Artificial Subtle Expressions: Intuitive Notification Methodology of Artifacts

Takanori Komatsu
Shinshu University
3-15-1 Tokida, Ueda, Japan
komatsu@shinshu-u.ac.jp

Seiji Yamada
National Institute of Informatics
2-1-2 Hitotsubashi, Tokyo, Japan
seiji@nii.ac.jp

Kazuki Kobayashi
Shinshu University
4-17-1 Wakasato, Nagano, Japan
kby@cs.shinshu-u.ac.jp

Kotaro Funakoshi
Honda Research Institute Japan Co., Ltd
8-1 Honcho, Wako, Japan
funakoshi@jp.honda-ri.com

Mikio Nakano
Honda Research Institute Japan Co., Ltd
8-1 Honcho, Wako, Japan
nakano@jp.honda-ri.com

ABSTRACT

We describe artificial subtle expressions (ASEs) as intuitive notification methodology for artifacts' internal states for users. We prepared two types of audio ASEs; one was a flat artificial sound (flat ASE), and the other was a sound that decreased in pitch (decreasing ASE). These two ASEs were played after a robot made a suggestion to the users. Specifically, we expected that the decreasing ASE would inform users of the robot's lower level of confidence about the suggestions. We then conducted a simple experiment to observe whether the participants accepted or rejected the robot's suggestion in terms of the ASEs. The results showed that they accepted the robot's suggestion when the flat ASE was used, whereas they rejected it when the decreasing ASE was used. Therefore, we found that the ASEs succeeded in conveying the robot's internal state to the users accurately and intuitively.

Author Keywords

Artificial subtle expressions (ASEs), Complementary, Intuitive, Simple, Accurate.

ACM Classification Keywords

H5.2 User Interfaces: Evaluation/methodology; J.4 Social and behavioral sciences: Psychology.

General Terms

Experimentation, Human Factors

INTRODUCTION

Although human communications are explicitly achieved through verbal utterances, paralinguistic information (e.g., pitch and power of utterances) and nonverbal information (e.g., facial expressions, gaze direction, and gestures) also play important roles [1]. This is because it is said that one's internal state is deeply reflected in her/his expressed paralinguistic and nonverbal information; that is, other people can intuitively and easily understand a person's internal state from such information when it is expressed. Recently, some researchers have reported that very small changes in the expression of such information (called subtle expressions [3]) significantly influence human communications, especially in the conveyance of one's internal states to others [2]. It is therefore believed that such subtle expressions can be utilized to help humans easily understand an artifact's internal state because humans can intuitively understand such subtle expressions. For example, Sugiyama et al. [4] developed a humanoid robot that can express appropriate gestures based on its situation recognition, and Kipp & Gebhard [5] developed a human-like avatar agent that can control its gaze direction according to the user's gaze direction. However, since these researchers tried to implement subtle expressions on artifacts (e.g., humanoid robots or dexterous avatar agents), their implementations were considerably expensive.

In contrast to the above approaches, Komatsu & Yamada [6] reported that simple beeping sounds from a robot with decreasing/increasing frequency enabled humans to interpret the robot's negative/positive states. Funakoshi et al [7] also reported that the robot's blinking LED could convey to users a robot's internal state (processing or busy) for the sake of reducing the occurrence of speech collisions during their verbal conversations. It then seemed that such simple expressions (beeping sounds or blinking LEDs) from artifacts could play a similar role to the subtle expressions of humans, so we named these expressions from artifacts "Artificial Subtle Expressions (ASEs)," referring to artifacts' simple and low-cost expressions that enable humans to estimate the artifacts' internal states accurately and intuitively. We stipulate that the ASEs should meet two requirements for designing and two requirements for functioning simultaneously. Specifically, the two requirements for designing are as follows:

CHI 2010, April 10–15, 2010, Atlanta, Georgia, USA.

- **Simple**: ASEs should be implemented on a single modality. It is then expected that the implementation cost would also be lower.
- **Complementary**: ASEs should only have a complementary role in communication and should not interfere with communication's main protocol. This means that the ASEs themselves do not have any meaning without a communication context.

The two requirements for functioning are as follows:

- **Intuitive**: ASEs should be understandable by humans who do not know about the ASEs beforehand.
- **Accurate**: ASEs should convey the designer's intended meanings accurately. Specifically, ASEs should convey the internal states of the artifact just as subtle expressions do in nonverbal information by humans.

In this study, we focused on audio ASEs. Related studies with audio ASEs include those that proposed simple and effective information to convey specific meaning to users, e.g., "earcon [8]" or "auditory icon [9]" These earcon and auditory icons play an effective role in informing users of specific meanings as communication's main protocol, while ASEs play a complementary role for the main protocol. This point is the significant difference between ASEs and earcon or auditory icons.

In this paper, we investigated whether the ASEs could convey the artifacts' internal state to the users accurately and intuitively; specifically, we created audio ASEs that were intended to meet the two requirements for designing them, and we investigated whether these ASEs met the two requirements for functioning by conducting a simple psychological experiment.

EXPERIMENT

Setting

1. Encountering three hills

2. Selecting the 2nd hill (but not knowing whether this selection was right or not)

3. Walking to the next three hills

Figure 1. Treasure hunting video game.

We used a "treasure hunting" video game as an experimental environment to observe the participants' behavior (Figure 1). In this game, a game image scrolls forward on a straight road, with small hills appearing along the way. A coin is inside one of three hills, while the other two hills have nothing. The game ends after the player encounters 20 sets of hills, and the approximate duration of this video game is about three minutes. The purpose is to get as many coins as possible. In this experiment, the participant was awarded 1 point for each coin that s/he found. The participants in this experiment were informed that 1 point was equivalent to 50 Japanese yen (about 50 US cents) and that after the experiment, they could use their points to purchase some stationery supplies (e.g., file holders or USB flash memory) of equivalent value.

The position of the coin in the three hills was randomly assigned. In each trial, an artifact placed next to the participants told them in which position it expected the coin to be placed. The artifact placed next to the participants was the MindStorms robot (LEGO Corporation). The robot told the expected position of the coin using their speech sounds. The participants could freely accept or reject the robots' suggestions. In each trial, even though the participants selected one hill among three, they did not know whether the selected hill had the coin or not (actually, the selected hill just showed a question mark and closed treasure box, as depicted in the center of Figure 1). The participants were informed of their total game points only after the experiment.

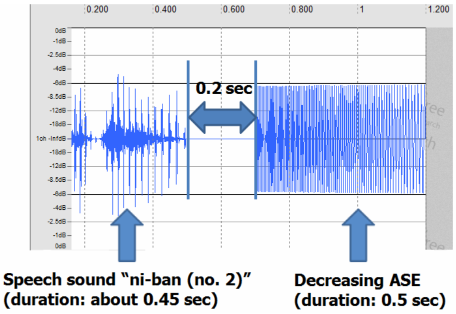

Figure 2. Speech sound "ni-ban (no.2)" and ASE.

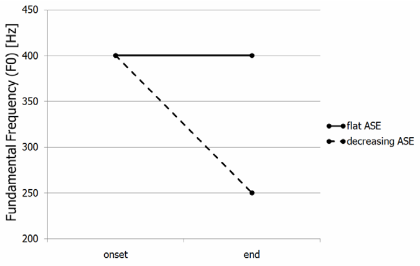

Figure 3. Flat and decreasing ASEs (duration: 0.5 second).

Utilized ASE

We implemented the audio ASEs in the robot's speech sounds. In this experiment, the robot expressed Japanese artificial speech sounds to tell the expected position of the coin; that is, "ichi-ban (no. 1)," "ni-ban (no. 2)," and "san-ban (no. 3)." These artificial speech sounds were created by the text-to-speech (TTS) function of "Document Talker (Create System Development Company)." Just 0.2 second after these speech sounds, one of the two simple artificial sounds was played as the ASE (Figure 2). These two ASEs were triangle wave sounds 0.5 second in duration, but their pitch contours were different (Figure 3); that is, one was a flat sound (onset F0: 400 Hz and end F0: 400 Hz, called "flat ASE"), and the other was a decreasing one (onset F0:

400 Hz and end F0: 250 Hz, called "decreasing ASE"). These ASE sounds were created by "Cool Edit 2000 (Adobe Corporation)." Komatsu & Yamada [6] reported that the decreasing artificial sounds expressed from the robot were interpreted as negative feelings by humans; therefore, we intended that the decreasing ASE would inform users of the robot's lower confidence in the suggestions as the robot's internal state.

Here, the main protocol of the robot was to tell the expected position of the coin, while the ASE protocol was to indicate the robot's confidence level in a complementary manner. The two ASE sounds were created quite easily by simply editing the consumer software. Thus, the ASEs met the two design requirements, that is, simple and complementary. Therefore, to confirm whether the ASEs were able to convey the robot's internal states to the users accurately and intuitively, we needed to investigate whether the utilized ASE met the two requirements for functioning, that is, being intuitive and accurate.

Procedure

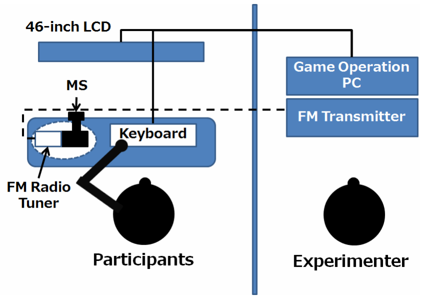

Figure 4. Experimental setting.

Nineteen Japanese university students (10 men and 9 women; 22 – 25 years old) participated. The treasure hunting video game was projected on a 46-inch LCD in front of the participants, and the robot was placed in front of and to the right of the participants, with the distance between them being approximately 50 cm. The sound pressure of the robot's speech sounds at the participants' head level was set at about 50 dB (FAST, A). The robot's speech sounds with the ASEs were remotely controlled by the experimenter in the next room using the Wizard of Oz (WOZ) method. Before the experiment started, the experimenter told the participant the setting and purpose of the game. However, the experimenter never mentioned or explained the ASEs. Therefore, the participants had no opportunity to acquire prior knowledge about the ASEs. Among the 20 trials, the robots expressed the flat ASE 10 times and the decreasing ASE 10 times. The order of expression for these two types of ASEs was counterbalanced across participants. Actually, the robot told the exact position of the coin in all 20 trials, but the participants did not know whether or not the robot was telling the right position because the participants were not able to find out whether the selected hill had the coin or not. If the participant actually knew whether or not the selected hill had the coin just after their selections, they would have associated the ASE with the robot's performance, e.g., whether or not the robot pointed to the right position. Thus, this experimental setting, where the participants were not notified of whether the selected hill was correct or not, was intended to reduce such associations and to clarify the effect of the ASEs on the participants' behavior.

The purpose of this experiment was to observe the participants' behavior as to whether they accepted or rejected the robot's suggestions in terms of the types of ASEs used. We assumed that *the participants would accept the robot's suggestion when the flat ASE was added to the speech sounds while they would reject the suggestion when the decreasing ASE was used.* If we could observe these phenomena, we could recognize that the utilized ASE had succeeded in conveying the robot's internal states to the participants accurately and intuitively; that is, the ASE had successfully met all four requirements. In addition, after the experiment, we conducted interviews to determine whether or not the participants had noticed the ASEs and, if so, how they had interpreted them.

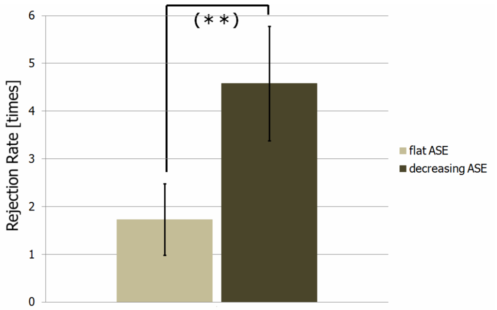

Figure 5. Rejection rate for all 19 participants.

RESULTS

To investigate the effect of the ASEs on the participants' behavior, we calculated the rejection rate, indicating how many of the robot's suggestions the participants rejected for 10 flat ASEs and 10 decreasing ASEs. For all 19 participants, the average rejection rate of the 10 flat ASEs was 1.73 (SD=1.51), while the rejection rate of the 10 decreasing ASEs was 4.58 (SD=2.43, see Figure 5). These rejection rates for the 10 flat ASEs and 10 decreasing ASEs were analyzed using a one-way analysis of variance (ANOVA) (within-subjects design; independent variable: type of ASE, flat or decreasing, dependent variable: rejection rate). The result of the ANOVA showed a significant difference between the two groups ($F(1,18)=13.38$, $p<.01$, (**)); that is, the robot's suggestions with the decreasing ASE showed a significantly higher rejection rate compared to the one with the flat ASE. Therefore, the ASEs affected the participants' behaviors significantly, and we found evidence supporting our assumption mentioned previously. The most interesting point was that the ASEs affected the behavior of the participants without their being informed of the meaning or even existence of the ASEs.

In the interview sessions, 5 out of the 19 participants said that they immediately realized the meanings of the ASEs after the robot's speech sounds and that they utilized these ASEs when it came to accepting or rejecting the robot's suggestions, e.g., "I felt that the decreasing artificial sounds meant that the robot had less confidence in its answer." However, the remaining 14 participants said that they did not notice the existence of the ASEs. Here, if there were significant differences between flat and decreasing ASEs in their rejection rate, the ASEs were interpreted by these 14 participants unconsciously. In this case, we strongly argue that the ASEs were able to convey the robot's internal state to the participants accurately and intuitively. For these 14 participants, the average rejection rate of 10 flat ASEs was 2.28 (SD=1.73), while the rejection rate of the 10 decreasing ASEs was 3.43 (SD=1.59, see Figure 6). These rejection rates were analyzed using a one-way ANOVA (within-subjects design; independent variable: ASE type, flat or decreasing, dependent variable: rejection rate). The result of the ANOVA showed a significant difference between them ($F(1,13)=4.98$, $p<.05$, (*)); that is, the robot's suggestions with the decreasing ASE had a significantly higher rejection rate compared to the one with a flat ASE, even though these participants did not notice the existence of the ASEs. To sum up, the results of this experiment clearly show that the utilized ASEs succeeded in conveying the robot's internal states to the participants accurately and intuitively; that is, the ASEs succeeded in meeting all four requirements.

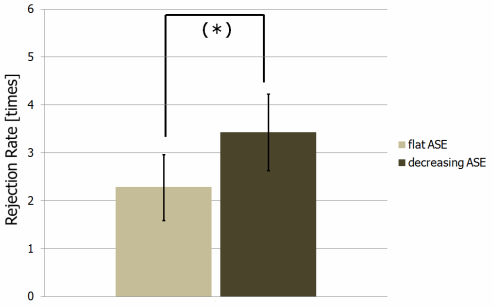

Figure 6. Rejection rate for 14 participants who did not notice ASEs.

DISCUSSION AND CONCLUSIONS

In this paper, we investigated whether the ASEs could convey the artifacts' internal state to the users accurately and intuitively; specifically, we created audio ASEs that were intended to meet the two requirements for designing them, and we investigated whether these ASEs met the two requirements for functioning by conducting a simple psychological experiment. As a result of this experiment, the robot's suggestions with the decreasing ASEs showed a significantly higher rejection rate compared to the ones with flat ASEs. Moreover, these ASEs were interpreted by the participants even though they were not instructed of the meaning or even the existence of the ASEs. Therefore, our experiment clearly showed that the utilized ASEs succeeded in conveying the robot's internal states to the participants accurately and intuitively; that is, the ASEs succeeded in meeting all four requirements. Thus, we confirmed that simple and low-cost expression ASEs could be utilized as intuitive notification methodology for artifacts to convey their internal states to users like paralinguistic or nonverbal information.

However, in this paper we have reported just the initial evidence of ASEs. Therefore, as a follow-up study, we are planning to implement the ASEs in various kinds of spoken dialogue systems such as ATMs and automatic telephone reservation systems. Specifically, we are now focusing on car navigation systems' speech sounds; the reason for this is that current car navigation systems still sometimes give poor driving routes to users. However, if this navigation system's confidence level regarding the route instruction is not very high, the instructions of speech sounds with ASEs could implicitly convey a lower confidence level. If the ASEs are still effective in such situations, they could be utilized in various situations in which artifacts have to convey their internal states to users.

REFERENCES

1. Kendon, A. Do gestures communicate? A Review. *Research in Language and Social Interaction 27*, 3 (1994), 175-200.
2. Cohen, P. R., Morgen, J., and Pollack, M. E. *Intentions in Communication*, The MIT Press, MA, USA, 1990.
3. Liu, K. and Picard, W. R. Subtle expressivity in a robotic computer. In *Proc. CHI2003 Workshop on Subtle Expressivity for Characters and Robots* (2003), 1-5.
4. Sugiyama, O., Kanda, T., Imai, M., Ishiguro, H., Hagita, N. and Anzai, Y. Humanlike conversation with gestures and verbal cues based on a three-layer attention-drawing model. *Connection Science 18*, 4 (2006), 379-402.
5. Kipp, M. and Gebhard, P. IGaze: Studying reactive gaze behavior in semi-immersive human-avatar interactions, In *Proc. IVA2008 (2008)*, 191-199.
6. Komatsu, T. and Yamada, S. How do robotic agents' appearances affect people's interpretation of the agents' attitudes? *Ext. Abstracts CHI2007*, ACM Press (2007), 2519-2524.
7. Funakoshi, K., Kobayashi, K., Nakano, M., Yamada, S., Kitamura, Y., and Tsujino H. Smoothing human-robot speech interactions by using blinking-light as subtle expression. In *Proc. ICMI 2008*, ACM Press (2008), 293-296.
8. Blattner, M. M., Sumikawa, D. A. and Greenberg, R. M. Earcons and Icons: Their Structure and Common Design Principles. *SIGCHI Bull. 21*, 1 (1989), 123-124.
9. Gaver, W. W. The SonicFinder: An Interface That Uses Auditory Icons. *Human-Computer Interaction 4*, 1 (1989), 67-94.