# Empathy Glasses

**Katsutoshi Masai**
Keio University
Yokohama, Japan
masai@imlab.ics.keio.ac.jp

**Maki Sugimoto**
Keio University
Yokohama, Japan
sugimoto@ics.keio.ac.jp

**Kai Kunze**
Keio University
Yokohama, Japan
kai@kmd.keio.ac.jp

**Mark Billinghurst**
University of South Australia
Adelaide, Australia
Mark.Billinghurst@unisa.edu.au

 CHI'16 Extended Abstracts, May 07-12, 2016, San Jose, CA, USA
ACM 978-1-4503-4082-3/16/05. http://dx.doi.org/10.1145/2851581.2892370

## Abstract
In this paper, we describe Empathy Glasses, a head worn prototype designed to create an empathic connection between remote collaborators. The main novelty of our system is that it is the first to combine the following technologies together: (1) wearable facial expression capture hardware, (2) eye tracking, (3) a head worn camera, and (4) a see-through head mounted display, with a focus on remote collaboration. Using the system, a local user can send their information and a view of their environment to a remote helper who can send back visual cues on the local user's see-through display to help them perform a real world task. A pilot user study was conducted to explore how effective the Empathy Glasses were at supporting remote collaboration. We describe the implications that can be drawn from this user study.

## Author Keywords
Remote Collaboration; Emotional Interface; Wearables; Facial Expression

## ACM Classification Keywords
H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous

## Introduction

Empathy is defined by Alfred Adler as "seeing with the eyes of another, listening with the ears of another, and feeling with the heart of another" [1]. In this paper, we present Empathy Glasses, a head worn prototype designed to create an empathic connection between remote collaborators. The glasses send to a remote helper a first person view from a local user with facial expression and gaze information. A remote helper has a pointer that is visible to a local user via a head-worn display. The goal of this research is to enable a remote person to indeed see and hear from another person's perspective and especially to understand how they are feeling. Although an early prototype, our research makes the following contributions; (1) the first interface to combine facial expression analysis and gaze tracking with a see-through Head Mounted Display (HMD), (2) the first user study of the effect of wearable facial expression sharing on remote collaboration, and (3) a set of research implications.

## Related work

Our research is based on previous work in the areas of emotion detection, remote collaboration, and wearable computing. Combining this research enables us to develop glasses that enable a user to share their emotions remotely.

Since Picard developed the concept of Affective Computing [11], many researchers have worked on detecting affect using facial expression, voice, body language and posture, physiological cues and so on. Calvo and D'Mello [3] provide an excellent summary of various methods for detecting affect, showing for example, that facial expression tracking is becoming a reliable way of detecting a user's emotions.Our earlier AffectiveWear glasses used photo reflective sensors mounted around a glasses frame to reliable recognize a number of facial expressions in an unobtrusive manner [9]. However this research focussed on recognizing a single user's emotional display and not on creating a shared experience.

Recently research has found that sharing emotion can improve performance on remote collaboration tasks. Eligio et al. [15] showed that remote collaborators who shared their emotions improved their understanding of each other's emotions and had a more positive experience during computer supported collaborations. Similarly Molinali et al. [10] found that emotion sharing was positively correlated to the perceived intensity of positive emotions after collaboration.Tan et. al. [14] developed a video conferencing system that automatically shared physiological cues with a remote user and found this significantly increased the positive affect score compared to audio only conferencing. However this research was all done on desktop systems and did not support remote shared viewing. The remote users were not able to see what the local users were seeing in their workspaces.

A key part of creating an empathic connection is being able to see from another person's perspective. There have been many studies that exploreed the use of head worn cameras (HWC) for remote collaboration, such as [2], [5]. This enables one user to see what another is doing, and when combined with a head mounted display (HMD), allows the remote user to provide visual feedback. For example, Bauer et. al. [2] showed how a HWC and HMD could be combined with remote pointing to significantly improve collaboration. However, these systems don't combine a head worn eye tracker with a HMD and HWC, so it is difficult to know exactly where the user is looking in the video sent from the HWC.

Fussell et al [6] have developed a system with a HWC with an attached eye tracker, which sent real time workspace video along with the user's eye gaze details to the mon-

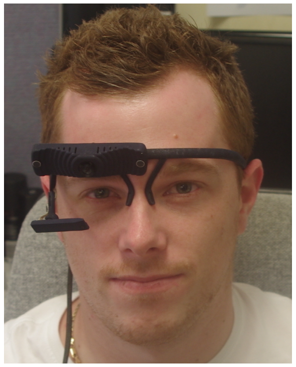

**Figure 1:** Pupil Hardware

**Figure 2:** Sensor Placement

**Figure 3:** AW Module Mounted on BT-200

itor of a remote helper. In a user study comparing using the HWC and eye-tracker to a wide-angle scene camera, the results showed a clear value for the scene camera, but no benefit from the HWC with eyetracking. However in this case the user with the HWC was not wearing a HMD and so the remote user could only provide audio feedback. To the best of our knowledge there have been no papers published that describe a system that combines a HMD, HWC and an eye tracker together.

This previous research shows that it is possible to detect user emotion, that sharing emotion improves collaboration, and that wearable technology can be used to share remote views. In our work we combine these three threads together to develop a wearable system that will share what a person is looking at and how they are feeling. The main novelty of our system is that it is the first to combine the following technologies together: (1) wearable facial expression capture hardware, (2) eye tracking, (3) head worn camera, and (4) see-through head mounted display, with a focus on improving remote collaboration.

## Prototype System

Our system consists of two components, (1) wearable hardware that captures and sends the wearer's viewpoint, gaze and facial expression and displays a visual feedback from a remote helper, and (2) a remote interface where the information sent is viewed and visual feedback is provided to local user by a remote helper.

*Hardware:Wearable System*

The wearable system combines three hardware systems; (1) Pupil eyetracker [8] (2) the AffectiveWear (AW) facial expression tracker [9], and (3) the Epson Moverio BT-200 head mounted display [4]. Together these systems allow the user's view, gaze point and facial expression to be sent to a remote helper, and virtual cues from the remote helper to be sent back.

*(1) Pupil (Gaze Information, HWC)*

Pupil is an open source platform for pervasive eye tracking and mobile gaze-based interaction. It uses two cameras: an eye camera to track the user's right eye gaze and a scene camera (HWC) to capture the user's view. The camera views are sent via USB to a desktop computer that runs the eyetracking software. The Pupil hardware can track the eye gaze with $0.6^\circ$ accuracy, at 120hz capture rate, and has a full HD scene camera with 5.7ms latency (figure 1).

*(2) AW module (Facial expression information)*

The AffectiveWear (AW) module is based on our previous work [9]. It uses a photo reflective sensor array to recognize the facial expressions of the wearer. It consists of an Arduino Fio, eight photo reflective sensors (SG-105), a transistor (IRLU3410PBF), Xbee, and li-po battery. The module is taped to a left lens of Epson BT-200 HMD with four sensors on top of the display and the other four sensors are placed below the display (see figure 2, 3). We use photo reflective sensors because the size is small enough to fit on a wearable device, it is unobtrusive, and the processing is fast enough for real-time prediction.

The AW Arduino module classifies four facial expressions (Neutral /Positive /Negative /Surprise) based on the signals coming from the photo reflective sensors. The AW module uses skin deformation in the area around the left eye caused by facial expression change. The sensors measure the distance between the module and the skin surface on the face. The data obtained from the sensors is used by a machine learning algorithm to classify the four facial expressions. Each user needs to calibrate and register a posed facial expression for each emotional label. The AW module sends the local user's facial expression to a remote

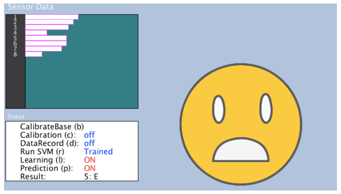

**Figure 4:** UI(Facial Expression)

**Figure 5:** Hardware being worn

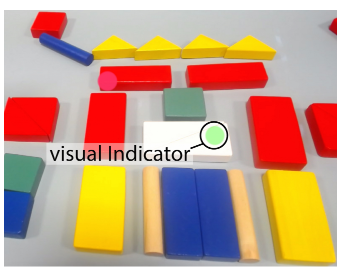


**Figure 6:** The Green Circle Indicator in the Pupil Software

interface (see figure 4). Data from the AW module is sent wirelessly via Xbee, at around 170 frames per second.

*(3) Epson Moverio BT-200 (HMD)*
The BT-200 is a commercial smart glass that has a stereo optical see-through display.In this case we are using the BT-200 to display a visual pointer from the remote helper.

The local user wears both the Pupil eye tracking hardware and the BT-200 (with AW sensors) at the same time, simulating a single integrated see-through display with eye-tracking ability.Figure 5 shows the hardware being worn.

*Software:Remote Interface*

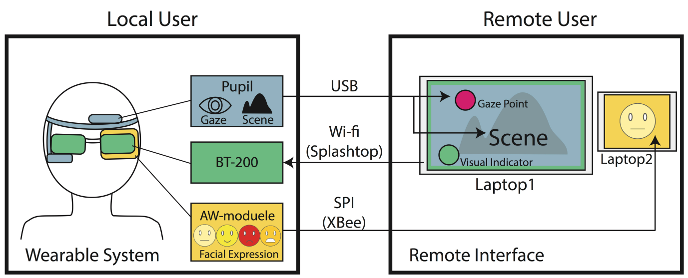


**Figure 7:** System Diagram

The hardware components are connected to two computers as shown in figure 7. The left computer (Laptop 1) shows the Pupil interface. This provides the local user's live scene camera view with his/her gaze information shown as a red circle superimposed on top of it. We modified the default Pupil interface to show visual pointer (green circle) following the user's mouse input (see figure 6). With this circle pointer, the remote helper can provide visual pointing feedback to the local user's see-through display BT-200 via the Splashtop software [13] at 30fps. The right laptop 2 shows a visualization of the local user's facial expression recognized by the AW module. This allows the remote helper to have an idea of the local user's emotional state. Since the Splashtop software can only send an entire screenit was necessary to use two laptops as we did not want local user to see their facial expression view.

## User Study

We conducted an initial pilot study inspired by some of the earlier work in HMD and HWC remote collaboration systems [2], [5]. The goal was to explore how the Empathy Glasses contributed to remote collaboration. We tested four different interface conditions:

- **V**: A video only condition, in which the remote user can only see video from the local user.
- **P**: Video plus pointer condition, which is the same as the V condition with the addition of gaze cues and pointing on the HWC video.
- **E**: Video plus expression condition, which is the same as the V condition with the addition of the facial expression monitor.
- **A**: All condition which adds both pointing, gaze tracking and facial expression to the V condition.

The users were asked to work together to construct 2D pictures of various objects out of wooden blocks. This is similar to earlier physical construction tasks used in remote collaboration studies [5]. The target objects included a sports car, castle, cruise liner and animal.

A within-subjects design was used where pairs of users would use each of the four different interface conditions with a different object. The order of the conditions and the objects were counterbalanced to reduce any order effects. Subject pairs were given five minutes to construct a picture for each condition and were told that they should try and use as many of the blocks as possible. The subjects

**Figure 8:** Experimental Set Up

(1) My partner and I worked well together on the task.
(2) It was easy to be aware of what my partner was doing.
(3) I felt connected with my partner.
(4) My partner and I communicated together well.
(5) I understood how my partner was feeling.
(6) My partner understood how I was feeling.
(7) I was satisfied with the output of the task.

**Figure 9:** The questions

sat side by side at different tables in the same room with a divider between them (see figure 8). They meant that they could talk to each other normally while sharing a remote view of the workspace.

Before the experiment began the subject wearing the head mounted hardware had to have calibration completed for their eye gaze and facial expression settings. After each condition they were asked a number of Likert scale questions shown in figure 9. These were asked on a scale of 1 to 7, where 1 = strongly disagree and 7 = strongly agree. After all the conditions were over they were also asked to rank each interface according to how well they communicated with their partner, and worked together, etc. These questions were taken from earlier research on remote collaboration with HMD [7],[12]. Observations of subject behavior was made and the subjects were interviewed after the experience.

## Results

A total of 5 pairs of subjects (6 men, 4 women) completed the pilot test with an age range of 20 - 45 years old. The subject pairs knew each other as friends or work colleagues and so collaborated together easily and there are no critical difference between all pairs.

Overall, subjects had no trouble completing the object construction task in the time allocated. Figure 9 shows the questions asked after they completed the task. There was no significant difference in the average Likert scale scores for each of the conditions for the questions. However there was a significant difference in the results of the forced ranking questions (see figure 10). After all the conditions were complete, subjects were asked to rank the four conditions in order from best (1) to worst (4) in response to the following questions; (Q1) Which condition did you work best with your partner in, (Q2) Which condition did you feel that you communicated best with your partner in, and (Q3) Which condition did you feel that you understood best how your partner was feeling. Figure 10 shows the average rankings for each condition (1 = best, 4 = worst).

A Friedman test was used and even with only five pairs of subjects significant differences were found. There was a significant difference between rankings by the local users (HMD) for Q2 ($\chi^2(3) = 8.3$, $p < 0.05$)) and near significance for the remote helpers(Comp) ( $\chi^2(3) = 7.3$, $p = 0.06$)). Similarly there was a significant difference between rankings by the local users for Q3 ($\chi^2(3) = 8.3$, $p < 0.05$)) and for the remote helpers ($\chi^2(3) = 9.2$, $p < 0.05$)). Finally there was a near significant difference in results for Q1 for the local users ($\chi^2(3) = 6.4$, $p = 0.09$)) and the remote helpers($\chi^2(3) = 5.9$, $p = 0.12$)).

## Implications of the Work

Although we have only completed an initial pilot test the results are promising. There are several implications drawn from the work: (1) In wearable collaborative systems it is important to provide a means for bi-directional visual communication. (2) Gaze cues can be used to establish shared understanding and confirm that users are referencing the same objects. (3) facial expression tracking can be used as an implicit cue to show comprehension.

Most interestingly, although the expression (E) and pointing (P) conditions were not rated particularly highly for communication (Q2), the combination of these two conditions (A) was extremely highly ranked, with almost every user rating this as the best. This may be because of the different communication channels offered by each modality. One remote helper stated "I ranked the (A) condition best, because I could easily point to communicate, and when I needed it I

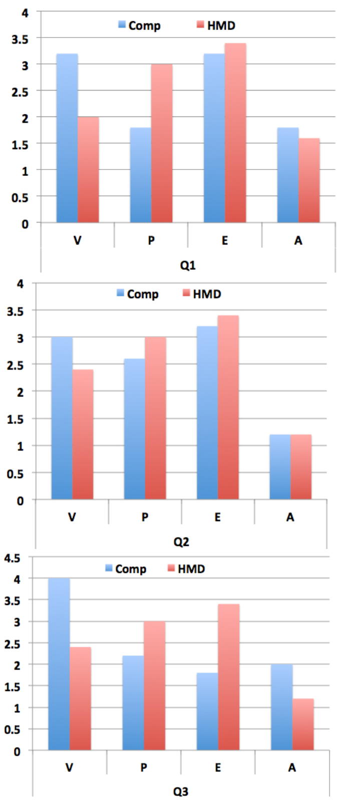


**Figure 10:** Average rankings for each of the ranking questions.

could check the facial expression to make sure I was being understood.". Also, it is interesting to note that (A) condition was highly ranked in how well user felt they understood their partner (Q3) by local users even though they prefer (V) to (E) or (P). This may suggest the synergy of using emotional feedback and a visual pointer at the same time.

Observations were also made of user behavior. In the Video only condition (V) users talked more and tended to describe the colour and shapes of the blocks to be moved. In conditions that supported pointing (P, A) the remote helper used more deictic language, for example pointing at a block and saying "Move this one". In conditions (E, A), the remote helper typically looked at the feedback less than 10% to 20% of the time. It was difficult for them to pay attention to it while watching the user actions on the HMC video footage.

After the experiment subjects were interviewed to understand their experience further. Their favorite conditions were remote pointing (P, A). Some remote helpers felt that the gaze cue was not so informative because they could know what blocks the local user was going to manipulate by looking at their hands. On the other hand, one remote user said that the gaze cue was useful because it showed the context of what the local user was talking about. For the remote helper, the pointer made them feel more connected to their partner. One remote helper said that "When I was pointing it felt like there was something for me to do", while one local user said that seeing the remote pointer made them feel like "There was a second pair of eyes helping me with the task".

However there are some limitations with this study. The facial expression viewer and HWC video interface were shown on separate screens forcing the remote helper to split their attention. The facial expression recognizer was trained on static facial expressions and so does not work so well when a user is having a conversation and their face moves through a range of different positions. In the future we could capture a wider range of physiological cues associated with emotion (e.g. heart rate, audio pitch tracking, etc). Finally, the task chosen was a construction task, which may not cause as much emotional display as other tasks such as negotiation or story telling. In the future we will explore a wider range of collaboration tasks.

## Conclusion

In this paper we have described the concept of Empathy Glasses that can enable a user to see, hear and feel from another person's perspective. We have combined several pieces of hardware and software together to create a rough prototype and tested it in an initial user study. The main novelty is combining together technology that allows a user to share their point of view, gaze information and facial expressions with a remote collaborator. The results are promising with subjects rating the combined interface (A) as the condition in which they felt that they communicated best with their partner in (Q2:Which condition did you feel that you communicated best with your partner in), and where they could best understand how their partner was feeling (Q3: Which condition did you feel that you understood best how your partner was feeling). This shows that it could be extremely valuable to further explore how gaze and emotional cues could be used to enhance head worn collaborative systems.

## Acknowledgements

This research is supported by JST, CREST.

## References

[1] Heinz L Ansbacher and Rowena R Ansbacher. 1956. The individual psychology of Alfred Adler. (1956).
[2] Martin Bauer, Gerd Kortuem, and Zary Segall. 1999.

"Where Are You Pointing At?" A Study of Remote Collaboration in a Wearable Videoconference System. In *Proceedings of the 3rd IEEE International Symposium on Wearable Computers (ISWC '99)*. IEEE Computer Society, Washington, DC, USA, 151–. http://dl.acm.org/citation.cfm?id=519309.856501

[3] R.A. Calvo and S. D'Mello. 2010. Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications. *Affective Computing, IEEE Transactions on* 1, 1 (Jan 2010), 18–37. DOI: http://dx.doi.org/10.1109/T-AFFC.2010.1

[4] EPSON. 2014. Moverio BT-200. Website. (2014). http://www.epson.com/moverio

[5] Susan R. Fussell, Robert E. Kraut, and Jane Siegel. 2000. Coordination of Communication: Effects of Shared Visual Context on Collaborative Work. In *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work (CSCW '00)*. ACM, New York, NY, USA, 21–30. DOI: http://dx.doi.org/10.1145/358916.358947

[6] Susan R. Fussell, Leslie D. Setlock, and Robert E. Kraut. 2003. Effects of Head-mounted and Scene-oriented Video Systems on Remote Collaboration on Physical Tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03)*. ACM, New York, NY, USA, 513–520. DOI: http://dx.doi.org/10.1145/642611.642701

[7] Anders Henrysson, Mark Billinghurst, and Mark Ollila. 2005. Face to Face Collaborative AR on Mobile Phones. In *Proceedings of the 4th IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR '05)*. IEEE Computer Society, Washington, DC, USA, 80–89. DOI: http://dx.doi.org/10.1109/ISMAR.2005.32

[8] Moritz Kassner, William Patera, and Andreas Bulling. 2014. Pupil: An Open Source Platform for Pervasive Eye Tracking and Mobile Gaze-based Interaction. (April 2014). http://arxiv.org/abs/1405.0006

[9] Katsutoshi Masai, Yuta Sugiura, Masa Ogata, Katsuhiro Suzuki, Fumihiko Nakamura, Sho Shimamura, Kai Kunze, Masahiko Inami, and Maki Sugimoto. 2015. AffectiveWear: Toward Recognizing Facial Expression. In *ACM SIGGRAPH 2015 Emerging Technologies (SIGGRAPH '15)*. ACM, New York, NY, USA, Article 4, 1 pages. DOI: http://dx.doi.org/10.1145/2782782.2792495

[10] G Molinari, G Chanel, M Bétrancourt, T Pun, and C Bozelle. 2013. Emotion Feedback during Computer-Mediated Collaboration: Effects on Self-Reported Emotions and Perceived Interaction. In *10th International Conference on Computer Supported Collaborative Learning*. Madison, Wisconsin, USA.

[11] Rosalind W. Picard. 1997. *Affective Computing*. MIT Press, Cambridge, MA, USA.

[12] Eva-Lotta Sallnäs. 2005. Effects of communication mode on social presence, virtual presence, and performance in collaborative virtual environments. *Presence - Teleoperators and Virtual Environments* 14, 4 (2005), 434–449. QC 20100630.

[13] Splashtop. 2015. Splashtop. Website. (2015). http://www.splashtop.com/personal

[14] Chiew Seng Sean Tan, Kris Luyten, Jan Van Den Bergh, Johannes Schöning, and Karin Coninx. 2014. The Role of Physiological Cues During Remote Collaboration. *Presence: Teleoper. Virtual Environ.* 23, 1 (Feb. 2014), 90–107. DOI: http://dx.doi.org/10.1162/PRES_a_00168

[15] Ulises Xolocotzin Eligio, Shaaron E. Ainsworth, and Charles K. Crook. 2012. Emotion Understanding and Performance During Computer-supported Collaboration. *Comput. Hum. Behav.* 28, 6 (Nov. 2012), 2046–2054. DOI: http://dx.doi.org/10.1016/j.chb.2012.06.001