

# Interpretations of Artificial Subtle Expressions (ASEs) in Terms of Different Types of Artifact: A Comparison of an on-screen Artifact with A Robot

Takanori Komatsu<sup>1</sup>, Seiji Yamada<sup>2</sup>, Kazuki Kobayashi<sup>3</sup>,  
Kotaro Funakoshi<sup>4</sup>, and Mikio Nakano<sup>4</sup>

<sup>1</sup> Shinshu University, International Young Researcher Empowerment Center,  
3-15-1 Tokida, Ueda 3868567, Japan

tkomat@shinshu-u.ac.jp

<sup>2</sup> National Institute of Informatics/ SOKENDAI/ Tokyo Institute of Technology,  
2-1-2 Hitotsubashi, Tokyo 1018430, Japan

seiji@nii.ac.jp

<sup>3</sup> Shinshu University, Graduate School of Engineering,  
4-17 Wakasato, Nagano 3808553, Japan

kby@shinshu-u.ac.jp

<sup>4</sup> HONDA Research Institute Japan,  
8-1 Honcho, Wako 3510188, Japan

{funakoshi,nakano}@jp.honda-ri.com

**Abstract.** We have already confirmed that the artificial subtle expressions (ASEs) from a robot can accurately and intuitively convey its internal states to participants [10]. In this paper, we then experimentally investigated whether the ASEs from an on-screen artifact could also convey the artifact's internal states to participants in order to confirm whether the ASEs can be consistently interpreted regardless of the types of artifacts. The results clearly showed that the ASEs expressed from an on-screen artifact succeeded in accurately and intuitively conveying the artifact's internal states to the participants. Therefore, we confirmed that the ASEs' interpretations were consistent regardless of the types of artifacts.

**Keywords:** Artificial subtle expressions (ASEs), interpretation, robot, on-screen artifact.

## 1 Introduction

Many studies concerning human communications have reported that small changes in the expressions of paralinguistic information (e.g., pitch and power of utterances) and nonverbal information (e.g., facial expressions, gaze directions, and gestures) can aid in facilitating smooth human communications especially in conveyance of one's internal states to others [1,2], and such simple information is called subtle expressions [3]. For example, Ward [4] reported that the subtle flexions 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 and Ayesh [5] offered a similar argument in terms of facial expressions. Some researchers then tried to implement such humans' subtle expressions in artifacts. For example, Kipp and Gebhard [6] developed a dexterous avatar agent that can slightly change its facial expression according to the user's gaze direction, and Sugiyama et al. [7] created a humanoid robot that can slightly change its behaviors based on its situational recognition. However, it can be easily imagined that these implementations costs were considerably expensive.

On the other hand, we have already found that the simple expressions from artifacts like beeping sounds or blinking LED's play similar roles to such human's subtle expressions [8,9]. Based on the results of these studies, we proposed "Artificial Subtle Expressions (ASEs)" as an intuitive notification methodology used to describe the internal states of artifacts for users [10]. In particular, we stipulated that ASEs are simple and low-cost expressions for artifacts that enable humans to accurately and intuitively estimate the internal states of artifacts, and then we were able to experimentally recognize that such ASEs had succeeded in accurately and intuitively conveying the internal states of a robot (i.e., confidence level of the suggestions) to the participants [10, 11]. We are now planning to implement these ASEs in various types of artifacts requiring communication with users; not only for robots but also for artifacts appearing on the display.

However, various studies have reported that different types of artifacts (e.g., robots vs. on-screen agents) evoke different attitudes or different impressions from users toward these artifacts. For example, Wainer et al. [12] reported that a robot was seen as most helpful, watchful, and enjoyable compared to an on-screen agent, while Shinozawa et al. [13] stated that the appropriate types of artifacts depended on their interactive situations. Thus, the issue of whether the ASEs can be utilized regardless of the types of artifacts still need to be investigated because we only confirmed that the ASEs expressed from a robot (MindStorms robot, LEGO Corporation) were effective in our former study [10].

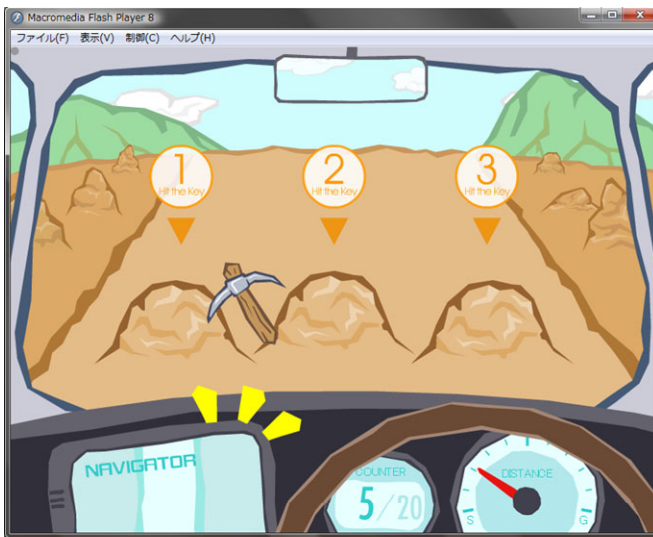
The purpose of this study is then to investigate whether the ASEs expressed from an artifact appearing on a display (we call this artifact "on-screen artifact") could accurately and intuitively convey its internal states to participants, and to compare the results of this experiment with the ones of our former study. If we could find that the ASEs expressed from an on-screen artifact also succeeded in accurately and intuitively conveying its internal states to the participants like in our former study, we could conclude that the ASEs' interpretation is consistent regardless of the types of artifacts and that the ASEs can be utilized in various types of artifacts.

## 2 Experiment

### 2.1 Setting

We used a "driving treasure hunting" video game as an experimental environment to observe the participants' behavior (Figure 1). In this game, the game image scrolls

forward on a straight road, which imitates the participant driving a car using a car navigation system (the left bottom of Figure 1), with small hills appearing along the way. A coin is inside one of the three hills, while the other two hills have nothing. The game ends after the participant 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 they 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., a ballpoint pen or mechanical pencil) of equivalent value.



**Fig. 1.** Screenshot of driving treasure hunting video game

The position of the coin in the three hills was randomly assigned. In each trial, a car navigation system next to the driver's seat on the screen told them in which position it expected the coin to be placed. The navigation informed them of the expected position of the coin using speech sounds. The participants could freely accept or reject the navigation's suggestions. In each trial, even though the participants selected one hill among the 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 2). The participants were informed of their total game points only after the experiment.



Fig. 2. Procedure of driving treasure hunting video game

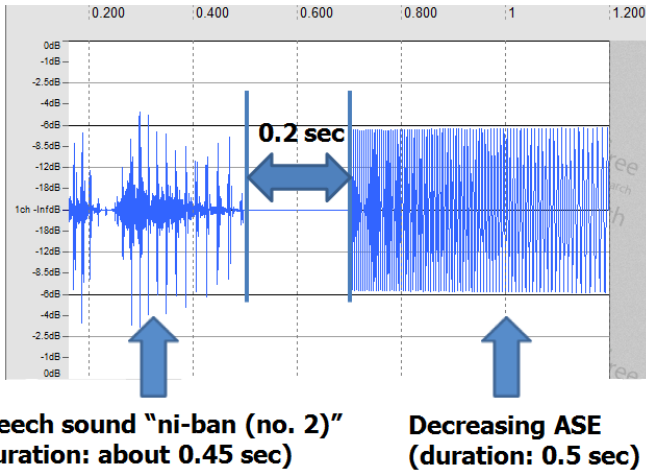


Fig. 3. Speech sound "ni-ban (no.2)" and ASE

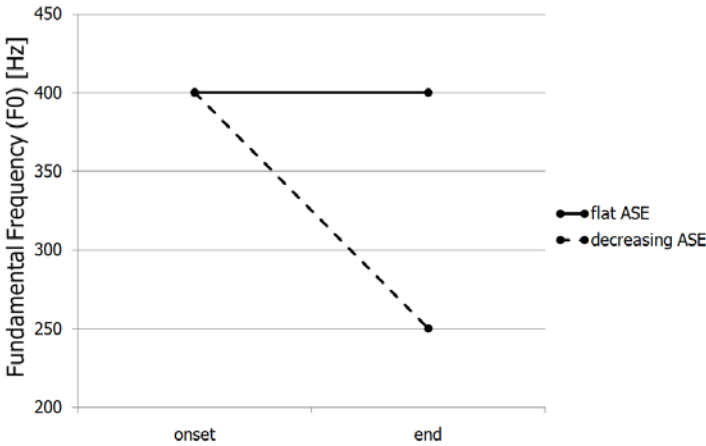


Fig. 4. Flat and decreasing ASEs (duration: 0.5 second)

## 2.2 Utilized ASE

We utilized the audio ASEs from the navigation system’s speech sounds. In this experiment, the navigation system expressed artificial Japanese speech sounds to express 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 the “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 triangular 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 using “Cool Edit 2000 (Adobe Corporation).”

Actually, these speech sounds and the ASEs were the same as the ones that were used in our former study [10]. In that study, we already confirmed that the speech sounds with decreasing ASEs informed users of the robot’s lower confidence in the suggestions given as the robot’s internal states.

### 2.3 Procedure

Twenty Japanese university students (14 men and 6 women; 21 – 24 years old) participated. The driving treasure hunting video game was projected on a 46-inch LCD in front of the participants at a distance between them of approximately 100 cm (Figure 5). The navigation’s speech sounds were played on the speaker equipped with this LCD, and the sound pressure of these speech sounds placed at the participants’ head level was set to about 50 dB (FAST, A).



**Fig. 5.** Experimental Scene

Before the experiment started, the experimenter told the participants the settings 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 navigation system used the flat ASE 10 times and the decreasing ASE 10 times. The order of expression for these two types of ASEs was counterbalanced among the participants.

For this experiment, the experimental stimuli and its procedure were completely the same as that in our former study [10], while the type of artifact for expressing the speech sounds with the ASEs was the only thing that was different. The purpose of this experiment was to observe the participants’ behavior whether they accepted or rejected the navigation’s suggestions in terms of the types of ASEs used; because, we

already confirmed that the speech sounds with decreasing ASEs informed users of the robot's lower confidence in the suggestions given as the robot's internal states in our former study. Therefore, if we could observe the phenomenon where the participants would accept the navigation's suggestion when the flat ASE was added to the speech sounds while they would reject the suggestion when the decreasing ASE was used, we would be able to recognize that the utilized ASEs had succeeded in accurately and intuitively conveying the navigation's internal states to the participants.

### 3 Results

#### 3.1 ASEs From an on-screen Artifact

To investigate the effect of the ASEs from the navigation as on-screen artifact on the participants' behavior, we calculated the rejection rate, indicating how many of the navigation's suggestions the participants rejected for the 10 flat ASEs and the 10 decreasing ASEs. For all 20 participants, the average rejection rate of the 10 flat ASEs was 1.75 (SD=1.61), while the rejection rate of the 10 decreasing ASEs was 4.35 (SD=2.76, Figure 6). These rejection rates for the 10 flat ASEs and 10 decreasing ASEs were then 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 stimuli ( $F(1,19)=8.16, p<.05, (*)$ ); that is, the on-screen artifact's suggestions when utilizing the decreasing ASE showed a significantly higher rejection rate compared to the ones with the flat ASE. Therefore, we could confirm that the ASEs expressed from an on-screen artifact also succeeded in accurately and intuitively conveying the artifact's internal states to the participants.

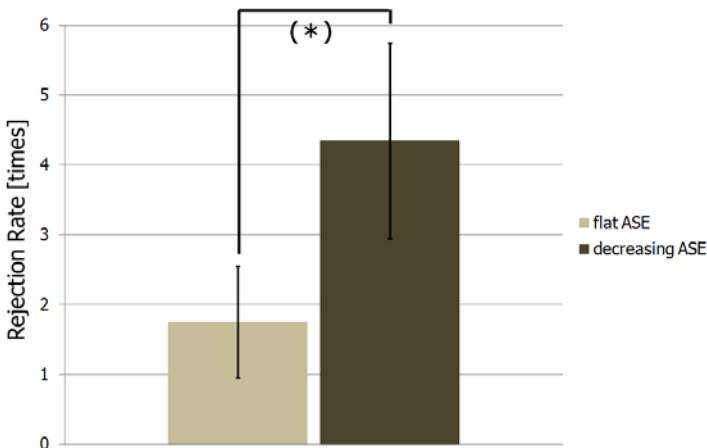


Fig. 6. Rejection rate for ASEs from on-screen artifact for all 20 participants

### 3.2 Comparison of an on-screen Artifact with a Robot

To investigate how the interpretations of the ASEs from an on-screen artifact were different with the ones from a robot, we compared the results from this experiment with the ones acquired in our former study [10]. In the former study, the average rejection rate of the 10 flat ASEs from the robot was 1.73 (SD=1.51), while the rejection rate of the 10 decreasing ASEs was 4.58 (SD=2.43). Since the participants in this paper did not participate in the former study and the utilized speech sounds and the ASEs were the completely same with ones utilized in the former study, these rejection rates acquired in this experiment and those in the former one were then analyzed using a 2 (independent variable in between-subjects factor: types of artifacts, on-screen artifact or robot)  $\times$  2 (independent variable in within-subjects factor: type of ASEs, flat or decreasing) mixed ANOVA (dependent variable: rejection rate). The results showed that there were no significant differences in the interaction effects ( $F(1,37)=0.04$ , n.s.) and in the main effect of “types of artifacts” ( $F(1,37)=0.08$ , n.s.), while there was a significant difference in the main effect of “types of ASEs” ( $F(1,37)=20.48$ ,  $p<.01$  (\*\*)) (Figure 7). Thus, we could confirm that the ways of interpretation of the ASEs from the on-screen artifact were the same as the ones from a robot. In other words, the ASEs’ interpretation was consistent regardless of the types of artifact.

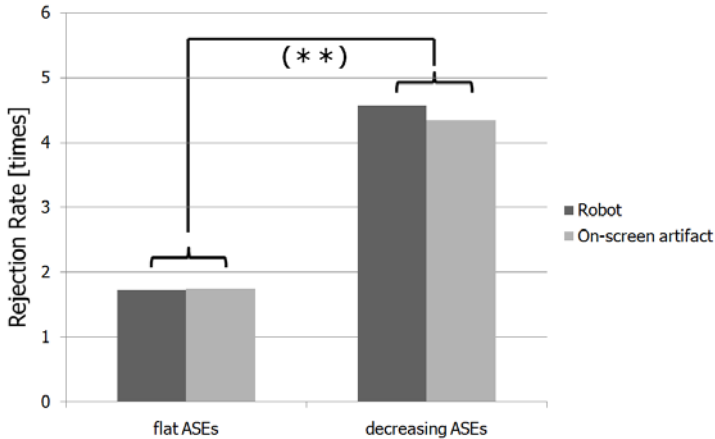


Fig. 7. Rejection rate for ASEs from an on-screen artifact and a robot

## 4 Discussion and Conclusions

We experimentally investigated whether the ASEs from the on-screen artifacts could convey the artifact’s internal states to participants in order to confirm whether the ASEs can be interpreted consistently regardless of the types of artifacts. The results of the experiment clearly showed that the ASEs from an on-screen artifact succeeded in accurately and intuitively conveying its internal states to the participants. In addition, a comparison of the results from this experiment with ones in a former study showed



that the ways of interpretations of the ASEs from an on-screen artifact were the same as the ones from a robot. In other words, the ASEs' interpretation was consistent regardless of the types of artifact. These results succeeded in strongly appealing the robustness and consistency of ASEs. Therefore, the experimental investigations in this paper strongly support this practical application of ASEs.

Now we are planning to implement the ASEs in various types of realistic applications requiring communication with their users; e.g., spoken dialogue systems such as ATMs or automatic reservation systems. In particular, we are now focusing on a car navigation system for the target application of ASEs, because current car navigation systems still sometimes present poor driving routes to their users. However, if this navigation system's confidence level regarding the route instruction is not very high, the speech instructions with the ASEs could implicitly convey a lower confidence level. If the ASEs are still effective in such situations, they could be utilized in realistic situations in which the 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), 175–200 (1994)
2. Cohen, P.R., Morgen, J., Pollack, M.E.: *Intentions in Communication*. The MIT Press, MA (1990)
3. Liu, K., Picard, W.R.: Subtle expressivity in a robotic computer. In: *Proc. CHI 2003 Workshop on Subtle Expressivity for Characters and Robots*, pp. 1–5 (2003)
4. Ward, N.: On the Expressive Competencies Needed for Responsive Systems. In: *Proc. CHI 2003 Workshop on Subtle Expressivity for Characters and Robots*, pp. 33–34 (2003)
5. Cowell, J., Ayes, A.: Extracting subtle expressions for emotional analysis. In: *Proc. 2004 IEEE SMC*, vol. (1), pp. 677–681 (2004)
6. Kipp, M., Gebhard, P.: IGaze: Studying reactive gaze behavior in semi-immersive human-avatar interactions. In: *Prendinger, H., Lester, J.C., Ishizuka, M. (eds.) IVA 2008. LNCS (LNAI)*, vol. 5208, pp. 191–199. Springer, Heidelberg (2008)
7. Sugiyama, O., Kanda, T., Imai, M., Ishiguro, H., Hagita, N., Anzai, Y.: Humanlike conversation with gestures and verbal cues based on a three-layer attention-drawing model. *Connection Science* 18(4), 379–402 (2006)
8. Komatsu, T., Yamada, S.: How do robotic agents' appearances affect people's interpretation of the agents' attitudes? In: *Ext. Abstracts CHI 2007*, pp. 2519–2524. ACM Press, New York (2007)
9. Funakoshi, K., Kobayashi, K., Nakano, M., Yamada, S., Kitamura, Y., Tsujino, H.: Smoothing human-robot speech interactions by using blinking-light as subtle expression. In: *Proc. ICMi 2008*, pp. 293–296. ACM Press, New York (2008)
10. Komatsu, T., Yamada, S., Kobayashi, T., Funakoshi, K., Nakano, M.: Artificial Subtle Expressions: Intuitive Notification Methodology of Artifacts. In: *Proc. CHI 2010*, pp. 1941–1944. ACM Press, New York (2010)
11. Funakoshi, K., Kobayashi, K., Nakano, M., Komatsu, T., Yamada, S.: Non-humanlike Spoken Dialogue: a Design Perspective. In: *Proc. SIGDIAL 2010* (2010) (to appear)
12. Wainer, J., Feil-Seifer, D.J., Shell, D.A., Mataric, M.J.: Embodiment and Human-Robot Interaction: A Task-Based Perspective. In: *Proc. IEEE ROMAN 2007*, pp. 872–877 (2007)
13. Shinowaza, K., Naya, F., Yamato, J., Kogure, K.: Differences in effects of robot and screen agent recommendations on human decision-making. *International Journal of Human-Computer Studies* 62, 267–279 (2004)