Enabling Human Micro-Presence through Small-Screen Head-Up Display Devices

Scott Greenwald
MIT Media Lab
75 Amherst St.
Cambridge, MA 02139 USA
scottgwald@media.mit.edu

Mina Khan
MIT Media Lab
75 Amherst St.
Cambridge, MA 02139 USA
minakhan@media.mit.edu

Pattie Maes
MIT Media Lab
75 Amherst St.
Cambridge, MA 02139 USA
maes@media.mit.edu

Abstract

One of the main ways that humans learn is by interacting with peers, in context. When we don't know which bus to take, how to prepare plantains, or how to use a certain app, we ask a nearby peer. If the right peer is not around, we can use a mobile device to connect to a remote peer. To get the desired answer, though, we need to find the right person, and have the right affordances to send and receive the relevant information. In this paper, we define a class of *micro-presence* systems for just-in-time micro-interactions with remote companions, and explore dimensions of the design space. Informed by theories of contextual memory and situated learning, we present TagAlong, a micro-presence system conceived for learning a second language in context with help from a remote native speaker. TagAlong uses a connected head-mounted camera and head-up display device to share images of the wearer's visual context with remote peers. The remote companion can convey knowledge using these images as a point of reference– images are sampled, annotated and then sent back to the wearer in real time. We report on a quantitative experiment to determine the effectiveness of using visual cues that involve spatial and photographic elements as a means of "pointing" at things to draw the wearer's attention to them.

CHI'15 Extended Abstracts, April 18–23, 2015, Seoul, Republic of Korea.
ACM 978-1-4503-3146-3/15/04.
http://dx.doi.org/10.1145/2702613.2732846

Author Keywords

Remote Collaboration, Situated Learning, Just-in-Time Information, Micro-Presence

ACM Classification Keywords

H.5.1 [Information Interfaces and Presentation (e.g., HCI)]: Multimedia Information Systems—*artificial, augmented, and virtual realities*; K.3.1 [Computers and Education]: Computer Uses in Education—*collaborative learning, distance learning*; H.4.3 [Information Systems Applications]: Communications Applications; H.5.3: Group and Organization Interfaces—*computer-supported cooperative work, synchronous interaction*.

Introduction and related work

In this paper, we conceptually investigate how mobile and wearable devices can facilitate learning through real-time micro-interactions with remote peers, and present the TagAlong system as a concrete instance of a system implementing these ideas. We call such systems that aid this type of interaction *micro-presence systems.* This concept uses wearable computing and augmented reality to reap benefits of contextual learning and situated learning. In memory theory, it has been shown that information is more easily recalled in context, meaning that we would expect for language learners to have an easier time recalling a word or phrase if it was taught to them in a context than if they had learned it from a book. One example of learning in context would be learning names for groceries while at the supermarket. In situated learning, learners in *communities of practice* take on roles and responsibilities that are at first peripheral and simple, and over time grow in importance and complexity. Situated learning was conceived primarily in the context of vocational/practical skills – one seminal work [4] makes reference to midwives, tailors, quartermasters, butchers and more. Bridging the gap between distributed groups and practical knowledge is key so that scarce expert practitioners could disseminate knowledge on a mass scale and raise the quality of practice everywhere. Imagine a world where every second-language learner is constantly accompanied by an expert speaker until maximal proficiency is achieved. A prior example of near-real-time micro-presence system is VizWiz, which assists blind users in distinguishing objects [2].

Micro-presence: Just-in-time micro-interactions with remote peers

Inspired by the insights above, we conceive of a class of *micro-presence* systems for just-in-time micro-interactions with remote peers. In advance of introducing our concrete example, the TagAlong system, we present some background considerations and dimensions of the micro-presence design space.

Information Flow

Schematically, micro-presence systems establish a flow of information back and forth between the wearer and a remote companion. Information sent from wearer to companion allows the companion to establish an awareness of the wearer's context, and thereby (i) infer what information is needed, and (ii) establish a vocabulary of common reference points to which such information can refer. Information sent from the companion to the wearer is just-in-time, contextual information. That is, it tells the wearer something relevant to her current environment and activity.

In general this information flow can consist of any type of information that the system can capture on the side of the wearer and represent on the side of the companion and vice versa. Audio, video, sensor and physiological data can all easily be captured by the wearer's device and sent to the companion. Information that passively flows from the wearer's device is referred to as the context stream. The type of feedback signals at the disposal of the companion depends on the capabilites of the wearer's device. In our experimental setup, we look at a wearer-side system that conveys information using a head-up display and visual cues containing photographic and spatial information. Wearing some other kind of output device, such as headphones, a vibration motor, an air-based tactile feedback device, or muscle stimulation device would give the companion different channels through which to interact with the wearer.

Context Awareness
Context awareness can be a powerful tool, automatically initiating interactions, changing the flow of information, and changing its presentation. For example, a request for presence can be made automatically when a wearer is in motion or is in a specific location. If the companion is also mobile, the wearer can be notified when she is standing still and able to respond to requests. The system may prompt the wearer to review prior information when it detects idleness.

Personal Connection
The choice of how personal or impersonal an interaction is afforded by a system determines a great deal about how users will engage, as well as the quality of information that can be obtained. A one-on-one phone call is quite personal, and will be advantageous in cases where for example personal or culturally-sensitive advice is sought. On the other hand, if the information sought is miniscule and targeted – such as what bus to take – it may not be important or advantageous to make interactions socially intimate. In another way, if the interaction is many-to-one it immediately becomes less personal, but may benefit from representing the aggregate knowledge of more people – such as a group that votes on your next chess move.

TagAlong

TagAlong is a micro-presence system conceived for aiding users with second language learning through the help of a remote native speaker. One or many remote *companion* users provide information for a mobile *wearer* with Google Glass. This information uses as its primary point of reference the visual and auditory information captured by the Glass device. Specifically, companions can extract rectangular samples from images captured by the Glass, annotate them, and send them back to the wearer. Annotations can be textual, sketched, or auditory. In this section, we describe the system in detail.

Figure 1: TagAlong wearer (left) and companion (right)

Wearer Affordances
Wait and watch. The wearer can passively receive information the companion decides to send, and can also request information. The information the wearer receives can be graphical and/or auditory. The graphical information may be a sample from a camera image taken from her device, and may also include text- or sketch-annotations.

Point and pin. To request information about an object or part of a scene, the wearer centers the object in question behind the crosshairs shown on the Glass display, and performs a gesture, which takes a picture, marks a point of interest with a pin, and sends it to the companion. The wearer can also include an audio recording to further specify what information is desired.

Companion Affordances
Once a TagAlong session has begun, a companion can either respond to a pin, or passively monitor images coming from the learner's Glass until he chooses to interject.

Implementation
The system consist of three main components – software for Glass, a Python server, and companion web client.

Glass software
The Glass software is written in JavaScript to run on the WearScript platform. We used a modified version of WearScript to allow for audio sent from the companion to be saved and played back.

Python server
The python server mediates communication between Glasses and companion web clients. It connects to both using WebSockets. Images sent from the Glasses to the server are saved, and the web clients are given urls where they can be requested over http.

Companion web client
The companion web client is written in HTML5 to run both on desktop clients and on mobiles. The client has two main views: the *Image Selector*, which allows the companion to choose from a selection of images, and the *Card Composer*, where cards are annotated and sent.

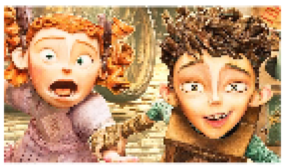
Pointing with small-screen HUD's

In order to enable learning that is situated and in context, companions need to be able to indicate or "point" at objects in the environment of the wearer, in order to give information about them or in reference to them (see also [1, 3]). Here we investigate the effectiveness of one method of doing this, approaching it in practical terms with hardware and networking capabilities that are widely available today.

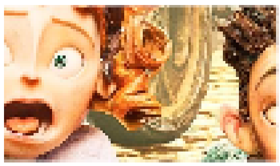
Summary of setup and variables
In our experiment we use a large 4K LCD touch screen to display the "environment" in front of the user, while simultaneously displaying an image on a Google Glass device. The screen is placed at roughly the same distance as the focal plane of the head-mounted display, to minimize the time necessary for the user to refocus his eyes between the head-mounted display and the screen. We measure the time it takes the user to find and touch the object in question.

We measure the effect of several variables on the time to complete the task. In the **environment stimulus** we vary *size of objects* and *location of objects,* as shown in Figure 2. We hypothesize that smaller objects will be more difficult to find, as will peripheral objects compared with central ones. In the **feedback stimulus**, we vary the *resolution* of cropped images as well as the presence or absence of a *mini-map*. Mini-maps represent the relative position of the cropped image in the wearer's environment, as shown in Figure 3. Due to limited camera resolution, small objects must be presented in low resolution. If it is the case that smaller objects take longer to find, we wish to separate the effects of low-resolution presentation from object size in the wearer's optical field of view. Trials with and without a mini-map test the hypothesis that a mini-map will be particularly helpful for finding small, peripheral objects.

Results and discussion
We had eight participants (N=8) in this experiment, and each performed 90 target acquisition tasks. We used within-subjects design for sizes, location and resolution (30 images each per user), and between subjects for mini-maps (three with, five without). Data for two participants was incomplete due to technical difficulties on the system side.



(a) Stimulus screen



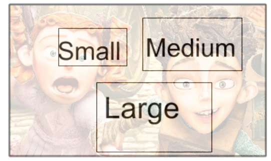
(b) Crop

(c) Mini-map

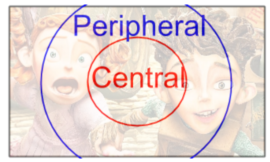
(d) HUD screen w/ mini-map

Figure 3: Screen and HUD stimuli

(a) Stimulus screen



(b) Crop size



(c) Crop location

Figure 2: Target sizes and location

The strongest expected results were (i) the inverse linear relationship between image size and time to acquire the target, and (ii) the longer time needed for peripheral vs central targets. The most interesting unexpected result was that mini-maps sped up acquisition for medium and small images, but slowed participants down on large images. Participants using a mini-map appeared to require a certain fixed time to interpret the map, regardless of the size of the target.

We interpret the large variance in our measurements to be a result of variations in image qualities that were not controlled for (e.g. low/high contrast, presence of repetitive elements). On the whole, these other variations have an influence of a comparable magnitude to the variables we did control for.

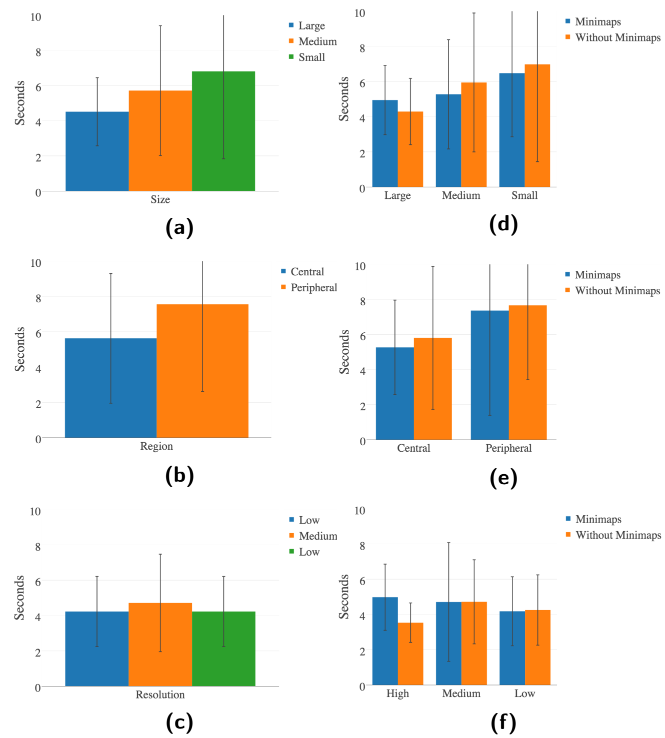


Figure 4: Results varying (a) image size, (b) region, and (c) resolution

Conclusion

We presented a framework based on wearable computing and augmented reality to designing micro-presence systems, in which the wearer is mobile and the companion is remote. It covers human factors as well as considerations related to system affordances. We introduced TagAlong, an end-to-end micro-presence system using Google Glass, that allows images annotated with text, sketches, and audio to be sent from the companion to the wearer. We showed through our experiment that our simple, robust, low-overhead system provides the core capability of establishing a correspondence between digital and physical information. This and the whole class of future micro-presence systems has potential benefits related to situated learning, contextual memory and social motivation.

Future Work

Pointing in motion. We have explored the question of how to point at objects in scenes while the device wearer is stationary. Given that the ultimate purpose is to use devices "in-the-wild," we need to explore methods for pointing while the user is mobile. A technique could either assists the user in returning to a location where the object is visible, or provides enough additional information that the reference can be useful without such physical backtracking.

Mobile learning with cognition awareness. Can we quantify and model the wearer's attention in such a way as to optimize her learning? It this best accomplished by giving the companion the decision of when to interject, or by using the system to regulate the delivery of information delivery to the wearer?

Social presence and motivation. While further exploring micro-presence use cases, where a remote human is involved in generating or mediating situation, real-time, or just-in-time information, the questions related to social presence will be important. When are there privacy concerns and how can they be dealt with? In what cases is remote presence attractive or motivating for the companion or the wearer – e.g. by creating a feeling of social closeness while sharing an experience? Does the addition of other data streams, such as audio, motion, or physiological signals enhance this effect?

Acknowledgements

We thank Blake Elias, Cory Kinberger, and Christian D. Vazquez for their work developing the TagAlong system. Special thanks to Chris Schmandt for his helpful suggestions and feedback.

References

[1] Baudisch, P., and Rosenholtz, R. Halo: A technique for visualizing off-screen objects. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, ACM (New York, NY, USA, 2003), 481–488.

[2] Bigham, J. P., Jayant, C., Ji, H., Little, G., Miller, A., Miller, R. C., Miller, R., Tatarowicz, A., White, B., White, S., and Yeh, T. VizWiz: Nearly Real-time Answers to Visual Questions. In *Proceedings of the 23Nd Annual ACM Symposium on User Interface Software and Technology*, UIST '10, ACM (New York, NY, USA, 2010), 333–342.

[3] Funk, M., Boldt, R., Pfleging, B., Pfeiffer, M., Henze, N., and Schmidt, A. Representing indoor location of objects on wearable computers with head-mounted displays. In *Proceedings of the 5th Augmented Human International Conference*, AH '14, ACM (New York, NY, USA, 2014), 18:1–18:4.

[4] Lave, J., and Wenger, E. *Situated Learning: Legitimate Peripheral Participation*. Learning in Doing: Social, Cognitive and Computational Perspectives. Cambridge University Press, 1991.