Available online at www.sciencedirect.com
ScienceDirect
Int. J. Human-Computer Studies 69 (2011) 362–374
International Journal of Human-Computer Studies
www.elsevier.com/locate/ijhcs

Using a game controller for relaying deictic gestures in computer-mediated communication

Swen Walkowski[a], Ralf Dörner[a,*], Mirja Lievonen[b], Duska Rosenberg[b]

[a]RheinMain University of Applied Sciences, Department of Design, Computer Science, Media, Kurt-Schumacher-Ring 18, D-65197 Wiesbaden, Germany
[b]Royal Holloway University of London, London, UK

Received 9 November 2009; received in revised form 26 October 2010; accepted 3 January 2011
Communicated by K. Isbister
Available online 11 January 2011

Abstract

In computer-mediated communication, participants often experience a loss of information that would have been conveyed by non-verbal means in a face-to-face communication. They need to compensate for this loss by being more verbose which may have a negative impact on the efficiency of communication or give rise to misunderstandings. In this paper, we present a computer system that augments the imagery perceived from the remote site with additional visual information and that can be controlled by using a Nintendo Wiimote game controller. Our prototype illustrates our approach how a system might be able to partially compensate for the information typically lost in conventional video conferencing where deictic gestures cannot be relayed properly to the remote site. Concerning user acceptance, our user tests also show that the visualization technique is crucial which is used for highlighting the objects or persons that are being pointed at. Adequate visualization techniques based on non-photorealistic rendering are proposed.

Keywords: Human computer interaction; Usability studies; Videoconferencing; Communication; Virtual pointing; Communicative function; Deictic reference; Deictic gestures; Shared understanding; Interface design; Wii; Highlighting; Computer-mediated communication

1. Introduction

Today, computer-mediated communication still has drawbacks when compared to face-to-face communication. In principle, novel technology carries the potential to mitigate these drawbacks. In this paper, we examine how a novel pointing device, the Wiimote game controller from Nintendo, can be employed in video conferencing settings. Building on advanced technology like infrared and gyroscopic sensors, a whole class of novel devices (such as the Wiimote, but also smart phones like the iPhone) became available recently. They significantly improved the quality of pointing devices, e.g. the accuracy. Moreover, these devices are wireless and are also reasonably priced since they are produced for large consumer markets.

One drawback of computer-mediated communication is the fact that deictic gestures, an important aspect of human communication, cannot be conveyed as well as in face-to-face communication. In particular, in video conferencing it is difficult to interpret deictic gestures correctly since the spatial layout of cameras and displays is often not known and even if it is, people may find it hard to visualize it mentally. Moreover, deictic gestures may not be interpretable at all because information is missing: only the parts of the remote site are visible that lie in the viewing frustum of the camera.

Our idea is that we introduce not only a novel pointing device in video conferencing in order to make information about the occurrence of deictic gestures available to the technical system. In addition, we modify the video image in order to visualize the highlighting of an object or a person that is the referent of a deictic gesture. As such, we present a novel way how to technically implement computer-mediated

*Corresponding author. Tel.: +4961194951216; fax: +4961194951210.
E-mail addresses: swen.walkowski@hs-rm.de (S. Walkowski), ralf.doerner@hs-rm.de (R. Dörner), M.Lievonen@rhul.ac.uk (M. Lievonen), D.Rosenberg@rhul.ac.uk (D. Rosenberg).

1071-5819/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ijhcs.2011.01.002

communication and offer new system features. We discuss the technical feasibility of such an approach. For this, we have built a tool called Enricos (Enriched Communication System) as a technical proof of concept.

Evaluating just the technical feasibility is not enough in order to assess the usefulness of Enricos. Designing Enricos, our aim was to reduce the need for verbalization and enable users to communicate with non-verbal means as in a face-to-face setting. An overall goal is to find IT-based substitutions of natural deictic gestures in order to achieve a more fluent, more efficient and thus more pleasant computer-mediated communication. Simultaneously, by increasing the deictic presence and the common ground between participants, we strive to reduce the probability of misunderstandings while increasing the sense of tele-presence. Does our approach, that is reflected in Enricos, achieve these goals? Does it really improve computer-mediated communication? This is a first important question. A second important question is concerned with user acceptance and usability. Is Enricos usable? Which implementation alternatives should be chosen to ensure user acceptance? We propose to answer question 2 before question 1 because technical decisions that lead to unsatisfactory user acceptance might hamper the improvement of computer-mediated communication.

Concerning question 2, we present the results of user tests conducted with Enricos. Our results provide guidance which implementation alternatives, e.g. which visualization and highlighting techniques, should be chosen when implementing our approach. Answering question 1 is beyond the scope of this paper. However, results presented in this paper provide a starting point for investigation and Enricos serves as a valuable tool for finding answers to question 1.

This paper is organized as follows. In the next section we give some background on the role of deictic gestures in communication in general and the differences concerning deictic gestures in face-to-face and computer-mediated communication in particular. In Section 3, we describe the hardware set-up of Enricos and the properties of the game controller to be used. We then give our rationale for the interaction design in Section 4. In Section 5 we focus on an issue that turned out to be of the special importance: the conceptual realization of visually highlighting objects in the video image. Section 6 gives implementation details of the Enricos tool. Before we conclude, user tests and their results are reported in Section 7.

2. Deixis and computer-mediated communication

When people negotiate shared meaning in natural, face-to-face settings, they rely on multiple modes of communication. Cues, such as gaze direction, hand gestures, facial expression, body position and orientation, volume of voice, and intonation of speech provide feedback, facilitate turn-taking, and convey subtle meanings (Kendon, 1990). People tend to pay attention in particular to the face and gesturing hand (Tan and Davis, 2004).

Deixis specifies the referent in a particular communicative context; it is typically communicated using different modes such as gaze direction, verbal utterances, and pointing gestures (Kendon, 1990). Deictic expressions (such as 'that one', 'you') rely completely on their context. Demonstrative pronouns (such as 'this', 'that') share with interrogatives (such as 'what' 'where', 'who') similar pragmatic function of initiating the search for a specific referent (Diessel, 2003). Due to its cross-referential nature, deixis bridges pragmatic and semantic aspects of discourse, e.g. March et al. (2009).

In spatial terms, deictic gestures function to indicate and thereby locate objects/ events in the physical world: what is pointed at is referred to. Pointing gestures are often used to disambiguate deictic reference (e.g. 'this one'). Pointing can be synchronous and parallel to verbal expressions but it often manages to convey a message, in concert with facial expressions and gaze direction, without any verbal utterances. Deictic gestures therefore contribute to our shared understanding (common ground): they make communication more efficient by reducing a need for procedural utterances; they also manage a long way across language barriers.

Tomasello and Carpenter (2007) point out the crucial role of shared intentionality in our cultural life: humans share psychological states with others; they for instance provide each other with helpful information; they form shared intentions and attention with others, and learn from them. In children's early cognitive development, gaze following is transformed by shared intentionality into joint attention, social manipulation into cooperative communication, group activity into collaboration, and social learning into instructed learning. That makes understandable why reciprocity of perspectives is regarded as a necessary foundation for socially organized interaction, and joint attention considered foundational for human communication. According to prior research, e.g. Heath and Luff (1991) and Gaver (1992), collaboration relies upon the participants' abilities to develop and sustain mutually compatible perspectives.

In human communication, intelligibility of referential actions is grounded in the activities (Hindmarsh and Heath, 2000). The basis on which people ground their communication in natural settings becomes spatially fragmented in videoconferencing. It impacts the perspective of the participants on their remote site colleagues, impedes feedback and the joint frame of reference for the participants. Due to information available to the participants of communication, the main difference between co-located and computer-mediated communication is modal fragmentation when a different basis for grounding is technologically configured using visual and audio channels. In addition, the communication technology influences people's usage of deictic gestures and also the way they build common ground (Clark and Brennan, 1991). This becomes apparent in the impact shared visual information has on reference in collaborative settings (Gergle et al., 2007).