

Fine-grained Hand-posture Recognition for Natural User-interface Technologies

Yuki Kubo

Abstract

NTT Service Evolution Laboratories is conducting research on user-interface technologies to improve system usability. As part of this research, we are focusing on new input methods that will be an important element of next-generation information devices, and in particular, we are researching hand-posture recognition technology to be used to implement input for glasses-based devices. This article describes initiatives related to this research.

Keywords: user interfaces, sensing technologies, hand-posture recognition

1. Introduction

Information devices are used to display information to users in all types of situations and have become indispensable in daily life. Most of these devices are either displays installed in public spaces or portable mobile devices such as smartphones. Many companies are also researching a new type of wearable information device that users will wear like glasses, referred to as smart glasses or augmented reality glasses. In contrast to previous devices, these glasses-type devices have characteristics that display information layered directly over the user's field of view.

Displaying information directly over the user's field of view enables this type of device to layer all types of information over the real world, presenting different information to each user. Information in public spaces, such as train stations and parks, has conventionally been presented through information displays that exist physically, such as signage installed in various locations in a train station to give directions, displaying the same information to everyone. However, with the ability of glasses-type devices to display different information to each user, all types of information that previously had a fixed existence in the real world can change dynamically to a form suitable for each user. The information each user

needs can be displayed to him/her, and the amount of information displayed can also be adjusted for the comfort of each user.

We are now connected to information more than ever before, so with information layered over the real world through glasses-type devices, it will be increasingly important to have an input method for such information, similar to how we input on a smartphone screen. This article introduces a technology for recognizing fine-grained hand postures, which is currently being investigated and will be necessary for developing natural user interfaces.

2. Related work

One input method that is promising for glasses-type devices involves gestures, i.e., using the state or motion of the hands. To make gesture input possible, the hand gesture has to first be recognized using a camera [1]. However, there are issues with using cameras: limited range where recognition is possible due to camera position and field-of-view and recognition difficulty if anything obstructs the camera's view (i.e., occlusion problem).

How hand gestures will be recognized is important, but the design of gestures used to perform input is also important. One can imagine using large arm

movements up, down, left, and right, and having the arms raised for long periods of time as types of hand gestures, but there are several problems with such gestures. Large movements or holding the arms up continuously will increase fatigue and place an increasing burden on the user over time. Large arm or body movements also have disadvantages in terms of social acceptance, such as being distracting to people nearby, being difficult when people or other obstacles are in close proximity, and allowing people nearby to see the gestures, thus allowing them to determine what the user is doing. As one solution for these problems, NTT Service Evolution Laboratories is focusing on input methods that use smaller hand gestures to reduce the burden on users and is researching a fine-grained hand-posture recognition technology that uses sensors attached to the back of the hand. Next, we introduce this technology, which is called AudioTouch [2], based on active acoustic sensing.

3. Technology overview

AudioTouch uses ultrasound waves through principles such as ultrasonography. It detects changes in the structure (e.g., bones, muscle, and sinew) in the back of the hand and uses them to estimate hand postures. Objects generally have their own resonance properties, which depend on features such as shape, material, and boundary conditions. A previous study [3] focused on how the resonance properties of an object change due to changes in its boundary conditions when the user touches it. They proposed a method for making touch interfaces from objects with sensors attached that recognize when the user touches the object. In contrast, AudioTouch focuses on changes in resonance properties due to changes in the shape of the hand itself due to movement of fingers and hands, including the internal structure of the hand. When the user changes his/her hand and finger postures, the resonance properties of the hand change due to the movements of parts of the hand, such as skin, bones, and muscles. By measuring the differences in these properties, the hand and finger postures can be recognized. Changes in resonance properties can be measured by sending oscillations (ultrasound waves) through the object and measuring the frequency response. AudioTouch observes the frequency response of the back of the hand on the basis of this concept to recognize changes in hand and finger postures. Specifically, it uses two piezoelectric sensors attached to the back of the hand (**Fig. 1**). The first emits ultrasonic waves both on the surface and inside

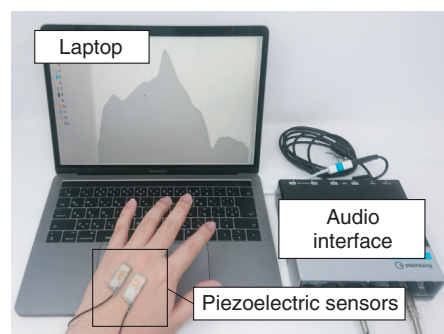


Fig. 1. AudioTouch components.

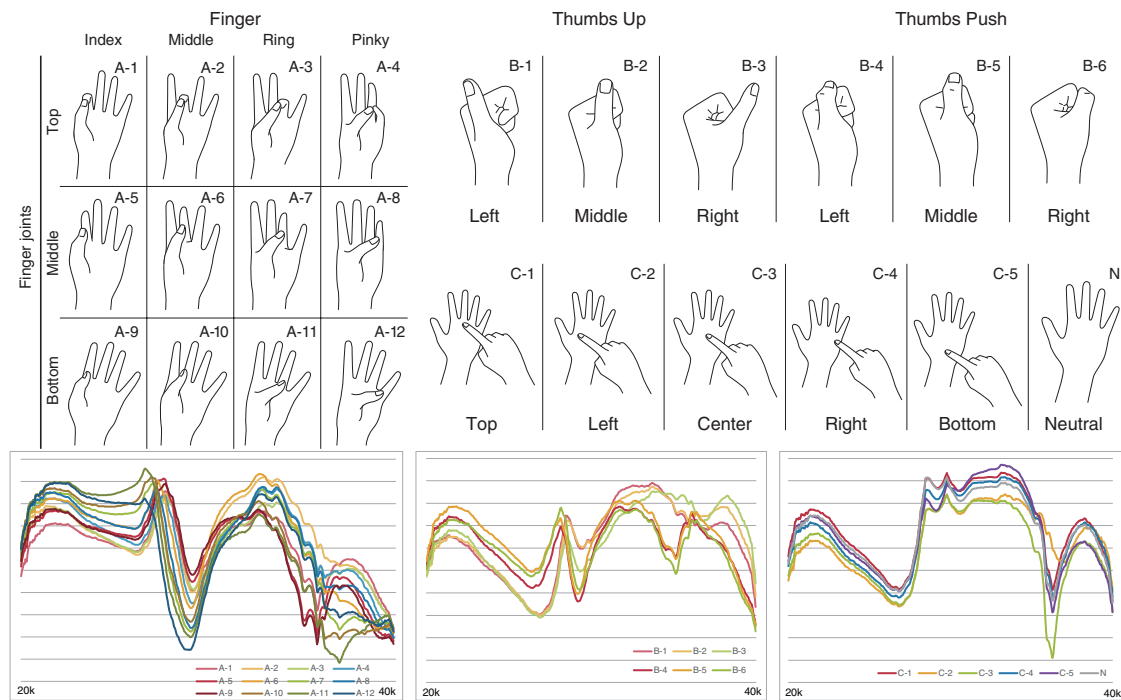
the hand, and the other measures the ultrasonic waves propagated over the surface and in the interior of the hand. The captured sound waves are analyzed to obtain the frequency response, and this is used to generate features, which in turn are used with machine learning to recognize hand and finger postures.

We conducted experiments to evaluate whether AudioTouch could recognize three gesture sets (**Fig. 2**). The three gesture sets were: one in which the thumb touches each of the finger joints, one in which the thumb is moved left, right, up, and down, and one in which a finger on one hand is used to touch the palm of the other. We also explored whether levels of pressure applied when pressing the thumb and other fingers together could be differentiated into two levels (i.e., strong and weak). Example applications for these gestures include using the palm as a touch-sensitive number keypad or for selecting menu items.

4. Conclusion

In this article, technologies for performing input with glasses-type devices were discussed and current initiatives in developing technology for recognizing fine-grained hand posture, as an element for such technologies, were presented. There are still technical and social issues to be resolved, and NTT Service Evolution Laboratories will continue to advance this research.

This work is partially supported by Japan Science Technology Agency (JST) Advanced Information and Communication Technology for Innovation (ACT-I) Project (JPMJPR16UA).



* Waveforms are examples of waveform-data averages for a given user and label.

Fig. 2. Examples of gesture sets and gesture waveforms.

References

- [1] D. Kim, O. Hilliges, S. Izadi, A. D. Butler, J. Chen, I. Oikonomidis, and P. Olivier, "Digits: Freehand 3D Interactions Anywhere Using a Wrist-worn Gloveless Sensor," Proc. of the 25th Annual ACM Symposium on User Interface Software and Technology (UIST 2012), pp. 167–176, Cambridge, MA, USA, Oct. 2012.
- [2] Y. Kubo, Y. Koguchi, B. Shizuki, S. Takahashi, and O. Hilliges "AudioTouch: Minimally Invasive Sensing of Micro-gestures via Active Bio-acoustic Sensing," Proc. of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI 2019), p. 13, Taipei, Taiwan, Oct. 2019.
- [3] M. Ono, B. Shizuki, and J. Tanaka, "Touch & Activate: Adding Interactivity to Existing Objects using Active Acoustic Sensing," Proc. of the 26th Annual ACM Symposium on User Interface Software and Technology (UIST 2013), pp. 31–40, St Andrews, Scotland, UK, Oct. 2013.



Yuki Kubo

Cybernetic Intelligence Research Project, NTT Service Evolution Laboratories.

He received an M.E. in systems and information engineering from University of Tsukuba in 2018. He joined NTT Service Evolution Laboratories in 2018. His current research interests include human-computer interaction and input techniques. He is a member of the Association for Computing Machinery, the Virtual Reality Society of Japan, and Information Processing Society of Japan.