Multimodal Signals for Designing an Awareness-Communication System

Dimitra Anastasiou[1], Wilko Heuten[2], Susanne Boll[3]

Media Informatics and Multimedia Systems
University of Oldenburg, 26121 Oldenburg, Germany[1 3]
OFFIS – Institute for Information Technology, Escherweg 2,
26121 Oldenburg, Germany[2]

Introduction

Nowadays synchronous (e.g., telephone, video-conference systems) and asynchronous (e.g., email, social networks) means help us communicate with our distantly separated peers. However, this distant communication is not as productive as face-to-face communication and it often happens abruptly and obtrusively. In our opinion, multimodal awareness signals before the beginning of communication would make the transition from the actual activity state to the communication state more fluent and in addition, would preserve privacy.
Our research is on facilitating spontaneous and informal communication in spatially distributed working groups by exploiting smart environments and ambient intelligence. In this abstract we particularly focus on the research question "what kind of multimodal signals are appropriate for awareness in workspace". We follow the definition of *workspace awareness* by Gutwin and Greenberg (1996): "the collection of up-to-the-moment knowledge a person uses to capture another's interaction with the workspace". A first step towards exploring which multimodal signals are more appropriate for awareness is to run a pilot user study. Some results on combining various interaction modalities in order to raise awareness for communication are presented in this abstract.

Multimodal signals for Representation Design

Our research goal is to design an awareness system (see Markopoulos et al., 2008), which can also serve as a communication system. The methodology to reach our goal is categorized into three distinct steps:

i) Detection of situations with the potential for spontaneous informal communication;
ii) Representation of these situations appropriately to distant users;
iii) Enabling them to engage in communication spanning multiple spatial locations.

The above steps include the perception of the current potential communication situations, the transparent and privacy-preserving detection of instances of situations, representation of formalized behavioural cues in distributed setting, and last but not least, human-computer interaction (HCI) methods. We focus particularly on the interaction methods and their requirements, such as implicit/implied communication, intuitiveness, and unobtrusiveness. Diagram 1 shows that multimodal signals, which can be for example, light, sound, touch, but also hand/body gestures, can raise unobtrusive awareness and through a fluent tranition, promote the spontaneous and yet explicit communication.

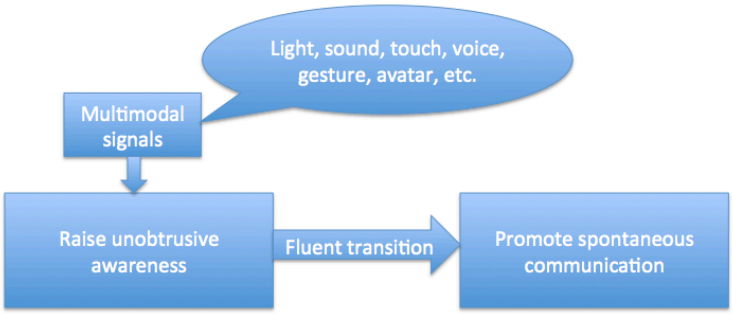

Diagram 1. Multimodal signals for social communication through HCI

We run a Wizard-of-Oz (WoZ) experiment in order to test the perception of some multimodal signals as well as the level of transition from the actual task of the participants to the communication step. For the evaluation we used a user experience questionnaire and a think-aloud protocol. We tested five conditions: i) close light, ii) distant light (lamp was in the periphery); sound effect; iv) combination of sound and light; v) absence of signals, and followed the heuristics by Mankoff et al. (2003). As the results showed, the close light was evaluated as the most easily perceptible signal with 94,12%. Regarding the overall preference of the signals, most participants preferred in general the situation-dependent option (light or sound) and not the combination of the signals (29,41%). One participant stated "one signal is actually sufficient, as the combination leads to stimulus satiation". The distant light distracted the participants less than the close light, thus they rated the transition through the distant light as more fluent compared to the other triggered signals.

As a next research step, we have to formalize potential communication situations. Let us imagine that at the foyer of our company there is an ambient display and we are currently passing by. In a branch office in another city a colleague does the same. Before a video communication software pops up, what kind of signals do we expect as an output from the display (or input from the other side), if any at all? As an output, it could be, for instance, a pulsing, background light, a soundscape with increasing volume as we come closer or even an avatar, which welcomes us and introduces to the display's functionalities. As an input, there could be an in-air hand gesture, voice, but also raw sensor data, such as spatial distance, etc. Some of these modalities are more implicit, some more explicit; the transition from implicit to explicit communication should be transparent but fluent. The future system should be unobtrusive, scalable, and customizable to the user's needs. For these and other reasons, the future interface should be multimodal in order to give the user the opportunity to intuitively choose the interaction mode and easily use this mode. Moreover, the representation design should not necessarily on an ambient display. With the actual developments of the Internet-of-Things (IoT), the ambient display might be replaced with any smart object that is available in a pervasive (working) environment: cup of coffee, desk, chair, whiteboard, flower pot, etc. For example, Wallbaum et al. (2015) developed an artificial social plant which enables users to keep track of a loved person throughout the day by unobtrusively visualizing the partner's current state of mind via different colors of the blossom.

Discussion

Non-verbal or implicit communication is very important in our frequent communication with our spatially distributed co-workers. This kind of communication includes the perception of the currently performed activity, behaviour or the presence of other people. In our pilot study, we observed the reaction of the participants after the trigger of multimodal signals and evaluated their perception and transition to communication. In the field of IoT and awareness systems, we would like to explore further possibilities of multimodal signals for awareness and communication systems. This is possible by interpreting of social signals through the recognition of behavioral cues (Meservy et al., 2008), such as facial expressions, head movement, body gestures, voice detection, and speech recognition.

References

Gutwin C., Greenberg, S. (1996). Workspace Awareness for Groupware: In: Proceedings of the Human Computer Factors in Computing Systems (CHI 1996), ACM Press, 208-209.

Mankoff, J., Dey, A.K., Hsieh, G., Kientz, J.A., Lederer, S., Ames, M. (2003). Heuristic evaluation of ambient displays. In: Proceedings of CHI, 169-176.

Markopoulos, P., De Ruyter, B., Mackay, W. (2009). *Awareness systems: advances in theory, methodology, and design*. Human-Computer Interaction Series. Springer London, London.

Meservy, T., Jensen, M., Kruse, W. (2008). Automatic extraction of deceptive behavioral cues from video. In: Terrorism Informatics Knowledge Management and Data Mining for Homeland Security, H. Chen, E. Reid, J. Sinai, A. Silke, and B. Ganor, Eds. Springer US, 198–208.

Wallbaum, T., Timmermann, J., Heuten, W., Boll, S. (2015). Forget Me Not: Connecting Palliative Patients and Their Loved Ones. In: Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems, 1403-1408.