Can we assign attitudes to a computer based on its beeps? －Toward an effective method for making humans empathize with artificial agents

Takanori Komatsu
Future University-Hakodate
School of Systems Information Science
Kamedanakano 116-2, Hakodate 041-8655, JAPAN
komatsu@fun.ac.jp

Abstract

Can we assign attitudes to a computer based on its beeps? If so, which kinds of beeps are perceived as specific attitudes, such as “disagreement”, “hesitation” or “agreement”? To examine this issue, I carried out an experiment to observe how participants perceive or assign an attitude to a computer according to beeps of different durations and F0 contour’s slopes. The results revealed that 1) beeps with increasing intonation regardless of duration were perceived by participants as “disagreement”, 2) flat sounds with longer duration were interpreted as “hesitation”, and 3) decreasing intonations with shorter duration were as “agreement.”

1 Introduction

The purpose of this study is to clarify an effective method for making users perceive artificial agents as having attitudes and empathize with them. Actually, many researchers have developed humanoid robots [for example, Breazeal and Velasquez, 1998] that can exist in the same physical space with users and life-like (or character) agents [for example, Morishima *et al*., 1995] that appear on computer screens. These anthropomorphic agents are believed to facilitate communication between these agents and humans. Some researchers, however, have reported that humans perceive the attitudes of others from the information (e.g., humming without any language information), even if this information is simpler than that derived from gestures or facial expressions [for example, Suzuki *et al*., 2004].

Therefore, one may reasonably assume that the simple information presented by an agent without a human-like shape or face would also cause people to perceive different attitudes and to assign specific attitudes to such agents. Moreover, this perception might induce the users’ sympathy toward the agent and eventually trigger smooth communication between them.

As a preliminary study aimed at achieving such a relationship between users and agents, I have focused on a normal laptop computer as the agent without a human-like shape or face, and on beeps (sound stimuli) as simple information presented to humans. Specifically, I conducted an experiment in which various beeps were presented to participants from a laptop computer. The purpose of the experiment was to observe and analyze how participants assign specific attitudes to the computer based on the presented beeps.

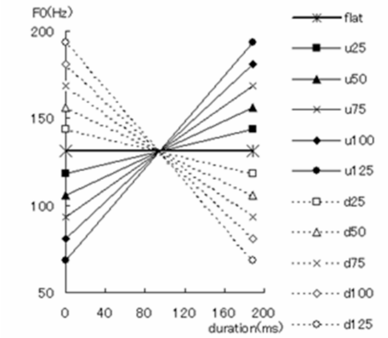

Figure 1. Eleven different F0 contours (duration; 189 ms): “u25” indicates that F0 transition range was 25 Hz with upward slope (increasing intonation).

2 Experiment

I prepared 44 different types of beeps with four different durations and 11 different F0 contours. Specifically, the four durations were 189, 418, 639, and 819 ms. The 11 F0 contours were set so that the transition range of F0 values that were linearly downward or upward (Figure 1). Participants were 23 Japanese (17 men and 6 women; 20 - 42 years old).

An experimenter asked participants to assume that these presented beeps were answering speeches in a telephone conversation (actually this was not true). As an experimental trial, one randomly selected beep among the 44 beeps was presented to participants. They were then asked to answer a question like *“Did you feel that [***] was this computer’s attitude based on this beep?”*; [***] was the randomly selected attitude among the following three that were complementary to each other. In addition, these three attitudes might correspond to valence evaluations; 1) Disagreement, e.g., surprising, doubting *(negative valence)*, 2) Hesitation, e.g., being lost for words *(neutral valence)*, 3) Agreement, e.g., acceptance *(positive valence)*. Specifically, each participant experienced 132 trials (44 beeps x 3 attitudes).

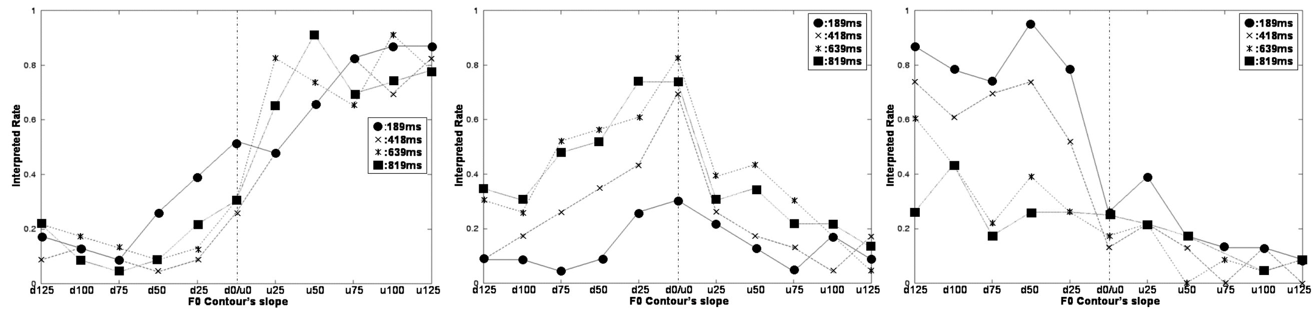

Figure 2. Relationship between F0 contours and rate of "disagreement" interpretation for each duration (left), rate of "hesitation" (center) and "agreement" (right).

3 Results

Case 1: Disagreement Figure 2 (left) shows which beeps were interpreted as the computer's attitude of "disagreement." In this figure, the horizontal axis lists the types of F0 contours and the labels on this axis, such as "u25" or "d100," are the same ones used in Figure 1. The vertical axis shows the interpretation rates (IR) that give the proportion of participants answering "yes" in the corresponding trials. There are four lines indicating the duration of sound stimuli.

In the results, no significant differences were found among different durations ($F(3,66)=.70$, n.s.) and different F0 transition ranges in falling tones ($F(4,88)= 1.68$, n.s.). On the other hand, rising tones showed a significant higher IR than falling tones ($F(1,22)=219.7$, $p<.01$), and wider F0 transition ranges in a rising tone showed higher IR ($F(4,88)= 9.81$, $p<.01$). To sum up, **when the beeps with upward slopes (increasing intonation) regardless of their duration were presented from the computer, this computer's attitude was interpreted as "disagreement."**

Case 2: Hesitation Figure 2 (center) shows which beeps were interpreted as "hesitation." Longer duration showed the higher IR ($F(3,66)=18.1$, $p<.01$), and falling tones showed the higher rate than the rising tones ($F(1,22)=11.64$, $p<.01$). Moreover, narrower F0 transition range in both rising and falling tones showed the higher IR than wider ones ($F(4,88)=11.23$, $p<.01$, $F(4,88)= 8.89$, $p<.01$). In sum, **when the slower slopes (flat sounds) with longer duration were presented, the computer's attitude was interpreted as "hesitation."**

Case 3: Agreement Figure 2 (right) shows which beeps were "agreement." Shorter duration showed the higher IR ($F(3,66)=19.42$, $p<.01$) and falling tones showed the significantly higher rate than rising tones ($F(1,22)= 67.5$, $p<.01$). And wider F0 transition range in falling tone showed rather higher rates than narrower ones ($F(4,88)= 3.12$, $p<.05$), and in rising tones, narrower transition range showed higher rates than wider ones ($F(4,88)= 5.48$, $p<.01$). To sum up, **when the downward slopes with shorter duration were presented, the computer's attitude was interpreted as "agreement."**

4 Discussion and Conclusions

The results have established that different kinds of simple information "beeps" can be interpreted as presenting the different attitudes of an information sender "a laptop computer." Therefore, sufficient possibilities exist to justify applying the results to an actual application. The same argument was presented by Lin and Picard [2003] in their "subtle expressivity" study. At first glance, these results seem to show that users can perceive only three agent attitudes; however, implementing the findings of this study into a laptop computer would create effective interactions with users. For example, when a computer responds to a certain action of users by presenting "beeps with increasing intonation," users would intuitively assign a particular attitude (in this case "disagreement") to the computer from this beep sound: "Umm... this computer seems to have some doubts about my last command..." And also implementing these results into a simple mobile robot would realize *a Star Wars R2D2 type robot*, which can not speak clearly but can express its attitudes by its electric sounds or behaviors. Thus, the results of this study can contribute to achieving an agent that enables users to assign or interpret intuitively and effectively particular agent attitudes.

References

[Breazeal and Velasquez, 1998] Breazeal, C., and Velasquez, J. Toward teaching a robot 'infant' using emotive communication acts. In *Proceedings of ISAB workshop on Social Situated Intelligence*, pages 25-40, 1998.

[Lin and Picard, 2003] Liu, K., and Picard. W. R. Subtle Expressivity in a Robotic Computer. In *Proceedings of the CHI2003 Workshop on Subtle Expressivity for Characters and Robot*, 2003.

[Morishima *et al*., 1995] Morishima, S., Iwasawa, S., Sakaguchi, T., Kawakami, F., and Ando, M. Better Face Communication. In *Visual Proceedings of SIGGRAPH95*, pages 117, 1995.

[Suzuki *et al*., 2004] Suzuki, N., Kakehi, K., Takeuchi, Y., and Okada, M. Social effects of the speed of hummed sounds on human-computer interaction. *International Journal of Human-Computer Studies*, 60, 455–468, 2004.