

Integrating talking faces in human-machine interaction: A long road from technological gadget to real application*

Lambert Schomaker (NICI, Nijmegen)

March 1999

1 Introduction

A recurring theme in theorizing about human-computer interaction is the duality of the following two views on the computer: (a) Computer-as-Tool, versus (b), Computer-as-Dialogue-Partner. Viewing the computer as a transparent tool (a), the intentionality of the dialogue is assumed to be present in the human user, whereas the machine is a passive tool, transparent and merely supportive with respect to the human goals. This is in fact the status of most current applications, and is also the acclaimed approach by well-known researchers in the field of human-computer interaction. In fact, an ongoing trend is still present from earlier anthropomorphizing error messages and dialogue fragments, towards more neutral, tool-like behavior. Alternatively however, there is a trend in which goal-oriented intentional behavior is explicitly encoded in the software of intelligent user-agents, which behave according to a logic based on beliefs about the world and about the goals of the user. At the same time there is a fast development of the technology by which anthropomorphic output can be produced by a computer in real time: Human movement output in general, speech and talking faces in particular. The critics of this latter approach pejoratively call it animism and point out

*Schomaker, L. (1999). Integrating talking faces in human-machine interaction: A long road from technological gadget to real application [Abstract] In: Zs. Ruttkay (Ed.), Face to Face Symposium, 25 March, CWI Amsterdam.

that user-friendly design is much more helped by giving the users the idea that they are fully in control of the situation, rather than having to negotiate with a potentially uncooperative anthropomorphic servant (How were you served, the last time you went to the bank or post office, lately?). An overview will be given of the findings of European research projects (mainly MIAMI, also EAGLES) in which talking-face technology played an essential role. The areas of interest concern the software-engineering aspects of building modular systems for talking-face applications and, secondly, user-interfacing problems. Although this is a growing and exciting field, it is evident that there are a number of serious problems with this technology. Many of these problems will have been solved in the near future due to hardware and software development, but there are still quite a few fundamental problems which cannot be solved by merely using more expensive hardware or employing more programmers. Leaving aside the problems at the peripheral side of the system (rendering, processing speed, synchronisation), the following insights have been gained:

- The main effect of improved rendering quality is that users demand still more naturalism
- In many applications the use of a human face is less desirable than using cute cartoon type animals such as dog or parrot
- For enjoying the benefits of visual speech output in case of speech synthesis, on-screen talking lips provide most of the needed cues
- Naive users will want to talk back, and expect a degree of intelligence at the other side of the screen that cannot be actually supported using current state-of-the art dialog control, including syntax and semantics analysis
- Whereas it may be acceptable to a user to organize the planning of daily activities in a dialogue with an anthropomorphic partner, it is equally unacceptable to use a human-style dialogue in mundane tasks like text editing or computer programming.
- Human-to-human interaction is characterized by a highly flexible protocol, in which visual and auditory analysis of non-verbal cues plays an essential role.

2 Conclusion

However, as the indicated by the title, the major stumbling block is the identification of applications and scenarios (interaction stages) in human-computer interaction that really profit from speech interaction with an anthropomorphic agent on screen. For the time being, fast integration into the office environment is unlikely, and many of the applications will be in terms of entertainment (computer gaming) and the information kiosk paradigm. Scientifically, this is not a problem. The challenge is there and remains to be exciting.

3 References

Esprit/BRA project MIAMI: Multimodal Integration in Advanced Multimedia Interfaces, <http://hwr.nici.kun.nl/~miami/>

EAGLES: <http://coral.lili.uni-bielefeld.de/EAGLES/>

Schomaker, L.R.B, Muench, S. & Hartung, K. (Eds). (1995). A Taxonomy of Multimodal Interaction in the Human Information Processing System. Report of the Esprit Project 8579 MIAMI (187 p.), Nijmegen: NICI.

Schomaker, L.R.B, Nijtmans, J., Camurri, A., Lavagetto, F., Morasso, P., Benoit, Guiard-Marigny, T., Le Goff, B., Robert-Ribes, J., Adjoudani, A., Defee, I., Muench, S., Hartung, K., Blauert, J. (1995). DI2 - Progress Report of the Esprit Project 8579 MIAMI (122 p.), Nijmegen: NICI.

(both Esprit/MIAMI reports are available as postscript at the MIAMI site: <http://hwr.nici.kun.nl/~miami/reports/reports.html>)