

WHY PUT AN AGENT IN A HUMAN BODY: THE IMPORTANCE OF
COMMUNICATIVE FEEDBACK IN HUMAN-HUMANOID DIALOGUE

Kristinn Thorisson and Justine Cassell
MIT Media Lab
20 Ames Street
Cambridge, MA 02139
kris@media.mit.edu, justine @media.mit.edu

Although many different human characteristics have been put forth as *the* key to making humanoid agents lifelike (eg. emotional expression, fluid body movement, face and hand gestures, realistic skin color) the young field of synthetic computer characters has not seen much research comparing these different putatively “most important” characteristics of lifelike computer characters. Of course, research on the effectiveness of natural language based, humanoid agent systems, and on the role of believability in the construction of such systems, has to date been hampered by the lack of real computer systems capable of sustaining and supporting spoken dialogue with a human user. In this paper we describe a comparison of two commonly discussed features: emotional facial icons and non-verbal communicative behavior. This comparison was made possible by a platform that supports the construction of humanoid agents and allows various features of those agents to be “turned off”.

We used a fully automated character generation system, capable of real-time, multimodal, face-to-face interaction with a user [Thórisson 1996], to assess users’ reactions to two commonly discussed human characteristics: facial emotional icons and non-verbal feedback about the interaction. We tested users’ reactions by way of a questionnaire assessing comfort with the interaction, but also by looking at the efficiency of the interaction, as measured by how many times users repeated themselves. Specifically, we compared users’ questionnaire responses to, and efficiency with a {1} content-only character (CONT), {2} a content + emotional facial icons character (EMO), and {3} a content + non-verbal communicative support character (ENV).

The characters all appear on a normal-sized monitor beside a big screen projector, on which a graphical model of the solar system is displayed. The users can ask the character questions about planets and have it show them the planets. The characters in each condition are equally knowledgeable about the solar system, and their responses are equally rapid, but they provide the following different feedback: In the CONT condition the character gives verbal feedback only relating to the content of the dialogue; the EMO character gives the same verbal feedback and also smiles occasionally, when it has finished some action, and looks puzzled if it doesn't understand what the user says; the ENV character provides the same verbal feedback as CONT with the addition of behaviors relating to the process of dialogue: turning to and/or looking at the big screen or the user at the right times, giving non-verbal feedback to show when it decides to take the turn (when the user has finished making a request), and hand gestures that support its utterances (beat gestures and pointing at the planets when speaking). It also blinks, and drums with its fingers when its hand is at rest. The experiment was a repeated-measures design, with twelve subjects. Thus, all subjects interacted with all three characters.

Two hypotheses were tested: {1} We expected to find no significant difference in ease of interaction or efficiency between the CONT and EMO conditions. That is, we didn't expect emotional facial icons to add anything to the interaction. {2} We expected to find a significant difference in ease and efficiency between the ENV condition and the other two conditions. In other words, we expected behaviors relating to the process of dialogue to

prove significantly more important to the users' acceptance of the character/interaction, as well as to the effectiveness of the dialogue, than either content feedback alone or content feedback and emotional facial displays.

Both hypotheses were confirmed ($p < .05$). This supports our claim that what really matters in face-to-face dialogue is, in addition to "classical information exchange", the supportive behaviors that often have been dismissed as incidental to effective interaction [Ochsman & Chapanis 1974]. Should designers of interactive computer agents (co-spatial, co-temporal speech-based interaction) ignore behaviors relating strictly to communication, they are likely to end up with less believable, less effective agents. We expect, however, that adding other behaviors such as emotional expression *on top of* such process-of-communication behaviors may be doubly effective.

References

- Ekman, P. (1979) "About Brows: Emotional and Conversational Signals." In M. von Crahach, K. Foppa, W. Lepenies & D. Ploog (eds.), *Human Ethology*, pp. 169-243.
- Hauptman, A.G. (1989) "Speech and Gesture for Graphic Image Manipulation." In *Proceedings of SIGCHI '89*, pp. 241-245.
- Maes, P. (1994) "Agents that Reduce Work and Information Overload." *Communications of the ACM*, 37(7), pp. 31-40, 146.
- Maulsby, D., D. Greenberg & R. Mandler. (1993) "Prototyping an Intelligent Agent through Wizard of Oz." In *Proceedings of InterCHI '93*, Amsterdam, April 24-29, pp. 277-284.
- Ochsman, R.B. & A. Chapanis. (1974) "The Effects of 10 Communication Modes on the Behavior of Teams During Co-operative Problem Solving." *International Journal of Man-Machine Studies*, 6, pp. 579-619.
- Thorisson, K.R. (1996) "Dialogue Control in Social Interface Agents." In *InterCHI Adjunct Proceedings '93*, Amsterdam, April 24-29, pp. 876-881.
- Thorisson, K.R. (1996) "Communicative Humanoids: A Computational Model of Psychosocial Dialogue Skills." Doctoral Dissertation, Massachusetts Institute of Technology, Media Laboratory, September 1996.