Designing Feedback for an Attentive Office

Teenie Matlock, Christopher S. Campbell, Paul P. Maglio, Shumin Zhai & Barton Smith

IBM Almaden Research Center, 650 Harry Rd, San Jose, CA 95120 USA

{tmatlock, ccampbel, pmaglio, zhai, basmith}@almaden.ibm.com

Abstract: In modern offices, information devices abound. How do people conceptualize interaction with a large number of such devices? What makes an effective design of a verbal information-access system? With these questions in mind, we observed users in a futuristic office setting, where information is accessed via only verbal commands. Results suggest a mismatch between behavior and attitude: when speaking to individual devices, people treat the devices socially, but they do not feel as though they are interacting socially. No mismatch was found when people spoke to the room as a whole. This implies that attentive user interfaces ought to be designed so that people interact with a single entity.

Keywords: Attentive interfaces, multimodal input, social interaction.

Technology is quickly moving beyond the personal computer (PC), becoming distributed in specialized “information appliances” that pervade our work and everyday environments (Norman, 1998). One way of organizing the technology around us is with attentive user interfaces (AUI)---systems or environments that monitor users through multimodal sensing, such as computer vision and speech recognition, and anticipate user desires and actions. There are three components to AUIs: user input (e.g. voice, gesture), feedback (e.g. speech, text, sounds), and information output.

Recent research suggests computer users sometimes treat computers like they treat people (e.g., Reeves and Nass, 1996). Many of these studies deal with situations in which computational systems present users with verbal information (speech or textual) in the course of problem-solving or some other activity, and then users make judgments about the systems. As far as we know, however, no studies deal with cases in which only the user presents a system with verbal information (e.g., speech) and the system responds non-verbally. In these cases, should we expect users to view the computers as they view people? The answer is important, for in an attentive environment, this is exactly the sort of interaction we would expect to occur most commonly.

We conducted a study of user actions and attitudes in a futuristic office mock up in which information is accessed on multiple displays via verbal commands. Our concern was how to design the nature of feedback. Feedback can supply complex information (e.g. text messages or speech) as well as simple information (e.g. a beep as acknowledgement). In addition, feedback may imply interaction with many entities (e.g. many feedback indicators) or just one. Do people expect to interact with many devices individually or do they expect to interact with the office as a single entity? Do people think of the devices as "social objects" when feedback is simple or must feedback be closer to normal speech? Answers to such questions can be used to inform the design of attentive environments.

1 EXPERIMENT

To separate conceptual issues from technology limitations, a Wizard-of-Oz design was used to provide accurate and timely reactions to user commands. The reactions of the office mock up were controlled by one of the experimenters behind a wall. Participants were given a set of office-related tasks to perform using verbal commands. A green blinking light served as feedback that the command was understood and being executed. There was one between-subjects factor with two levels, distributed feedback and non-distributed feedback. In the distributed condition (DC), feedback in the form of a green flashing light was seen on each device. In the non-distributed condition (NC), feedback appeared in a single location on the wall representing the "room". We were interested in whether and how people's attitudes change with the kind of feedback provided. Additional details of our experimental setup and other results can be found in (Maglio, Matlock, Campbell, Zhai, and Smith, 2000). Here, we report the results of a post-experiment questionnaire, which was not reported previously.

In the attentive office, three small screen displays were labeled "Calendar", "Map/Directions", and "Address", and a plastic, futuristic-looking orb was labeled "Dictation". There was also a printer. In the DC, a small black box with a green light (feedback module) was attached to the top of each screen. For the dictation device, no screen was used, so the feedback module was placed behind the ball. No manual input devices were in the room. Thirteen volunteers were randomly placed into one of two conditions—DC or NC—and each was given a list of seven tasks, including get an address, dictate memo, print memo, find date from calendar, get directions, and print directions. These were to be completed using the four devices in the room. After performing the tasks, participants were given a questionnaire that elicited their reactions to the environment.

2 RESULTS

As reported in Maglio et al. (2000), we found that verbal and gaze interactions with the environment did not differ between DC and NC. Participants did not verbally address the various devices, but the majority of looks at devices occurred before the verbal requests. Looking before speaking may indicate that participants were specifying the recipient of the request with gaze as in person-to-person communication. This observed pattern in the attentive office suggests participants default to a natural social interaction with individual devices.

Contrary to the gaze data, however, participants felt that talking to the room was more like talking to a person and was more natural than talking to distributed devices (see Figure 1). This suggests that while people may default to a social type of interaction with computers, they feel as though they are still dealing with a computer. In sum, people are happy using voice interaction in the distributed condition but the interaction in this condition is less natural than the non-distributed condition.

Taken together, the gaze and attitude data suggest that simple blinking feedback was sufficient for social interaction and does not detract from it. This result is not entirely surprising given that people already use simple indicators for interacting socially, such as, car horns and lights, and hand gestures.

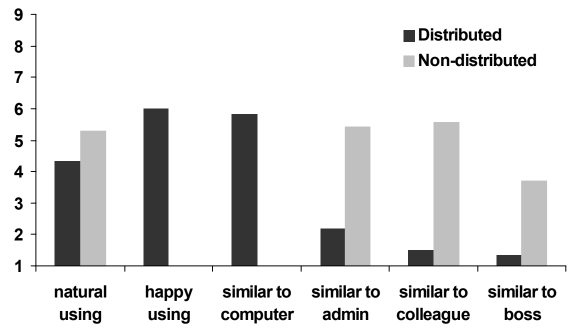

Figure 1: Interactions were judged to be similar to a computer in the distributed condition, but more like a person in the non-distributed condition.

Future work will be aimed at examining a wider range of feedback (e.g., auditory speech, text speech, simple sound feedback, no feedback, multimodal feedback) and whether these different types of feedback provide natural interaction in distributed/non-distributed situations. Additional work will look at the design of AUI in other contexts, such as kitchens, entertainment rooms, bedrooms, and automobiles.

3 CONCLUSION

Our primarily concern was how people's expectations and experience influence the design of AUIs. By observing people working in a mock office of the future, we found that a simple blinking light was sufficient for natural social interaction. Though speech feedback has also been shown to produce social interactions (Reeves and Nass, 1996), simple visual feedback may be less disturbing, require less cognitive load, and less computational power. Moreover, the acceptable level of distraction resulting from certain feedback may vary by context (i.e., blinking lights would be distracting in an automobile). Finally, we found that only the room as a single entity was treated and viewed as a social object.

References

Maglio, P. P., Matlock, T., Campbell, C. S., Zhai, S., & Smith, B. A. (2000). Gaze and speech in attentive user interfaces, in Proceedings of the International Conference on Multimodal Interfaces 2000.

Norman, D. A. (1998). The invisible computer. Cambridge, MA: MIT Press.

Reeves, B. & Nass, C. (1996). The media equation. Cambridge, England: Cambridge University Press.