# ECSGlasses and EyePliances: Using Attention to Open Sociable Windows of Interaction

**Jeffrey S. Shell[1,2], Roel Vertegaal[1,2], Daniel Cheng[1,2], Alexander W. Skaburskis[1,2], Changuk Sohn[1,2], A. James Stewart[2], Omar Aoudeh[1] and Connor Dickie[1,2]**

[1]Human Media Lab, Queen's University,

Kingston, ON, Canada K7L 3N6
http://www.hml.queensu.ca

[2]School Of Computing, Queen's University,

Kingston, ON, Canada K7L 3N6

{shell, roel, cheng, skaburs, changuk, jstewart, aoudeh, connor} @cs.queensu.ca

## ABSTRACT
We present ECSGlasses: wearable eye contact sensing glasses that detect human eye contact. ECSGlasses report eye contact to digital devices, appliances and EyePliances in the user's *attention space*. Devices use this attentional cue to engage in a more sociable process of turn taking with users. This has the potential to reduce inappropriate intrusions, and limit their disruptiveness. We describe new prototype systems, including the Attentive Messaging Service (AMS), the Attentive Hit Counter, the first person attentive camcorder eyeBlog, and an updated Attentive Cell Phone. We also discuss the potential of these devices to open new windows of interaction using attention as a communication modality. Further, we present a novel signal-encoding scheme to uniquely identify EyePliances and users wearing ECSGlasses in multiparty scenarios.

## Keywords
Attentive User Interfaces, Ubiquitous Computing, Context-Aware Computing, Eye Tracking, Eye Contact Sensing.

## 1 Introduction and Motivation
Ubiquitous computing has brought a welcome shift in computer design [24]. The one-size fits all approach to interface design is gradually being replaced with user, task and situation specific solutions [11]. This has resulted in users having *many* digital devices, fundamentally changing our relationship with computers.

We now live *among* computers, which both mediate many of our communications, and initiate communications of their own. A person may have a mobile telephone, a personal data assistant (PDA), a desktop or a laptop computer. These devices supply phone calls, email messages, schedule reminders and instant messaging notifications, which often produce distracting notifications at inappropriate times. This fragments a user's attention, creating not only an information, but also an *interaction overload*, where users are interrupted before they have a chance to complete tasks.



In this paper, we will discuss a method to address these ubiquitous patterns of interruption. We apply the same process of turn taking inherent in human group communication to interactions with groups of computers. This allows a similarly sociable process of information exchange. Towards this end, devices must recognize the non-verbal turn taking cues humans provide, and respond in a timely and appropriate manner. Using human attentional cues to actively manage a user's attention space is fundamental to the vision of Attentive User Interfaces (AUIs). AUIs use human attention as input to drive the human computer interface. Applications include:

- Improving speech recognition by adding environmental context [17, 19]
- Conveying eye contact among participants in videoconferencing interfaces [22]
- Providing dynamic and malleable input mechanisms that are triggered by proximity [18]
- Conveying the computer's attention for a user [5]
- Coordinating requests for attention from digital devices using implicit attentional cues [16]
- Communicating a user's attentive context to remote conversants in order to foster a more natural process of human turn taking during mediated communication [16]

### 1.1 ECSGlasses
Eye Contact Sensing Glasses (ECSGlasses), the central topic of this paper, are wearable glasses that recognize whether a user is receiving eye contact from another person, or paying attention to an EyePliance. An EyePliance is a device or appliance augmented with an embedded eye contact sensor [17]. The information obtained by ECSGlasses is used to negotiate the volume and timing of interruptions by, and through digital devices. For example, as discussed further in section 4.2.2, the Attentive Messaging System (AMS) uses information obtained by ECSGlasses to communicate the wearer's availability to his or her contact list. This gives remote contacts real-time information which can be used to judge whether it is worth interrupting the recipient. It also provides the messaging service with context which is used to choose the most appropriate method to contact the user.

The encoding scheme proposed in section 5 will allow EyePliances and ECSGlasses to identify multiple users, and which people or EyePliances they are interacting with. This will facilitate a broader, and more robust knowledge of the attention negotiation that takes place between people and devices interacting in the real-world.

**Figure 1a**. LC Technologies Eye Tracker **1b**. Eye Contact Sensor Rev 1 **1c**. Eye-aRe

## 2 Attentive User Interfaces: Taking Turns for Attention

### 2.1 Introduction

The Attentive User Interface paradigm [16] uses the human group communication scenario as a guiding metaphor to model mixed interactions between humans and devices. This is because of the striking parallels between the respective coordination problems. In both situations, human attention is a limited and competitive resource. By issuing distracting interruptions, digital devices continually vie for the attention of their user. This is analogous to human group communication, where many people might simultaneously have an interest in speaking. The primary difference is that humans, more often than not, know whether and when it is appropriate to interrupt, and can sense which mechanism for interruption is least disruptive to others. If a person simply speaks over others, he would violate the implicit rules of social convention. This could result in him being labeled rude, and excluded from the group. However, unlike their human counterparts, conventional devices can not sense user activity, thus they often inappropriately intrude on the user.

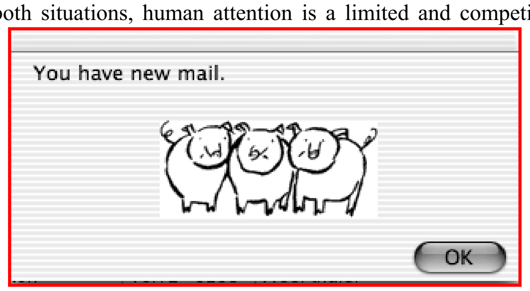


**Figure 2**. Email Notification

### 2.2 The Human Approach

Because humans are only equipped to listen to, and absorb the message of one verbal stream at a time [3], they have developed attentive mechanisms to regulate the flow of interactions. The Cocktail Party Effect allows listeners to focus their attention on a single speaker among a cacophony of conversations and background noise [3]. However, this is a cognitively expensive task and less reliable than simply allowing only one person to speak at a time. This is accomplished by using nonverbal cues to convey attention. Using these, humans achieve a sociable and efficient process of speaker exchange, or turn taking. Humans use eight nonverbal cues to determine when to speak, and when to yield the floor [17]. Of these, eye contact is the only cue that cross-culturally indicates to whom the speaker wishes to yield the floor [22]. Eye fixations provide the most reliable indication of the target of a person's attention. When a speaker falls silent, and looks at a particular listener, this is interpreted as an invitation to take the floor. Experimental research has shown that 49% of the amount of speech a person produces may be explained by the amount of eye contact she receives [21]. Because it is a nonverbal visual signal, eye contact can be used to negotiate turns without interrupting the verbal auditory channel.

If employed as a separate channel of input, eye gaze can give devices an awareness of who, or what a person is attending to. This can be used to establish less intrusive requests for attention from digital devices.

## 3. Detecting Eye Gaze

Eye gaze can be detected using a number of methods, each providing benefits in specific interaction scenarios. In this section, we will briefly describe standard eye trackers, Eye-aRe, the PupilCam and the Eye Contact sensor. Finally, we will discuss research that examines the value of eye contact in human-device interactions and describe how we have applied these results to create EyePliances.

### 3.1 Eye Tracking

Eye trackers are spatially fixed cameras, generally residing below the monitor in a standard desktop setting (see figure 1a). They report specifically where a user is looking within a highly constrained mapping area. These systems can track eye gaze at up to a ½ inch of precision at a viewing distance of approximately two feet. Because of the great accuracy of these systems, they are the most applicable solution to augment the traditional desktop experience. However, they are poor candidates for use in ubiquitous computing scenarios because they are expensive, intolerant to lateral movement, and require calibration.

### 3.2 Eye-aRe

Eye-aRe (figure 1c) is a wearable eyeglass style device developed in the Context Aware Computing Group at MIT [14]. Eye-aRe sends and receives infrared (IR) signals and measures eye movement to determine when the user is fixating. If the user is fixating, the IR receiver is polled to check whether another device or user wearing Eye-aRe glasses is emitting in its range. Eye-aRe does not detect eyes. All objects and people in the environment must have an IR transmitter attached to it in order to be recognized by the system.

### 3.3 PupilCam

The PupilCam was developed at IBM Almaden research center as part of the Blue Eyes project [10]. It exploits the property held by most mammalian eyes that light shined into an eye from a source returns in the direction of that source. Two sets of light sources synchronized with the camera clock alternate flashes. One produces a bright pupil, which is the red eye effect commonly found in flash photography, while the other does not. Eyes are detected by comparing the two images, and applying computationally inexpensive image processing.

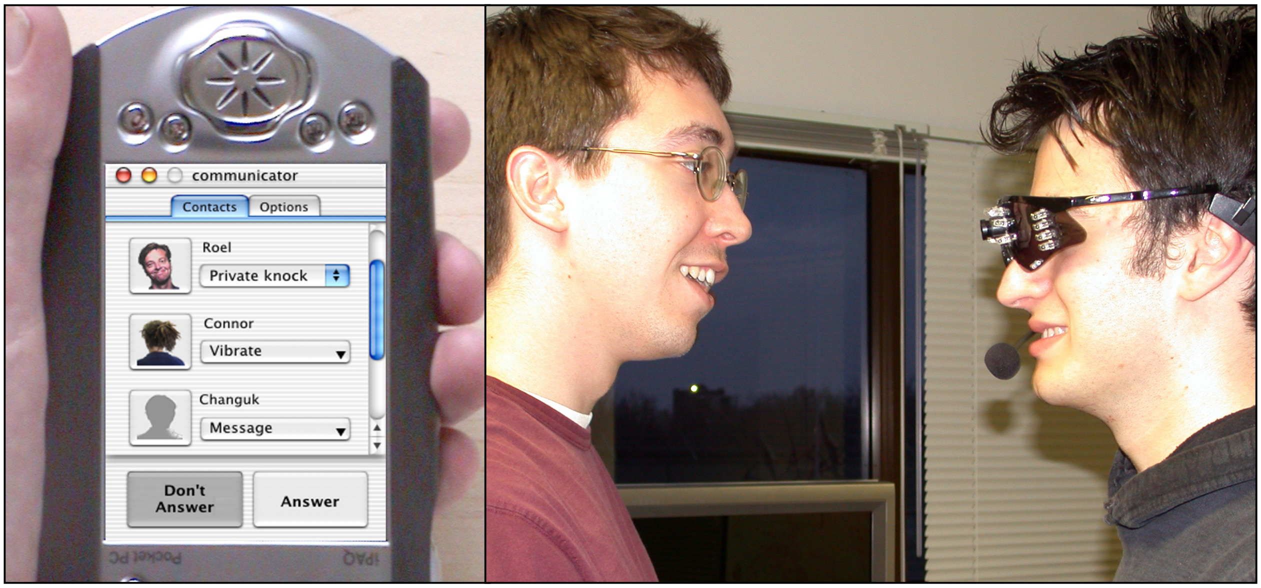


**Figure 3a**. Attentive Cell Phone. **3b**. ECSGlasses in Action

### 3.4 Eye Contact Sensor

The Eye Contact Sensor (figure 1b) was developed at the Human Media Lab at Queen's University. This system can be produced for 2% of the cost of an eye tracker. The current version works at up to a two-meter distance and is tolerant to lateral head movement. The hardware is based on the PupilCam design, but the eye contact sensor's computer vision algorithm employs a novel approach that removes the need for calibration. Eye contact sensors determine both the pupil location, as well as the location of a glint produced by the on-axis infrared light reflecting off the cornea. When this glint appears inside the pupil reflection, it means the pupil is aligned with the camera axis, signifying that someone is looking at the camera lens. When two eyes are found, and the glints do not align, the eye contact sensor recognizes presence, and attention in its general region. This is often valuable in of itself, as this tells us that the device is in the periphery of the user's attention space (see Attentive TV sec. 3.5).

With the analog ECS, the on-axis, bright pupil image is interlaced with that of the off-axis, dark pupil image to form a single interlaced NTSC source image. Extraction of this encoded information requires a deinterlacing operation, effectively halving the vertical resolution of the source image. Image subtraction between the deinterlaced images yields the pupil.

Eye Contact Sensor Rev. II, which is currently in testing, will replace the current generation eye contact sensors. The new system uses a High-Definition 2MPixel CMOS camera, capable of providing over forty times greater resolution at 10 frames per second. With the digital ECS, the source image is not interlaced. Rather, the camera controls the strobing of the on and off-axis LED's, synching their activation with the beginning of each new image frame. A sliding subtraction algorithm is used to find the difference between successive frames (e.g. |frame1-frame2|, |frame2-frame3|, |frame3-frame4|, etc.) to identify the pupil. Due to the lack of the deinterlacing operation found in the analogue version, the effective resolution of the source image is preserved. The additional resolution allows more precise detection and increased range of up to 4 meters.

### 3.5 Look-To-Talk and EyePliances

Wizard of Oz experiments at MIT [12], and IBM [8] evaluated the role of gaze in speech-enabled environments. Both studies found that subjects looked at a device before issuing a spoken command. This process is known as *Look-To-Talk*. We applied this to develop EyePliances: devices and appliances with embedded eye contact sensors. The sensors monitor the environment in their field of view, searching for eye contact. When eye contact is detected, the EyePliance responds by paying attention to the user. For speech enabled EyePliances, paying attention involves listening, or activating its speech lexicon. Thus the target of speech in single user scenarios is the device that the user is looking at. For example, the Attentive Lava Lamp [17,19] (figure 6) responds to the phrases "turn [on/off]" and "turn [this/that] [on/off]" only when it receives eye gaze from a user. By toggling the active speech vocabulary, and selecting the target of speech commands using eye gaze, speech recognition accuracy may be improved and presented in a more natural manner. Because the dynamic lexicon keeps the vocabulary small, partials and colloquialisms can also be added to increase speech recognition performance.

EyePliances can also meaningfully respond to the user given only eye contact as input. For example, as shown in figure 7, the Attentive Television [17,19] pauses its feed when nobody is watching it. Upon resumption of eye gaze, the television continues the program at the point where the viewer stopped watching. The Attentive Television will continue to play as long as the television is in the foreground or periphery of the user's attention space.

When aroused by eye contact, EyePliances open up interaction windows with the user, and respond accordingly, depending on the device's capabilities, and the interaction scenario. However, when there are many people in the room, the target of speech commands is no longer obvious. We will address this problem in section 5.

## 4. ECSGlasses

Eye Contact Sensing (ECS) Glasses are a new wearable input device that detects eye contact from other people. They are a standard pair of glasses (sun glasses as shown in figure 4), with an eye contact sensor embedded in them. ECSGlasses can determine whether someone is present and looking at the wearer within a meter distance. However, the distance will substantially increase when next generation eye contact sensors replace the current model.

Unlike the Eye-aRe approach, ECSGlasses do not simply look for IR beams transmitted from other ECSGlasses. They search for human eye gaze oriented toward the viewers eyes, which implies mutual gaze. The result is that only the wearer, not everyone in his environment, needs to have glasses to obtain useful information about his attentive status.

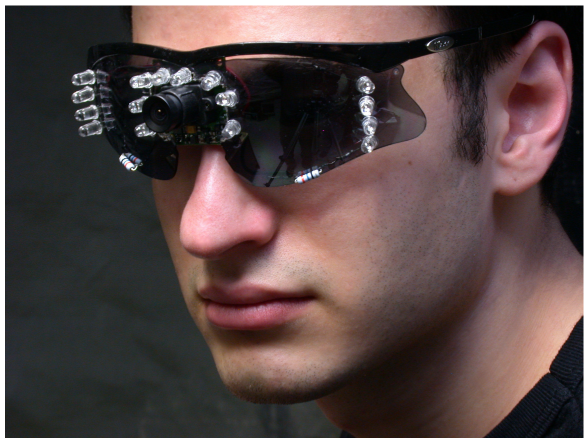

**Figure 4.** ECSGlasses

This information can be transmitted to ubiquitous devices that can determine whether the user is available for interruption. This provides a mechanism for devices to automatically switch between foreground and peripheral notification modes, depending on the user's sensed attentive context. As a result, the need for explicit input is reduced. As further described in section 4.2, instant messaging tools, and attentive cell phones can relay the user's attentive status to members of his trusted personal "buddy list". Thus both people and devices can consider the user's activities before interrupting him in mediated situations.

### 4.1 Construction

The current version of ECSGlasses has an RJ45 cable running into a compact 3.3 x 2.2 x 1.1 inch box, worn at the waist level. The box supplies power using a 12 volt DC battery, and has a Light Emitting Diode (LED) driver circuit board which alternates the flashing between the on axis and off axis LED's (see figure 5). On axis flashing produces the bright pupil effect, and off axis does not. The Image Processing Unit isolates the pupils and glints, and determines whether they are oriented at the glasses.

### 4.2 Applications

Applications for the ECSGlasses include the *Attentive Cell Phone* [23] (figure 3a) which informs potential callers when the desired recipient is involved in a conversation, and the *Attentive Messaging Service* (AMS), which communicates the user's availability based on his attentive context. Both of these applications extend the human turn taking metaphor through mediated interactions. They permit remote individuals within the user's explicitly managed trust network to use sensed attentive cues to determine when, and how to request a turn from the user.

The *Attentive Hit Counter* is essentially a vanity application that keeps track of visual attention the user has received from others. Finally, eyeBlog is a system that transmits parallax free video, capturing what the user sees, which can be used as an aid to facilitate shared mediated experiences. It can be used for collaborative group work, social purposes, and as a first person camcorder which can optionally be set to only capture interactions involving other people, as determined by eye contact.

#### 4.2.1 Attentive Cell Phone V2

Standard telephones do not provide users with any information to use in order to decide whether it is an appropriate time to interrupt the recipient and place a call. This is why many phone calls begin with questions such as "Is this a bad time for me to call?" To reduce distracting interruptions, the Attentive Cell Phone senses and communicates whether the user is involved in a face-to-face conversation, or using a personal EyePliance. This system is restricted to only sensing interactions with EyePliances which have a one to one relationship with users, such as an augmented PDA. This is because EyePliances which may have many users, for instance, the Attentive Television, can not currently identify which person is using them. This problem will be addressed in section 5.

In designing the Attentive Cell Phone, we applied the common sense notion that, all else equal, embodied interactions are more important than average priority mediated communications. By sampling both speech energy provided by a wearable headset, and incoming eye contact sensed by ECSGlasses, the system hypothesizes whether the cell phone owner is engaged in a face-to-face conversation. Unfortunately, the cell phone does not know *whom* the user is conversing with, so it is not properly equipped to strictly determine whether the incoming call is more important than the face-to-face interaction. However, even if the phone did know who was involved in face-to-face conversation, refusing phone calls on the user's behalf might not be desirable, so the Attentive Cell Phone never decides outright to block an incoming call. It does adjust the 'ring' to a less interruptive notification mode, for example vibrate, based on sensed user activity. This produces peripheral requests for attention, rather than audible demands. In the tradition of Horvitz's Priorities system [6], the percentage and frequency of calls that the user has received from the sender is used as a heuristic to estimate the priority of an incoming telephone call. This is applied to automatically adjust the ring mode of the incoming phone call.

The perceived attentive state of the user is transmitted to his personalized 'buddy' list. Recipients can use this information to decide whether, when, and how to contact the user, or whether to forward the message to the personal EyePliance in use, if supported. Buddies also have the option to force an 'override', triggering a notification of their choosing. If the call is not an explicit override, the Attentive Cell phone waits for a break in the conversation before allowing a ring. Although interruptions are still permitted, they are only triggered by people who have knowledge of the user's attentive status and have been empowered to do so by the user. We consider these 'informed' interruptions necessary, to ensure that the user is available in case of emergencies.

#### 4.2.2 Attentive Messaging Service (AMS)

The AMS provides much more detailed and immediate status information than conventional messaging clients. For example, instead of making simple inferences based on mouse and keyboard activity thresholds, presence is determined immediately, by an eye contact sensor positioned in front of the computer. An icon of the user facing 'buddies' on their screen conveys presence, and facing away relays absence, similar to the Attentive Cell phone (figure 3a).

Optionally, users can explicitly tag members of their buddy list as being interruption worthy within a certain time bound. This forces an interruptive notification, instead of the standard system of peripheral notifications. The default setting graduates the volume of the interruption request based on the previous response time to notifications from the user, the personal EyePliance the recipient is using, or the face-to-face conversation she is involved in. The AMS system also provides the facility to forward a high-priority incoming message to the personal EyePliance in use.

#### 4.2.3 Attentive Hit Counter

The Attentive Hit Counter is a system that measures the number of times somebody makes eye contact with the user. This information can be stored locally, or optionally transmitted over the internet. Once the person identification process is implemented, the functionality of the Attentive Hit Counter will improve. Mutual eye contact could trigger an exchange of contact information, identify conversational topics of mutual interest, or even act as a dating service. ECSGlasses could be modified to support an augmented reality display, where this information is exchanged in real time, or forwarded to a digital device, such as a PDA. Devices such as the Attentive Hit Counter have the potential to encourage social behavior in increasingly anonymous urban environments.

#### 4.2.4 eyeBlog

eyeBlog addresses one of the major problems in first person video narratives – where to place the camera to effectively capture the user's experiences from her vantage point. Because ECSGlasses are light and relatively unobtrusive, the user can record her experience with limited displacement from the social circumstance. Thus people speak with the videographer instead of talking to a large camera. Placing the camera between the videographer's eyes also solves the parallax problem. People are extremely sensitive to a change in the direction of an object, caused by an adjustment in the observational position that provides a new line of sight, to the extent that they will have trouble perceiving eye contact if the angle is perturbed more than 5º [2]. Thus first person filming using ECSGlasses can facilitate shared experience by literally showing the world through the videographer's eyes. To cut down on unnecessarily wasting film, in interaction mode, video footage is only taken when the videographer receives eye contact from another person.

The other main expected application of eyeBlog is as a mechanism to facilitate empathy during remote interactions. If mutual understanding of a collective problem is quickly achieved, then the efficiency of mediated group work should improve accordingly. This is expected to be especially useful in physical work situations, for example, during auto repair, where an expert can remotely relate specific instructions relative to the user's transmitted field of view.

## 5. Identifying People and EyePliances

"*Unfortunately, in many contexts, detecting that one or more people are present is not often as useful as identifying* ***which*** *people are present*." – Joseph F. McCarthy [9]

As can be seen from the applications discussed so far, determining which people or EyePliances are involved in interactions with the user would greatly improve the number and scope of scenarios these devices are equipped to handle. Although it is possible to use magnetic field, radio frequency (RF) and GPS devices, we are only interested in the relationship among people and devices, and not their absolute positioning. Thus, we aim to apply the Look-To-Talk principle to follow the user's eye gaze and uniquely detect the target. Additionally, magnetic field, and RF tracking devices are often imprecise, easily corrupted and do not scale well in real world scenarios [15] and GPS devices are not implicitly directional, and would require additional hardware infrastructure to support. In this section, we describe two approaches to identifying specific EyePliances, and ECSGlasses without changing the current infrastructure. The first approach, which is currently implemented, involves placing and detecting glyphs that identify devices. The second approach describes how we intend to use the additional frame rate provided by the next generation eye contact sensors to transmit and receive unique identification information.

### 5.1 ECSGlasses, meet EyePliances

When 'looking' at an EyePliance, ECSGlasses detect the multiple light source flashing from the Eye Contact sensor. Fortunately, the number of glints produced from Eye Contact sensors greatly exceeds the number produced by human eyes, and they follow a recognizable and predictable spatial pattern. Thus minor enhancements allow it to identify when the light source is an Eye Contact sensor, permitting EyePliances and ECSGlasses to detect each other. However, mutual detection does not imply mutual identification, so when there are multiple EyePliances and ECSGlasses wearers in a room, these systems cannot determine which person is interacting with what device or individual.

### 5.2 Glyph Detection

When an EyePliance or an ECSGlasses detects the presence of another, a glyph detection process allows mutual identification. The device searches its frame space for glyphs constructed from geometrical shapes arranged in a sequence. Unique graphical identifiers are used to differentiate between various classes of EyePliances, as well as specific devices within a given EyePliance class. Glyph detection is accomplished by analyzing each image from the input source and applying pattern recognition algorithms. The number of possible glyphs in a sequence is constrained only by the resolution of the camera, but for many applications, an impractically high resolution camera would be required to ensure uniqueness.

### 5.3 Future Directions: IDing EyePliances using IR

Using multiple light source flashing, we will be able to create enough unique identification codes to accommodate even the most aggressive applications. Cameras are now available that allow for a frame rate of up to 500 Hz, at the same resolution as our original analog eye contact sensor. The next generation of eye contact sensors will permit a tradeoff in frame rate versus resolution. Additional resolution allows more precise detection and increased range, while increased frame rate produces more responsive detection. The identification scheme which we will implement involves sending unique ID codes by toggling the flashing of on axis and off axis LED's. Referring to figure 5, the on axis flashes will be interpreted as a 1, and the off axis flashes as a 0. As a consequence of the additional frame rate, we can encode a signal using sixteen successive frames in which we only have to send one transition between on axis and off axis flashing to maintain real time performance, as real time performance can be achieved at 30 hertz. However, we still have to identify the transitions, and prime the signal because the eye contact sensors are not synchronized with each other.

The Nyquist theorem maintains that a transmitted signal can be accurately reconstructed if the receivers sampling rate is at least double that of the transmission rate. So if we sample a signal that transmits at 250 hertz at 500 hertz, halving the length of the signal, we can send eight frames while guaranteeing only one transition. Three successive off axis flashes will prime the signal, identifying to the recipient that an identification code will be transmitted. Because we use three 0's to prime the signal, identification codes must not have three consecutive 0's.

After the 3 off axis flashes have been transmitted as a primer, we still have a binary code of length five. When we account for the six cases where at least 3 0's occur in succession we are left with twenty-six unique identification codes. Because there is no limit to the length of the signal because of the no consecutive three off axis flashes clause, theoretically we can transmit ID strings of any length using the alternating flash process. However, a long code reduces the number of samples that can be taken, affecting the accuracy of the reception process. By repeatedly sampling the transmission, errors can be identified and corrected, by using the mode of the detected identification code, and dynamically refreshing the signal by discarding each sample two seconds after reception.

### 5.4 Person-to-Person Identification using ECSGlasses

If many users choose to wear ECSGlasses, then they could similarly identify each other based on the transmitted code. This would be more difficult to achieve, because both parties would be in motion, and bi directional line of sight is required. However, it does not have to be sustained line of sight. Based on the code sequence, the first bit can identify a person as opposed to a device, and the sampling procedure can vary accordingly. As digital camera technology improves, frame rates will increase, allowing longer codes to be transmitted in less time, potentially simplifying this problem and increasing the practical number of identification codes. Alternatively, a hybrid of the glyph detection and multiple light source flashing approaches could be applied to increase the length of the code and improve the robustness by adding redundancy.

## 6. DISCUSSION

Once the encoding system is successfully deployed, EyePliances will be able to detect users wearing ECSGlasses, and the ECSGlasses will likewise detect EyePliances, and other users wearing ECSGlasses in their field of view. As a result, Attentive User Interfaces will obtain a much more precise knowledge of the user's present activities, greatly improving their ability to manage a user's attention space. If, in addition, users wear wireless microphones (shown in figure 3b), deictic references in speech could be accurately interpreted by EyePliances, improving the quality and capability of speech recognition. This will allow people to use their eyes as pointing devices, and their mouths as keyboards.

### 6.1 Security and Privacy

The identification infrastructure proposed in this paper, and its usage implications, like any other new technological system, would have to be applied with due care to privacy considerations. We anticipate that of primary concern will be the stored repository of identification codes which must reside somewhere. If they are hard coded into the vision software, we will be restricted to identifying devices by type, rather than by owner. Most likely, these systems will have to consult an external source to find out who or what the code identifies. This would require a secure storage and retrieval infrastructure, which is in theory no different from standard product registration data banks.

We must also be mindful of the effect of this sort of technology on total surveillance systems [7] presented in science fiction stories such as Orwell's *1984* [13], and more recently *Minority Report* [4]. In Minority Report as a form of social control, people are identified and tracked based on the physical properties of their eyes. In principle, this tracking process this is no different from ECSGlasses transmitting identification codes from eye level. While current applications only use identification to prioritize notifications, applications such as the proposed Attentive Hit Counter will require frequent exchange of identification codes and personal data. Although people using this system can explicitly tag who has access to what information, wireless security must be a priority. We must also consider which parties would control the identification database, and question their interest in the process.

This system was, however, designed with privacy safeguards in mind. Cross referencing the user's activities with their daily schedule, and physical whereabouts were purposely excluded from the systems infrastructure, because of the problematic issue of how to share this data, and its questionable value. The actual attention management system must only know the user's activities relative to other users wearing ECSGlasses and EyePliances. Because the knowledge of whom the user is interacting with is only used locally to prioritize interactions, there is no reason to transmit specific information beyond the user's general attentive status to anyone, including to the user's trusted 'buddy' list.

The infrastructure, as described, can also be used to promote safety and security within the household. Access to potentially dangerous appliances can be restricted to adults by checking the identification code. This can be validated by verifying that height and voice correspond to the rightful owner of the device. Additionally, personal data security can be refined by requiring identification to access email. Personal preferences such as desktop interface presentation and favorite television program listings can be loaded upon presence by various EyePliances. Presence can also be used with contact list knowledge to automatically redirect telephone calls, and only allow the relevant phone to ring instead of disrupting the whole household. Collective preferences, such as the musical taste of all people in the room can be used to choose which songs to play, similar to the MusicFX system [9].

Although we have attempted to proactively address privacy considerations, latent applications of the surveillance required to effectively manage interruption requests that we have not anticipated will arise. We hope to continually learn from prototyping and evaluating these systems in order to better address human factors in ubiquitous computing environments in the future.

### 6.2 Future Work

We are currently investigating a next generation eye contact sensor which will permit the identification encoding scheme discussed in section 5. We are also expanding the potential of the many devices and services we are currently prototyping. We also intend to improve the portability of ECSGlasses by adding an onboard miniature computer to handle image processing on site and communicate with other devices using either Bluetooth, or WiFi. As wireless Internet connectivity becomes increasingly ubiquitous in public spaces, we will have greater opportunities to test our prototypes in real world scenarios.

## 7. CONCLUSIONS

In this paper, we presented ECSGlasses: wearable eye contact sensing glasses that detect and report human eye contact. Interfaces that recognize and respond to a user's attentive cues using EyePliances and ECSGlasses have great potential to address the ubiquitous patterns of interruptions users are increasingly faced with. By recognizing the non-verbal attentional cues provided in human-to-human group interactions, and responding accordingly, we are developing interfaces that can engage in a more sociable and efficient process of turn taking with users.

The proposed identification encoding scheme will provide Attentive User Interfaces with rich information to use in order to mediate interactions between people and groups of computers. It also provides a mechanism to greatly augment the scope and functionality of the prototype attention sensitive applications described in this paper. Using EyePliances and ECSGlasses, we wish to open up new windows of interaction with users, reducing inappropriate intrusions, improving communication coordination from devices, and multiparty interactions in untethered environments.

### Acknowledgements

We would like to thank all of the students who comprise the Human Media Lab. In particular, our main engineer on the ECS project, Changuk Sohn.

## 8. REFERENCES

1. Argyle, M. and M. Cook. Gaze and Mutual Gaze. London: Cambridge University Press, 1976.
2. Chen, M. Leveraging the Asymmetric Sensitivity of Eye Contact for Videoconferencing. In *Proceedings of CHI 2002* Minneapolis MN: ACM Press, 2002, pp. 49-56.
3. Cherry, C. Some Experiments on the Reception of Speech with One and with Two Ears. *Journal of the Acoustic Society of America* 25, 1953, pp. 975-979.
4. Dick, P. K. The Minority Report. Pantheon Books, 2002.
5. Jabarin, B., Wu, J.,Vertegaal, R., and L. Grigorov. Establishing Remote Conversations Through Eye Contact With Physical Awareness Proxies. In *Extended Abstracts of CHI 2003* Ft.Lauderdale: ACM Press, 2003, pp. 948-949.
6. Horvitz, E., Jacobs, A., and D. Hovel. Attention-sensitive alerting. In *Proceedings of UAI'99*. Stockholm: Morgan Kaufmann, 1999, pp. 305-313.
7. Lyon, D. The Electronic Eye: *The Rise of Surveillance Society*. Minneapolis: University of Minnesota Press, 1994.
8. Maglio, P., Matlock, T., Campbell, C., Zhai, Z., Smith, B. Gaze and speech in attentive user interfaces. In *Proceedings of the International Conference on Multimodal Interfaces*. Berlin: Springer-Verlag, 2000.
9. McCarthy, J. F. Active Environments: Sensing and Responding to Groups of People. *Personal and Ubiquitous Computing* Vol. 5 No. 1 (February 2001), Springer-Verlag, London, pp. 75-77.
10. Morimoto, C. H., D. Koon, A. Amir, and M. Flickner. Pupil detection and tracking using multiple light sources. *Image and Vision Computing* 18, 2000, pp. 331-334.
11. Norman, D. A. The Invisible Computer. Cambridge: MIT Press, 1998.
12. Oh, A., Fox, H., Van Kleek, M., Adler, A,. Gajos, K., Morency, L-P, and T. Darrell. Evaluating Look-to-Talk: A gaze-aware interface in a collaborative environment. In *Extended Abstracts of CHI 2002*. Seattle: ACM Press, 2002, pp. 650-651.
13. Orwell, G. 1984. London: Penguin, 1989.
14. Selker, T, Lockerd, A. and J. Martinez. Eye-R, a Glasses-Mounted Eye Motion Detection Interface. In *Extended Abstracts of CHI 2001*. Seattle: ACM Press, 2001. pp. 179-180.
15. Schmidt, A. and K. Van Laerhoven. How to Build Smart Appliances. *IEEE Personal Communications* Vol. 8 No. 4 (August 2001). pp. 66-71.
16. Shell, J. S., Selker, T., and R. Vertegaal. Interacting with Groups of Computers. *Communications of the ACM* Vol. 46 No. 3 (March 2003), ACM Press, New York. pp. 40-46.
17. Shell, J. S., Vertegaal, R. and A. Skaburskis. EyePliances: Attention-Seeking Devices that Respond to Visual Attention In *Extended Abstracts CHI* 2003 Ft. Lauderdale: ACM Press, 2003. pp. 770-771.
18. Shell, J. S., Bradbury, J., Knowles, C., Dickie, C., and R. Vertegaal. eyeCOOK: A Gaze and Speech Enabled Attentive Cookbook. In *Extended Abstracts of* UbiComp 2003. Seattle: Springer-Verlag, 2003.
19. Shell, J. S., Vertegaal, R., Mamuji, A., Pham, T., Sohn, C. and A. Skaburskis. EyePliances and EyeReason: Using Attention to Drive Interactions with Ubiquitous Appliances. In *Extended Abstracts of UIST 2003*. Vancouver: ACM Press, 2003.
20. Short, J., Williams, E., and B. Christie. The Social Psychology of Telecommunications. London: Wiley, 1976.
21. Vertegaal, R. and Ding, Y. Explaining Effects of Eye Gaze on Mediated Group Conversations: Amount or synchronization? In *Proceedings of CSCW 2002*. New Orleans: ACM Press, 2002. pp. 41-48.
22. Vertegaal, R. The GAZE Groupware System: Mediating joint Attention in Multiparty Communication and Collaboration. In *Proceedings of CHI'99*. Pittsburgh: ACM Press, 1999. pp. 294-301.
23. Vertegaal, R., Dickie, C., Sohn, C. and M. Flickner. Designing Attentive Cell Phones Using Wearable Eyecontact Sensors. In *Extended Abstracts of CHI 2002*. Minneapolis: ACM Press, 2002. pp. 646-647.
24. Weiser, M. The Computer for the 21st Century. *Scientific American* 265(3), 1991, pp. 94-104.