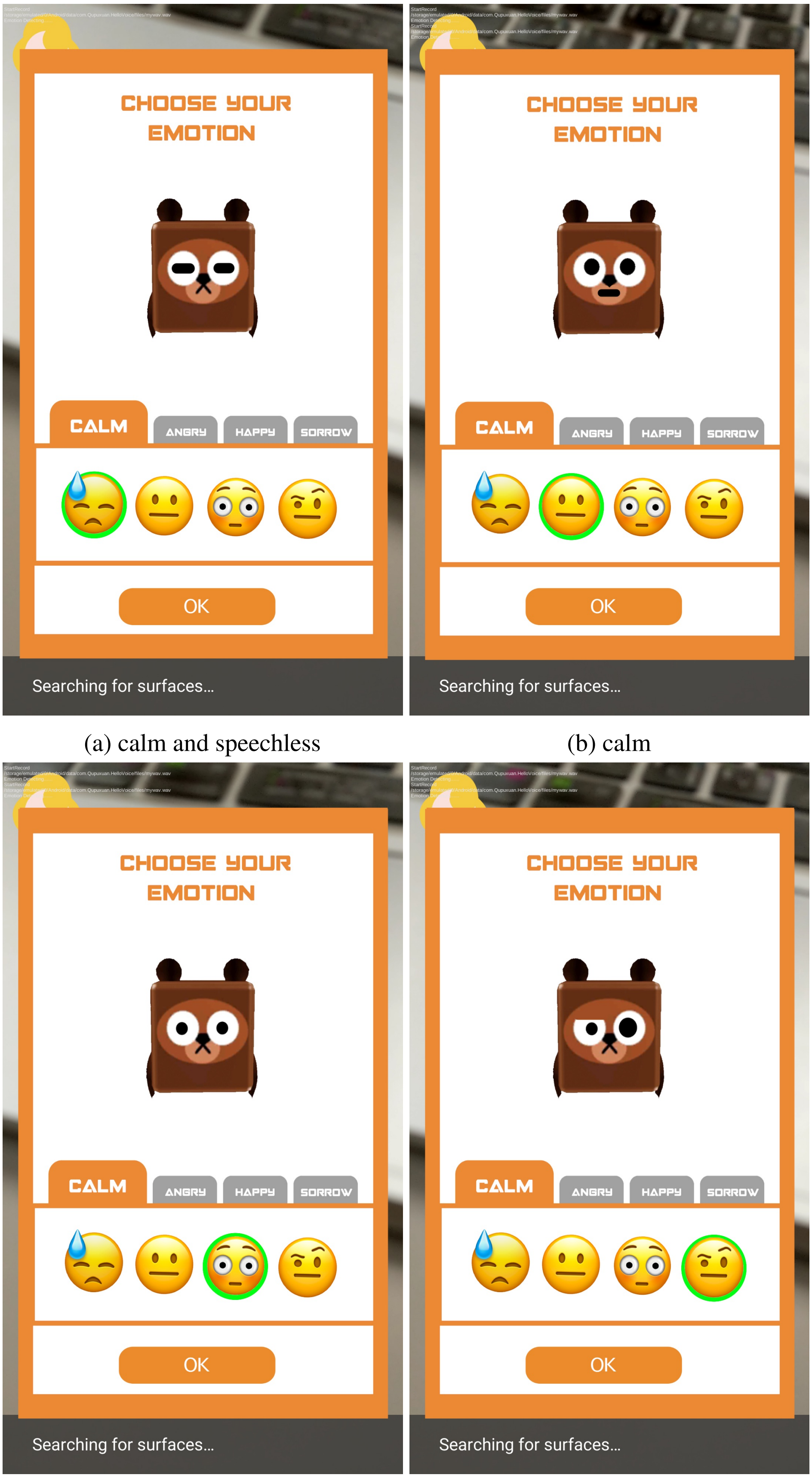
(a) Default virtual agent without change

(b) Change level to a little angry

(c) Change level to much angry

(d) Change level to extremely angry

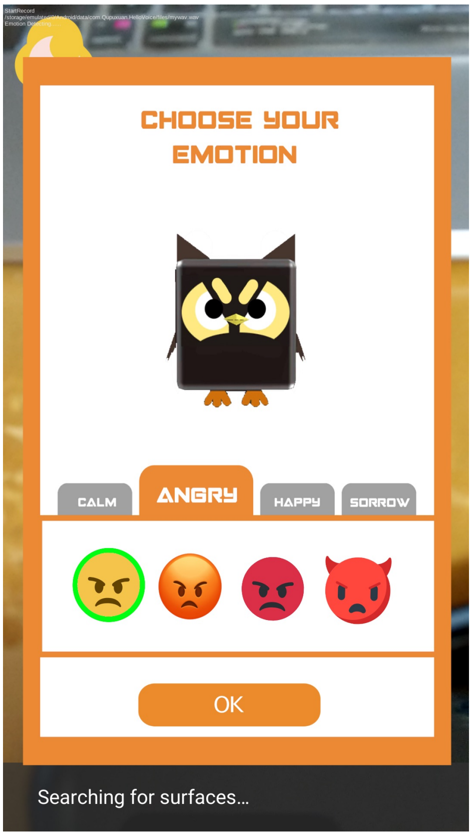
Fig. 5.9 Emotion levels change



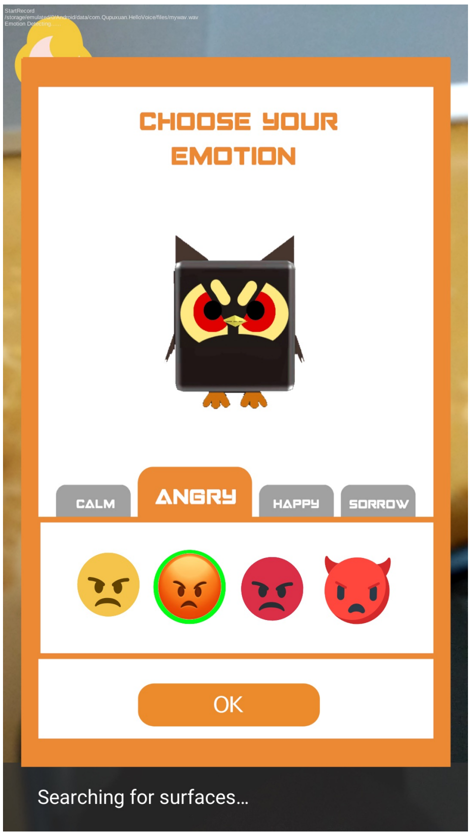
(a) calm and speechless (b) calm

(c) calm and curious (d) calm and confused

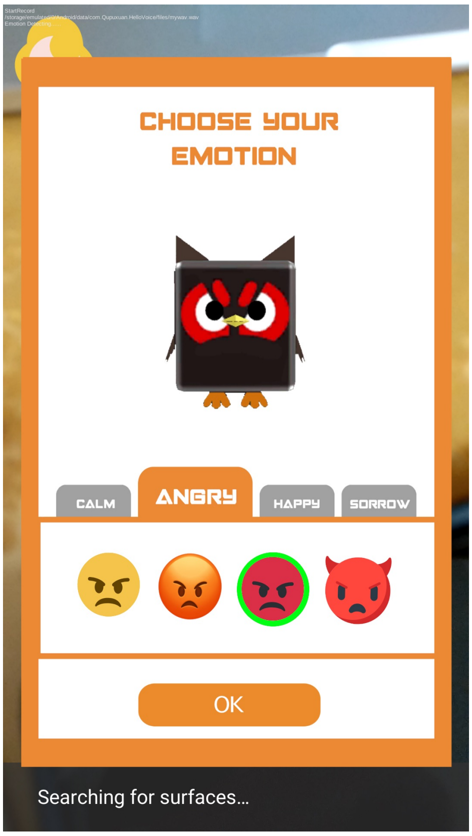
Fig. 5.10 Calm Emotion



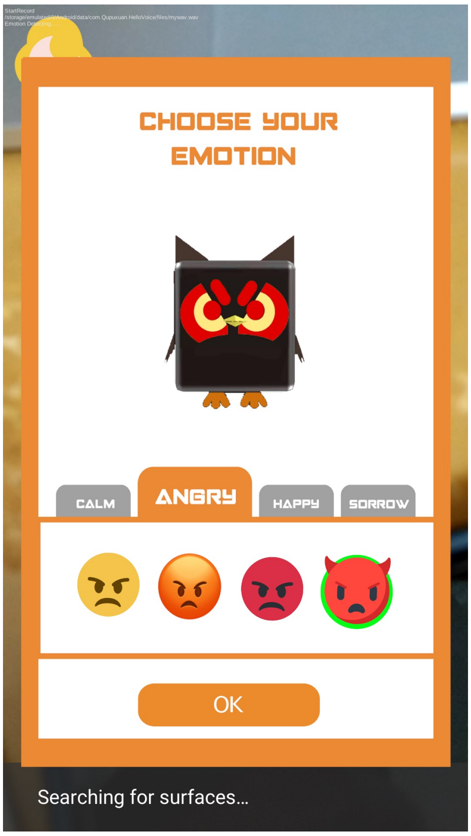
(a) A little angry



(b) Much angry

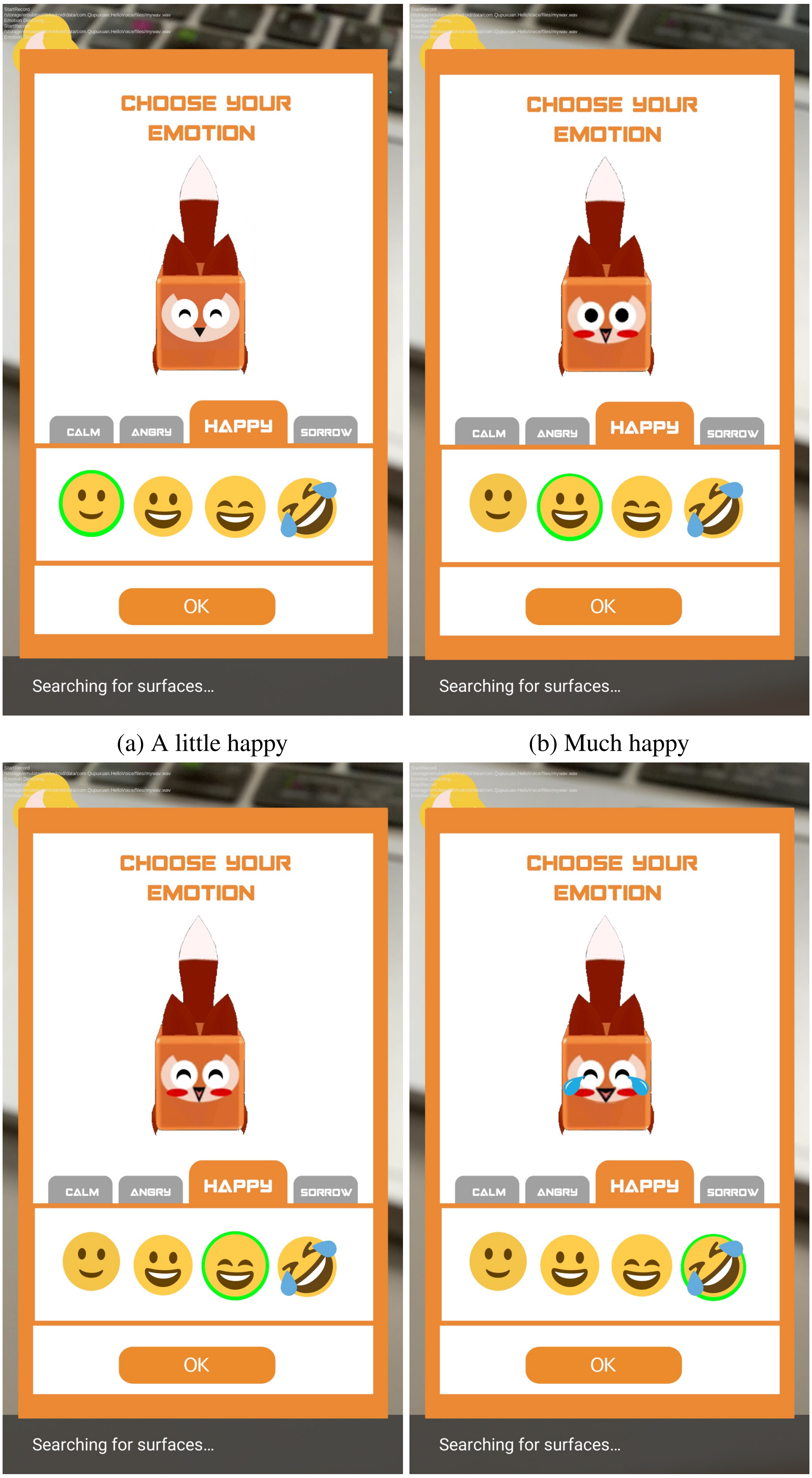


(c) Very much angry



(d) Extremely angry

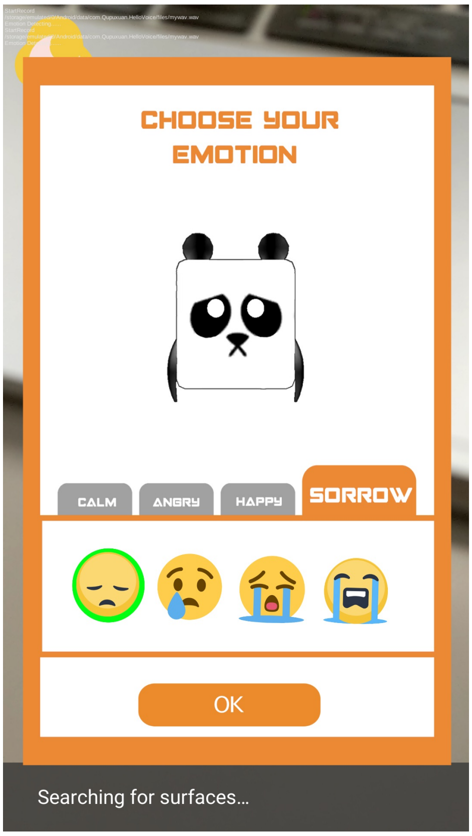
Fig. 5.11 Angry Emotion



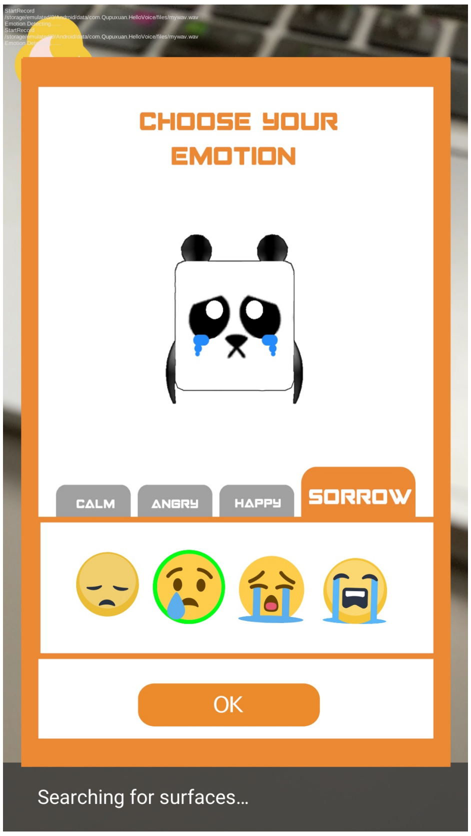
(a) A little happy (b) Much happy

(c) Very much happy (d) Extremely happy

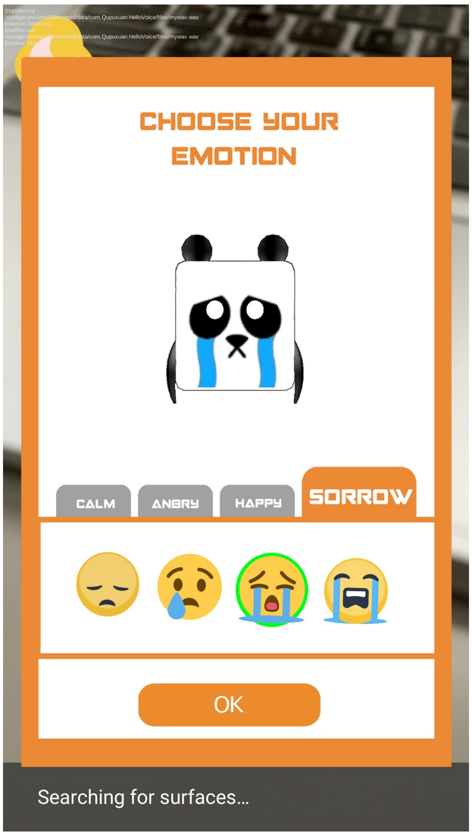
Fig. 5.12 Happy Emotion



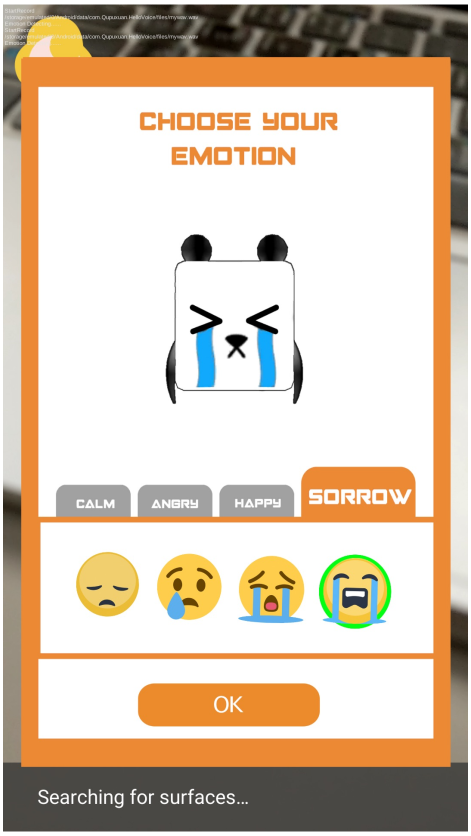
(a) A little sorrow



(b) Much sorrow



(c) Very much sorrow



(d) Extremely sorrow

Fig. 5.13 Sorrow Emotion

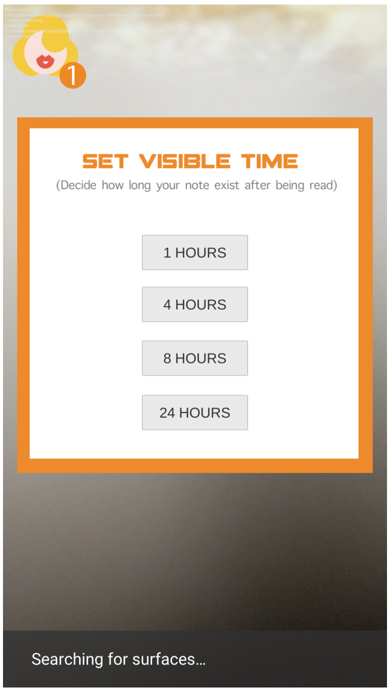


Fig. 5.14 Time setting interface

Fig. 5.15 Receiver Selection Interface

Chapter 6

Related Work

With the popularization of the Augmented Reality (AR), more and more researchers determined themselves in proposing new method, new idea in Augmented Reality. Augmented Reality enables the direct or indirect view of the physical, real-world environment whose elements are augmented by computer. Jun Rekimoto[26] presents a system called CyberCode. The CyberCode is a visual tagging system based on a 2D- barcode technology and provides several features not provided by other tagging systems. CyberCode tags can be recognized by cameras found in more and more mobile devices, and it can also be used to determine the 3D position of the tagged object as well as its ID number. Mihai Bace[27] proposed a novel wearable ubiquitous method which is described as ubiGaze to augment any real-world object with invisible messages through gaze gesture that lock the message into the object. This enables a context and location dependent message services. Mistry[28] presents WUW-Wear Ur World, which is a wearable gestural interface that allows projecting information out into the real-world. This is a system composed of a projector and a tiny camera, which allows the system to see what users see and through projection, information can be displayed on any surface, planes and physical objects around users. Nassani[29] proposed a system called Tag-It. It is a wearable system that allows people to place and interact with 3D virtual tags placed around them. They use two wearable technologies: a head-worn wearable computer (Google Glass) and a chest-worn depth sensor (Tango). The Google Glass is used to generate and display virtual information to the user, while the Tango is used to provide robust indoor position tracking for the Glass.

Tonchidot proposed Sekai Camera, an augmented reality application that allows users to share tags for any place on the planet based on GPS. Sekai Camera makes it possible to create text messages, photos, and audio recordings (dubbed air tags) and "drop" them on the spot in the form of floating bubbles and icons. Other users in the vicinity who open the app and spin the camera around can then click on and interact with those geo-tagged virtual Post-It notes. Sekai Camera is essentially a flashy-looking, location-based, augmented reality-powered social network. Sekai camera gives the inspiration in using AR to send environment-based messages and making the messages more emotional.

Tarumi[30] proposed an overlaid virtual system called SpaceTag. SpaceTag is an object system where each object called SpaceTag can only be accessed from a limited location and for a limited period of time. Its applications include entertainment systems, advertising services, bulletin board systems and personal communication systems. For one-way communication, they are broadcast from the server; for two-way communication, the user can cut and paste the space tag between the portable PC and real space. The SpaceTag system is a location-aware information system and an augmented reality system because it attaches information to real space. However, they classify it as a superimposed virtual system because it does not link directly to real objects. It can achieve public services without causing drastic changes in society and without much cost. And in this system, the virtual world consists of virtual architectural objects and virtual creatures. Virtual creatures are dynamic objects that can move or interact with other objects, or with users visiting the virtual world. In other words, a virtual creature is an active agent that can react to stimuli from the environment and dynamically execute methods like giving messages to the user. They can also exchange messages with other agents. Sometimes we call virtual creatures simply agents. From the user's mobile phone, user can see a perspective view of the virtual world. Far objects are drawn as small images, while closer objects are displayed as larger images. If the face of the virtual creature can be seen from the north side of the virtual creature, its back can be seen from its south side. The user can detect the location of the user through

GPS embedded on the mobile phone. Therefore, the user can walk in the virtual world while he/she walks in the real world. The correspondence between these two worlds is based on location. They also have two versions of the virtual world system: a browser-based version and a BREW-based version. BREW-based versions require BREW-based special software on the terminal side. With BREW-based versions, graphics can be automatically redrawn. However, since it requires a pre-subscription contract with each terminal. The browser-based version does not require any special software on the mobile phone. Only use the built-in browser. All necessary processing of the virtual world system is performed on the server side. However, it is a pull information system, so when he/she moves to a new location, the user should manually send a request to the server to download a new description or an image of the virtual world. The concept of this system gives us an inspiration in creating virtual agents in an AR system to help and entertain people in the daily life.

Chapter 7

Evaluation

7.1 Experiments

7.1.1 Participants

As Table 7.1 and Table 7.2 shows, in order to evaluate usability and efficiency of using our system, we plan to recruit 3 families and each family has 3 participants. Totally we have 3 participants of children, aging from 22 to 24 and 6 participants of parents, aging from 49 to 52. All participants have general knowledge of computer and have the experience of using smart-phone.

Table 7.1 Subject Demographic Information of children

Elements	Description
Participants	1 male, 2 females
Age	22-24; Mean: 23
Profession	Students

7.1.2 Method

One family consists of 3 persons, so there are two cases.

a) Sending a message to one person

Table 7.2 Subject Demographic Information of parents

Elements	Description
Participants	3 males, 3 females
Age	49-52; Mean: 51
Profession	Workers

b) Sending a message to all members, i.e., two persons.

The method we use for evaluation is described in the following n steps:

1. Each participant will be asked to send 10 a) messages (5 messages per person) and 5 b) messages for practice.

2. Each participant will be asked to listen messages and reply, if needed.

3. Each participant will be asked to change emotion at least once.

4. Each participant will be asked for using the application to leave notes in their home to exchange information for at least 1 day.

5. Each participant is asked to do the questionnaire after finishing the previous two parts as some feedback. Each participant need to write the number of messages he/she left today and answer 5 question about using the system showing in the Table 7.3.

6. Each participant is invited to ask questions to me about the system or anything related to the experiment. The answer is grading from 1 to 5 (1 = very positive, 5 = very negative).

Table 7.3 Investigative Questions after Using the System

Question	1	2	3	4	5
Do you think it is easy to leave a message?					
Do you think it is easy to find messages and listen?					
Do you think it is correct of the emotion detection result?					
Do you think it is easy to change emotion?					
Do you think your family connection is enhanced by using this system?					

7.1.3 Questionnaire

The questionnaire is showed in Fig 7.1. We plan to investigate the basic information of each participant and get their feedback.

Questionnaire

Thanks for using the system and help us with this questionnaire. We appreciate your kindness and patience.

Personal Information:

Your Role of family: _________(Mom/Dad/Child) Date: ________________

How many messages you left today: ______________

Questions:

Please choose your answer of the question:

1. Do you think it is easy to leave a message?
   Very easy O—O—O—O—O Difficult
2. Do you think it is easy to find messages and listen?
   Very easy O—O—O—O—O Difficult
3. Do you think it is correct of the emotion detection result?
   Correct O—O—O—O—O Wrong
4. Do you think it is easy to change emotion?
   Very easy O—O—O—O—O Difficult
5. Do you think your family connection is enhanced by using this system?
   Very much O—O—O—O—O Not at all

Do you have any suggestion about the improvement of the system? If any, please write below:

____

____

____

____

____

____

Fig. 7.1 Questionnaire

7.2 Evaluation

After collecting the results given by the participants (showing in Table 7.4 and Table 7.5), the evaluation of using the spatial note system to enhance family connection can be carried out. All the participants are asked to rate on a Likert Scale ranging from 1 to 5.

It is necessary to introduce the Likert Scale. A Likert scale is a psychometric scale commonly involved in research that employs questionnaires.

The Table 7.4 shows all the answers from children. There are 3 participants choose Grade 1 in Q1, meaning that they all think it easy to leave a message. In Q2, 1 child chooses Grade 1 (very easy) and 2 children choose grade 2 (easy). In Q3, 2 children choose grade 1 (very easy) and 1 child chooses grade 2 (easy). In Q4, 3 children choose grade 2 (easy) and in Q5, 3 children choose grade 1 (very easy).

Table 7.4 Answers Statistics of Investigative Questions from Children

Question	1	2	3	4	5
Q1:Do you think it is easy to leave a message?	3				
Q2:Do you think it is easy to find messages and listen?	1	2			
Q3:Do you think it is correct of the emotion detection result?	2	1			
Q4:Do you think it is easy to change emotion?		3			
Q5:Do you think your family connection is enhanced by using this system?	3				

The table 7.5 shows the statistical answers from parents. There are 4 participants choose grade 1 in Q1, meaning that they think it easy to leave a message and 2 participants choose grade 2 (easy). In Q2, 1 participant chooses grade 1 (very easy) and 2 participants choose grade 2 (easy) and 3 participants choose grade 3 (normal). In Q3, 2 participants choose grade 1 (very easy), 2 participants choose grade 2 (easy) and 2 participants choose grade 3 (normal). In Q4, 1 participant chooses grade 2 (easy) and 5 participants choose grade 3 (normal. In Q5, 1 participant chooses grade 1 (very easy), 3 participants choose grade 2 (easy) and 2 participants choose grade 2 (normal).

From the Fig 7.2, we can see that all participants' give grades under 3 to each question. Because the higher the grade, the harder using the system, so the figure tells that the system is easy to operate. There are 7 participants choose grade 1 in Q1, meaning that they think it

Table 7.5 Answers Statistics of Investigative Questions from Parents

Question	1	2	3	4	5
Q1:Do you think it is easy to leave a message?	4	2			
Q2:Do you think it is easy to find messages and listen?	1	2	3		
Q3:Do you think it is correct of the emotion detection result?	2	2	2		
Q4:Do you think it is easy to change emotion?		1	5		
Q5:Do you think your family connection is enhanced by using this system?	1	3	2		

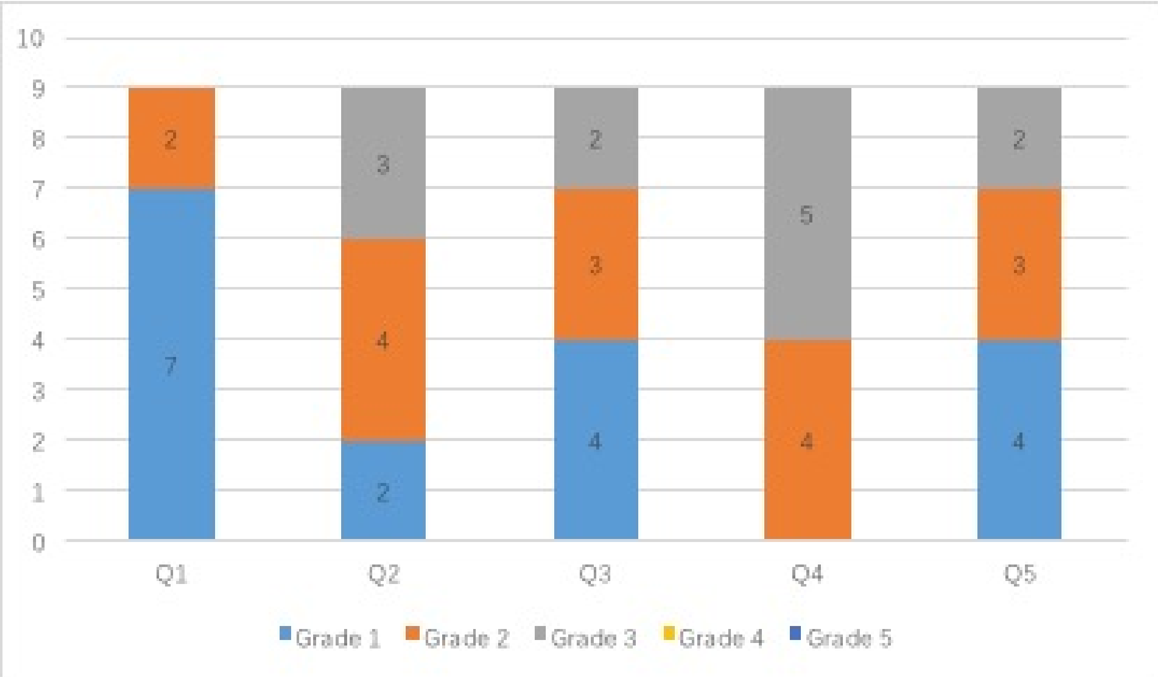


Fig. 7.2 Statistical graph of answers from all participants

easy to leave a message and 2 participants choose grade 2 (easy). In Q2, 2 participant choose grade 1 (very easy) and 4 participants choose grade 2 (easy) and 3 participants choose grade 3 (normal). In Q3, 4 participants choose grade 1 (very easy), 3 participants choose grade 2 (easy) and 2 participants choose grade 3 (normal). In Q4, 4 participants choose grade 2 (easy) and 5 participants choose grade 3 (normal. In Q5, 4 participant choose grade 1 (very easy), 3 participants choose grade 2 (easy) and 2 participants choose grade 2 (normal).

The Fig 7.3 shows the average of grade to each question of all participants. We can also see that Q4 has highest average grade. It means that changing emotion are harder than the other functions. But all values float up and down around the value 2 and less than 3. It shows

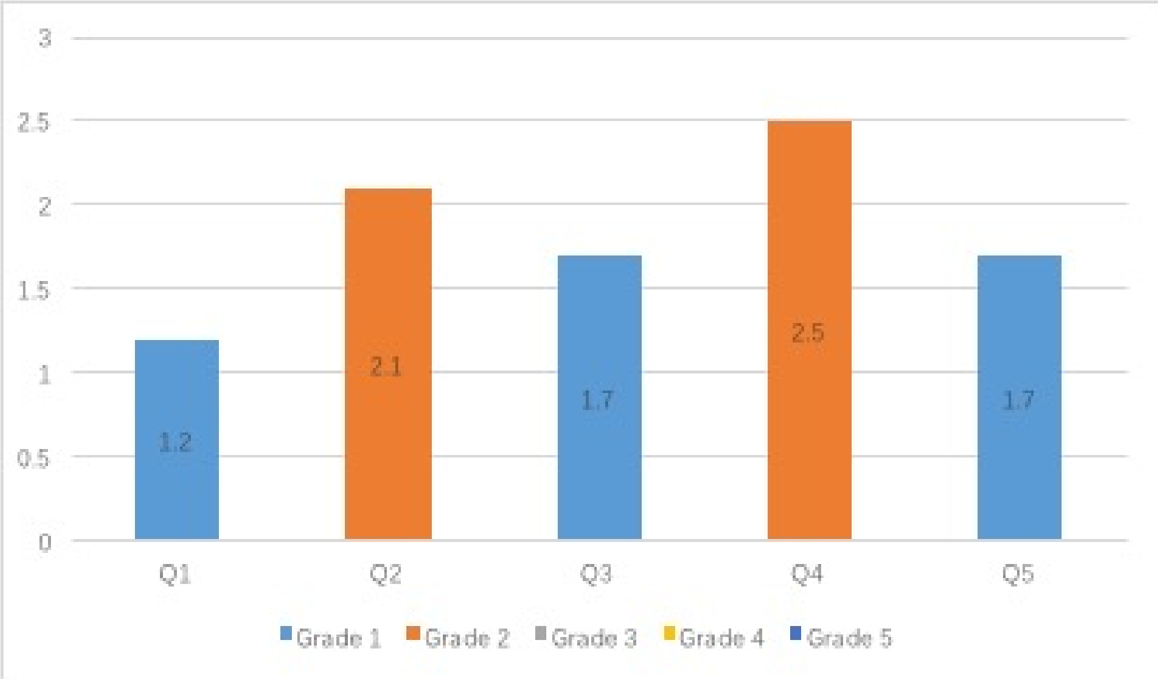


Fig. 7.3 Grade average of each question

the function of this system, such leaving notes, responding notes, and changing emotion are not hard. Especially, for the question 5, the question is "Do you think your family connection is enhanced by using this system?" and we got 4 grade 1 (very much), 3 grade 2 and 2 grade 3. Also the average of this question is 1.7, which can tells that the system do enhance the family connection of our participants.

We also get some comments and questions from participants:

1. "You can add more emotions in your system, if the system can detect more complex emotions, maybe the virtual agents will be more vivid."

2. "Can we check the notes we left days ago?" Ans: You can't in this version of the system, this is a function we want to add in the future work. We want to think more about how to save notes, maybe users can remain all the notes they have listened or they can choose notes to save. This is our next work.

3. "The system is great, our family really like it. But if you can make more virtual agents with more looks, and this will make more fun."

Chapter 8

Conclusion

8.1 Summary

In summary, this paper introduces some problem with the modern family, and in order to solve these problem, we propose an augmented reality system in smart-phone to leave emotional notes in the real environment using Virtual-Agent to enhance family connection.

We assume that there are 3 people in the modern family as Dad, Mom and child. The system can run in smartphones of each family member. Spatial note system is based on two parts, one is the virtual agent services and the other is the AR (augmented reality) system. The virtual agent services allow users to make voice messages by recording users' voice and is able to detect the users' voice after recording it. Then generate a Virtual-Agent with the appropriate facial expression to express users' emotion. Users can also change the facial expresses as they like to express feeling. There are 4 emotions of the virtual agents: happy, sorrow, angry, calm. Each emotion has 4 levels to express. The AR (augmented reality) system can detect the nearest plane in the real environment through the phone's camera. Users can put the Virtual-Agent anywhere they want in the reality according to the AR system.

Through this system, family members can exchange information, share feeling, discuss topics, and what's more, bond their relationships. For the note-making user, he can make any note he wants in the real world and conveys his emotion to the other users easily. For the note-receiving user, he can freely choose whether to hear it according to the agent's facial expression. He can assume the others' feelings today by seeing agents' faces intuitively.

In the evaluation part, we conducted a experiment to evaluate the usability of the system. We let volunteers to use the system in their family and then did the questionnaires to give us feedback. We collected and analyzed the answers of these volunteers and positive feedback has been received.

8.2 Future Work

In the future, there are still some function need to be done, such as more roles for family to choose, more accurate emotions to add to make the virtual agents more vivid, and also add some place to store all notes users left. Each work needs effort. The virtual agent can have more facial expression in the future. In short, making more vivid virtual agent with the Augmented Reality is the point of the future work.

References

[1] Ronald T Azuma. A survey of augmented reality. *Presence: Teleoperators & Virtual Environments*, 6(4):355–385, 1997.

[2] Mark Billinghurst, Adrian Clark, Gun Lee, et al. A survey of augmented reality. *Foundations and Trends® in Human–Computer Interaction*, 8(2-3):73–272, 2015.

[3] Patrick Schueffel. The concise fintech compendium. *Fribourg: School of Management Fribourg/Switzerland.*, 2017.

[4] Louis B Rosenberg. Virtual fixtures: Perceptual tools for telerobotic manipulation. In *Virtual Reality Annual International Symposium, 1993., 1993 IEEE*, pages 76–82. IEEE, 1993.

[5] Dimitris Chatzopoulos, Carlos Bermejo, Zhanpeng Huang, and Pan Hui. Mobile augmented reality survey: From where we are to where we go. *IEEE Access*, 5:6917–6950, 2017.

[6] Paul Milgram and Fumio Kishino. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, 77(12):1321–1329, 1994.

[7] Mark Billinghurst and Hirokazu Kato. Collaborative mixed reality. In *Proceedings of the First International Symposium on Mixed Reality*, pages 261–284, 1999.

[8] Yuichi Ohta and Hideyuki Tamura. *Mixed reality: merging real and virtual worlds*. Springer Publishing Company, Incorporated, 2014.

[9] Steve Benford, Chris Greenhalgh, Gail Reynard, Chris Brown, and Boriana Koleva. Understanding and constructing shared spaces with mixed-reality boundaries. *ACM Transactions on computer-human interaction (TOCHI)*, 5(3):185–223, 1998.

[10] Rachel Metz. Augmented reality is finally getting real technology review. *Internet WWW-page, URL: http://www. technologyreview. com/news/428654/augmentedreality-is-finally-getting-real/(10.07. 2013)*, 2012.

[11] Yohan Baillot, L Davis, and J Rolland. A survey of tracking technology for virtual environments. *Fundamentals of wearable computers and augumented reality*, page 67, 2001.

[12] Ronald Azuma, Yohan Baillot, Reinhold Behringer, Steven Feiner, Simon Julier, and Blair MacIntyre. Recent advances in augmented reality. *IEEE computer graphics and applications*, 21(6):34–47, 2001.

[13] Gentaro Hirota, David T Chen, William F Garrett, Mark A Livingston, et al. Superior augmented reality registration by integrating landmark tracking and magnetic tracking. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 429–438. ACM, 1996.

[14] Mark Billinghurst, Hirokazu Kato, and Ivan Poupyrev. The magicbook: a transitional ar interface. *Computers & Graphics*, 25(5):745–753, 2001.

[15] Gun A Lee, Andreas Dünser, Seungwon Kim, and Mark Billinghurst. Cityviewar: A mobile outdoor ar application for city visualization. In *Mixed and Augmented Reality (ISMAR-AMH), 2012 IEEE International Symposium on*, pages 57–64. IEEE, 2012.

[16] Jonathan Gratch, Ning Wang, Jillian Gerten, Edward Fast, and Robin Duffy. Creating rapport with virtual agents. In *International Workshop on Intelligent Virtual Agents*, pages 125–138. Springer, 2007.

[17] Amy L Baylor. Promoting motivation with virtual agents and avatars: role of visual presence and appearance. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3559–3565, 2009.

[18] Margaret Rouse. Virtual agent (intelligent virtual agent or virtual rep). https://searchcrm.techtarget.com/definition/virtual-agent. Accessed May 21, 2018.

[19] Stephen Haag, Maeve Cummings, and James Dawkins. Management information systems. *Multimedia systems*, 279:280–297, 1998.

[20] Vajid H Jafri and Sajid H Jafri. Reservation software employing multiple virtual agents, November 3 1998. US Patent 5,832,454.

[21] Ramin Yaghoubzadeh, Marcel Kramer, Karola Pitsch, and Stefan Kopp. Virtual agents as daily assistants for elderly or cognitively impaired people. In *International Workshop on Intelligent Virtual Agents*, pages 79–91. Springer, 2013.

[22] Jamy Li. The benefit of being physically present: A survey of experimental works comparing copresent robots, telepresent robots and virtual agents. *International Journal of Human-Computer Studies*, 77:23–37, 2015.

[23] Alberto Del Bimbo and Enrico Vicario. Specification by-example of virtual agents behavior. *IEEE Transactions on Visualization and Computer Graphics*, 1(4):350–360, 1995.

[24] Timothy W Bickmore, Laura M Pfeifer, and Michael K Paasche-Orlow. Health document explanation by virtual agents. In *International Workshop on Intelligent Virtual Agents*, pages 183–196. Springer, 2007.

[25] Volker Wiendl, Klaus Dorfmüller-Ulhaas, Nicolas Schulz, and Elisabeth André. Integrating a virtual agent into the real world: The virtual anatomy assistant ritchie. In *International Workshop on Intelligent Virtual Agents*, pages 211–224. Springer, 2007.

[26] Jun Rekimoto and Yuji Ayatsuka. Cybercode: designing augmented reality environments with visual tags. In *Proceedings of DARE 2000 on Designing augmented reality environments*, pages 1–10. ACM, 2000.

[27] Mihai Bâce, Teemu Leppänen, David Gil de Gomez, and Argenis Ramirez Gomez. ubigaze: ubiquitous augmented reality messaging using gaze gestures. In *SIGGRAPH ASIA 2016 Mobile Graphics and Interactive Applications*, page 11. ACM, 2016.

[28] Pranav Mistry, Pattie Maes, and Liyan Chang. Wuw-wear ur world: a wearable gestural interface. In *CHI'09 extended abstracts on Human factors in computing systems*, pages 4111–4116. ACM, 2009.

[29] Alaeddin Nassani, Huidong Bai, Gun Lee, and Mark Billinghurst. Tag it!: Ar annotation using wearable sensors. In *SIGGRAPH Asia 2015 Mobile Graphics and Interactive Applications*, page 12. ACM, 2015.

[30] Hiroyuki Tarumi, Ken Morishita, Megumi Nakao, and Yahiko Kambayashi. Spacetag: An overlaid virtual system and its applications. In *Multimedia Computing and Systems, 1999. IEEE International Conference on*, volume 1, pages 207–212. IEEE, 1999.