EyePliances: Attention-Seeking Devices that Respond to Visual Attention

Jeffrey S. Shell, Roel Vertegaal, Alexander W. Skaburskis
Human Media Lab
Queen's University
Kingston, ON, Canada K7L 3N6
{ shell, roel, skaburskis } @cs.queensu.ca

ABSTRACT
We present EyePliances: appliances and devices that detect and respond to human visual attention using eye contact sensors. EyePliances receive *implicit input* [3] from users, in the form of eye gaze, and respond by opening communication channels. By allowing devices to recognize the attentional cues people already provide, requests for explicit input from users can be reduced. Further, eye contact sensing gives devices a mechanism to determine whether a user is available for interruption, and can provide the missing environmental context to improve speech recognition.

Keywords Attentive User Interfaces, Eye Tracking, Eye Contact Sensing, Alternative Input, Speech I/O, Non Command User Interfaces, Ubiquitous Computing.

INTRODUCTION
Recently, the speed, number, and power of mobile communication devices has increased dramatically, resulting in increased fragmentation of the already limited resource of human attention. This paper presents a method to augment attention-seeking devices, like mobile phones [8] and televisions, with eye contact sensors. Eye contact sensors, based on the IBM PupilCam [2] report whether a person is looking at them. Using this attentional cue, EyePliances respond in a manner appropriate to the context of the event, reducing distracting and inappropriate requests for user attention [6].

The problem of improving and augmenting users' ability to navigate environments armed with ubiquitous technology has been investigated in many ways, notably through wearable computing [7,8]. Wearable computing often requires an element of portable computer vision, with the ability to identify real world objects. By using static eye contact sensors, object identification is implicit – the object selected is the object with the eye contact sensor embedded in it. The difficult problem of object identification using computer vision is reduced to simply finding pupils.

Visual Attention
Eye contact functions as a *nonverbal visual* signal that peripherally conveys attention, without interrupting the *verbal auditory* channel. In human group conversation, visual attention conveyed by eye contact is a reliable indicator of whom one speaks or listens to [9]. It socially relates when it is time for a speaker to relinquish the floor, and empowers who is expected to speak next. Using eye gaze, humans achieve a remarkably efficient process of conversational turn taking. In human-to-device verbal communication, research has shown that people look at the devices they wish to communicate with before issuing spoken commands [1,4]. This is why we can use eye contact to initiate interactions with attention seeking devices. Visual attention is an often involuntary action [10], which conveys interest in the target, implicitly relating the visual search to the task oriented goal. EyePliances interpret eye contact to determine when, whether and how to attend to users [6].

Eye Contact Sensing
The eye contact sensor (Figure 1) consists of a camera that uses computer vision to find pupils in its field of view. The presence of pupils implies that a user is looking at the sensor. Unlike most commercially available eye trackers, it is cheap, unobtrusive, tolerant to user head movement and requires no calibration. By embedding eye contact sensors in household appliances and other digital devices, awareness of visual attention can be obtained without encumbering the user. Unlike Eye-R [5], eye contact sensors can determine not only that eyes exist in the cameras field of view, they can also accurately relate where in the environment the user is looking.

Figure 1. Eye Contact sensor.

CHI 2003, April 5-10, 2003, Ft. Lauderdale, Florida, USA.
ACM 1-58113-637-4/03/0004

Figure 2. Attentive TV (Photo: Stephen Wild)

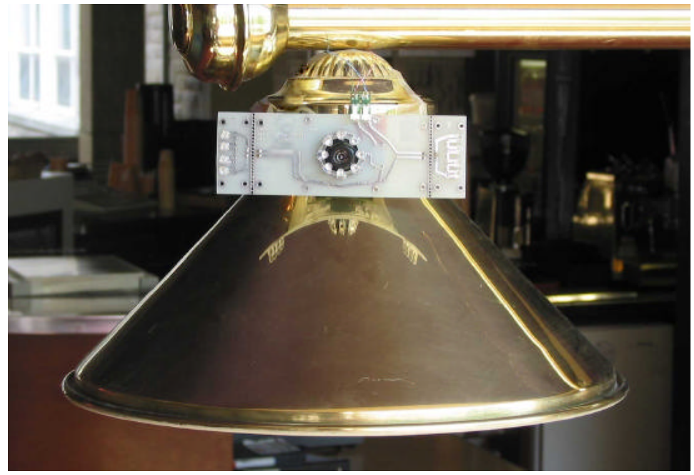

Figure 3. Light fixture with eye contact sensor.

EyePliance

By looking at an EyePliance, a user conveys attention for the device, which opens up a communication channel. A user interacts with the device using speech, remote or manual controls. By signaling attention with eye gaze, users empower devices to communicate, analogous to how participants in human group conversation encourage others to speak by making eye contact. If the device has not recently received visual attention from the user, then it must choose an unobtrusive method to signal the user (i.e. by vibrating), equivalent to a gesture or a non word utterance in human communication [6]. A device remains in the periphery of user activity until the user has acknowledged the device's request for attention. Alternately, a lack of visual attention may also constitute a meaningful event. For example, the Attentive Television (Figure 2), pauses the movie because it receives no visual attention, and concludes that nobody is watching it.

Improving Speech Recognition

By having only one device listen at a time, speech recognition is simplified because generic terms like "on" and "off" can be reused for different devices. The problem of naming devices and appliances is avoided, because the target device is implicit via visual attention in a single user scenario. Verbally naming the subject of the voice command is redundant as this information is implicitly visually communicated. In the case that multiple users are within the environment, the number of candidate devices is bounded by the number of people in the room. Colloquial, imprecise speech commands such as 'this' and 'that' can be interpreted because user eye gaze continually selects objects of interest. Eye contact sensors, embedded in EyePliances allow people to use their eyes as pointing devices, and their mouths as keyboards.

FUTURE WORK

We are currently working on expanding the range and field of view of eye contact sensors. We are also planning an intelligent environment [4] to evaluate EyePliances in ecologically valid scenarios.

CONCLUSION

We have presented EyePliances, attention-seeking devices augmented with eye contact sensors. These devices respond to visual attention as communicated by human eye contact. We have also described how speech recognition may be improved, and how a human group conversational metaphor can be applied to model interactions with such devices. By using eye gaze as an implicit attentional cue to regulate communication, more sociable and efficient interfaces may be designed.

REFERENCES

1. Maglio, P. et al. (2000) Gaze and Speech in Attentive User Interfaces. In *Proceedings of the Third International Conference on Multimodal Interfaces*. Bejing, China.
2. Morimoto , C.H. et al. Pupil Detection and Tracking Using Multiple Light Sources**.** *Image and Vision Computing*, 18 (2000) 331-335
3. Nielsen, J. Noncommand User Interfaces. *Commun ACM* 36(4), April 2003. pp. 83-99.
4. Oh, A. et al. (2002) Evaluating Look-to-Talk. In *Extended Abstracts of CHI 2002*, Minneapolis: ACM Press pp 650-651
5. Selker, T et al. (2001). Eye-aRe: a Glasses-Mounted Eye Motion Detection Interface. In *Extended Abstracts of CHI 2001* Seattle: ACM Press, pp. 179-180.
6. Shell, J., et al. Interacting with Groups of Computers. *Commun. ACM* 46(3) March, 2003.
7. Starner, T. et al. Augmented Reality Through Wearable Computing. *Presence 6(4)* April 1997 pp. 386-398
8. Vertegaal, R. et al. (2002). Designing Attentive Cell Phones Using Wearable Eyecontact Sensors. In *Extended Abstracts of CHI 2002*. Minneapolis: ACM Press, pp. 646-647.
9. Vertegaal, R., et al. (2001). Eye Gaze Patterns in Conversations. In *Proceedings of CHI 2001* Seattle: ACM Press, pp. 301-308.
10. Zhai, S. What's in the Eyes for Attentive Input. *Commun ACM* 46(3) March, 2003.