

Human Agent Interaction for Learning Service-Minded Communication

Kaoru Sumi¹ and Ryuji Ebata¹

¹Future University Hakodate

Abstract: This paper describes a dialogue system for learning how to serve customers by communicating with a character agent using a facial expression recognition system that employs brain wave measurement equipment, a speech recognition system, a speech synthesis system, and a dialogue control system using AIML (artificial intelligence markup language). The purpose of this system is to communicate the Japanese concept of service-mindedness, as typified by paying attention to customers. The system was evaluated through an experiment with subjects who used it to learn how to serve customers.

1 Introduction

This paper introduces a system for providing educational training in hospitality via dialogue with a character agent. This system focuses on the Japanese concept of service-mindedness, which is typified by paying attention to individual customers. As is often said, Japanese airline service is ranked among the highest in the world. This “service-mindedness” means a technique or an attitude for making a customer happy. Smiling is said to be as important for servicing customers [1]. Our system displays a 3D model character agent, and it has the function of facial expression recognition using brain wave measuring equipment, a dialogue via speech recognition and a speech synthesis system.

According to Media Equation [2], people tend to respond to media as they would to another person. Persuasive Technology [3] is a technology using which media can persuade users to do something. Our research group has studied persuasive technology via human agent interaction using facial expressions and dialogues, with the goal of developing a virtual agent that can persuade a human to act in a certain way [4][5][6].

Nowadays, we are coming closer to voice interaction, e.g., Google’s speech recognition for Android, and Siri which uses a natural language user interface for the iOS. In recent years, speech dialogue systems have been developed. Galatea[7] is a speech dialogue system which displays a character agent using face image synthesis that can make a 3D facial image from a photograph of

someone’s face. Galatea also uses speaker adaptive speech recognition, a speech synthesis system, and a changeable dialogue control system. MMDAgent[8] is a speech dialogue system building tool which uses a speech recognition system, a speech synthesis system, a 3D agent display system and a voice control system. As an evaluation of communication between a human and a machine, there is the Turing test [9][10] which is a test of a machine’s ability to exhibit intelligent behavior equivalent to, or indistinguishable from, that of an actual human. Alicebot[11] is a speech dialogue system building tool which uses AIML (artificial intelligence markup language) [11]. Alicebot was awarded the Loebner Prize which is the prize of the Turing test in 2000, 2001, and 2004. We use AIML for a dialogue control system. The main characteristic of our system is that we can communicate with the agent interactively through facial expressions because brain wave measuring equipment is used to recognize our facial expressions.

2 Character Agent System

Our character agent system consists of a 3D character agent display system, a speech recognition system, a speech synthesis system, a dialogue control system and a facial expression recognition system (Figure 1).

The 3D character agent display system displays a 3D character agent which can show the facial expressions of “smile”, “laugh”, “anger”, “sadness”, “disgust”, “fright”, and “surprise.” These facial expressions can be

controlled in intensity digitally. The system superimposes the mouth shapes of the vowels onto each facial expression, lip-synching with the sounds. Fluid movement of the facial expressions and lip-synching is made possible using Microsoft XNA morphing technology. Our system uses Google speech API as the speech recognition system, and AITalk as the Japanese speech synthesis system. As a dialogue control system, we revised Artificial Intelligence Markup Language (AIML) which is based on Extensible Markup Language (XML) to use the Japanese language. Using templates, we can express a dialogue freely.

```
<pattern>* tell me your name *</pattern>
```

```
<template>My name is Ayaka</template>
```

For example, if there is a question “Tell me your name?” from a user, then the system answers “My name is Ayaka” in this case. The important pattern here is “*tell me your name*”, so the system can answer the sentence “Please tell me your name” or “Could you tell me your name?” because these sentences include the phrase “tell me your name.”

When we analyze Japanese language, we have to add spaces between words because there are no spaces in Japanese sentences. We use MeCab, which is a fast and customizable Japanese morphological analyzer, to add spaces between words

As the facial expression recognition system, this system uses the facial recognition application of brain wave measuring equipment called “Emotive EPOC”. It recognizes the intensity of facial expressions digitally.

Using our system, a user can talk to the character agent, and the character agent teaches the user how to interact with customers. The character agent is displayed on the screen and sometimes the system displays lines in a scene which the user should practice, along with an appropriate facial expression. The system judges whether the user has spoken the lines appropriately by comparing them with the speech recognition system using AIML templates. The system judges whether the user’s facial expression is appropriate or not by comparing it with the facial expression recognition system.

3 Service-minded communication

The material to be learned about service-minded communication is divided into several topics. In this research, we set two topics, which are “how to compliment a customer” and “how to settle the complaints of a customer” by reference to the customer



Figure 1. System snapshot

service book [1]. Users can practice the expressions to be used in customer service in the scene by talking to the character agent. Users learn how to deal with a customer using a smiling face in the topic of “how to compliment a customer”, and using a sad face, which is the facial expression of caring about others, in the topic of “how to settle the complaints of a customer.”

Basically, the process of this training is that first the character agent shows a user talking lines emotionally as an example of how to compliment a customer or how to settle the complaints of a customer, and then the user tries to say the lines, with the same emotion that the agent has shown.

According to the settings of the brain wave measuring equipment or depending on the person, facial expression recognition sometimes does not work well for some people. In consideration of that, the system has been designed not to repeat training indefinitely. According to the user’s pre-test scores for expressing a smiling face and the score taking the average of these scores as a threshold value, the system automatically regulates the threshold value to a lower level if the user’s facial expressions fail to be recognized.

The dialogue flow of our system is as follows.

1. The user chooses a topic.
2. The agent prompts the user to speak if the line spoken by the user is not recognized by the speech recognition system using the AIML templates.
3. The agent prompts the user to improve his facial expression if it is weak judging by the facial expression recognition system. The expression needs to exceed the standard value 0.2 of the EPOC application.

4 Experiment

We conducted an experiment using the developed system on twenty subjects, comprising males and females aged from 21 to 25 years old (AV. age 22.15, SD age 0.73). We conducted an experiment to examine the effectiveness of a system designed for learning about service-mindedness, and at the same time, because the facial expression can be varied in intensity digitally using our system, we were able use such variations to investigate how the strength of the facial expression affected the system. We prepared a strong facial expression condition which expressed a clear emotional display, and a weak facial expression condition which conveyed less well-defined emotions. Each subject was assigned one of the two conditions, and answered a questionnaire.

The experiment procedure was as follows.

1. Getting the subject to read the system manual.
2. Checking the operation of the brain wave measuring equipment and conducting a pre-test.
3. Getting the subject to wear an earpiece and click the start button.
4. Getting the subject to learn about two topics and stop the system.
5. Getting the subject to answer the questionnaire.

The subjects were asked 11 questions (Figure 2). We describe the comment from a subject in this paper.

For the question “Did you enjoy the system?” there were many very favorable comments. For example, the answers were “I enjoyed it because I had never seen such a system”, “The system is very helpful for training in use of facial expressions”, “I enjoyed talking to the system”, “It is good to understand that if the way of speaking is different, the way of conveying the meaning is also different”, “I thought it was great that the system told me to smile more when I didn’t smile”, “I learned what kind of facial expression to use when I serve a customer”, and so on. We believe that the reasons there were so many positive comments were that the users could learn by talking and they enjoyed showing the facial expressions that were suggested by the brain wave measuring equipment. On the other hand, there were comments about difficulty, such as “It is difficult to show facial expressions” and “The recognition of sound was difficult.”

For the question “Do you want to use this system again?” there were many very favorable comments. For example, “I want to try some other facial expressions”, “It is helpful training to serve customers”, “It was a fresh

Q. 1	Do you have experience in customer service? If yes, tell me the work and working period.
Q. 2	Did you enjoy the system?
Q. 3	Do you want to use this system again?
Q. 4	Does this system seem to be helpful for service training?
Q. 5	Do you have an aversion to attaching a headset?
Q. 6	What was your impression of the agent? 6.1: good-bad, 6.2: kind-unkind, 6.3: polite-rude, 6.4: gentle-roughly
Q. 7	Was the system easy to talk?
Q. 8	Was the voice from the system easy to understand?

Figure 2. Questions from the questionnaire

experience for me”, “The training helps me with waiting on customers”, “I can confirm whether I am giving a real smile with the system”, “I want use it again if I can learn other things”, and so on. We believe that the reasons there were so many comments like this were that the users enjoyed using the system and recognized that it was helpful. On the other hand, there were comments such as “It was easy to use the system if I knew what kind of facial expression I was making” and “I do not know whether the criterion of expression recognition of the system is right.” It is thought that these comments are caused by the fact that the user’s expression is not displayed on the screen and the system was not trusted.

For the question “Does this system seem to be helpful for service training?” there were many very favorable comments. For example, “I can have the system evaluate my expression objectively”, “The system is a chance to review my expression”, “I learned how to speech”, “The system can teach knowledge I did not have previously and it is useful for training in customer service”, “I cannot do it only in a mirror, but using the system is helpful for making expressions” and so on. We believe that the reasons there were so many positive comments were that it was a novelty to have the practice determined by a machine and users could be trained to deal with issues they were not aware of. On the other hand, there were comments such as “The system advances to the next subject even if I cannot form an expression well.” It is thought that the consideration shown to users of allowing them to advance to the next subject after they had made several mistakes causing them to fail with speech or facial recognition was actually considered a problem.

The subjects were divided over the answer to the

question “Do you have an aversion to attaching a headset?” For example, there were positive answers such as “I feel it was like a hat”, and “I feel really comfortable with the headset”. On the other hand, there were negative answers such as “I had a pain behind the ear”, and “It was a bit hard to attach it.” It is thought that the users had a hard time adjusting the headset.

For the question “What was your impression of the agent?”, there were many comments that were critical when the strength of the agent’s expression was weak. For example, “It showed a slightly cold expression”, “It looked a little scared”, and so on. On the other hand, there was a comment that “I had a gentle impression of the agent.” The comment “I was slightly scared because the eyes were open all the time” is thought to mean that the agent did not have the human quality of being able to blink.

For the question “Was the system easy to talk?” there were many comments that were critical. For example, “It was hard for my words to be recognized.” Improvement of the speech recognition system is needed. There was a comment that “Because the face of the agent was in front of me, I was a little tense.” On the other hand, there were the comments that “I have learned the service by being tense” and “It was more comfortable than talking with a person in a meeting.” In this way, we might say that the system utilized the advantage of being a machine.

For the question “Was the voice from the system easy to understand?” there were many very favorable comments such as “It is easy to hear it”, “The sound is very easy to understand.”

5 Discussion

We conducted an experiment to examine the effectiveness of a system designed to teach about the service-mindedness that is found in Japanese hospitality and is exemplified by consideration for the customer. The system operates using dialogues between a user and a character agent. The subjects were divided into two groups in which the agent displayed either strong or weak facial expressions.

Overall, the answers to the questionnaire revealed many positive opinions. Considering the free answers to each question, (1) subjects felt the brain wave measuring equipment and the speech recognition and synthesis system were novelties, (2) the system’s specialty was that it facilitated learning through dialogues and it was highly

motivating for learners to be aware of the facial expression of the agent, and (3) it judged voices and appearances objectively so that users could teach themselves about service-mindedness.

However, there was some feedback that the system proceeded even if the user’s facial expression was poor when the user practiced lines. This happened when the system’s loop avoidance function came into operation for some people for whom facial expression recognition would not work, as mentioned before. Consideration could be given to using image recognition and brain wave measuring equipment concurrently, or using values of facial and emotional recognition on EPOC concurrently. Other feedback indicated that subjects could not evaluate whether their facial expression was really being understood because they could not see their actual face on the display. The system needs further improvement to the human interface. There were some opinions that the character agent should blink its eyes. Blinking, nodding, age and sex have an effect on the impression of the character agent, and need further consideration.

The brain wave measuring equipment and the speech recognition and synthesis system which are used in our system were highly motivating; however subjects were divided over their use, and some felt them to be awkward to use. Some of them made comments on the speech recognition system such as “It was difficult to recognize words”, “It was difficult”, etc. Consideration should be given to improving the system’s ease of use by developing or using a higher-accuracy speech recognition system. Also, regarding the brain wave measuring equipment, some made comments on it such as “It was hard to wear” and “It was painful”, so it could be considered that it was not easy to accept. Consideration should be given to improving the human interface, including using image processing and the headset concurrently.

6 Conclusion

This paper introduces a dialogue system for learning about service-minded communication through talking with a character agent, including how to deal with customers by copying lines and facial expressions provided for various situations. The effect of the system is evaluated using test subjects.

References

- [1] Hiromichi Morishita: Shinri Sekkyaku Jutsu (Cosinitive Service for Customers), Soshimu(2009), In Japanese.
- [2] Reeves, Byron and Clifford Nass The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places, Cambridge University Press, 1996.
- [3] B.J.Fogg: Persuasive Technology –Using Computers to Change What We Think and Do-, Elsevier, 2003.
- [4] Kaoru Sumi and Mizue Nagata: Evaluating a Virtual Agent as Persuasive Technology, Psychology of Persuasion, Janos Csapó and Andor Magyar eds., Nova Science Publishers.2010.
- [5] Kaoru Sumi: Human Interface of Robots or Agents via Facial and Word Expression, International Symposium on Artificial Life and Robotics(AROB), (2012).
- [6] Kaoru Sumi, Mizue Nagata: Characteristics of Robots and Virtual Agents as a Persuasive Talker, HCII, Lecture Notes in Computer Science, Springer (2013), To appear.
- [7] <http://hil.t.u-tokyo.ac.jp/~galatea/index-jp.html#documents>
- [8] <http://www.mmdagent.jp/>
- [9] Alan Turing: Computing Machinery and Intelligence, 433–460, Mind LIX (236), (1950).
- [10] Brian Christian: The Most Human Human: What Artificial Intelligence Teaches Us About Being Alive, Anchor (2012)
- [11] AIML: <http://www.alicebot.org/documentation/>