Spatial Note System Using Virtual Agent to Enhance Family Connection

Puxuan Qu

Graduate School of IPS
Waseda university
Fukuoka, Japan
Email: qpx1995331@outlook.com

Jiro Tanaka

Graduate School of IPS
Waseda university
Fukuoka, Japan
Email: jiro@aoni.waseda.jp

***Abstract*—The pace of life is becoming faster. As a result, some people are too busy to communicate with their families. In this paper, we propose an Augmented Reality (AR) system using smartphone. The system allows us to leave emotional notes in the real environment using Virtual-Agent. The crucial part of the system focuses on the emotional short voice message exchange with the Virtual-Agents. Spatial note system is based on two parts, one is the virtual agent services and the other is the AR system. The virtual agent services allow users to make a voice message by recording user's voice. Then, a Virtual-Agent with the appropriate facial expression is generated. User can also change the facial expressions as he likes. There are 4 emotions of the virtual agents: happy, sorrow, angry and calm. Each emotion has 4 levels to express. The AR system will detect planes from the smartphone view of the real world. Users can put the Virtual-Agent anywhere on a plane.**

Keywords–Augmented Reality; Virtual Agents; Emotional Notes

I. INTRODUCTION

Some people are too busy to communicate with their families. When parents are busy with their work, the children are studying at school. It is also difficult for children to chat with their parents in their home because their parents return home late in the evening.

The best way to enhance the family connection is to provide the means which will increase the communication between children and parents. We try to find a way for both children and parents to share each other's life without taking up their time.

Message exchange will be an appropriate method in sharing information. The current ways of exchanging messages are:

1. Leaving a sticky note at home. This approach makes it easy for others to understand what the user wants to express because the notes are placed in the real world. One drawback is that these notes can easily be lost.
2. Sending a message in LINE using smartphone. In this case, users can read messages instantly, regardless of time and place. It also delivers the true voice of the user. But on the other hand, Messages are completely separated from the real world.

So, we propose an Augmented Reality (AR) system using smartphone. Our system can leave emotional notes in the real environment. The crucial part of the system is that we focus on the emotional short voice message exchange using the Virtual-Agent. Our system allows users to make voice message by recording user's voice. Additionally, the system can detect the emotion, i.e., calm, happy, angry or sorrow, from their voice. Then, the system generates a Virtual-Agent with the appropriate facial expression. Users can place the Virtual-Agent anywhere they want in the real world by using their smartphones. This mechanism can help families start small conversations, talking about daily life without facing each other.

AR is a new technology to make computer interfaces invisible and enhance user interaction with the real world [1]. 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 [2]. AR 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].

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. They can be embodied as software, such as chat bots [4]. Virtual agents can be autonomous and can also work with other agents and personnel [5].

II. GOAL AND APPROACH

The goal of this research is to create a useful and novice system that helps busy families share, communicate and exchange information.

In this paper, we propose Spatial Note System using AR and Virtual Agent. With the help of this system, family members can exchange information, share feelings, discuss topics, and eventually, strengthen family connection and keep good relationships

Through this system, we can:

1. Put messages into real environment to create environment-supportive messages (like sticky notes),
2. Make messages compelling so that it won't be ignored (like sticky notes),
3. Leave messages based on the real voice (like LINE), and
4. Make messages convey emotions to improve empathy.

So, we propose an Android AR system that uses Virtual-Agent to leave emotional notes in space to enhance family connection

 ISBN: 978-1-61208-686-6

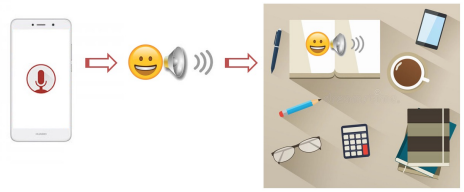

Figure 1. Family member A gets home, he can leave a voice message using smartphone and put the emotional voice message in the real environment.

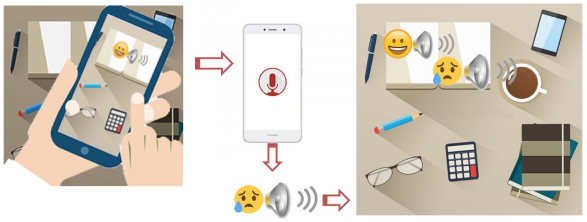

Figure 2. Family member B gets home, he can see that message. And after listening message of A, B leave back a new one.

and convey emotions. The crucial part of the system focuses on the emotional short audio message exchange with Virtual-Agent.

Spatial note system allows the user to create a voice message by recording the user's voice. The following is an example of how the system works. In Figure 1, when a family member A returns home, he can use smartphone to leave a voice message and put the emotional voice message in the real environment. In Figure 2, when other family member B gets home, he can see that message. And after listening to the message of A, B can figure out what A means and what A feels. And then B can leave a new message. Additionally, the system can detect the users' voice after recording and then generate a Virtual-Agent with the appropriate facial expression. These facial expressions express user's emotions. The virtual agents can detect four main emotions from the user's voice, they are calm, happy, angry and sorrow. Users can use a smartphone to place Virtual-Agent anywhere they want.

III. Spatial Note System

In this section, we introduce the usage scenario of the system and explain how to use this system.

A. Usage Scenario

- Notes-leaving Scenario: Mom prepares a delicious breakfast in the morning for the family and eats with her husband. Before she leaves, she puts her daughter's breakfast on the table. The daughter is still sleeping. She then leaves a voice note to her daughter using the spatial note system in her smartphone. She said:" I bought your favorite bread, please remember eat it all, mom loves you!" Then she uses the smartphone to put the note on the table near the breakfast. The note is represented by a smiling virtual agent and can be seen through smartphone camera. The usage scenario is shown as Figure 3. After that, the parents leave home to go to work.

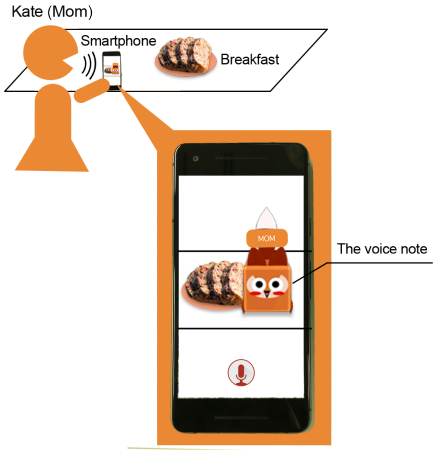

Figure 3. Mom leaves notes in the morning to her daughter about breakfast.

- Responding Scenario: After daughter gets up, she knows that her parents are out. She sees the breakfast on the table. When she walks near the table, there is a new note alert on her phone. She turns on the spatial note system in the smartphone to see that note. When she turns on the phone's camera and holds the phone towards the breakfast, a smiling virtual agent appears on the table. She finds that her mom left her a note, so she clicks on the agent to listen. After hearing what her mom said, she records a new note with the words "Thanks mom, the bread is delicious, but it is too much." She puts this new note next to her mom's note, responding to her mom, shown as Figure 4. Finally, she goes to school.

B. System Overview

Spatial note system is based on two parts, one is the virtual agent services and the other is the AR system (see Figure 5). The virtual agent services allow users to make voice messages by recording user's voice. A virtual agent with the appropriate facial expression is then generated to express the user's emotions. User can also change facial expressions according to his own preferences. Virtual agents have four emotions: happy, sorrow, angry and calm. Each emotion has 4 levels to express. The AR system can detect planes in the real environment through the camera of the mobile phone. Users can put the Virtual-Agent anywhere on a plane using the AR system.

C. Role Classification

In general, there are three types of roles in the family, i.e., Mom, Dad and Child, as shown in the Figure 6. Communication happens between two of them. Once the user chooses one of the roles, she/he will enter the role's interface. It is important to categorize different roles when using the system. Without this feature, it will be difficult to identify who is the note sender and who will be the note receiver. The users of

 ISBN: 978-1-61208-686-6

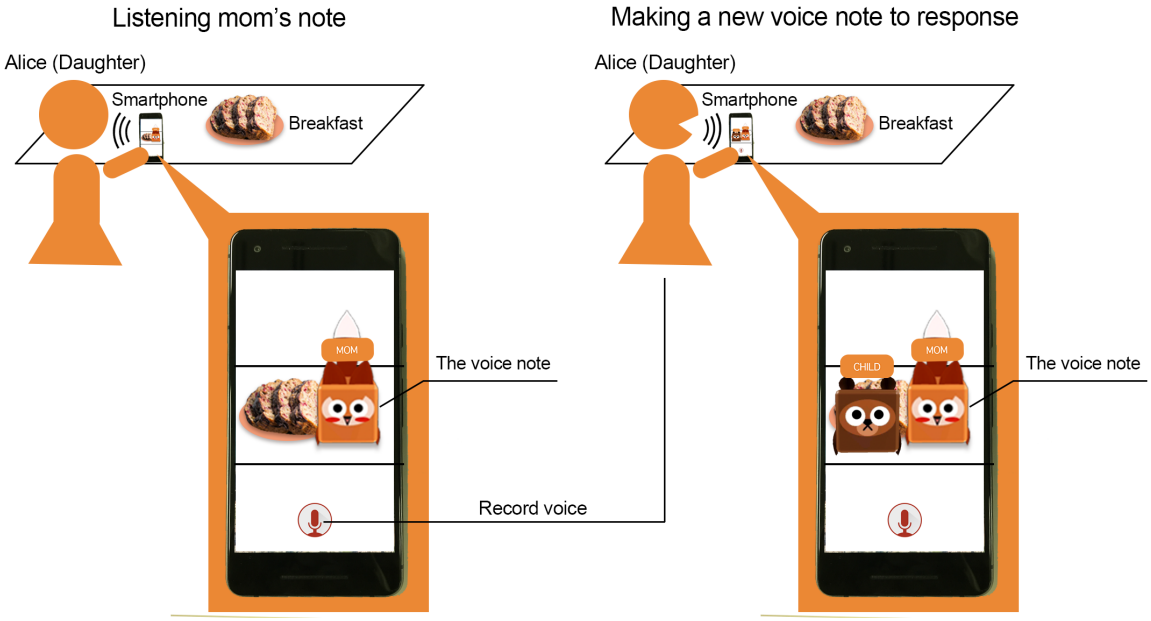

Figure 4. Daughter responds her mom about the breakfast when she gets home.

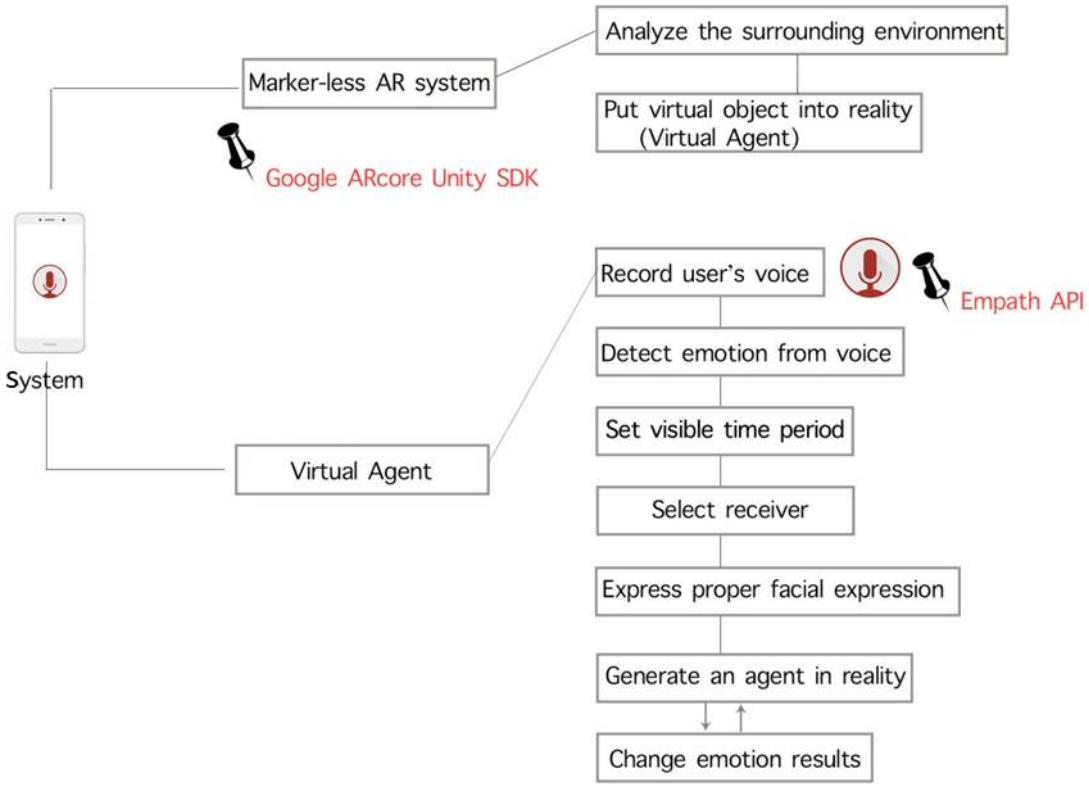

Figure 5. Spatial note system is based on two parts, one is the virtual agent services and the other is the AR system.

 ISBN: 978-1-61208-686-6

Figure 6. From left to right: Mom, Dad, Child.

Figure 7. Mom got a new note from daughter, and from left to right is : before listening to the new note, after listening to the new note.

different roles can communicate with each other and content can only be seen by the addressed user. It is also possible to send messages to all members.

D. New Notes Alert

Once a family member A leaves a note to another family member B, the family member B will receive a new notice. Specifically, in the role's interface, a figure will appear in the lower right corner of the role 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 is decremented by 1. The figure in the lower right corner of the roles icon disappears when the user hears all new notes. This is to remind users not to forget any messages left to him and remind user to reply to the message.

E. Checking New Notes and Listening

Users can hold smartphones to find new notes at home, and if found, the virtual agent will appear on the screen, users can click the virtual agent and listen to the voice notes left by the other family members, as shown in Figure 7.

F. Making Voice Notes

1) Environment scanner: The system uses ARCore [6]. In addition to identifying key points, ARCore can detect flat surfaces, such as tables or floors, and estimate the average illumination of the surrounding area. The combination of these

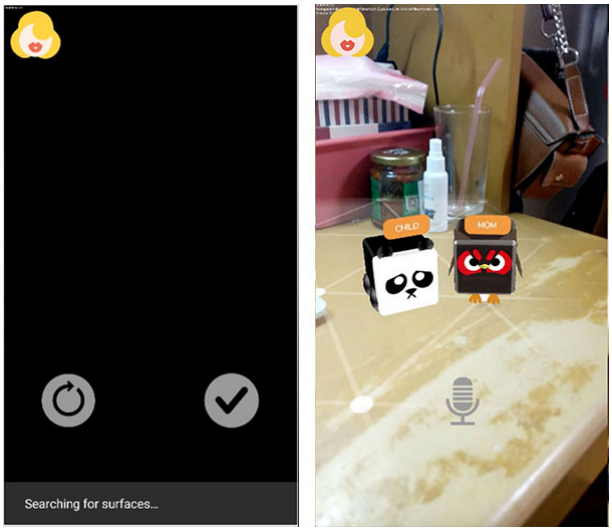

Figure 8. From left to right: Recording voice, system generate mom's virtual agent as voice note and mom puts it near the daughter's.

features allows ARCore to build its own understanding of the world around it. ARCore's understanding of the real world allows users to place objects, notes or other information in a way that integrates seamlessly with the real world. Users can place a napping kitten at the corner of the coffee table or annotate a painting with the artist's biography. This is the fundamental part of putting virtual agent into reality.

2) Recording Voice and Emotion Detection: User can click the Record button to start recording the voice. The user presses stop button after completing the sentence. The Record button is always displayed on the main page, so user can leave a message by recording voice anytime, anywhere. When recording a voice note, the system can perform voice emotion recognition. When the user records his voice to leave a note, the system will call the Empath API [7] and get the emotion detection results. The Empath API recognizes emotion by analyzing physical properties of user's voice, such as pitch, tone, speed and power. It can detect emotion in every language. After the user records his voice and the system completes emotion detection, the spatial virtual agent with the emotional facial expression will be displayed at the location selected by the user. Figure 8 shows the interfaces of voice recording and virtual agent. If the user presses the virtual agent for a long time, emotion results interface will be displayed. Emotional outcomes include 4 emotions, such as calm, angry, happy and sorrow. Each emotion has 4 levels, a little, much, very much and extreme. They are represented by 4 different emojis. Figure 9 shows the images of emotion detection interface. In the interface, there are 4 emotion elements and their values, which are detected from the user's voice. The level of emotion is determined by the energy value, which is also given from Empath API. For the emotional elements, each element is measured independently from speech. The detected values vary from 0 to 50. Only one element will be selected, and this element determines the final emotional outcome.

3) Emotion Result Change: Users can change emotional outcomes by clicking the virtual agent on the detection interface, which can change emotional outcomes and emotion

 ISBN: 978-1-61208-686-6

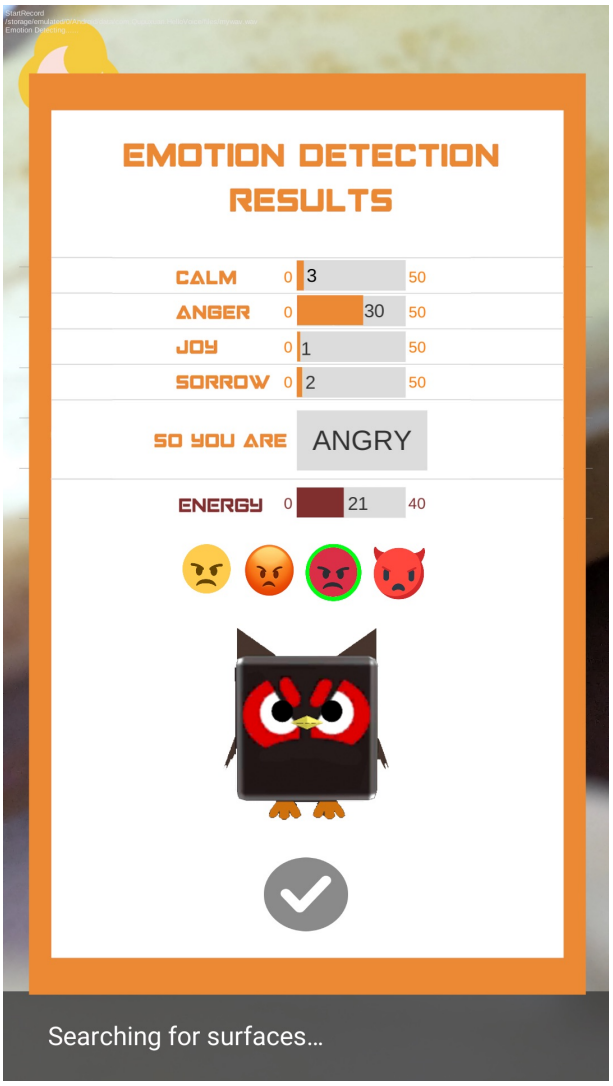

Figure 9. Emotion Detection Result.

levels. Selecting different emotional levels will change the facial expression of the virtual agent. The Figure 10 shows the change of the emotion.

4) Visible Time Setting and Receiver Setting: The user can decide how long this agent exists after being listened to. Options include 1 hour, 4 hours, 8 hours and 24 hours. User can also decide who can receive notes, individuals or entire families. If the user only sends notes to an individual, the notes can only be seen by the recipient, not anyone else. But if the user chooses the entire family, then all family members can see that note.

IV. PRELIMINARY EVALUATION

We conducted a preliminary user study to verify whether users can communicate more with their family by using our proposed system and assess the usability of the system. We asked 3 families to do the experiments.

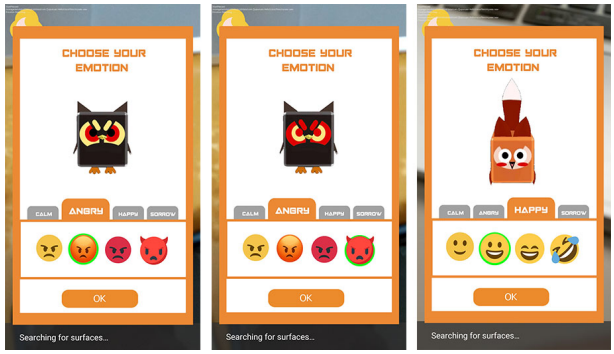

Figure 10. Emotions changes from left to right: a little angry, extremely angry, a little happy.

A. Participants

As shown in Table 1 and Table 2, in order to assess the usability and efficiency of using our system, we plan to recruit 3 families from friends, each with 3 participants. A total of 3 children participated, aged between 22 and 24, with 6 parents, aged 49 to 52. All participants have a general knowledge of computers, and have experience using smartphones.

TABLE I. SUBJECT DEMOGRAPHIC INFORMATION OF CHILDREN.

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

TABLE II. SUBJECT DEMOGRAPHIC INFORMATION OF PARENTS.

Elements	*Description*
Participants	3 males, 3 females
Age	49-52; Mean: 51
Profession	Working in company

B. Method

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

a) Sending a message to a person

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

The method we use for evaluation is described in the following 6 steps:

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

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

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

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

5. Each participant is asked to fill a questionnaire survey after finishing their tasks. Each participant needs to write down the number of messages he/she wrote today and answer 5 questions about using the system. The answer is from 1 to 5 (1 = very positive, 5 = very negative).

 ISBN: 978-1-61208-686-6

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				

Figure 11. Answers Statistics of Investigative Questions from Children.

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		

Figure 12. Answers Statistics of Investigative Questions from Parents.

These question are:

- 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?

C. Results

After collecting the results given by the participants, the evaluation of using the spatial note system to enhance family connection can be carried out.

The Figure 11 shows all the answers from children. There are 3 participants chose Grade 1 in Q1, which means that they all thought that it was easy to leave a message. In Q2, 1 child chose Grade 1 (very easy) and 2 children chose grade 2 (easy). In Q3, 2 children chose grade 1 (correct) and 1 child chose grade 2 (almost correct). In Q4, 3 children chose grade 2 (easy) and in Q5, 3 children chose grade 1 (very much).

The Figure 12 shows all the answers from parents. There are 4 participants chose grade 1 in Q1, which means that they thought it easy to leave a message and 2 participants chose grade 2 (easy). In Q2, 1 participant chose grade 1 (very easy) and 2 participants chose grade 2 (easy) and 3 participants chose grade 3 (normal). In Q3, 2 participants chose grade 1 (correct), 2 participants chose grade 2 (almost correct) and 2 participants chose grade 3 (normal). In Q4, 1 participant chose grade 2 (easy) and 5 participants chose grade 3 (normal. In Q5, 1 participant chose grade 1 (very much), 3 participants chose grade 2 (much) and 2 participants chose grade 2 (normal).

Overall, we received positive feedback through the preliminary user study.

V. Related Work

With the popularization of the AR, more and more researchers determined themselves in proposing new method, new idea in AR. AR enables the direct or indirect view of the physical, real-world environment whose elements are augmented by computer.

Rekimoto et al. [8] 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, and can determine the 3D position of the tagged object as well as its ID number.

Bace et al. [9] 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. Mistry et al. [10] presents WUW-Wear Ur World, which is a wearable gestural interface that allows projecting information out into the real-world.

Nassani et al. [11] 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.

Tonchidot [12] proposed Sekai Camera, an augmented reality application that allows users to share tags for any place on the planet based on GPS.

Tarumi et al. [13] proposed an overlaid virtual system called SpaceTag. The SpaceTag system is a location-aware information system and an augmented reality system because it attaches information to real space. In this system, the

 ISBN: 978-1-61208-686-6

virtual world consists of virtual architectural objects and virtual creatures.

VI. Conclusion and Future Work

In this paper, we addressed the family communication problems of modern families. In order to solve these problems, we have proposed an AR system using smartphone. The system allows us to leave emotional notes in the real world using Virtual-Agent.

We have assumed 3 kinds of people, i.e., Dad, Mom and Child, in our system. Our system runs on the smartphones of each family member. Our system consists of two parts, one is the virtual agent services and the other is the AR system. The virtual agent services allow the user to make a voice message by recording the user's voice. Then, a virtual agent with an appropriate facial expression will be generated to express the user's emotions. User can also change the facial expressions according to his own preferences. There are 4 emotions, i.e., happy, sorrow, angry and calm. Each emotion has 4 levels to express. The AR system can detect planes in the real environment through the camera of the smartphone. User can put the Virtual-Agent anywhere he wants in the real world using the AR system.

Through this system, family members can exchange information, share feelings, discuss topics and keep good relationships. For the note-making user, he can make any notes he wants in the real world and easily convey his emotion to other users. For the note-receiving user, he can freely choose whether to hear it according to the facial expression of the agent. He can understand the feelings of note-making user by looking at the agent's facial expression intuitively.

In the preliminary evaluation part, we conducted an experiment to evaluate the usability of our system. We asked volunteers to use the system in their homes and asked them to fill the questionnaires to give us feedback. We have collected and analyzed their answers.

There exists several things that need to be done, such as making the emotions of the virtual agents more accurate, and also adding place to store all the notes left by users. Virtual agents can have more facial expressions in the future. In short, using AR to create more vivid virtual agents will be the focus of our future work.

Acknowledgment

The authors deeply thank Boyang Liu for his comments and support for refining this paper.

References

[1] M. Billinghurst, A. Clark, and G. Lee, "A survey of augmented reality," Foundations and Trends® in Human–Computer Interaction, vol. 8, no. 2-3, 2015, pp. 73–272.

[2] R. T. Azuma, "A survey of augmented reality," Presence: Teleoperators & Virtual Environments, vol. 6, no. 4, 1997, pp. 355–385.

[3] P. Schuettel, "The concise fintech compendium," Fribourg: School of Management Fribourg/Switzerland, 2017.

[4] J. Gratch, N. Wang, J. Gerten, E. Fast, and R. Duffy, "Creating rapport with virtual agents," in International Workshop on Intelligent Virtual Agents. Springer, 2007, pp. 125–138.

[5] A. L. Baylor, "Promoting motivation with virtual agents and avatars: role of visual presence and appearance," Philosophical Transactions of the Royal Society B: Biological Sciences, vol. 364, no. 1535, 2009, pp. 3559–3565.

[6] J. Glover, Unity 2018 Augmented Reality Projects: Build four immersive and fun AR applications using ARKit, ARCore, and Vuforia. Packt Publishing Ltd, 2018.

[7] E. Fast, B. Chen, and M. S. Bernstein, "Empath: Understanding topic signals in large-scale text," in Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. ACM, 2016, pp. 4647–4657.

[8] J. Rekimoto and Y. Ayatsuka, "Cybercode: designing augmented reality environments with visual tags," in Proceedings of DARE 2000 on Designing augmented reality environments. ACM, 2000, pp. 1–10.

[9] M. Bâce, T. Leppänen, D. G. De Gomez, and A. R. Gomez, "ubigaze: ubiquitous augmented reality messaging using gaze gestures," in SIGGRAPH ASIA 2016 Mobile Graphics and Interactive Applications. ACM, 2016, p. 11.

[10] P. Mistry, P. Maes, and L. Chang, "Wuw-wear ur world: a wearable gestural interface," in CHI'09 extended abstracts on Human factors in computing systems. ACM, 2009, pp. 4111–4116.

[11] A. Nassani, H. Bai, G. Lee, and M. Billinghurst, "Tag it!: Ar annotation using wearable sensors," in SIGGRAPH Asia 2015 Mobile Graphics and Interactive Applications. ACM, 2015, pp. 12:1–12:4.

[12] B. Butchart, "Architectural styles for augmented reality in smartphones," in Third International AR Standards Meeting, 2011, pp. 1–7.

[13] H. Tarumi, K. Morishita, M. Nakao, and Y. Kambayashi, "Spacetag: An overlaid virtual system and its applications," in Multimedia Computing and Systems, 1999. IEEE International Conference on, vol. 1. IEEE, 1999, pp. 207–212.

 ISBN: 978-1-61208-686-6