HandshakAR: Wearable Augmented Reality System for Effortless Information Sharing

Mihai Bâce, Gábor Sörös, Sander Staal, Giorgio Corbellini
Department of Computer Science
ETH Zurich, Switzerland
mihai.bace@inf.ethz.ch, gabor.soros@inf.ethz.ch, staals@student.ethz.ch, gcorbellini@gmail.com

ABSTRACT

When people are introduced to each other, exchanging contact information happens either via smartphone interactions or via more traditional business cards. Crowded social events make it more challenging to keep track of all the new contacts. We introduce HandshakAR, a novel wearable augmented reality application that enables effortless sharing of digital information. When two people share the same greeting gesture (e.g., shaking hands) and are physically close to each other, their contact information is effortlessly exchanged. There is no instrumentation in the environment required, our approach works on the users' wearable devices. Physical proximity is detected via inaudible acoustic signals, hand gestures are recognized from motion sensors, the communication between devices is handled over Bluetooth, and contact information is displayed on smartglasses. We describe the concept, the design, and an implementation of our system on unmodified wearable devices.

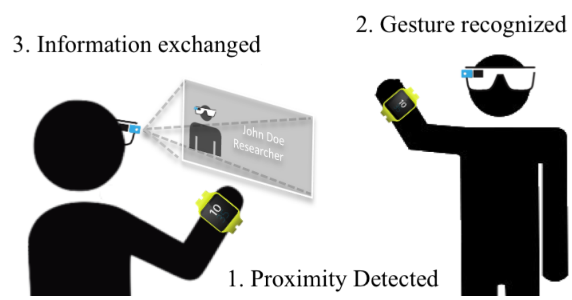

Figure 1: When two people are close to each other (1) and share the same greeting gesture (2), their contact information is effortlessly exchanged (3) and displayed on the smartglasses.

CCS Concepts

•Human-centered computing → Ubiquitous and mobile computing systems and tools;

Keywords

Wearable; Gesture recognition; Proxemics; Contact exchange

1. INTRODUCTION

Wearable devices are a specific class of miniature computers that have become increasingly popular. Smartwatches and fitness trackers are pervasive in our lives, and smartglasses like Google Glass, Microsoft HoloLens, Meta[1], or Daqri[2] are getting closer to real products than to lab prototypes. Wearables have also gained significant interest in the research community. These devices offer continuous sensing, information access, and interaction with the users that wear them. They do not need to be turned on or off and they allow true multi-tasking, i.e., users do not need to stop doing one activity to engage in another.

From a social point of view, technology plays an important role in our lives and enables us to interact with people from all over the world. For most cases, the smartphone can act as a universal interaction device both with smart appliances and between people. Besides voice and video calls, people can also share photos, videos, or exchange messages. In his theory of proxemics, Edward Hall showed that closer physical distances between people lead to higher engagement [4]. However, the way in which smart devices currently interact does not change when their owners are physically far away or close to one another.

Imagine the following simple, well-known scenario. When two people are introduced to each other, they get acquainted and want to exchange contact information. Business cards have been traditionally the most common way to achieve this task. More recently, people can exchange contact information via their smartphones: by creating a new contact entry or by using a social networking application like Facebook or LinkedIn. These methods require explicit interaction from the user and, for the latter, proximity does not play any role. What if we could make the process of exchanging information in specific contexts effortless?

Contributions of the paper

We propose HandshakAR, a wearable augmented reality system for effortless information sharing between people who are physically close to one another. Our system leverages

[1]http://www.metavision.com/
[2]http://www.daqri.com/products.html

AH '17, March 16-18, 2017, Mountain View, CA, USA

DOI: http://dx.doi.org/10.1145/3041164.3041203

only existing off-the-shelf unmodified wearable devices. When two people share the same greeting gesture and are physically close to each other, their contact information is effortlessly exchanged. Gestures are recognized from motion sensors, physical proximity is detected via inaudible acoustic signals, and the exchanged information is displayed on smartglasses. The underlying communication between the devices is handled over Bluetooth.

With HandshakAR, we augment the interaction experience between people. We qualitatively and quantitatively evaluate the components of the system in different settings running on different unmodified smartphones, smartwatches, and smartglasses.

2. RELATED WORK

There are several sensing technologies to track proxemic relationships between people and devices in smart environments, e.g., cameras, inertial sensors, radio signals, acoustic sensing, and combinations of those. An overview of related technologies is given by Marquardt [13]. Our work is indirectly related to toolkits and middlewares for proxemic user interactions like the *ProximityToolkit* [12] or *ProxemicUI* [1], but rather than providing a general framework, we focus on what is currently feasible with ubiquitous unmodified resource-constrained devices.

Connecting devices that share a common context (e.g., location) has been explored in several projects in the past like *Smart-Its Friends* [6] or *RFID-Shakeables*[10]. Examples for concepts, systems, and devices for connecting people include [16, 17, 11] (cameras), *iBand* [8], *CommonTies* [3] (custom wearables), *SocialSensing* [5] (WiFi sensing), *RoomSense* [15] (audio sensing), *High5* [9] (skin electric potential sensing), etc. The *Bonjour!* [7] paper explores the properties of greetings in different cultures, but only hints at how these greetings could be recognized with wearable technology. *The office smartwatch* [2] aims to improve interaction in work environments by enabling locking/unlocking of doors, virtual knocks and retrieving room information using a smartwatch. We present an effortless contact sharing method that requires neither instrumentation in the environment nor custom wearable hardware.

3. SYSTEM OVERVIEW

HandshakAR leverages unmodified off-the-shelf wearable and mobile devices. Figure 1 illustrates the main components and their purpose in the system. Users are equipped with a wrist-worn motion tracking device (e.g., smartwatch) and a head-mounted display (HMD) (e.g., smartglasses). The wrist-worn device is responsible for detecting proximity between people and for recognizing in-air greeting gestures. The HMD is responsible for showing contextual information (e.g., an electronic business card) at the right moment.

Next, we describe the main components of the system: proximity detection via acoustic ranging, gesture recognition via motion sensors, and information sharing between users and devices.

Proximity detection

Radio waves have been applied in many indoor localization technologies to estimate the distance between multiple smart devices. Unfortunately, for short ranges, a distance measurement based on the received signal strength indicator (RSSI) is not accurate enough because the signals are affected by environmental factors.

Our goal is a widely applicable, low cost, and accurate ranging approach, therefore, we chose acoustic ranging for close proximity detection. This method only requires the devices to be equipped with a microphone and a speaker which are given in most wearables.

HandshakAR builds upon two-way acoustic ranging, a method initially introduced by Peng et al. in the BeepBeep system [14]. The main advantage of two-way ranging is that the participating devices do not need to be synchronized. The devices sequentially emit an acoustic signal, and each device records its own signal and the signal produced by the remote device. The time difference between the two signals is used to estimate the distance. Out-of-channel coordination between devices happens via Bluetooth.

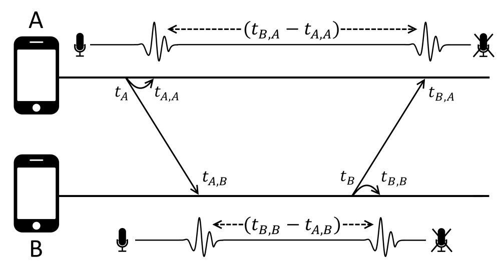

Figure 2: Acoustic ranging between two devices. Both devices are simultaneously recording, where device A first emits a signal and then B emits another signal. The distance between devices A and B can be approximated from the difference of signal arrival times.

Figure 2 presents the ranging procedure in more detail. Device A emits a sound at timestamp t_A and records its own signal, with its own microphone, at timestamp $t_{A,A}$. Device B records A's signal at timestamp $t_{A,B}$. Afterwards, device B emits a signal at timestamp t_B and records it at time $t_{B,B}$. Similarly, device A records device B's signal at timestamp $t_{B,A}$. The round trip time of flight can be calculated from the time intervals between $(t_{B,A} - t_{A,A})$ and $(t_{B,B} - t_{A,B})$. $d_{A,A}$ and $d_{B,B}$ represent the distance between the microphone and speaker on a specific device and c is the speed of sound.

$$D = \frac{c}{2} * ((t_{B,A} - t_{A,A}) - (t_{B,B} - t_{A,B})) + \frac{1}{2} * (d_{A,A} + d_{B,B}) \quad (1)$$

The reference signals have to be robust against ambient noise. BeepBeep [14] uses linear chirps with a Gaussian envelope to obtain a strong autocorrelation property. The method's reliability is dependent on several design parameters: the length of the recording timeframe, the length of the signal, and the frequency band (e.g., audible, inaudible). The reference chirp is detected using cross-correlation. This operation is computationally expensive but can be accelerated via the fast Fourier transform. To find the correlation peaks, we adopt the method proposed by Peng et al. [14]. The parameters used in our prototype are listed in Section 4.

Figure 3: HandshakAR prototype. (A) The system is inactive. (B) The devices are in close proximity to each other and, at the same time, a handshake greeting gesture was recognized. The devices exchange contact information, in this case a Facebook friend request. (C) The contact information appears on the user's HMD (the Google Glass in our prototype).

Gesture recognition

We recognize in-air gestures from motion sensors directly on the wearables using Dynamic Time Warping (DTW). Our primary goal is not to advance existing gesture detection and recognition algorithms, but rather to show that even a simple approach like DTW can be used to detect hand gestures.

DTW consists of two steps: data quantization and warping. The measurements from three-axis accelerometers within a sliding time window (length w, step size v) are compressed. Due to their limited processing power, wearables can benefit from data quantization which reduces the size of the time series and thereby the computational complexity. It can remove accelerometer noise, however, the parameters have to be chosen carefully. By choosing w too large, we can loose important features of the motion which may lead to lower recognition accuracy. By choosing w too small, there might be no compression (if w lower than the sampling rate) which only adds computational overhead. The parameters which yield a balanced trade-off between run time and recognition accuracy are presented in the evaluation section. Gesture recognition using DTW utilizes a predefined template dictionary for each gesture. Our method can distinguish between eight different greeting gestures. At test time, the new measurements are warped against all the templates in the dictionary and a ranked list of candidates is returned, from which the best matching pair is selected.

Information sharing

The wearables form a body area network (BAN). Bluetooth technology enables wireless communication within a BAN and across multiple BANs from multiple users. It also ensures interoperability between a wide range of wearable devices.

Each BAN is based on a client-server architecture. The wrist-worn device is responsible for managing the Bluetooth connection to the HMD, which is a Google Glass in our prototype. The server is also responsible for managing the connection to other BANs from different users. The messages exchanged between the devices are JSON formatted strings. The header contains information about the sender, the type of the message, and a timestamp. In HandshakAR, when two people shake hands, the BANs exchange contact information, which is further relayed to the HMD.

Sharing contact information can raise privacy concerns. We argue that such a system only has to be used in specific contexts, for example, at a cocktail party or a networking event. The wearable's OS could offer a specific feature to enable or disable such applications.

4. EVALUATION

We evaluated the feasibility of HandshakAR using unmodified mobile and wearable devices. We experimented with two different smartphones (LG G3 and Google Nexus 5X), two different smartwatches (Sony Smartwatch 3 and Motorola 360 Sport 2^{nd} Generation), and a Google Glass as display. Due to the lack of loudspeakers on our smartwatches, we were unable to test the acoustic ranging component on these wearables. This is why in our prototype implementation, presented in Figure 3, we use a smartphone. We report results from evaluating the two main components of the system: proximity detection and gesture recognition. We discard the details and the evaluation of the underlying communication framework since Bluetooth is an already established standard.

Proximity detection

Proximity is estimated using the two-way ranging method described previously. HandshakAR uses inaudible (> 20 kHz) acoustic signals, as they are not intrusive to the human ear. The maximum distance that we could measure in this frequency range was around 20 m in an outdoor environment. Our system is focused on close proximity, which is why we limit the evaluation to distances up to 5 m. The characteristics that influence the ranging mechanism are the recording timeframe, the frequency bandwidth, and the length of the chirp.

We evaluated the ranging component both indoors and outdoors, both in quiet and dynamic noisy environments (e.g., a cafeteria at lunchtime). We use a sampling rate of 48 kHz for both the microphone and the speaker, which is the maximum on current Android devices. This implies that we can reconstruct frequencies without aliasing up to 24 kHz (Nyquist frequency). Most human ears can only hear sounds up to 20 kHz, which is why we evaluate frequencies between 20 kHz and 24 kHz. The length of the chirp has been set empirically to 75 ms.

We varied the distance (0-5 m), the frequency bandwidth (250 Hz to 2000 Hz) and the recording timeframe (350 - 750 ms). For each configuration, we collected over 100 sample points.

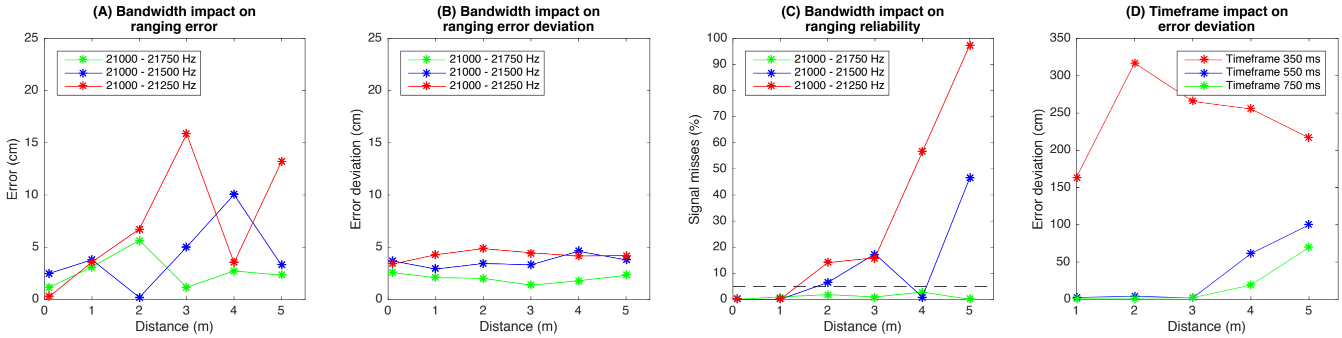

Figure 4: Acoustic ranging measurements indoors. Signal misses represents the % of unsuccessful measurements, where one device cannot detect the signal from the other device. The dotted line indicates a probability of 5%.

Figure 4 illustrates the most significant results in an indoor environment. Figure 4A shows that the measurement error is below 20 cm which is not significant for the considered scenario. The deviation shown in Figure 4B is quite similar, however, we can see a significant impact on the method's reliability in Figure 4C because the chirp was often not detected by the peer device. The frequency bandwidth with the best overall performance is 21,000-21,750 Hz.

The length of the recording timeframe also has an impact on the measurement error. We found the best compromise between time and accuracy to be 550 ms which up to 3 m has an error deviation under 10 cm (Figure 4D).

The method also has a few assumptions. As the 21'000 to 21'750 Hz band is in the inaudible spectrum, we require devices to have a microphone which can operate in this range. The current method does not support simultaneous ranging of more than two devices. If multiple devices are present, ranging happens in a round-robin fashion which impacts the update frequency for a new distance measurement.

Gesture Recognition

The on-board gesture recognition is using DTW. We implemented a multi-dimensional DTW, where the distance measure represents the cumulative distances of all dimensions independently measured under DTW. A dependent multi-dimensional DTW implementation is computationally more expensive, a limitation for wearables.

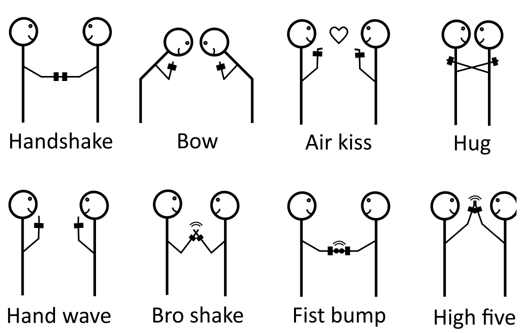

Figure 5: We evaluate the gesture recognition component on eight different greeting gestures.

The HandshakAR prototype stores a total of 80 samples, 10 samples for each of the eight predefined gestures (Figure 5). These samples were collected with the Nexus 5X, the device used in the prototype (Figure 3). Each user has to record their own set of gestures which will serve as a template library when the system is running.

Additionally, using the Sony Smartwatch 3, we collected a total of 800 greeting gesture samples (5 participants, 8 different gestures, 20 samples for each gesture, 3 s per gesture). The users tested the recognizer with their own template library and DTW achieved close to 99% accuracy (without any data quantization) for each participant. This outlines that DTW works well in distinguishing gestures performed by the same person. A user-independent gesture recognition method is out of the scope of this paper.

We also evaluated the data quantization parameters, the window size w and step size v. A good trade-off between runtime and recognition accuracy is $w = 250\,ms$ and $v = 200\,ms$. With these parameters, the recognition accuracy is close to 98%, but much faster than without quantization. It takes about 7 ms to classify a new sensor reading on a desktop computer and 700 ms on the Sony Smartwatch 3. The difference is significant, but we expect wearables to become computationally more powerful.

Further design aspects

For future work, we are interested in exploring the energy consumption of our system. Extensive use of the microphone and speaker can have an impact on the battery. Furthermore, communication over Bluetooth and peer discovery consume significant amounts of energy. This limitation can be addressed by switching to a beaconing mode, however, devices must support BLE peripheral mode.

5. CONCLUSION

We presented HandshakAR, a wearable augmented reality system for effortless information sharing. Our system combines proximity detection via acoustic ranging with gesture recognition from motion sensors. HandshakAR is a demonstrator for this concept which enables people to share contact information and display it on the smartglass. No explicit interaction from the users or any instrumentation of the environment is required. Additionally, we demonstrated the feasibility of the concept on unmodified devices and provided a set of practical design parameters to facilitate further research.

6. REFERENCES

[1] M. Alnusayri, G. Hu, E. Alghamdi, and D. Reilly. 2016. ProxemicUI: Object-oriented Middleware and Event Model for Proxemics-aware Applications on Large Displays. In *Proc. ACM Symposium on Engineering Interactive Computing Systems (EICS '16)*.

[2] Yannick Bernaerts, Matthias Druwé, Sebastiaan Steensels, Jo Vermeulen, and Johannes Schöning. 2014. The Office Smartwatch: Development and Design of a Smartwatch App to Digitally Augment Interactions in an Office Environment. In *Proc. Companion Publication on Designing Interactive Systems (DIS Companion '14)*.

[3] J. Chen and A. Abouzied. 2016. One LED is Enough: Catalyzing Face-to-face Interactions at Conferences with a Gentle Nudge. In *Proc. ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*.

[4] E. T. Hall. 1966. The hidden dimension. (1966).

[5] Y. Halperin, G. Buchs, S. Maidenbaum, M. Amenou, and A. Amedi. 2016. Social Sensing: A Wi-Fi Based Social Sense for Perceiving the Surrounding People. In *Proc. Augmented Human International Conference (AH '16)*.

[6] L. E. Holmquist, F. Mattern, B. Schiele, P. Alahuhta, M. Beigl, and H.-W. Gellersen. 2001. Smart-Its Friends: A Technique for Users to Easily Establish Connections Between Smart Artefacts. In *Proc. Springer-Verlag International Conference on Ubiquitous Computing (UbiComp '01)*.

[7] S. Houben, S. Perrault, and M. Serrano. 2015. Bonjour! Greeting Gestures for Collocated Interaction with Wearables. In *Proc. International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct (MobileHCI '15)*.

[8] M. Kanis, N. Winters, S. Agamanolis, A. Gavin, and C. Cullinan. 2005. Toward Wearable Social Networking with iBand. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems (CHI EA '05)*.

[9] Yuhwan Kim, Seungchul Lee, Inseok Hwang, Hyunho Ro, Youngki Lee, Miri Moon, and Junehwa Song. 2014. High5: Promoting Interpersonal Hand-to-hand Touch for Vibrant Workplace with Electrodermal Sensor Watches. In *Proc. ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '14)*.

[10] L. Kriara, M. Alsup, G. Corbellini, M. Trotter, J. D. Griffin, and S. Mangold. 2013. RFID Shakables: Pairing Radio-frequency Identification Tags with the Help of Gesture Recognition. In *Proc. ACM Conference on Emerging Networking Experiments and Technologies (CoNEXT '13)*.

[11] M. Kurze and A. Roselius. 2011. Smart Glasses Linking Real Live and Social Network's Contacts by Face Recognition. In *Proc. 2nd Augmented Human International Conference (AH '11)*. 31:1–31:2.

[12] N. Marquardt, R. Diaz-Marino, S. Boring, and S. Greenberg. 2011. The Proximity Toolkit: Prototyping Proxemic Interactions in Ubiquitous Computing Ecologies. In *Proc. ACM Symposium on User Interface Software and Technology (UIST '11)*.

[13] N. Marquardt and S. Greenberg. 2015. *Proxemic Interactions: From Theory to Practice*. Morgan & Claypool.

[14] C. Peng, G. Shen, Y. Zhang, Y. Li, and K. Tan. 2007. BeepBeep: A High Accuracy Acoustic Ranging System Using COTS Mobile Devices. In *Proc. ACM International Conference on Embedded Networked Sensor Systems (SenSys '07)*.

[15] M. Rossi, J. Seiter, O. Amft, S. Buchmeier, and G. Tröster. 2013. RoomSense: An Indoor Positioning System for Smartphones Using Active Sound Probing. In *Proc. Augmented Human International Conference (AH '13)*.

[16] Y. Utsumi, Y. Kato, K. Kunze, M. Iwamura, and K. Kise. 2013. Who Are You?: A Wearable Face Recognition System to Support Human Memory. In *Proc. Augmented Human International Conference (AH '13)*.

[17] W. Xu, Y. Shen, N. Bergmann, and W. Hu. 2015. Poster: Robust and Efficient Sensor-assisted Face Recognition System on Smart Glass. In *Proc. ACM Conference on Embedded Networked Sensor Systems (SenSys '15)*.