Designing Attentive Cell Phones Using Wearable EyeContact Sensors

Roel Vertegaal, Connor Dickie, Changuk Sohn
Human Media Lab, CISC
Queen's University
Kingston, ON K7L 3N6, Canada
{roel, connor, changuk}@cs.queensu.ca

Myron Flickner
IBM Almaden Research Center
650 Harry Road
San Jose, CA
flick@almaden.ibm.com

ABSTRACT
We present a prototype attentive cell phone that uses a low-cost EyeContact sensor and speech analysis to detect whether its user is in a face-to-face conversation. We discuss how this information can be communicated to callers to allow them to employ basic social rules of interruption.

KEYWORDS: Attentive Interfaces, Mobile Phone, Gaze, Eye tracking, Alternative Input.

INTRODUCTION
With the increased ubiquity of mobile phone usage comes an increase in the number of ways in which technology can interrupt our everyday activities. We have all been in situations where a mobile phone inappropriately interrupted a meeting. This problem has even prompted the installation of cell phone jamming technology in some buildings. It is time that mobile phones start behaving more responsibly towards their user and the people around her, and implement some of the basic social rules that surround human face-to-face conversation. Face-to-face interaction is different from the way we interact with most technological appliances in that it provides a rich selection of both verbal and nonverbal communication channels [5]. This richness is characterized by a flexibility in choosing alternate channels of communication to avoid interference, a continuous nature of the information conveyed, and a bi-directionality of communication. E.g., when person A is in a conversation with person B, there are a number of ways in which person C may interrupt without interfering. Firstly, person C may position himself such the interlocutors are aware of his presence. Proximity and movement may peripherally indicate urgency without verbal interruption. This allows the interlocutors to wait for a suitable moment in their conversation to grant the interruption, e.g., by establishing eye contact with person C. When the request is not acknowledged, person C may choose to move out of the visual field of the interlocutors and withdraw the interruption. This subtlety of interruption patterns is completely lost when using mobile phones. Firstly, users of mobile phones tend not to be aware of the status of interruptability of the person they are trying to call. Secondly, they have limited freedom in choosing alternative channels of interruption. Thirdly, the channels that do exist do not allow for any subtlety of expression.

CHI 2002, April 20-25, 2002, Minneapolis, Minnesota, USA.
ACM 1-58113-454-1/02/0004.

In this paper, we describe how we might augment mobile devices with the above capabilities. We discuss a prototype attentive cell phone based on a Compaq iPAQ handheld augmented with a low-cost wearable eye tracker capable of detecting when a user is in a face-to-face conversation.

PREVIOUS WORK
Our work is based upon two separate strands of thinking about attentive technologies, the Priorities System by Horvitz [2] and GAZE by Vertegaal [5]. With Priorities, Horvitz developed software that reasons about notification strategies given observed priorities of e-mail messages. It developed what we consider to be the cognitive dimension of an Attentive Interface: a user interface that dynamically prioritizes the information it presents to its users, such that processing resources of both user and system are optimally distributed across a set of tasks [6]. Attentive Interfaces cannot exist without knowing the status of attention of their user. With the GAZE project [5], Vertegaal demonstrated that eye tracking and speech energy provide reliable channels of input for determining the attentive state of a user. The present work extends that of Hinckley and Horvitz [1] towards a more sensitive cell phone by adding the capability to detect when a user is in a face-to-face meeting. It extends Selker's EyeR [4] system to include the tracking of an onlooker's pupils.

DETERMINING ATTENTIVE STATE
Wearable microphone headsets are becoming increasingly common in mobile phones. The signal from such microphones is available with high fidelity even when the user is not making a call. We modified the iPAQ to accept such input, allowing it to monitor user speech activity to estimate the chance that its user is engaged in a face-to-face conversation. Wireless phone functionality is provided by voice-over-ip software connected through a wireless LAN to a desktop-based call router. An attentive state processor running on the same machine samples the energy level of the voice signal coming from the iPAQ. To avoid triggering by non-speech behavior we used a simplified version of a turn detection algorithm described by Vertegaal [5]. When more than half the samples inside a one-second window indicate speech energy, and those samples are evenly balanced across the window, the probability of speech activity by its user is estimated at 100%. For each second that the user is silent, 5% is subtracted from this estimate, until zero probability is reached. Thus we achieved a short-term memory of 20 seconds for speech activity by its user.

Figure 1. Our attentive cell phone detects that its user is in a meeting using an EyeContact sensor worn on clothing.

Figure 2. Attentive cell phone display showing attentive state of contacts. Calls are made by touching a contact's picture.

EyeContact Sensors Detect Listening Behavior

Speech detection works well in situations where the user is the active speaker in a conversation. However, when the user is engaged in prolonged listening, speech detection alone does not suffice. Given that there is no easy way to access the speech activity of an interlocutor without violating privacy laws, we developed an alternative source of input. According to Vertegaal [5], eye tracking provides an extremely reliable source of information about the conversational attention of users. In dyadic conversations, speakers look at the eyes of their conversational partner for about 40% of the time. To access this information we developed a low-cost EyeContact sensor capable of detecting eye gaze at a user by an interlocutor. Our current prototype, based on the IBM PupilCam [3], is light and small enough to be worn on a baseball cap (see Fig. 1). The sensor consists of a video camera with a set of infrared LEDs mounted on-axis with the camera lens. Another set of LEDs is mounted off-axis. By syncing the LEDs with the camera clock a bright and dark pupil effect is produced in alternate fields of each video frame. A simple algorithm finds any eyes in front of a user by subtracting the even and odd fields of each video frame [3]. The LEDs also produce a reflection from the cornea of the eyes. These glints appear near the center of the detected pupils when the onlooker is looking at the user, allowing the sensor to detect eye contact without calibration. By mounting the sensor on the head, pointing outwards, the sensor's field of view is always synchronized with that of the user. Sensor data is sent over a TCP/IP connection to the attentive state processor, which processes it using an algorithm similar to that used for speech to determine the probability that the user received gaze by an onlooker in the past 20 seconds.

CHOOSING A NOTIFICATION

The attentive state processor determines the probability that a user is in a conversation by summating the speech activity and eye contact estimates. The resulting probability is applied in two ways. Firstly, it sets the default notification level of the user's cell phone. Secondly, it is communicated over the network to provide information about the status of a user to potential callers.

Communicating Attentive State to Callers

When a user opens his contact list to make a phone call, our attentive phone updates the attentive state information for all visible contacts. Fig. 2 shows how attentive state information is communicated. Below the contact's name a menu shows the preferred notification channel. Notification channels are listed according to their interruption level: message; vibrate; private knock; public knock; and public ring. Users can set their preferred level of interruption for any attentive state. They can also choose whether to allow callers to override this choice. When contacts are available for communication, their portraits display eye contact. A typical preferred notification channel in this mode is a knocking sound presented privately through the contact's head set. When a contact is busy, her portrait shows the back of her head. A preferred notification channel in this mode is a vibration through a pager unit. When their request times out, callers may choose a different notification strategy, if allowed. However, in this mode the contact's phone will never ring in public. Users can press a "Don't Answer" button to manually forestall notifications by outside callers for a set time interval. This is communicated to callers by turning the contact's portrait into a gray silhouette. Offline communication is still possible in this mode, allowing the user to leave voicemail or a text message.

CONCLUSIONS

We presented an attentive cell phone design that uses an EyeContact sensor and speech detection to communicate when its user is in a face-to-face conversation. Awareness of this helps callers use more sociable interruption rules.

REFERENCES

1 Hinckley, K. & Horvitz, E. Toward More Sensitive Mobile Phones. In Proceedings of UIST '01. Orlando: ACM, 2001
2. Horvitz, E., Jacobs, A., and Hovel, D. Attention-Sensitive Alerting. In *Proceedings of UAI '99.* Morgan Kaufmann, 1999.
3. Morimoto, C. et al. Pupil Detection and Tracking Using Multiple Light Sources. Image and Vision Computing, vol 18, 2000.
4. Selker, T. et al. EyeR, a Glasses-Mounted Eye Motion Detection Interface. In Extended Abstracts of CHI '01. Seattle: ACM, 2001.
5. Vertegaal, R. The GAZE Groupware System. In Proceedings of CHI'99. Pittsburgh: ACM, 1999.
6. Vertegaal, R. Designing Attentive Interfaces. In Proceedings of Eye Tracking Research and Applications '02. New Orleans: ACM, 2002.