Chapter XXXV
Gaze-Aided Human-Computer and Human-Human Dialogue

Pernilla Qvarfordt
FX Palo Alto Laboratory, USA

Shumin Zhai
IBM Almaden Research Center, USA

ABSTRACT

Eye-gaze plays an important role in face-to-face communication. This chapter presents research on exploiting the rich information contained in human eye-gaze for two types of applications. The first is to enhance computer mediated human-human communication by overlaying eye-gaze movement onto the shared visual spatial discussion material such as a map. The second is to manage multimodal human-computer dialogue by tracking the user's eye-gaze pattern as an indicator of user's interest. The authors briefly review related literature and summarize results from two research projects on human-human and human-computer communication.

The eyes of men converse as much as their tongues, with the advantage that the ocular dialect needs no dictionary, but is understood all the world over.

—Ralph Waldo Emerson, 1860

INTRODUCTION

Increasingly, people are working and socializing in distributed groups that seldom, if ever meet face to face. When conversing face to face, we use a range of non-verbal behavior, such as eye-gaze, to complement and enhance our speech. When using communication technology of today, these non-verbal behaviors are either completely lost, or distorted, as in the case of gaze in video conferences, so that they become hard to interpret. This chapter describes two related novel paradigms for

using one channel of non-verbal behavior, eye-gaze, to enhance effective communication.

In the first paradigm, designed for computer mediated human-human communication, the conversation partner's eye movement is directly superimposed onto the visual-spatial material being discussed, such as a map, so that one party can not only hear what the other party says, but also where the other party is looking. In the second paradigm, designed for improving multimodal human-computer dialogue, the computer utilizes the information of its user's eye-gaze pattern on the computer screen to initiate or manage the human-machine dialogue. The first paradigm can be useful in itself, but the knowledge gained in studying it can also inform the design of the second paradigm, as well as other collaborative systems. This chapter is organized accordingly. We first present the development and experimentation of a simulated tourist consulting service, RealTourist, which allows a tourist to talk to a remote tour consultant to plan a conference trip. The tourist and the consultant see the same map displayed on their monitors respectively. On the consultant's side the system also superimposes the tourist's eye-gaze onto the map, so the consultant could use it to determine the tourist's interest. Later we present the design, implementation, and a user study of an automatic tourist information system, iTourist, which automatically provides the user with city tour information in the form of a map, photos of different places, and synthesized speech. iTourist directs its information output based on the user's interests and needs analyzed from the user's eye-gaze pattern. Finally, we discuss the implications of the findings from these two paradigms for future collaborative system to enhance communication between groups of people and for future research directions.

GAZE AND CONVERSATION

In face-to-face conversation, much can be intuitively felt from the conversational partners' eye-gaze—whether they are interested or bored, attentive or preoccupied, engaged or unmindful, in doubt or in agreement, wanting to continue or trying to finish the conversation. Indeed, research has confirmed that eye-gaze plays an important role in face-to-face conversation. It enables us to assess a conversational partner's understanding, what he or she is looking at, and his or her feelings (Argyle & Cook, 1976).

Gaze plays a particularly important role in face-to-face communication when it comes to regulating the turn-taking behavior in a conversation. Gaze is used to signal if the speaker is about to hand over the turn, if he or she will continue after a pause, or if the speaker expects some feedback from the listener (Bavelas, Coates, & Johnson, 2002; Kendon, 1967). When two people attempt to take the turn simultaneously, gazes are used to resolve who should have the turn (Duncan & Niederehe, 1974). Beyond turn-taking, gaze is also used for emphasizing particular words or phrases, and aversion of gaze indicates lack of interest or disapproval (Argyle & Cook, 1976). The use of gaze is also related to the content of speech. Cassell, Torres, and Prevost (1999) has showed that when the speaker starts a new topic with a new utterance, he or she looks at the listener. When the speaker is pursuing an old topic, the speaker looks away at the beginning of the turn. When the utterance is a request, gaze is used to make sure that the addressee understands that he or she is supposed to listen (Goodwin, 1980, 1981). Similar pattern have been found when a person gives commands to an interactive object (Maglio, Matlock, Campbell, Zhai, & Smith, 2000).

The benefit of seeing a communication partner's eye-gaze has motivated many well known Computer Supported Collaborative Work (CSCW) design solutions. For example, Buxton (1990) used a half-silvered mirror to optically align a camera with a video screen to enable eye contact ("video tunnelling"). Vertegaal (1999) used virtual faces (avatars) that could rotate depending "who is talking to whom." Although research in face-to-face communication indicates that gaze serves an important function in communication, it has not been easy to empirically demonstrate the impact of preserving