EarTouch: Turning the Ear into an Input Surface

Takashi Kikuchi
Keio University
Yokohama, Japan
tkiku393760@gmail.com

Yuta Sugiura
Keio University
Yokohama, Japan
sugiura@keio.jp

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

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

Bruce H. Thomas
University of South Australia
South Australia, Australia
bruce.thomas@unisa.edu.au

ABSTRACT
In this paper, we propose EarTouch, a new sensing technology for ear-based input for controlling applications by slightly pulling the ear and detecting the deformation by an enhanced earphone device. It is envisioned that EarTouch will enable control of applications such as music players, navigation systems, and calendars as an "eyes-free" interface. As for the operation of EarTouch, the shape deformation of the ear is measured by optical sensors. Deformation of the skin caused by touching the ear with the fingers is recognized by attaching optical sensors to the earphone and measuring the distance from the earphone to the skin inside the ear. EarTouch supports recognition of multiple gestures by applying a support vector machine (SVM). EarTouch was validated through a set of user studies.

Author Keywords
Earphone; Skin Deformation; Photo Reflective Sensor.

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

INTRODUCTION
Earphones allow people to listen to music in various situations. They also allow the user to operate their smartphones in a hands-free manner, such as answering a phone call and operating applications of their phones with voice commands. More complicated information, such as map guidance can also be provided through earphones. As these examples show, the usage of earphones is diversified. Earphones may be connected via a wire or by a wireless connection such as Bluetooth. Moreover, as exemplified by Apple AirPods, they are becoming increasingly smaller. As a result of these advantageous features, people will be able to wear earphones in their daily lives for longer periods of time and exchange an increasing amount of information through their earphones.

MobileHCI '17, September 04-07, 2017, Vienna, Austria
© 2017 Association for Computing Machinery.
ACM ISBN 978-1-4503-5075-4/17/09□$15.00
http://dx.doi.org/10.1145/3098279.3098538

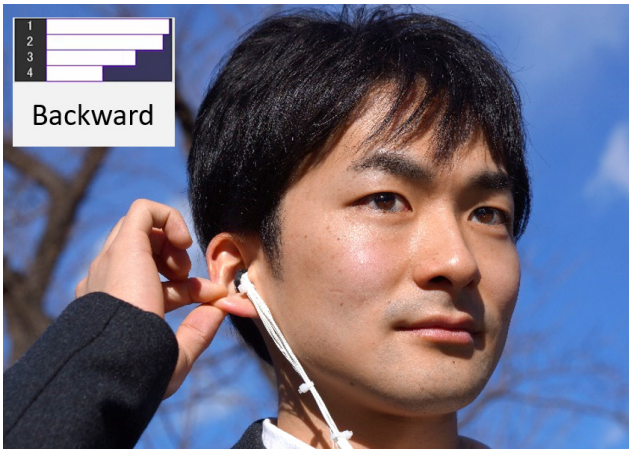

Figure 1: Concept of EarTouch: The earphone device recognizes the input made by the user pulling the ear (e.g. pulling the ear backwards).

An earphone is a device that outputs sound, and it is often operated through a mobile device. Some earphones can be operated via a switch embedded in the wire of the earphones as an input device connecting to the mobile device. However, when the earphone is small, the wiring-embedded switch is usually not used, so a new type of input method is required.

Input by voice commands is a powerful method, but it is difficult to apply in noisy environments or environments where the user cannot speak. As a different idea concerning input, attaching a touch sensor to the earphone itself has been considered. However, the sensor is so small that this input method is limited. The Apple AirPod supports simple tapping gestures, but more complex commands require a new approach.

Therefore, we propose EarTouch, a new input method that uses the ear as an input surface and allows various input gestures with an earphone of conventional size (Figure 1).

As for this method, several optical sensors are attached to the earphone. Deformation of the ear is estimated by measuring the skin deformation with an optical sensor. The optical sensor used is a reflection type, which can measure the distance to an object by emitting infrared light and measuring the intensity of the light reflected from the object. Since the sensors are attached to the earphone, and the distance to the skin inside the ear is measured at multiple points, the deformed shape can be detected. Using the sensor-measured data, gestures identification is performed by a support vector machine (SVM). The EarTouch can recognize gestures like gently pulling the ears up, down, forward, and backward.

Using the ear as an input interface in this manner has several advantages. First, since the ear deforms softly, it provides natural tactile feedback to the touching finger, thereby enabling the user to input intuitively. Second, the ear is a part of a person's body that they touch naturally; therefore, the user can input commands naturally without worrying about provoking stares from other people. The proposed input method can operate even if the ears are covered by the user's hair or a hat.

RELATED WORK

Interacting with Ears

Several attempts to measure human behavior and extend input methods by adding functions to earphones have been proposed. For instance, Cord Input is a command-input method based on twisting or pulling the wire of an earphone [16]. In addition, another developed system utilizes reflective optical sensors attached to a set of earphones, and the sensors recognize which earphones are fitted to which ear and appropriately provide sounds to the left and backward ears [9]. Another proposed method uses earphones as an input operation via an electrode attached to an earphone for measuring movements of the eyes [7]. "Earable" has succeeded in measuring expression and chewing state by measuring the distance between the eardrum and the earphone with a light sensor embedded in the earphone [1]. Another proposed method uses ultrasonic frequency from a phone to determine which earbud was removed [5]. This method allows interactions such as answering a phone call by simply removing an earbud. Unlike these previously proposed methods, the current proposed method, EarTouch, utilizes the ear's skin as an explicit input interface without changing the appearance of the earphone.

It is noteworthy that methods based on gestures around the ear or interaction by directly touching the ear have also been proposed. "FreeDigiter" measures aerial gestures performed at a slight distance from the ears [10]. And "EarPut" is a device that measures the interaction when the rim of the ear is touched by a finger wearing a sensor [6]. These measuring devices, however, are bulky and might malfunction if the ears are covered by the hair or a hat. In contrast to these methods, the proposed method utilizes devices that do not change the shape of the original earphone.

Skin as an Input Surface

The proposed method utilizes the ear as an input interface. Many methods using human skin as an input interface have been proposed. One such method uses the back of the hand as an input interface [12]. As for this method, a wristband-type device houses an array of optical sensors, and it recognizes the position of the fingertip touching the back of the hand. Moreover, SenSkin is a band-type device equipped with multiple optical sensors [13]. Wrapped around the user's arm, it measures skin deformation of the forearm. The user can input a gesture generated on the forearm such as tweaking or pressing. Another developed system recognizes a tapping operation on the skin by measuring the vibration propagating in the skin surface [3]. A film-type sensor that is stretchable and adapts to the skin motion has also been developed [17]. Interactive operations for mobile devices using the skin surface of the forearm are also being designed [18].

To understand peoples' daily state of mind, measuring the motions of the face is important. For instance, the movement of the jaw can be measured by light sensor [4]. Another developed device measures the distance between the skin and the frame of eyeglasses by an array of optical sensors on the frame of the eyeglasses [8]. Moreover, a system that recognizes facial expressions by that measurement device has been proposed. And the deformation between the eyebrows has been measured by using a light sensor and applied to information manipulation [11]. Compared with these studies, the proposed method uses the ear as an input surface.

A device that helps a user navigate by gently pulling their ears has been proposed [2]. This device pulls the user's ear left and backward with a clip in order to navigate the user to their destination. Moreover, using hand-to-face interaction to control an HMD has been investigated [15]. In these studies, the ear was also defined as a gesture area, and optical cameras and markers were used for tracking.

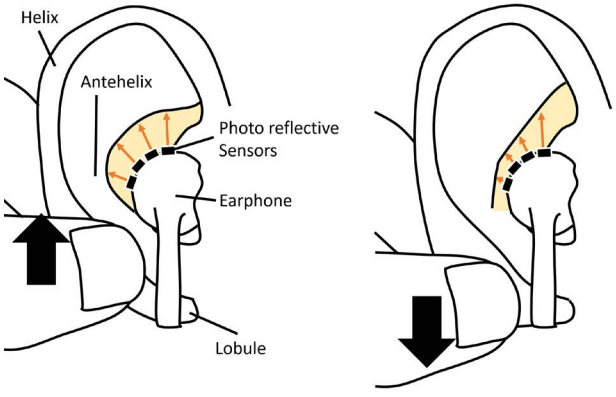

Figure 2: Principle of proposed method: The sensors attached to the earphone measure the skin deformation of the ear.

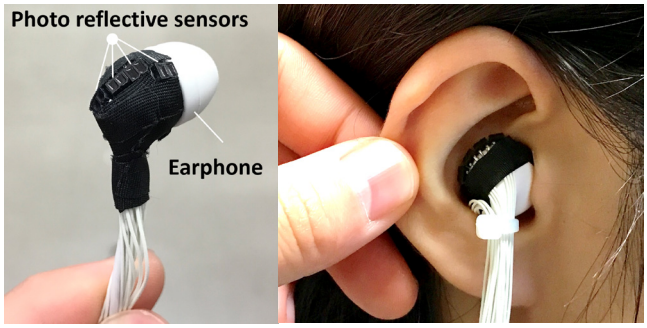

Figure 3: EarTouch device (left) and sensor layout when inserted in ear (right).

However, compared to these methods, the method proposed in the current study is more compact.

EARTOUCH

Principle

Ear deformation is measured by four optical photo reflective sensors, which are a combination of an infrared LED and a phototransistor, attached to earphones. This form of sensor is generally used for measuring the distance between a sensor and an object. The set of optical sensors is embedded in the earphone, and the distance from different points on the earphone to the skin of the ear is measured (Figure 2). When the user touches their ear with a finger (on the lobule or helix, see Figure 2), a force is applied to the ear. The skin inside the ear (antehelix) is deformed by that force. As a result, it is possible to recognize with several optical sensors that the ear has been touched and deformed slightly into different shapes. This recognition is determined from the distances from the earphone sensors to the points at which the skin changes.

Hardware

A device with four optical sensors mounted on an earphone (Figure 3). The sensors were installed on the backward earphone only. (Note that the sensors may be placed on both earphones, thereby allowing interactions with both ears.) The optical sensors (SG-105 models manufactured by KODENSHI Co., Ltd.) were connected to a microcontroller (Arduino Pro Mini, 3.3 V), and data acquired by the sensors was transmitted to a PC (Intel Core i7-4770 processor, 8GB memory) through XBee. The PC performs gesture recognition, which will be switched to a smartphone platform in future. The sensors are relatively small (2.7× 3.2×1.4 mm). As for the prototype, Arduino Pro Mini was used; however, in the future, it is planned to utilize the Bluetooth connectivity of wireless earphones in order to produce a wireless and compact device.

Directional-gesture Recognition

The prototype EarTouch device recognizes gestures by using the data acquired by the photo sensors. A support vector machine (SVM), was used for the gesture recognition. The "Support Vector Machine for Processing" (PSVM) library was used for the implementation the SVM [14].

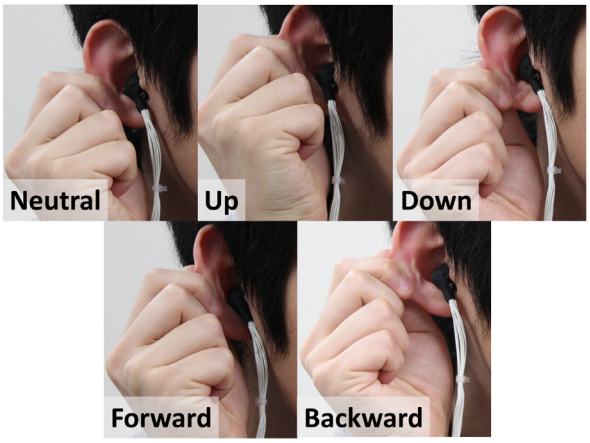

Figure 4: Five directional gestures.

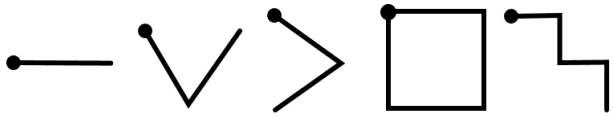

Figure 5: Five symbolic gestures.

A direction dataset for four gestures, i.e., pulling the earlobe up, down, forward and backward, was first prepared. When the user is wearing the device, learning data is accumulated by recording the data acquired by the sensors when the earlobe is pulled up and down and forward and backward with the fingers a hundred times. Before the SVM was applied, an IIR filter was applied to the raw sensor data to stabilize it. After learning, by repeating the same gesture, the EarTouch system can recognize the four different gestures (Figure 4). The PSVM provides the probability of how close the input is to each basic gestural direction. On the basis of that probability, each direction is weighted. This weighting enables the Eartouch system to recognize not only the basic four directions but also calculate directions on the 2D plane that the gestures are formed on. For example, a gesture that combines "up" and "forward" can be recognized.

Symbolic-gesture Recognition

Using the directional gestures as a base, the user can input more complicated gestures. How much (and in what direction) the ear is moved from the origin is calculated from the direction data. By sampling the directions, the system can calculate the 2D points of the trajectory of the ear movement. By sampling the calculated 2D points, the user can draw symbols which can be used as input.

The prototype EarTouch system also allows recognition of symbolic-gestures. The SVM and $1 Unistroke Recognizer were used for implementing gesture recognition [19]. First, the SVM is used to recognize the direction in which the ear is pulled (as described in the preceding section). Then, from the directional input, a stroke input is created by plotting 2D points, which were calculated by adding the unit vector of the directional input to the previous 2D point. When the "neutral" (no pulling direction, see Figure 4) condition is

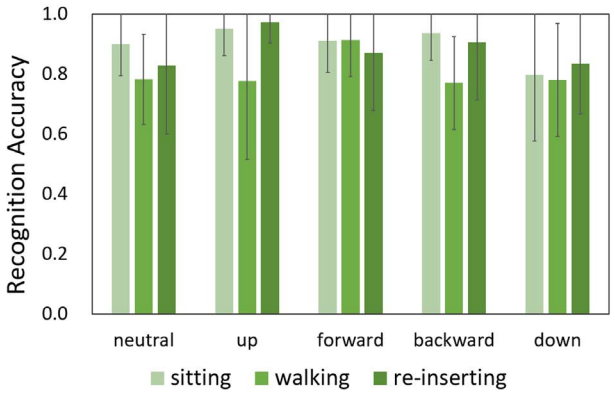

Figure 6: Recognition accuracy of user studies 1 & 2.

Figure 7: Confusion matrix of recognition accuracy when sitting.

recognized for a set duration, the system determines this to be the end of the gesture, and sends all the 2D points of the trajectory to the $1 recognizer as stroke input. Using stroke input makes it possible to use the sensor data chronologically and thereby expand the variety of gestures as well as directions.

The EarTouch prototype can recognize the following five gestures: "line," "check mark," "inverted caret," "square," and "stairs" (Figure 5). In this paper, we tested only five gestures but the EarTouch device can recognize complex gesture that consists of a combination of simple strokes.

EVALUATION

Study 1: Accuracy Concerning Directional Gestures

A user study was conducted to investigate the recognition accuracy of the directional gestures. Participants were instructed to hold their ear and pull it in the four directions (up, down, forward, and backward) and just hold it (neutral direction) without deforming the ear itself. The participants conducted the experiment while in two postures (sitting or walking). The walking experiment was performed indoors in a room, in which the participants walked around the room in arbitrary direction and speed. The participants were instructed to walk as usual without staying in the same place.

Before data was collected, the participants were instructed where to hold their ear and allowed to practice moving their ear in each direction only once. They were not instructed to pull their ear with a certain force, and they were allowed to insert EarTouch in their ear as they would insert any earphone. They were told to pull the ear in one of the directions for five seconds. The sampling rate of the sensors was set to 30 fps, and sensor data was collected for 100 frames per subject. The 100 frames in the middle of those 5s were used for training. These procedures were defined as one trial. In total, 5000 sets of data were collected per participant (i.e., 5 directions × 100 frames × 5 trials × 2 postures = 5000 sets of data). The order of conditions was randomized per participant.

The participants included six men and two women in their 20s. One dataset collected from each participant was subjected to five-fold cross validation. Sensor data of four trials was used as a training dataset, and one trial was used as a test dataset. The training dataset for each participant was subjected to SVM with a linear kernel. The training and the cross-validation were performed on each participant. It is necessary to perform training on each participant because the shape of peoples' ears differs.

According to the results of study 1 (see Figures 6 and 7), average recognition accuracy was 89.56% (with SD of 5.24) when the participants were sitting. That recognition accuracy was higher than that in the case of walking (80.78% with SD of 12.16). One reason for this discrepancy is decreased recognition accuracy in the case of the neutral condition while walking compared to that in the case of sitting. This was because the user tended to pull their ear unconsciously through vibration caused by the movement of the body.

Study 2: Precision after Re-inserting EarTouch

Another user study was conducted to investigate the recognition accuracy for the directional gestures when re-inserting EarTouch. The same experiment as in Study 1 was conducted, except that after each trial, the participants were instructed to briefly take EarTouch out of their ear and re-insert it. The participants included six men and two women in their 20s. All of the participants were the same from Study 1. For each participant, sensor data of four trials were used as a training dataset, and that of one trial was used as a test dataset. Add datasets were subjected to five-fold cross validation.

According to the results of Study 2, average recognition accuracy for directional gestures was 88.21% (with SD of 11.86) (see Figure 6). This result shows that it is unnecessary to collect more data for the training dataset after the earphone is repeatedly re-inserted. It also shows that the position in which EarTouch feels comfortable is constant. For some participants, recognition accuracy after they re-inserted EarTouch was higher than that determined in Study 1. For those participants, the earphone often tended to slip out of the ear. That problem was caused by

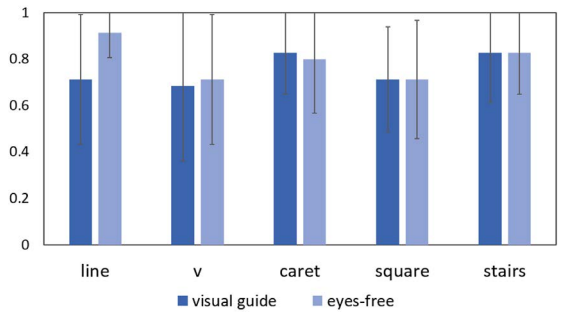

Figure 8: Recognition accuracy for symbolic gestures.

the weight of the wiring of the earphone, and it can be solved by building the sensor into the earphone.

Study 3: Accuracy Concerning Symbolic Gestures

Another user study was conducted to investigate the recognition accuracy for symbolic gestures. Participants were instructed to input the symbolic gestures shown in Figure 5. They sat in front of a display and performed the same gestures as those shown randomly on the display. This experiment was conducted under two conditions: with visual aid and "eyes free." For the first condition, the trajectory of the participants' gesture input was showed, and the participants were allowed to redo their gesture input if they were not satisfied. For the second condition, the visual aid was hidden, and the participants had to trust their intuition and were not allowed to redo their gesture. Before the experiment, the participants practiced for a few minutes without being shown the recognition result of their input. For this experiment, the SVM was first trained with 100 sensor data per direction. In the training step, the participants were to pull the ear in the direction instructed and stay in the position while collecting the sensor data. The frame rate of the system was set to 30fps. The participants included six men and two women in their 20s. All of the participants were the same from Study 1 and 2.

Mean recognition accuracy for symbolic gestures was 77.43% (with SD of 10.83) (see Figure 8). Recognition accuracy was higher without visual aids (79.43% with SD of 12.95) compared to that with visual aids (75.43% with SD of 8.77). This discrepancy occurs because many of the participants were trying to improve their gesture input and were going back and forth making the input more complicated. On the other hand, in the eyes-free condition, the user could not see the trajectory that made them want to create a simpler input.

LIMITATIONS AND FUTURE WORK

After the studies, factors that caused false positives were identified. We have confirmed low recognition error when the participants were talking. Slight changes in facial expression did not have much effect on the accuracy. This is because the ear is a part of the face that cannot be deformed easily by facial expressions only. Although some people can move their ears without touching them, the movement is only slight, like twitching.

However, when the user is moving intensively, such as running or jumping, the frequency of false positives will rise because the earphone itself might move. For these situations, it is possible to use other sensors, for example, acceleration sensors, to recognize the state of the user and reduce the frequency of false positives. In addition, as it can be seen from the Study section, we used a low number of trials for training. We showed that our system can recognize gestures with certain accuracy with few examples collected. For the future work, we would collect more training data to see whether the accuracy will improve and also reduce some effects on false positives.

Also, the proposed method measures the distance from the earphone to the inside part of the ear (antehelix). Therefore, it cannot be used with earphones that are placed over the ear, for example, earphones designed to be used during sports. Another limitation is the photo sensors measure distance using infrared light, so the recognition accuracy might decrease in places under strong sunlight.

Using photo sensors, however, increases the power consumption of the earphones. Power consumption is an important factor when using mobile devices. Accordingly, for future work, it is planned to switch the IR LEDs on and off. To switch the power on and off, one sensor can be used to determine the start of a gesture, which signals the rest of the LEDs to turn on. In addition, it is planned to investigate the difference in power consumption of a wireless earphone fitted with or without EarTouch.

CONCLUSION

A new interaction method, called EarTouch, which turns the ear into an interface input device was proposed and experimentally evaluated. Photo sensors attached to an earphone measure the deformation of the ear when slightly pulled. The method maintains the conventional size of an earphone and allows natural eyes-free interaction. EarTouch recognizes five basic directional gestures by using an SVM. And the recognized directions are used to allow the user to input symbolic gestures. In three user studies, the recognition accuracy of the gestures in several situations was measured. The results of the studies revealed that EarTouch achieves a suitable level of recognition accuracy.

ACKNOWLEDGMENTS

This work was supported by JSPS KAKENHI Grant Numbers JP26700017 and JP16H01741.

REFERENCES

1. earable. http://www.earable.jp/
2. Yuichiro Kojima, Yuki Hashimoto, Shogo Fukushima, and Hiroyuki Kajimoto. 2009. Pull-navi: a novel tactile navigation interface by pulling the ears. In *ACM SIGGRAPH 2009 Emerging Technologies (SIGGRAPH '09)*. ACM, Article 19, 1 pages. DOI=http://dx.doi.org/10.1145/1597956.1597975

3. Chris Harrison, Desney Tan, and Dan Morris. 2010. Skinput: appropriating the body as an input surface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, 453-462. DOI=http://dx.doi.org/10.1145/1753326.1753394
4. Naoya Koizumi, Hidekazu Tanaka, Yuji Uema, and Masahiko Inami. 2011. Chewing jockey: augmented food texture by using sound based on the cross-modal effect. *In Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology (ACE '11)*, ACM, Article 21 , 4 pages. DOI=http://dx.doi.org/10.1145/2071423.2071449
5. Gierad Laput, Xiang 'Anthony' Chen, and Chris Harrison. 2016. SweepSense: Ad Hoc Configuration Sensing Using Reflected Swept-Frequency Ultrasonics. In *Proceedings of the 21st International Conference on Intelligent User Interfaces (IUI '16)*. ACM, 332-335. DOI: http://dx.doi.org/10.1145/2856767.2856812
6. Roman Lissermann, Jochen Huber, Aristotelis Hadjakos, and Max Mühlhäuser. 2014. EarPut: augmenting ear-worn devices for ear-based interaction. In *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: the Future of Design (OzCHI '14)*. ACM, 300-307.

   DOI=http://dx.doi.org/10.1145/2468356.2468592
7. Hiroyuki Manabe and Masaaki Fukumoto. 2006. Full-time wearable headphone-type gaze detector. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems (CHI EA '06)*. ACM, 1073-1078. DOI=http://dx.doi.org/10.1145/1125451.1125655
8. Katsutoshi Masai, Yuta Sugiura, Masa Ogata, Kai Kunze, Masahiko Inami, and Maki Sugimoto. 2016. Facial Expression Recognition in Daily Life by Embedded Photo Reflective Sensors on Smart Eyewear. In *Proceedings of the 21st International Conference on Intelligent User Interfaces (IUI '16)*. ACM, 317-326.

   DOI=http://dx.doi.org/10.1145/2856767.2856770
9. Kohei Matsumura, Daisuke Sakamoto, Masahiko Inami, and Takeo Igarashi. 2012. Universal earphones: earphones with automatic side and shared use detection. *In Proceedings of the 2012 ACM international conference on Intelligent User Interfaces (IUI '12)*. ACM, 305-306. DOI=http://dx.doi.org/10.1145/2166966.2167025
10. Christian Metzger, Matt Anderson, and Thad Starner. 2004. FreeDigiter: a contact-free device for gesture control, In *Proceedings of the 8th International Symposium on Wearable Computers (ISWC '04)*. IEEE, 18-21.

    DOI=10.1109/ISWC.2004.23
11. Hiromi Nakamura and Homei Miyashita. 2010. Control of augmented reality information volume by glabellar fader. In *Proceedings of the 1st Augmented Human International Conference (AH '10)*. ACM, Article 20, 3 pages. DOI=http://dx.doi.org/10.1145/1785455.1785475
12. Kei Nakatsuma, Hiroyuki Shinoda, Yasutoshi Makino, Katsunari Sato, and Takashi Maeno. 2011. Touch interface on back of the hand. In *ACM SIGGRAPH 2011 Emerging Technologies (SIGGRAPH '11)*. ACM, Article 19, 1 pages. DOI=http://dx.doi.org/10.1145/2048259.2048278
13. Masa Ogata, Yuta Sugiura, Yasutoshi Makino, Masahiko Inami, and Michita Imai. 2013. SenSkin: adapting skin as a soft interface. In *Proceedings of the 26th annual ACM symposium on User interface software and technology (UIST '13)*. ACM, 539-544. DOI: http://dx.doi.org/10.1145/2501988.2502039
14. PSVM: Support Vector Machines for Processing. http://makematics.com/code/psvm/
15. Marcos Serrano, Barrett M. Ens, and Pourang P. Irani. 2014. Exploring the use of hand-to-face input for interacting with head-worn displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, 3181-3190. DOI=http://dx.doi.org/10.1145/2556288.2556984
16. Julia Schwarz, Chris Harrison, Scott Hudson, and Jennifer Mankoff. 2010. Cord input: an intuitive, high-accuracy, multi-degree-of-freedom input method for mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, 1657-1660. DOI=http://dx.doi.org/10.1145/1753326.1753573
17. Martin Weigel, Vikram Mehta, and Jürgen Steimle. 2014. More than touch: understanding how people use skin as an input surface for mobile computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, 179-188.

    DOI=http://dx.doi.org/10.1145/2556288.2557239
18. Martin Weigel, Tong Lu, Gilles Bailly, Antti Oulasvirta, Carmel Majidi, and Jürgen Steimle. 2015. iSkin: Flexible, Stretchable and Visually Customizable On-Body Touch Sensors for Mobile Computing. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, 2991-3000.

    DOI=http://dx.doi.org/10.1145/2702123.2702391
19. Jacob O. Wobbrock, Andrew D. Wilson, and Yang Li. 2007. Gestures without libraries, toolkits or training: a $1 recognizer for user interface prototypes. In *Proceedings of the 20th annual ACM symposium on User interface software and technology (UIST '07)*. ACM, 159-168. DOI=http://dx.doi.org/10.1145/1294211.1294238