Proposing Artificial Subtle Expressions as an Intuitive Notification Methodology for Artificial Agents' Internal States

Takanori Komatsu (tkomat@shinshu-u.ac.jp)
International Young Researcher Empowerment Center, Shinshu University,
3-15-1 Tokida, Ueda 386-8567, Japan

Seiji Yamada (seiji@nii.ac.jp)
National Institute of Informatics/ SOKEDAI,
2-1-2 Hitotsubashi, Tokyo 101-8430, Japan

Kazuki Kobayashi (kby@cs.shinshu-u.ac.jp)
Graduate School of Science and Technology, Shinshu University
4-17-1 Wakasato, Nagano 380-8553, Japan

Kotaro Funakoshi (funakoshi@jp.honda-ri.com) and Mikio Nakano (nakano@jp.honda-ri.com)
Honda Research Institute Japan Co., Ltd,
8-1 Honcho, Wako 351-0188, Japan

Abstract

We describe artificial subtle expressions (ASEs) as an 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 in its suggestion. We then conducted a simple experiment to observe whether the participants accepted or rejected the robot's suggestion based on 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. We thereby concluded that the ASEs succeeded in conveying the robot's internal state to users accurately and intuitively.

Keywords: Artificial subtle expressions (ASEs); Complementary; Intuitive; Simple; Accurate.

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 (Kendon, 1994). This is because one's internal state is deeply reflected in one's paralinguistic and nonverbal information. In other words, other people can intuitively and easily understand a person's internal state from such information when it is expressed (Cohen et al., 1990). Recently, some researchers have reported that very small changes in the expression of such information, called subtle expressions (Liu & Picard, 2003), significantly influence human communications, especially in the conveyance of one's internal state to others. For example, Ward (2003) reported that the subtle flections of the pitch information in speech sounds reflect one's emotional states even when contradicted by the literal meanings of the speech sounds, and Cowell & Ayesh (2004) offered a similar argument in terms of facial expressions.

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. (2006) developed a humanoid robot that can express appropriate gestures based on a recognition of its situation, and Kipp & Gebhard (2008) 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), it resulted in considerably high implementation costs.

In contrast to the above approaches, Yamada & Komatsu (2006) and Komatsu & Yamada (2007) 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. (2008) 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 in artifacts "Artificial Subtle Expressions (ASEs)," referring to artifacts' simple and low-cost expressions that enable humans to estimate the artifacts' internal state accurately and intuitively. We stipulate that the ASEs should

simultaneously meet two design and two functional requirements.

Specifically, the two design requirements are as follows:

- **Simple**: ASEs should be implemented on a single modality. This is expected to lower the implementation cost.
- **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 functional requirements are as follows:

- **Intuitive**: ASEs should be understandable by humans who have no prior knowledge of the ASEs.
- **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 (Blattner, 1989)" or "auditory icon (Gaver, 1989; Gaver, 1997)" These earcons 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 is the significant difference between ASEs and earcons 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 design requirements and investigated whether they also met the two functional requirements by conducting a simple psychological experiment.

Experiment

Setting

We used a "treasure hunting" video game as an experimental environment to observe 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.

Figure 1: Treasure hunting video game.

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, see Figure 2). The robot told the participant the expected position of the coin using its speech sounds. The participants could freely accept or reject the robots' suggestions. In each trial, even though the participants selected one hill from 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 a closed treasure box, as depicted in the center of Figure 1). The participants were informed of their total game points only after the experiment.

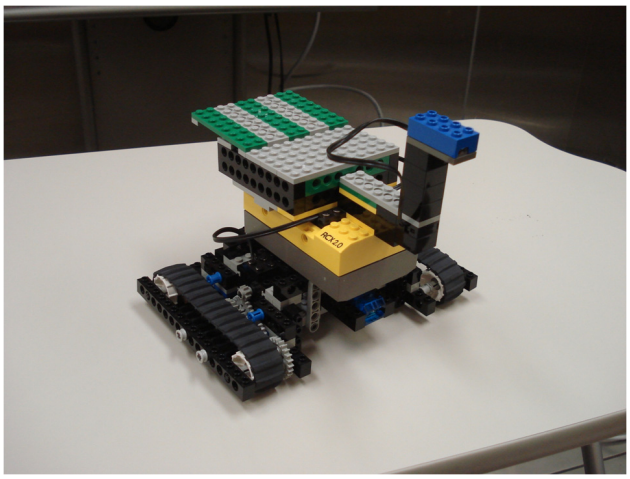

Figure 2: MindStorms Robot.

Utilized ASEs

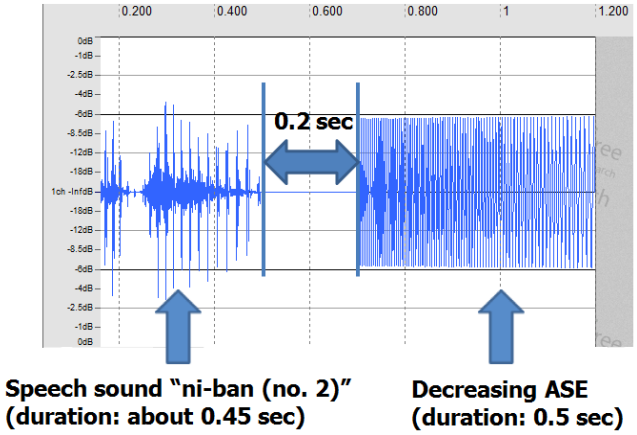

Figure 3: Speech sound "ni-ban (no.2)" and decreasing ASE.

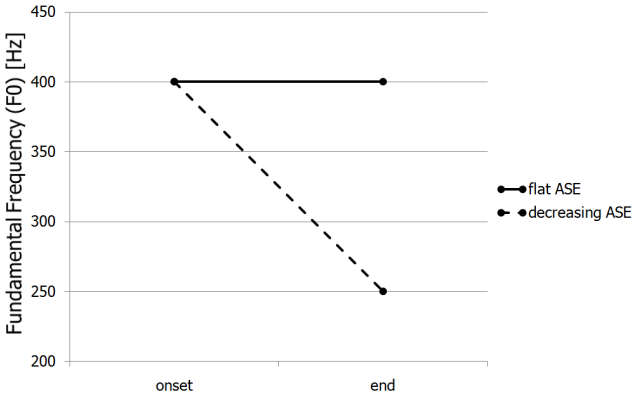

Figure 4: Flat and decreasing ASEs (duration: 0.5 second).

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 seconds after these speech sounds, one of the two simple artificial sounds was played as the ASE (Figure 3). These two ASEs were triangle wave sounds 0.5 seconds in duration, but their pitch contours were different (Figure 4); 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 (2007) 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

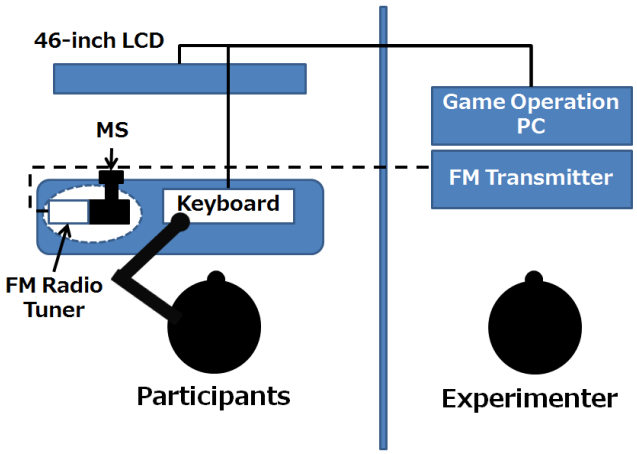

Figure 5: Experimental setting.

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 (see Figures 5 and 6). 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.

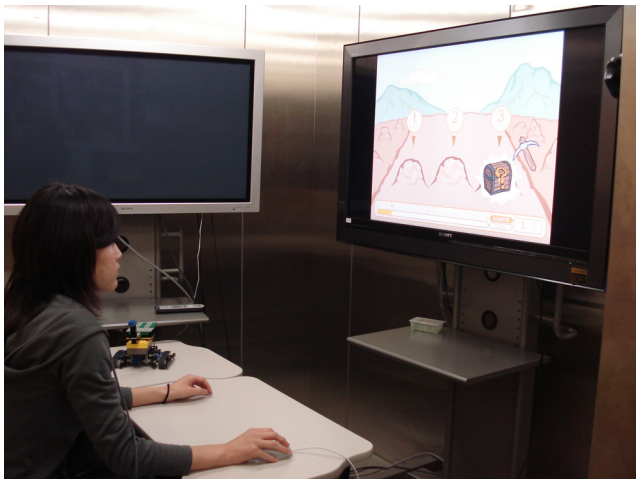

Figure 6: Experimental Scene

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 state 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.

Results

To investigate the effect of the ASEs on 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 7). 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 those with the flat ASE. Therefore, the ASEs significantly affected the participants' behavior, and we found evidence supporting our previously mentioned assumption .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.

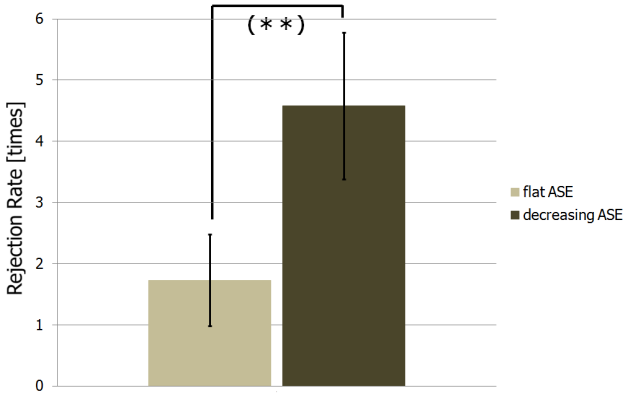

Figure 7: Rejection rate for all 19 participants.

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 8). 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 those 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.

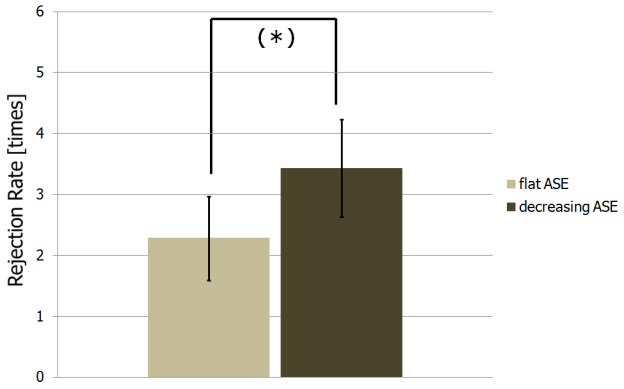

Figure 8: Rejection rate for 14 participants who did not notice ASEs.

Discussion

Future Applications

As a result of the experiment, we could confirm that the robot's suggestions with the decreasing ASEs showed a significantly higher rejection rate compared with those with flat ASEs. Moreover, these ASEs were interpreted by the participants even though they were not informed 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.

Currently, 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.

In our experiment, we only focused on the internal state of the artifact in order to convey to users its level of confidence in its own expressed information. However, we are planning to investigate which kinds of internal states could be conveyed to the users by means of ASEs. For example, it is expected that the artifacts should also convey other kinds of internal states, such as feelings or conditions, and the confidence level in interpreting the user's expressions. This consecutive study would also contribute to expanding the applicability of ASEs to various interactive situations.

Advantage of utilizing ASEs

It is said that the most significant advantage in utilizing ASEs is the lower implementation cost compared to utilizing human-like expressions. Therefore, it is expected that many applications in human-computer interaction or human-robot interaction will be able to include the ASEs quite easily. In addition to the lower cost, we believe that the advantage of utilizing ASEs includes the possibility of solving several problems such as those reported in the above research areas.

So far, it has been strongly believed that most robots or on-screen agents required to interact with users should have a human-like appearance and produce human-like expressions. However, we feel that these research directions have had two difficulties; one is the implementation cost mentioned above, and the other is that users have unexpected attitudes or impressions toward human-like artifacts; i.e., artifacts having a human-like appearance have a higher possibility of diving them into the "uncanny valley" (Mori, 1970). Moreover, users are likely to overestimate the artifacts' ability when it has a human-like appearance or expressions, so they would be disappointed if these artifacts were to demonstrate unpredictable or poor behavior (Komatsu & Yamada, 2010).

Therefore, our approach that the artifact should not produce human-like expressions but artifact-like ones to convey its internal state to the users has succeeded in proposing a novel research approach in the research area of human-computer interaction or human-robot interaction in order to resolve the above issues. Now we are planning to conduct a consecutive study to compare ASEs to human-like expressions in terms of users' cognitive load or cost-benefit relationships. Comprehending the advantages and disadvantages of these two expressions (ASEs and human-like expressions) would constitute a design methodology for artifacts' expressions in order to achieve smooth interaction between users and artifacts.

Conclusions

In this paper, we investigated whether the ASEs could convey artifacts' internal states accurately and intuitively to users; specifically, we created audio ASEs intended to meet the two requirements for design, and we investigated whether these ASEs met the two requirements for function by conducting a simple psychological experiment. As a result of this experiment, the robot's suggestions accompanied by decreasing ASEs showed a significantly higher rejection rate compared with those accompanied by flat ASEs. Moreover, these ASEs were accurately interpreted by participants even though they were not informed 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 an intuitive notification methodology for artifacts to convey their internal states to users through paralinguistic or nonverbal information.

Acknowledgments

This study was partially funded by the Special Coordination Funds for Promoting Science and Technology granted by the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan.

References

Blattner, M. M., Sumikawa, D. A. & Greenberg, R. M. (1989). Earcons and Icons: Their Structure and Common Design Principles. *SIGCHI Bulletin. 21*, *1*, 123-124.

Cohen, P. R., Morgen, J., & Pollack, M. E. (1990). *Intentions in Communication*, The MIT Press, MA, USA.

Cowell, J. & Ayesh, A. (2004). Extracting subtle expressions for emotional analysis, In *Proceedings of 2004 IEEE International Conference on Systems, Man, Cybernetics (IEEE SMC 2004)*, pp. (1) 677-681.

Funakoshi, K., Kobayashi, K., Nakano, M., Yamada, S., Kitamura, Y., & Tsujino H. (2008). Smoothing human-robot speech interactions by using a blinking-light as subtle expression. In *Proceedings of the 10th International Conference on Multimodal Interface (ICMI 2008)*, pp. 293-296.

Gaver, W. W. (1989). The SonicFinder: An Interface That Uses Auditory Icons. *Human-Computer Interaction 4*, *1*, 67-94.

Gaver, W. W. (1997). *Auditory Interfaces. Handbook of Human-Computer Interaction*, Elsevier Science.

Kendon, A. (1994). Do gestures communicate? A Review. *Research in Language and Social Interaction 27*, *3*, 175-200.

Kipp, M. & Gebhard, P. (2008). IGaze: Studying reactive gaze behavior in semi-immersive human-avatar interactions, In *Proceedings of the 8th International Conference on Intelligent Virtual Agent (IVA2008)*, pp. 191-199.

Komatsu, T. & Yamada, S. (2007). How do robotic agents' appearances affect people's interpretation of the agents' attitudes? In *Extended Abstracts of CHI2007*, pp. 2519-2524.

Komatsu, T. & Yamada, S. (2010). Effects of Adaptation Gap on Users' Differences in Impressions of Artificial Agents, In *Proceedings of the 14th. World Multiconference on Systemics, Cybernetics and Informatics (WMSCI 2010)*, to appear.

Liu, K. & Picard, W. R. (2003). Subtle expressivity in a robotic computer. In *Proceedings of CHI2003 Workshop on Subtle Expressivity for Characters and Robots*, pp. 1-5.

Mori, M. (1970). Bukimi no tani (The uncanny valley, K. F. MacDorman & T. Minato, Trans.). *Energy 7, 4*, 33–35. (Originally in Japanese).

Sugiyama, O., Kanda, T., Imai, M., Ishiguro, H., Hagita, N. & Anzai, Y. (2006). Humanlike conversation with gestures and verbal cues based on a three-layer attention-drawing model. *Connection Science 18*, *4*, 379-402.

Yamada, S. & Komatsu, T. (2006). Designing Simple and Effective Expression of Robot's Primitive Minds to a Human, In *Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'06)*, pp. 2614-2619.

Ward, N. (2003). On the Expressive Competencies Needed for Responsive Systems, In *Proceedings of CHI2003 Workshop on Subtle Expressivity for Characters and Robots*, pp. 33-34.