Affective Human Computer Interaction

Kaoru Sumi
Future University Hakodate,
Hakodate, Hokkaido, Japan

E-mail: kaoru.sumi@acm.org

Abstract

This paper introduces a study of spoken dialogue agent systems using emotional expressions as affective human computer interaction. The paper describes an experiment investigating the effect that the expression and words of the agent have on people, introduces a spoken agent for customer services using expressive facial expressions and a spoken agent for mental care using expressive facial expressions and positive psychology as application systems for affective human computer interaction, and presents a discussion and a conclusion.

Keywords: Affective computing, facial expression, human computer interaction

1. Introduction

According to Media Equation [1], people treat computers, television, and new media as real people and places, thereby making the users uncomfortable if an agent behaves in a disagreeable manner. In the field of persuasion technology research [2] it is said that if a user recognizes the presence of something in a computer, he or she will respond to it according to the normal social rules. However, there are still many things that we do not know about how an agent's response affects a user during their interaction.

In the development of intelligent systems, it is important to consider how best a feeling of affinity with the system and show the presence of the system that has human-like intelligent functions such as recommendation or persuasion. Therefore, evaluating the interpersonal impressions conveyed by agents is very important.

Our research group has developed intelligent dialogue system that proactively interacts with a user according to the user's circumstances. As a first step, we examined how to react with the user under several emotional situations through the experiment.

We performed an experiment to evaluate how the facial expressions of an agent and the words used by the agent affected users during agent-user interaction.

In the past, ELIZA [3] was developed to imitate a Rogerian psychotherapist, and was an early example of primitive natural language processing. SimCoach [4] is a spoken dialogue agent system used in the mental care of returned soldiers. It displays a speaking counselor agent and analyses the psychology of the user using a question and answer technique.

The paper describes an experiment investigating the effect that the expression and words of the agent have on people, introduces a spoken agent for customer services using expressive facial expressions and a

spoken agent for mental care using expressive facial expressions and positive psychology[5][6] as application systems for affective human computer interactions, presents a discussion and a conclusion.

2. Experiment investigating impressions and behavior change caused by replies from the agent

We conducted an experiment to examine how the impression that the user gets from the agent's answer is affected by the combination of facial and word expressions. It was intended to clarify the impression that the agent gave the user by answering when interacting with the user in an emotion-arousing scenario.

We chose six kinds of feelings. From the total of 216 combinations, covering multiple feelings that the user felt (6 patterns) and the facial expressions for the agent's interaction with the user (6 patterns) and word expressions used by the agent (6 patterns), we selected 96 patterns in this experiment. These covered 16 patterns in each feeling: empathetic words and consistent facial expressions, nonempathetic words and consistent facial expressions, word consistent and facial inconsistent, and word inconsistent and facial consistent. This is because the conditions of nonempathetic and both inconsistent word and facial expressions are nonsensical in normal communication. This is the condition where the word and facial expressions are inconsistent, which is the condition for double bind communication, but it can be considered as either word or facial being empathetic to the user. The case of nonempathetic condition and inconsistent word and facial expressions can be considered as pathological.

A total of 1236 people, 568 male and 668 female (AV. 38.0, SD 11.5), were assigned 96 contents. More than ten users were assigned to each content [7].

As a result, people are easy to be persuaded when the agent was favorable impression. This result is very natural; however, the combination of favorable was different from our prediction. First, we predicted that words and facial expressions reflected on the emotions aroused by the scenario would lead to the most favorable impression, so we set these data as the control group. In fact, there were more favorable impressions than those obtained for the control group. For example, the words and facial expressions were "joy" when the user's emotion was "joy" for the control group. It is very interesting that when the user's emotion was "joy", the agent's words for "joy" with facial expressions of "surprise", "sadness", or "fright" were most favorable. On the other hand, when the user's emotion was "fright", the agent's words for "fright" with facial expressions of "disgust" or "sadness" were the most favorable.

These facial expressions were recognized as the emotion conveyed by the words and were more empathetic and somewhat meaningful emotions. For example, when the user's emotion was "joy", the agent's words of "joy" with facial expressions of "surprise" or "fright" might have been recognized as the agent being exaggeratedly surprised at the "joy" scenario. When the user's emotion was "joy", the agent's words of "joy" with facial expressions of "sadness" might have been recognized as the agent being highly pleased from the heart at the "joy" scenario. When the user's emotion was "fright", the agent's words of "fright" with facial expressions of "sadness" might have been recognized as the agent grieving deeply at the user's "fright" scenario. When the user's emotion was "fright", the agent's words of "fright" with facial expressions of "disgust" might have been recognized as the agent feeling deep hate at the user's "fright" scenario.

Through these observations, we concluded that there is a rule for facial expressions: in a certain scenario, synchronizing foreseen emotion of the user caused by the situation will make a favorable impression. For example, when the user has the emotion of "joy", he/she wants someone to be surprised or highly pleased. Then, showing surprised or highly pleased face expression make the user feels favorable impression. When the user has the emotion of "fright", he/she wants someone to grieve deeply or disgust. Then, showing grieved or disgust face expression make the user feels favorable impression.

Users want the agent to ooze synchronized their foreseen emotion by hearing the news instead of simply showing synchronized reaction according to emotion at present time.

3. A Spoken Agent System for Learning Customer Services

As an application system for affective human computer interaction, this first spoken agent system consists of a 3D spoken agent display system, a speech recognition system, a speech synthesis system, a dialogue control system and a facial expression recognition system. This system is for providing educational training in hospitality through dialogue with a spoken agent. This system focuses on the Japanese style of service-mindedness, which is typified by paying attention to individual customers [8][9].

The 3D spoken agent display system displays a 3D spoken agent, which can show the facial expressions of ""smile", "laugh", "anger", "sadness", "disgust", "fright", and "surprise". 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 [10]) 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 spoken agent, and the spoken agent teaches the user how to interact with customers. The spoken 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.

4. A Spoken Agent System for Mental Care using Expressive Facial Expressions and Positive Psychology

We developed a spoken agent system for mental care of hikikomori (a Japanese term to refer to the phenomenon of reclusive adolescents or young adults who withdraw from social life) persons improving previous system (Figure 1). An increasing number of people are in need of mental care, because associated social issues include problems of bullying in schools, mental disability and hikikomori or social withdrawal of young people, suicide of young people, people being pressured to quit work in middle age, and PTSD (posttraumatic stress disorder) of the victim. According to the result of the experiment of first spoken agent system, it was revealed we must improve interface of the system, especially for
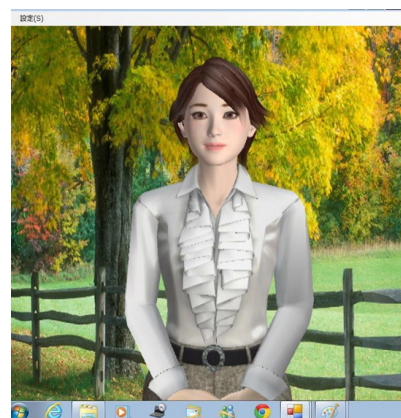

Figure 1. A spoken agent system for mental care

the attachment of Emotiv. We improved a spoken agent system. The system consists of a 3D spoken agent display system, a dialogue control system, Japanese word spacing system, a speech recognition system, a speech synthesis system, and a facial/gesture expression recognition system. The improved points are as follows. 1) KINECT sensor recognizes a user's facial expressions and gestures, 2) An agent displays gestures , 3) The facial expression/gestures can be controlled in intensity digitally, 4) An agent can be controlled its size, direction and position, 5) A background of the system is changeable , and 6) An agent blinks its eyes.

Interactive talks using gestures are enabled by recognizing a user's motion from the skeleton using a KINECT sensor. In addition, a simple facial expression is enabled by recognition of a user's face using KINECT sensor without an Emotiv. The system can be switched between a KINECT and/or an Emotiv. An agent displays gestures of "nodding", "nodding no", "waving her hand", "bowing her head" and "bowing her head deeply". An agent can be controlled its size (full body/ upper body/ face or parentage of full body) , direction and position. The background of the system can be changed using the dialogue editor. The agent blinks its eyes for making the agent more human. Because, according to the experiment, there are many comments that the agent should blink its eyes.

In addition, this system offers the function of answering positively when a user speaks negatively, for creating a positive mood. This system answers with positive words to a user's negative words using Negapo Dictionary [11], which is a dictionary for translate negative words to positive words. Such counseling is provided in a field of positive psychology. For example, such as follows is talked.

The agent: How was your day? The user: I went to my university.

The agent: What kind of thing happened?

The user: I get frustrated because only a friend might be good.

The agent: I felt sorry for you. (Facial expression is "Sad") I think you should be able to

make the feeling of frustrated with a motivation. (Positive words from Negapo Dic.)

5. Discussion

Even if a partner is not a human being, we are persuaded by an agent with superficial information such as facial expressions or words. Media Equation [1] and Persuasive Technology [2] researches reveal the same.

From the result of the first experiment in section 2, facial expressions and words are very important because people are easy to be persuaded when the agent was favorable impression. It is thought that an agent can make in-depth contact with the user by foreseeing the user's emotion and empathizing with the user.

Using the results, we have developed spoken agent systems for learning customer services and mental care using expressive facial expressions. These are working well as the system for an agent as an instructor or a guidance counselor using affective human interfaces.

According to the interviews, users felt the brain wave measuring equipment and the speech recognition and synthesis system were novelties, and the system's specialty was that its use was facilitated through dialogues and it was highly motivating. The brain wave measuring equipment and the speech recognition and synthesis system used in our system were highly motivating; however the 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 speech recognition system of greater accuracy. 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.

According to the interviews we must improve the interface of the system, especially for the attachment of the brain wave measuring equipment. Then we developed next version of the system for mental care of

hikikomori social withdrawal persons or using KINECT, which recognizes the facial expression/gestures of the user.

I think this system is helpful for mental care of hikikomori social withdrawal persons if it can be used easily in house. Because even hikikomori or social withdrawal persons are at home and stay in front of a computer most of the time. I think that supporting these people by a personal computer is a key to solution.

Study of a spoken dialogue agent system needs to be investigated further including emotions, technology and design domains. Our research group is going to advance our own studies of a spoken agent system using superficial information in the future.

6. Conclusion

This paper introduces an experiment and the spoken agent system developed according to the results. We evaluated the first version of the system and revised. Even if a partner is not a human being, we are perceived by an agent with superficial information such as facial expressions or words.

Acknowledgements

References

1. 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.
2. B.J.Fogg: *Persuasive Technology –Using Computers to Change What We Think and Do-*, Elsevier, 2003.
3. Weizenbaum, Joseph: "ELIZA—A Computer Program For the Study of Natural Language Communication Between Man And Machine", *Communications of the ACM* 9 (1): 36–45, (January 1966).
4. A. Rizzo, K. Sagae , E. Forbell, J. Kim, B. Lange, J. G. Buckwalter, J. Williams, T. D. Parsons, P. Kenny, David Traum, J. Difede, B. O. Rothbaum: SimCoach: an intelligent virtual human system for providing healthcare information and support, *The Interservice/Industry Training, Simulation & Education Conference* (I/ITSEC), 2011.
5. Christopher Peterson : *A Primer in Positive Psychology*, Oxford University Press;
6. Martin E. P. Seligman: *Using the New Positive Psychology to Realize Your Potential for Lasting Fulfillment*, Atria Books (2004)
7. 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.
8. Kaoru Sumi, Ryuji Ebata: A Character Agent System for Promoting Service-Minded Communication, Intelligent Virtual Agents, *Lecture Notes in Computer Science*, LNAI8108, pp.438, Springer (2013.8).
9. Kaoru Sumi, Ryuji Ebata: Human Agent Interaction for Learning Service-Minded Communication, *iHAI2013, 1st international conference on Human-Agent Interaction*, (2013.8).
10. AIML: http://www.alicebot.org/documentation/
11. Negapo *Dictionary*, Syufuno-Tomo-Sya (2012). In Japanese.