ubiGaze: Ubiquitous Augmented Reality Messaging Using Gaze Gestures

Mihai Bâce
Department of Computer Science
ETH Zurich
mihai.bace@inf.ethz.ch

Teemu Leppänen
Center for Ubiquitous Computing
University of Oulu
teemu.leppanen@ee.oulu.fi

Argenis Ramirez Gomez
School of Computing and Communication
Lancaster University
a.ramirezgomez@lancaster.ac.uk

David Gil de Gomez
School of Computing
University of Eastern Finland
dgil@uef.fi

Abstract

We describe ubiGaze, a novel wearable ubiquitous method to augment any real-world object with invisible messages through gaze gestures that lock the message into the object. This enables a context and location dependent messaging service, which users can utilize discreetly and effortlessly. Further, gaze gestures can be used as an authentication method, even when the augmented object is publicly known. We developed a prototype using two wearable devices: a Pupil eye tracker equipped with a scene camera and a Sony Smartwatch 3. The eye tracker follows the users' gaze, the scene camera captures distinct features from the selected real-world object, and the smartwatch provides both input and output modalities for selecting and displaying messages. We describe the concept, design, and implementation of our real-world system. Finally, we discuss research implications and address future work.

Keywords: Augmented Reality; Eye Tracking; Gesture Interaction; Gaze Gestures; Messaging

Concepts: •**Human-centered computing** → **Mixed / augmented reality;** *Ubiquitous and mobile devices;*

1 Introduction

Augmented Reality (AR) enables the direct or indirect view of a physical, real-world environment whose elements are augmented by a computer. This is an important paradigm as it allows us to enrich our physical world with digital content without having to alter the environment.

While getting smaller, cheaper and interconnected, computing technology not only pervades physical objects, but comes closer to humans. This shift in technology has materialized mostly due to wearable devices of various forms. Nowadays, due to advances in technology and manufacturing, many companies are involved in releasing new iterations of their wearables, ranging from smartwatches to fitness trackers, smartglasses including eye trackers, and even virtual reality or AR headsets. Wearable devices have one particular aspect that make them so attractive. They offer an egocentric perspective: smart glasses can see what we see and smart watches know how we move [Mayer and Sörös 2014]. We expect that in the close future, smartwatches and smartglasses will become ubiquitous and together with the smartphone they will form a universal user interface. Compared to the already classical mobile devices and smartphones, wearable devices offer new interfaces which are more suitable and more intuitive for AR applications.

We present *ubiGaze*, a wearable AR system that enables users to augment any real-world object with invisible messages (Figure 1). With our system, users are able to embed context dependent invisible messages into any real object using eye-gaze gestures. Gaze gestures are deliberate and unnatural movements of the eye that follow a specific pattern. By fixating at an object, users select their intended object and then, through gaze gestures, they are able to embed a message into that object. Message recipients receive the message by looking at the same object and performing the same gaze gesture. This provides a discrete and effortless novel interaction technique for AR, extending the concept of annotations or tags to a messaging service. The novelty of this work lies in the coupling of these three elements together: any gaze gesture, any real-world object, and a message. We consider a real-world scenario, where users are equipped with two wearable devices: an eye tracker and a smartwatch, which enable this method.

2 Related Work

Some researchers have defined AR as a paradigm that requires Head Mounted Displays (HMD). Ronald Azuma mentions in his survey [Azuma 1997] that AR should not be limited to certain technologies and further defines three characteristics: AR combines real and virtual, it is interactive in real time, and it is registered in three dimensions. The author mentions AR as an interesting topic because it enhances the user's perception of and interaction with the world. Further, it is augmented with information that users cannot see with their own senses. Ronald Azuma defines six classes of potential AR applications that have been explored: medical visualization, maintenance and repair, annotation, robot path planning, entertainment, and military applications. Almost 20 years later, a new study [Billinghurst et al. 2015] gives an overview of the developments in the field. While the taxonomy for applications has changed, the authors focus on marketing, education, entertainment, and architecture scenarios, we can see that there is still an ever growing interest in enhancing normal, everyday objects with additional information.

The annotation problem for AR has been studied before. We present some of the existing approaches in this direction. WUW - Wear Ur World [Mistry et al. 2009] is a wearable gestural interface that allows projecting information out into the real world. It is a wearable system composed of a projector and a tiny camera, which allows the system to see what the user sees and through projection, information can be displayed on any surface, walls, and physical objects around us. SixthSense [Mistry and Maes 2009] further ex-

SA '16 Symp on Mobile Graphics and Interactive Applications, December 05-08, 2016, Macao
ISBN: 978-1-4503-4551-4/16/12...$15.00
DOI: http://dx.doi.org/10.1145/2999508.2999530

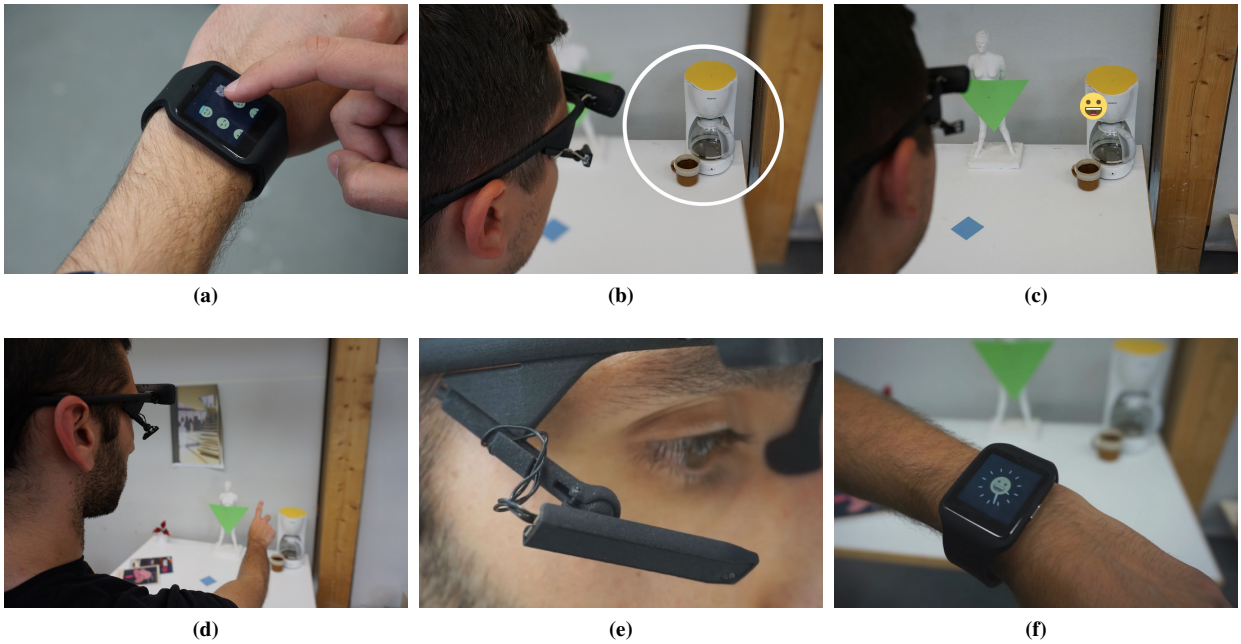

Figure 1: *Application scenario. (a) Users select a message from their smartwatch; (b) Users select the real-world object, which they want to augment with a message. (c) Users perform a gaze gesture to lock the message into the object. (d) A peer comes and fixates on the same object. (e) The peer performs the corresponding gaze gesture to unlock the message from the object. (f) The message will become visible on the peer's smartwatch.*

tends the previous concept and enables users to interact with natural hand gestures by manipulating the augmented information. The Ambient aNnotation System [Quintana and Favela 2013] is another AR annotation system that aims at assisting persons suffering from Alzheimer's disease. Caregivers can create ambient tags, while patients can use a mobile device to recognize tags in their environment and retrieve the relevant information. SkiAR [Fedosov et al. 2016] is a sports oriented AR system with a focus on skiing and snowboarding. Their system allows users to share rich content in-situ on a printed resort map while on the slope. SkiAR has two components: an input device (a smartwatch) and an output device (a mobile phone used as a display). [Nuernberger et al. 2016] have developed a method to display 2D gesture annotations in 3D augmented reality. The authors do not focus on recognizing objects and embedding virtual tags, but rather allowing users to draw annotations in a 2D environment and display them overlaid on real 3D objects.

An approach similar to ours is Tag It! [Nassani et al. 2015]. The authors describe a wearable system that allows users to place 3D virtual tags and interact with them. While the general goal of annotating real objects with virtual messages is the same, there are two significant differences that set our works apart. First, there is a difference in the underlying method. [Nassani et al. 2015] propose a method based on 3D indoor tracking with a chest-worn depth sensor and a head-mounted display, while in our approach we have developed a method based on object recognition using a head-mounted eye tracker, equipped with a scene camera, and a smartwatch. Our approach does not require any tracking. Second, we propose an extension to annotation (or tagging) systems by enabling a context and location dependent messaging service. Messages can be read only if the users manage to authenticate themselves with the appropriate gaze gestures.

3 Ubiquitous AR messaging

ubiGaze is a location- and object-based AR application, in which the main use case is to allow users to embed invisible messages into any real-world object, and, moreover, retrieve such messages discreetly and effortlessly. Gaze gestures also operate as an authentication protocol to access the messages. Users are equipped with two wearable computers: an eye tracker and a smartwatch. The head-worn eye tracker is used for estimating the gaze direction of the user. The gaze direction together with the scene camera can indicate where the user is looking in the surrounding physical environment. The smartwatch is used as an input and output modality. Users can select the message that they want to embed by using the touch interface of this wearable. The smartwatch is also used to display the otherwise invisible messages that the users have retrieved from augmented objects. Figure 2 illustrates the system architecture.

3.1 Gaze gestures

Gaze tracking is a technology that has been around for many years. A more recent alternative to using fixations and dwell times are gaze gestures. Generally, the concept of gestures is well known in the HCI community and it is something that most humans are familiar with. The benefit of gaze gestures in comparison to other eye tracking techniques is that they do not require any calibration, as only the relative eye movement is tracked, and they are insensitive to accuracy problems caused by the eye tracker. A common problem of gaze-based interaction is Midas-Touch, which is due to the fact that our eyes are never "off", so we must use a clutch mechanism for differentiating between intentional and accidental interactions. Gaze gestures overcome this limitation by making accidental gestures unlikely.

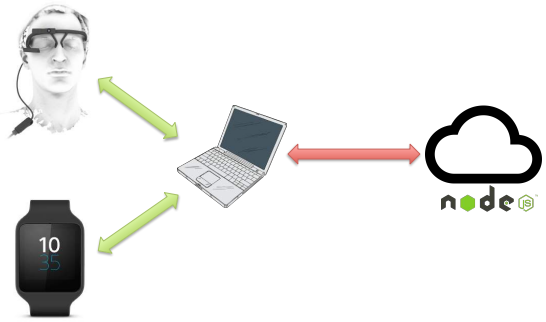

Figure 2: *System architecture with two wearables: an eye tracker equipped with a scene camera and a smartwatch. These two wearables are communicating through a gateway device, which communicates to a server that stores the messages.*

Drewes and Schmidt defined a gaze gesture as consisting "of a sequence of elements, typically strokes, which are performed in a sequential time order" [Drewes and Schmidt 2007]. The authors were among the first to propose a gaze gesture recognition algorithm. A stroke is a movement across one of the 8 possible directions and a gesture is defined as a sequence of such strokes. The user study showed that, on average, a gesture took 1900 ms, but this is dependent on the number of strokes. Gestures can also consist of only one single stroke [Møllenbach et al. 2009]. Such gestures have reduced cognitive load, they are easy to remember, and can be integrated with dwell time to create gaze controlled interfaces. A further investigation shows that selection times using single stroke gaze gestures are influenced by various factors (e.g., the tracking system, vertical or horizontal strokes) [Møllenbach et al. 2010].

Gaze gestures have been applied and evaluated in different application scenarios. EyeWrite is a system that uses alphabet-like gestures to input text [Wobbrock et al. 2008]. Users can use gaze gestures and dwell time to interact with mobile phones [Drewes et al. 2007]. A user study highlighted that the participants found interactions using dwell time more intuitive, but gaze gestures were more robust since it is unlikely that they are executed unwillingly. Gaze gestures were also evaluated as an alternative to PIN-Entry, with the goal to overcome shoulder surfing [De Luca et al. 2007]. This is a common fraud scenario where a criminal tries to observe the PIN code. A user study reveals similar findings that using gaze gestures the number of input errors is significantly lower.

Games can benefit from gaze gestures as an additional input modality. Istance et al. investigated gaze gestures as a means of enabling people with motor impairments to play online games [Istance et al. 2010]. A user study with 24 participants reveals that gestures are particularly suited for issuing specific commands, rather than for continuous movement where more traditional input modalities (e.g., mouse) work better. Hyrskykari et al. compared dwell-based interaction to gaze gestures in the context of gaming [Hyrskykari et al. 2012]. Their findings show that with gestures, users produce less than half of the errors when compared to the alternative.

Eye tracking capable devices are becoming more popular and gaze can be used as input. One aspect relevant to interaction is user feedback. Kangas et al. [Kangas et al. 2014] have evaluated gaze gestures in combination with vibrotactile feedback. Their findings show that participants made use of the haptic feedback to reduce the number of errors when performing gaze gestures. In our application, we also use vibrations to inform the users when a message has been received.

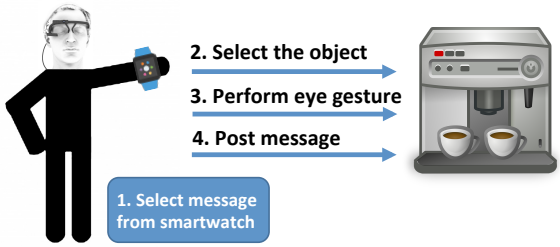

Figure 3: *Interaction flow for augmenting a coffee machine with an invisible message.*

3.2 Posting and reading messages

To embed a message into an object, first, the users have to find the object that they want to augment. Figure 3 illustrates the case when a user wants to attach a message to a coffee machine. The system relies on the eye tracker and the scene camera, which offers a first-person-view of the world, so the system sees what the user sees. When users fixate for a longer period of time on a specific object, the system sets a lock on it. User fixations are based on dwell time, which represents the amount of time users are looking in a specific direction. Using computer vision techniques we extract relevant features from the region of interest, which is given by the gaze direction. Next, users select a message from their smartwatch and perform a gaze gesture to couple the message and the object. This way, the message gets embedded into the object (see Figure 4).

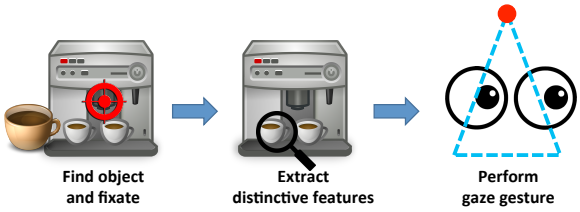

Figure 4: *First-person-view interaction flow. A user selects an object through fixation. Distinctive features are extracted from the selected object. The user performs a gaze gesture and the message is locked to the object.*

Peers can retrieve messages from the object in a similar way. There are several possibilities. First, the simplest case is when the user knows in advance the object which contains the message. Second, a visual marker could be shown in the user's camera view to indicate that messages are available in this location. Third, the user can try the gesture on different objects of the same type to check if there are messages. Fourth, the user could have no knowledge about the objects or presence of messages. In such a case, the user would need a general location authentication, which would then reveal if any messages are available. In the depicted coffee machine scenario, we assume that users know in advance which objects have been augmented with invisible messages. Users must also know what gesture unlocks that message. This coupling of elements also enables different authentication keys, i.e. gestures, for locations, objects, individual users and user groups. When users are in the proximity of the augmented object, they must first fixate on the desired object. This is viewed as a selection mechanism.

For identifying the correct object, we rely on extracting computer vision based features from the region of interest. This combines the user and object localization method and removes the need for other

infrastructure based methods, such as GPS or Wi-Fi. Afterwards, users have to perform the same eye gaze gesture (e.g., a circle or a triangle) and then the message is unlocked. Users will receive a notification with the content of the message on their smartwatch. Figure 5 illustrates this interaction scenario.

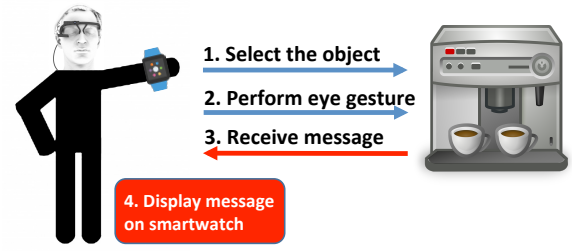

Figure 5: *Interaction flow for retrieving a message from an augmented coffee machine.*

3.3 Alternative application scenarios

Other application scenarios could benefit from the combination of eye trackers, cameras, and smartwatches, such as wearable gaming. Children around the playground could use such a technology for interacting and unlocking hidden messages from everyday objects. Geocaching and treasure hunts are another possible gaming application, where users have to follow a specific course and unlock clues until they reach their final destination. Our system could be used for hiding and finding such invisible clues in the environment.

Besides games and entertainment, eye tracking and gaze gestures unlock additional interaction techniques. Compared to hand gestures, gaze gestures are discrete, effortless, and do not attract unwanted attention. Industrial scenarios can also benefit from such an interaction technique. In factories or warehouses, workers use their hands to control equipment and gaze gestures could prove a viable alternative to hand gestures.

4 Real-world implementation

Our real-world messaging system prototype consists of the following software and hardware components. Users wear a head-mounted Pupil[1] eye tracker with a scene camera and a Sony Smartwatch 3 on their wrist. These devices are interconnected through a portable computer that processes the eye tracking data (implementation with Python and Processing[2]) and acts as an Internet gateway to the messaging server. The server (implemented with Node.js[3]) stores the couplings between gestures, objects, and messages in a database that is accessed through RESTful interfaces atop HTTP.

Color markers, i.e. a single color clearly visible patch, are used in the implementation as the distinct feature of the object to be detected. The algorithm retrieves a region of interest (ROI) from the image given the gaze direction. If the average RGB color values from all the pixels within the ROI were within a certain range (e.g., yellow), then we detected it as a marker. Other alternative methods are presented in the following section.

[1] https://pupil-labs.com/pupil/
[2] https://processing.org
[3] https://nodejs.org/en/

4.1 Challenges and future research directions

Considering our prototype, there are still challenges and issues that need to be addressed in the future.

One of the first challenges when using the Pupil eye tracker was the need for calibration and registration with respect to the scene camera. Calibration is not yet automatic, users must manually calibrate the eye tracker by finding several calibration points. We also experienced that the calibration becomes rather inaccurate once the user is mobile. Another limitation with regards to the eye tracker is that it requires a connection to a computer to process the data, which is a significant limitation for wearable applications. In the future, we believe that such eye trackers could evolve into devices similar to the Google Glass. They will be self-contained, have their own sensors and processing power. We also assume wearables will be able to connect to the Internet by themselves, thus infrastructure components (e.g., gateway) will not be required.

For gaze gestures, we have relied on a library that is delivered with the Pupil SDK. Our research objective was not to develop and improve gaze gestures recognition algorithms, but use existing technology. The feasibility of using gaze gestures as an authentication mechanism for augmenting/reading invisible messages is still unclear. The question whether users will adapt easily to gaze gestures requires further investigation. In our preliminary evaluation, users have expressed that this input modality is rather cumbersome due to the inaccuracy of the gaze gesture recognition.

Using a smartwatch as an input modality was straightforward. In our prototype, we developed an application with pre-defined messages (a set of Emoticons) that the users can select from a list. Given the small form factor, text input is difficult, as there is no keyboard available. However, this limitation can be addressed through spoken commands and voice interaction. This way, users could dictate the content of their message.

Our method is dependent on computer vision techniques for object recognition. Object detection and recognition is an active research area of computer vision and there is a wide body of research in that direction. Our application can only be as good as the existing techniques. In our implementation, to facilitate prototyping, we have used simple color markers for recognizing the objects that users want to augment.

Any approach for object detection and recognition could be used, as long as it can provide results in real time. Learning algorithms (e.g., supervised learning) are a good fit. Users could provide examples of objects and the system would train a classifier to detect and recognize such objects. This classifier would have to be shared among the sender, who creates it, and the receiver, who uses it to try to identify the object with the hidden message. Having detectors for the shape or characteristics of the objects in advance means that the system would not work with any object, but with a restricted set. A comprehensive survey on object tracking explains in more detail such challenges [Yilmaz et al. 2006].

5 Conclusion

This paper presents *ubiGaze*, a wearable AR system that enables users to augment any real-world object with invisible messages. This idea extends the concept of AR tags or annotations to a ubiquitous messaging system based on gaze gestures, where messages are locked into a set of distinctive features of real-world objects. Our system is composed of two wearable devices, an eye tracker equipped with a scene camera (Pupil) and a smartwatch (Sony Smartwatch 3). Users are able to post messages through a combination of gaze gestures and input from their smartwatch. Similarly,

users are able to read invisible messages from augmented objects by performing gaze gestures and use their smartwatch as a display. By combining different wearable devices we proposed a discrete and effortless interaction technique for embedding AR messages into any real-world object. We believe that this new technique can lead to novel interaction scenarios in wearable computing.

Acknowledgements

Parts of this work were developed at the Eyework workshop during UBISS 2016, Oulu, Finland. We thank Hans Gellersen (Lancaster University) and Eduardo Velloso (University of Melbourne) for their support for this work. We would also like to thank Gabor Sörös (ETH Zurich) for his ideas and help throughout this project.

References

AZUMA, R. T. 1997. A survey of augmented reality. *Presence: Teleoperators and Virtual Environments 6*, 4 (Aug.), 355–385.

BAY, H., ESS, A., TUYTELAARS, T., AND VAN GOOL, L. 2008. Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding 110*, 3 (June), 346–359.

BILLINGHURST, M., CLARK, A., AND LEE, G. 2015. A survey of augmented reality. *Foundations and Trends in Human-Computer Interaction 8*, 2-3 (Mar.), 73–272.

DE LUCA, A., WEISS, R., AND DREWES, H. 2007. Evaluation of eye-gaze interaction methods for security enhanced pin-entry. In *Proceedings of the 19th Australasian Conference on Computer-Human Interaction: Entertaining User Interfaces*, ACM, OZCHI '07, 199–202.

DREWES, H., AND SCHMIDT, A. 2007. Interacting with the computer using gaze gestures. In *Proceedings of the 11th IFIP TC 13 International Conference on Human-computer Interaction - Volume Part II*, Springer-Verlag, INTERACT'07, 475–488.

DREWES, H., DE LUCA, A., AND SCHMIDT, A. 2007. Eye-gaze interaction for mobile phones. In *Proceedings of the 4th International Conference on Mobile Technology, Applications, and Systems and the 1st International Symposium on Computer Human Interaction in Mobile Technology*, ACM, Mobility '07, 364–371.

FEDOSOV, A., ELHART, I., NIFORATOS, E., NORTH, A., AND LANGHEINRICH, M. 2016. SkiAR: Wearable augmented reality system for sharing personalized content on ski resort maps. In *Proceedings of the 7th Augmented Human International Conference 2016*, ACM, AH '16, 46:1–46:2.

HYRSKYKARI, A., ISTANCE, H., AND VICKERS, S. 2012. Gaze gestures or dwell-based interaction? In *Proceedings of the Symposium on Eye Tracking Research & Applications*, ACM, ETRA '12, 229–232.

ISTANCE, H., HYRSKYKARI, A., IMMONEN, L., MANSIKKAMAA, S., AND VICKERS, S. 2010. Designing gaze gestures for gaming: An investigation of performance. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, ACM, ETRA '10, 323–330.

KANGAS, J., AKKIL, D., RANTALA, J., ISOKOSKI, P., MAJARANTA, P., AND RAISAMO, R. 2014. Gaze gestures and haptic feedback in mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, CHI '14, 435–438.

MAYER, S., AND SÖRÖS, G. 2014. User interface beaming - seamless interaction with smart things using personal wearable computers. In *Proceedings of the 11th International Conference on Wearable and Implantable Body Sensor Networks (BSN 2014)*, 46–49.

MISTRY, P., AND MAES, P. 2009. SixthSense: A wearable gestural interface. In *ACM SIGGRAPH ASIA 2009 Sketches*, ACM, SIGGRAPH ASIA '09, 11:1–11:1.

MISTRY, P., MAES, P., AND CHANG, L. 2009. WUW - wear ur world: A wearable gestural interface. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems*, ACM, CHI EA '09, 4111–4116.

MØLLENBACH, E., HANSEN, J. P., LILLHOLM, M., AND GALE, A. G. 2009. Single stroke gaze gestures. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems*, ACM, CHI EA '09, 4555–4560.

MØLLENBACH, E., LILLHOLM, M., GAIL, A., AND HANSEN, J. P. 2010. Single gaze gestures. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, ACM, ETRA '10, 177–180.

NASSANI, A., BAI, H., LEE, G., AND BILLINGHURST, M. 2015. Tag it!: AR annotation using wearable sensors. In *SIGGRAPH Asia 2015 Mobile Graphics and Interactive Applications*, ACM, SA '15, 12:1–12:4.

NUERNBERGER, B., LIEN, K. C., HÖLLERER, T., AND TURK, M. 2016. Interpreting 2D gesture annotations in 3D augmented reality. In *2016 IEEE Symposium on 3D User Interfaces (3DUI)*, 149–158.

QUINTANA, E., AND FAVELA, J. 2013. Augmented reality annotations to assist persons with Alzheimers and their caregivers. *Personal and Ubiquitous Computing 17*, 6 (Aug.), 1105–1116.

WOBBROCK, J. O., RUBINSTEIN, J., SAWYER, M. W., AND DUCHOWSKI, A. T. 2008. Longitudinal evaluation of discrete consecutive gaze gestures for text entry. In *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications*, ACM, ETRA '08, 11–18.

YILMAZ, A., JAVED, O., AND SHAH, M. 2006. Object tracking: A survey. *ACM Computing Surveys 38*, 4 (Dec.).