Dim NK, Ren X. Designing motion gesture interfaces in mobile phones for blind people. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 29(5): 812–824 Sept. 2014. DOI 10.1007/s11390-014-1470-5

Designing Motion Gesture Interfaces in Mobile Phones for Blind People

Nem Khan Dim and Xiangshi Ren* (任向实), *Senior Member, ACM, IEEE*

School of Information, Kochi University of Technology, Kochi 782-8502, Japan

E-mail: nemkdim@gmail.com; ren.xiangshi@kochi-tech.ac.jp

Received February 28, 2014; revised July 23, 2014.

Abstract Despite the existence of advanced functions in smartphones, most blind people are still using old-fashioned phones with familiar layouts and dependence on tactile buttons. Smartphones support accessibility features including vibration, speech and sound feedback, and screen readers. However, these features are only intended to provide feedback to user commands or input. It is still a challenge for blind people to discover functions on the screen and to input the commands. Although voice commands are supported in smartphones, these commands are difficult for a system to recognize in noisy environments. At the same time, smartphones are integrated with sophisticated motion sensors, and motion gestures with device tilt have been gaining attention for eyes-free input. We believe that these motion gesture interactions offer more efficient access to smartphone functions for blind people. However, most blind people are not smartphone users and they are aware of neither the affordances available in smartphones nor the potential for interaction through motion gestures. To investigate the most usable gestures for blind people, we conducted a user-defined study with 13 blind participants. Using the gesture set and design heuristics from the user study, we implemented motion gesture based interfaces with speech and vibration feedback for browsing phone books and making a call. We then conducted a second study to investigate the usability of the motion gesture interface and user experiences using the system. The findings indicated that motion gesture interfaces are more efficient than traditional button interfaces. Through the study results, we provided implications for designing smartphone interfaces.

Keywords design, motion gesture, user-defined study

1 Introduction

There are 39 million blind people and 246 million low-vision people around the world①. For these disabled people, mobile devices have become indispensable, empowering them for both leisure and more independent living. Mobile phones are the most commonly carried devices by blind people in their daily lives[1]. However, most people in this user group prefer old-fashioned mobile phones with familiar layouts and tactile buttons to modern smartphones. Touch screen interfaces in smartphones primarily require users to look-for-interaction sensitive areas on the screen and that is a major challenge for blind people because the interfaces on the screen are invisible to them. Existing assistive solutions such as screen readers, vibration, sound and speech output, are still not adequate or efficient enough to deliver all smartphone affordances to blind users. These interfaces require the blind user to memorize and browse soft-buttons on the screen. Nowadays, motion gestures have been gaining attention as more natural and intuitive interfaces that support distracted inputs and require less visual attention. We believe that motion gestures designed to logically map the users' mental model can offer more learnable and accessible interfaces for blind people.

Most blind people are not smartphone users and they are aware of neither the affordances available in smartphones nor the potential interactions available through motion gestures. Furthermore, they have a slim or no chance to see motion gestures performed by other people. We were thus motivated to find the best practices to design the usable motion gestures for blind people. With respect to this research goal, we conducted a user-defined study where the participants were asked to define their own gestures to invoke some common tasks in a smartphone, and to mention the rationale and heuristics for the gestures they performed. We set three research questions for the user-defined study. 1) Are motion gestures usable as mobile interactions for

Regular Paper

This study has been partially supported by the Grant-in-Aid for Scientific Research of Japan under Grant Nos. 23300048, 25330241, and the National Natural Science Foundation of China under Grant No. 61228206.

*Corresponding Author

①http://www.who.int/mediacentre/factshects/fs282/en/, Aug. 2014.

blind people? 2) What unique motion gestures do the special characteristics of blind-users inspire for mobile interactions? 3) What are the heuristics of gestures produced by blind people, and how can these heuristics be described as principles of gesture design for blind users?

Using gestures suggested by blind participants in the user-defined study, we developed motion gesture interfaces with speech and vibration feedback for browsing phone books and making a call. The second study is concerned with how the users actually experience the use of motion gesture interfaces. In the study, the participants used both a smartphone with motion gesture interfaces and a feature phone with button interfaces to browse contacts and make a call. We then analyzed the task completion time and errors using the two kinds of interfaces. We also collected subjective assessments and comments from the participants.

2 Literature Reviews

Studies related to our first study include the classification of human gestures, motion gestures for mobile interactions, and user-defined studies. For the second study, we presented related studies regarding mobile eyes-free interfaces for blind people.

2.1 Classification of Human Gestures

Gestures are spontaneous movements of hands and other body parts that people do to express their mental images and thoughts. Gestures can convey communicative information as speech[2], and they are performed in parallel with verbal expression at the conceptual level of communication[3]. Gestures serve as some cognitive functions for communication, and they are performed regardless of individual abilities or the impacts produced by the gestures generated. Even individuals who have been blind from birth and have thus never seen anyone else's gestures spontaneously express themselves in gestures. Studies have shown that blind people gesture even when they are conversing with other blind people[4].

Poggi[2] proposed a procedure for the generation of gestures. Gestures are generated by taking into account the meaning intended and the cognitive construction of the gesture to be made. When people gesture to communicate some meaning, *codified gesture* is first considered. Codified gestures are gesture-meaning pairs constantly represented in the mind and standardized by repeated use, regulations and social conventions. Such gestures are spontaneously shared and understood by everyone. If a codified gesture is not readily found in memory, creative *metaphoric gestures* are resorted to. Evidently, these metaphoric gestures are generated by mimicking daily human actions (i.e., biological) or by the similarity in visual resemblances (i.e., iconic). If the intended meaning is the information in a human's mind, *gestural mind makers* that can represent beliefs, goals and emotions of the human mind are invoked. Finally, gestures are arbitrary when signals and meaning are linked by neither a relationship of similarity nor any other kind of relationship.

In the literature of human computer interaction, gesture-based interactions have become a leading trend in natural user interface development[5]. Studies including [6] and [7] investigated user gestures for different applications. Despite the general classification of human gestures discussed above, the gesture taxonomies presented by user studies vary depending on the applications and users. Thus, in order to design motion gestures for mobile interactions that are usable for blind people, it is vital to understand the gesture producing mechanisms and gesture taxonomy of this user group.

2.2 Motion Gestures for Mobile Interactions

The earliest system that uses device tilt as an input mechanism was proposed by Rekimoto[8]. Using device tilt as a 3D (3-dimentional) motion gesture, he presented interaction techniques for several functions ranging from menus and scroll bars to more complicated functions such as map browsing. Today mobile phones are integrated with a set of motion sensors such as accelerometers and gyroscopes to detect 3D movements of the device for interactions. Motion-based mobile interactions have thus become an emerging research interest to many researchers. Many studies have proposed the use of motion gestures for distracted inputs, namely, eyes-free control of media players[9], map navigation[8], text input[10-11], cursor control[12] and user verification[13]. We believe that these natural and intuitive mobile interactions can be a benefit for blind people. This study was thus motivated to elicit the most preferable and ergonomic gestures for such interactions.

2.3 User-Defined Studies

User-defined study has been a recommending and maturing practice in human-computer interaction research. The core idea of this human-based approach is that users must be understood so that the system can be adapted to the users instead of requiring the users to adapt to a given interface. The major benefit of a user-defined study is the higher likelihood of designing interfaces that are easy to perform and to remember.

Many user-defined studies have been conducted especially for gesture-based natural interactions. Wobbrock *et al.*[6] presented a user-defined study where participants were shown the effect of a gesture and then were asked to perform gestures for commands in surface computing. Inspired by the contributions of this study that include gesture taxonomy and implications for surface computing, many similar studies have been performed for various applications and computing environments. These include user-defined studies for device to device interactions[14-15], surface and motion gestures for 3D manipulation of objects through mobile devices[16], human-robot interaction[17], free-hand TV control[18], deformation-based interface[19], augmented reality[20], and gesture sets for people with communication difficulties[21].

Kane *et al.*[22] presented a gesture elicitation study where touch screen gestures performed by blind and sighted people were compared. This study found that blind people have different gesture preferences from sighted people, and reported design guidelines for accessible touch screen interfaces. The inclusive design guidelines presented by this study are specifically accommodated to touch screen devices. On the other hand, Ruiz *et al.*[7] presented an elicitation study of motion gestures for mobile interaction. This study reported a consensus of motion gestures to invoke commands on a smartphone and presented a taxonomy of motion gestures. However, the study did not cover people with visual disability even though blind people are one of the largest potential user groups of motion gesture interfaces. Blind and sighted people have different visual experiences and daily activities that can affect the expression of their mental images as gestures. We identified the differences between gestures performed by the two user groups through gesture taxonomy and user-defined gesture sets, based on the previous study[7].

2.4 Mobile Eyes-Free Interfaces for Blind People

Some past research studies have attempted to provide blind people with more access to mobile devices, including smartphones and touch screen-based systems. Kane *et al.*'s slide rule[23] provides a specialized touch interface optimized for non-visual interaction. The slide rule is a set of audio-based multi-touch interaction techniques that enable blind users to access smartphone functions including making phone calls, mailing and music functions. The talking tactile tablet[24] uses speech and tactile overlay to provide audio and tactile feedback to users. This system allows users to explore a 2-dimensional (2D) space and provides feedback as the user probes with a finger or stylus. Pirhonen *et al.*[25] and O'Neill *et al.*[26] proposed eyes-free mobile interfaces that use directional gestures to perform basic operations on mobile devices. Zhao *et al.*[27] developed EarPod using touch input and relative output for eyes-free menu selection. Audio-based text entry systems were also developed by Sánchez and Aguayo[28] and Yfantidis and Evreinov[29]. These systems allow users multi-tap and directional gestures, and to enter text on touch screens using audio feedback provided to the user's entry. Azenkot *et al.*[30] also presented the input finger-detection (IFD) text entry method for blind people. IFD uses the 6-bit Braille encoding with audio feedback, for eyes-free text entry.

All these studies attempted to provide eyes-free access to mobile devices using touch gestures. Indeed, touch-based interfaces are neither the only way nor always the best way. They still require the users to have good spatial ability and awareness of the device and its interfaces. Motion gestures can offer simpler and more efficient ways to interact in many cases, for example, making or answering a call. By contrast with previous studies, our system provides eyes-free interfaces for browsing the contact list and making calls, using motion gestures, haptic feedback and speech output. Fig.1 illustrates the summary of related work on mobile eyes-free interfaces for blind people, and the position of our study. Most of the previous studies were built around tactile or touch gesture input. There is very little work on motion gesture for basic mobile phone functions.

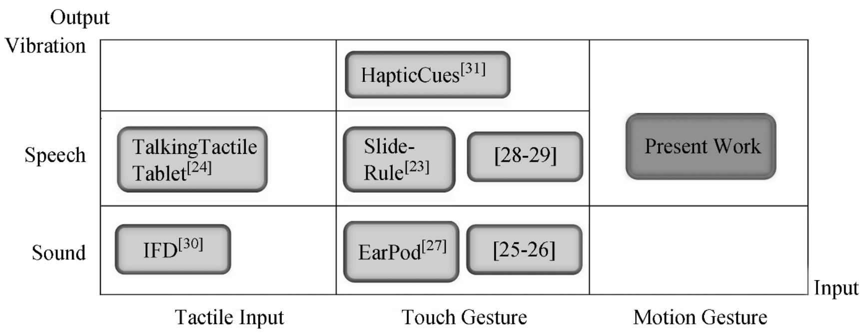

Fig.1. Mobile eyes-free interfaces for blind people.

3 Study 1: User-Defined Motion Gestures

User-defined studies[2] were conducted to elicit user behaviors to enhance the design process. During the study, we explored motion gestures from 13 blind participants. The participants were asked to perform motion gestures that could be used to command the available functions on a smartphone. We presented 15 tasks to the participants. The participants were asked to use a think-aloud method and to perform gestures in think-aloud protocol. They were also asked to supply subjective preference ratings for each gesture they performed.

All the sessions were video-recorded and each session took approximately one hour to complete. We made clustered quotes of each participant from the video transcript. A careful analysis was then performed to classify each motion gesture designed by the participants. Each gesture was labeled with a corresponding rationale and each was generalized to identify common characteristics shared by the participants. For this purpose, we adopted a bottom-up inductive analysis approach where we analyzed the specific cases of each participant to identify common themes.

3.1 Participants and Apparatus

For the experiment, we recruited 13 blind people from a local blind association (9 males and 4 females). The ages ranged from 25 to 77 with the mean value (*mean*) of 61 and the standard deviation (*SD*) of 16.91. Three of them could see light and two could see objects, but none of these were able to distinguish between objects. The rest were totally blind. Two of our participants were smartphone users. All the participants were right-handed, and each was paid $10 for their participation.

We used a Samsung Galaxy smartphone to define the participant's motion gestures. Participants were video-recorded while performing gestures and two experimenters took detailed notes for the think-aloud data.

3.2 Experimental Tasks

The tasks were categorized into action and navigation. Each was subcategorized into phone application (e.g., answering a phone call or switching to a previous application) and particular applications (e.g., navigating a map). We paid specific attention to explaining some tasks like zooming functions in map navigation. For example, just saying the function zoom in as for enlarging objects would be unreasonable for our participants because they cannot see objects. Instead, we made them understand that enlarging an object (an onscreen menu or a location point on a map) on a screen can help them more easily target or select that object. Furthermore, two of our participants with low vision used a magnifying glass. This encouraged us to include these tasks. The experimental tasks we presented to each participant are described in Table 1.

Table 1. Experimental Tasks Presented to the Participants

Category	Sub-Category	Task Name
Action	Phone	Answer call
		Hang up call
		Ignore call
		Voice search
		Place call
	Application	Act on selection
Navigation	Phone	Home screen
		Next contact
		Previous contact
	Application	Pan left
		Pan right
		Pan up
		Pan down
		Zoom in
		Zoom out

3.3 Procedures

We started each experimental session by explaining the purpose and procedures of the study and also the think-aloud method to the participants. Then the participant was handed a smartphone and was asked to perform a gesture for each task. The tasks were grouped into three sets of similar tasks. For example, tasks for normal use of the phone such as calling, answering and muting a call were grouped into the same set. Because our participants could not be presented with any visual description, an experimenter read aloud the descriptions and explanation about each task carefully.

Because repeating every defined gesture for confirmation was tiring, we did not ask the participants to repeat their gestures. Instead, we carefully captured the expressed gesture and tried to confirm each gesture and rationale. After each group of tasks, the participants were asked to rate the gestures they performed using a 7-point Likert scale to indicate their agreement on the criteria (1 stands for strongly disagree and 7 for strongly agree):

- The gesture I made is a good match for its intended use.
- The gesture I made is easy to perform.

In order to assess the impression on using motion gesture interactions of our participants, we added an

[2]Users are asked to define the way (with its reasons) to invoke a system function. The users are first depicted the effect of the system function, and then asked to perform the operation that causes the function.

item, "I would often use this gesture if it existed" on a 7-point scale ranging from never (1) to very often (7). To conclude each session, we asked the participants to suggest additional use cases. If they had any suggestion, we encouraged them to perform gestures for the tasks suggested. The interview ended with the experimenter asking the participants if they had any questions, suggestions or comments. We recorded every comment or suggestion of the participants for later analysis.

3.4 Results and Analysis

After detailed analysis of the data we collected from our study, we presented the study results including a user-defined gestures set, motion gestures of the blind, physical characteristics of the gestures, subjective responses and open-ended use cases.

3.4.1 User-Defined Gesture Set

From the gestures collected, we grouped identical gestures for each task and selected the largest group as the user-defined gesture for the task. We adopted Wobbrock *et al.*'s method[6] in order to investigate the extent of agreement for each task. We calculated the agreement score for each task using the formula:

$$A_t = \sum\nolimits_{P_i} \left(\left|\frac{P_i}{P_t}\right|\right)^2, \tag{1}$$

where t is a task in the set of all tasks T, P_t represents a set of gestures proposed for task t, and P_i is a subset of identical gestures from P_t. The value of agreement score of task t (A_t) ranges from 0 to 1. Fig.2 depicts the agreement score for each task. For example, the agreement score for "Answer Call" can be calculated as:

$$A_{\text{AnswerCall}} = \left(\frac{8}{13}\right)^2 + \left(\frac{3}{13}\right)^2 + \left(\frac{1}{13}\right)^2 + \left(\frac{1}{13}\right)^2 = 0.443\,7.$$

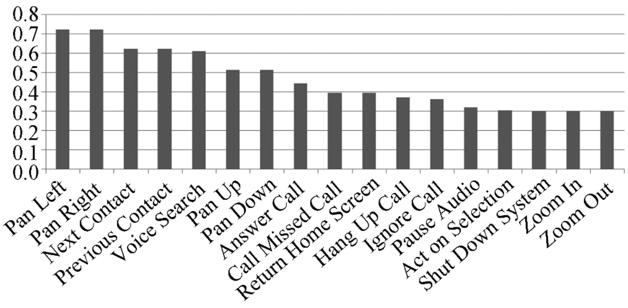

Fig.2. Agreement score of each task sorted in descending order.

The user-defined gesture set is illustrated in Fig.3. Among use-cases suggested during the interview, we included gestures for "Application Pause" and "Power Off" in the gesture set because most of our participants performed gestures for these commands and there was obvious consensus among the participants. On the other hand, among the experimental tasks, gestures for zooming (i.e., Zoom In and Zoom Out) could not be presented in the user-defined gesture set. Although the participants used similar rationales for creating gestures, the gestures were still performed differently. Unlike sighted people, blind people cannot have visual references that are shared with others. Thus, different gestures were generated as being influenced by daily experiences.

Being encouraged to freely express the most inspired gestures, our participants had the same preferences of non-motion gestures for some tasks. For example, gestures for "Ignore Call", "Pause" and "Power Off" were seen to cover or sweep the phone screen by hand with the common reason being "stop the sound" or "finish". This informed us that the most ergonomic interactions need combinations of different models optimized for a specific context.

3.4.2 Motion Gestures of Blind People

We analyzed the motion gestures collected and grouped the gestures into 4-fold taxonomic themes. The themes include natural and intuitive gestures, real-world metaphors, natural consistent mappings, and arbitrary gestures.

Natural and Intuitive Gestures. As noted in the literature reviews[2-3], some codified gestures are constantly presented to peoples' minds. Here the natural gestures we called are subsets of codified gestures. These gestures are more natural than every other gesture for the intended meaning, and the purposed meaning can be inferred without learning. For example, bringing the mobile phone to the ear for answering a call, or bringing the phone to the mouth for making a voice search is standardized by repeated use, and thus becomes intuitive motion to everyone. Evidently, most of the participants designed the same motion gestures for tasks that have such codified gestures. In our study, for making a voice search, 8 out of 13 participants designed their motion gestures by bringing the smartphone to the mouth. The common reason for choosing that gesture was described as "natural".

Real-World Metaphors. For the tasks where codified gestures cannot readily be found, the participants tried to generate creative gestures. It is obvious that generating creative gestures is primarily influenced by real-world metaphors which occur in daily lives. Blind people are primarily influenced by what they *do* on a daily basis. Gestures performed by blind people are linked to their meaning by mechanical determinism (i.e., daily

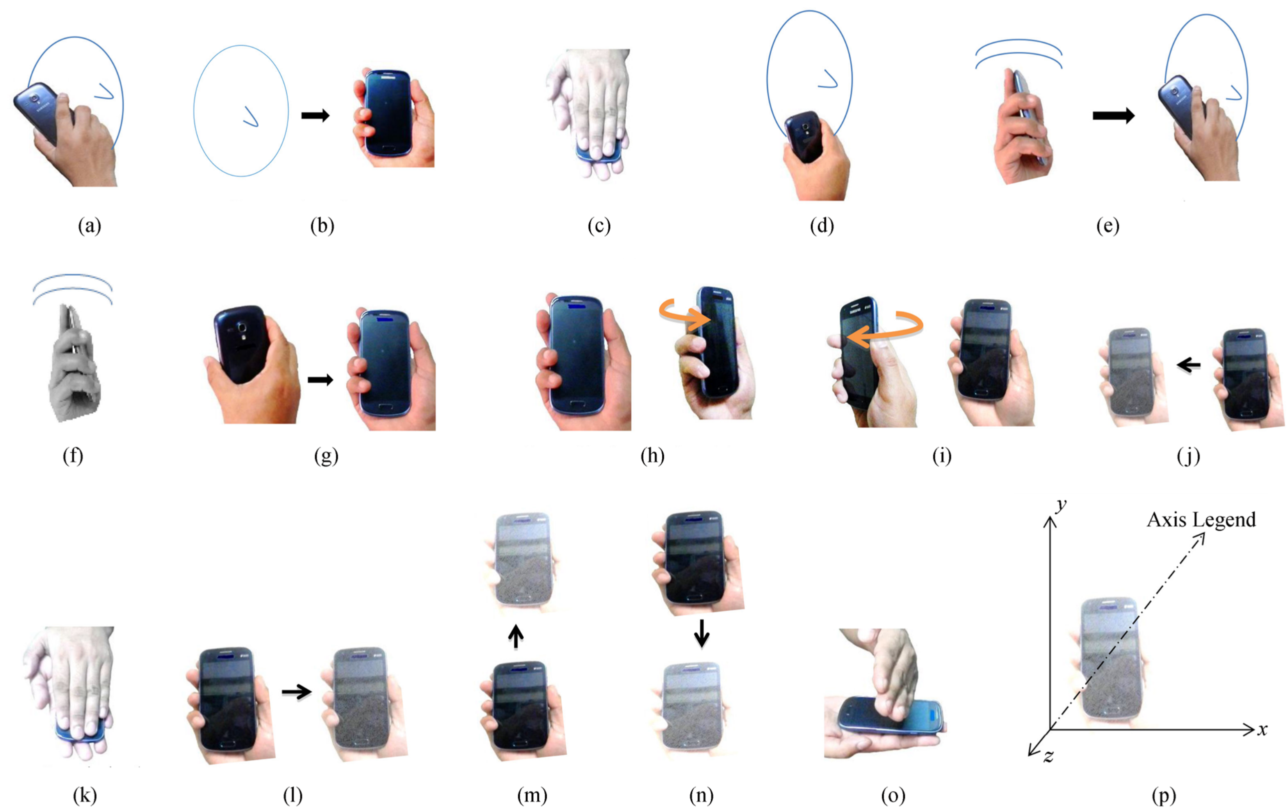

Fig.3. User-defined gesture set. Commands for zoom in and zoom out are not included due to the lack of agreement among the participants. Blue curve shows front-back movement of shake, black arrow indicates state change of the phone or direction of movement, and brown-bold arrow indicates the direction of rotations. (a) Answer call: bring phone to ear. (b) Hang up call: remove phone from ear. (c) Ignore call: cover by hand. (d) Voice search: bring to mouth. (e) Call missed call: shake front-back and then ring phone to ear. (f) Act on selection: shake front and back. (g) Home screen: turn back and front the phone. (h) Next: rotate flip along y-axis to right. (i) Previous: rotate flip along y-axis to left. (j) Pan left: move to left. (k) Pause: cover by hand. (l) Pan right: move to right. (m) Pan up: move up. (n) Pan down: move down. (o) Power off: sweep from the top to the bottom. (p) Axis legend.

action) rather than by the similarity of visual resemblance (i.e., iconic). For example, to navigate to the home screen, half of our participants designed their gestures to flip the phone screen back to front (i.e., undo). The most common reason for that motion gesture was "returning to the original place." Also, being asked to perform gestures for zoom in, some of our participants designed their gestures to continuously rotate the phone up along the y-axis. Some others performed gestures by raising the smartphone, tending to increase the height of the device. For design reason, a majority of the participants shared the opinion of one of the participants: "Enlarging an object is like increasing the level of volume in a music player. It is increasing a level". We also noted that blind people tend to create gestures when mimicking the normal use of different devices that they use daily. For example, for panning tasks in map navigation, one of our participants mentioned his design rationale was to mimic the use of the cardinal directions in his compass.

Consistent Mapping. In general, consistent mapping is where opposed actions are achieved by reversed movements. For example, if a rotate flick to the right is gestured for "Application Next", the gesture for "Application Previous" is a rotate flick to left. In our study, we found that consistent mapping was mentioned very often by the participants as a design rationale for the gestures they made. For example, while performing gestures for "Hang Up Call", one of the participants flipped the smartphone forwards and stated his design rationale as: "I flip the phone backward and bring to the ear for answering a call. Then I will remove the phone from the ear and flip forward to hang up the call". Again, some other participants were found to generate related gestures for "Ignore Call" and "Place Call", "Pause" and "Power Off", etc. We speculated that blind people are more likely to use consistent mapping than sighted people to arrange tasks in more memorable and accessible ways. This occurred more obviously when suggesting open-ended use cases and performing gestures for the suggested tasks. It is arguable that consistent mapping is a way of grouping or relating gestures but not a specific type of gesture in and of itself. However, appreciating this often-used methodology for creating gestures by our participants, we presented consistently mapped gestures as a specific gesture type for this context.

Arbitrary Gestures. Arbitrary gestures are those not linked to the meaning by either the similarity of real-world metaphor or any other relationship. During the study, we found that most of the gestures performed by our participants were labeled with a relevant rationale that came from their daily experiences. However, in some cases, some of the gestures performed with a rationale were still difficult to infer without learning the reason. We treated those gestures as arbitrary gestures.

3.4.3 Percentages of Gesture Types

Fig.4 illustrates the taxonomic decomposition of 208 gestures collected during our study. We classified the user gestures by the gesture types discussed above, namely, natural gestures presented in long-term memory, metaphoric gestures generated from daily-life experiences, consistently mapped gestures and arbitrary gestures. As seen in the figure, most gestures are metaphoric gestures generated from daily experiences. Consistently mapped gestures also take up a considerable percentage of the gestures produced. Most task suggestions during open interviews came from consistent mapping, including use cases to "pause" the screen reader or music player, "resume", "close application" and "power off" the system.

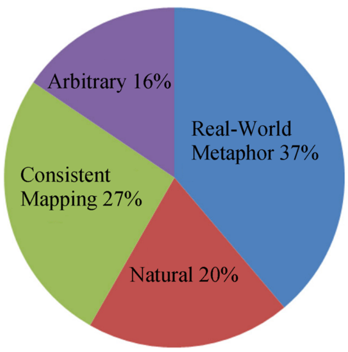

Fig.4. Percentages of gestures included in each gesture type.

3.4.4 Physical Characteristics of the Gestures

Regarding the physical characteristics of gestures, we found that our blind participants used large movements of the hands to produce gestures, tending to have high kinematic impulses. We paid attention to the physical characteristics of gestures performed by our participants and used the videos recorded during the study for close analysis. We defined motions performed using only the wrist as small gestures, and defined those performed using both the wrist and the elbow as large gestures. Our participants were found to mostly perform large gestures. This can be partially due to the lack of feedback in our study which would have confirmed that their gestures had been recognized. Despite this, we speculated that blind people are more likely to use large movements as they treat the movements themselves as feedback to their actions instead of visual feedback. In any case, it is still questionable whether the participants will still use large movements to create gestures when feedback is provided. We answer this question in the second study.

Because the physical characteristics of gestures are one of the main concerns when designing interfaces and supporting sensors, we also analyzed the physical characteristics of the gestures in terms of dimension and complexity. Here dimension means the number of axes involved in the movement while performing gestures. Single-axis motions include simple gestures like a flip or a flick. Motions that include a single rotation or translation are tri-axis gestures, and those including both translations and rotations of the device around its six degrees of freedom are defined as 6-axis gestures. The complexity is concerned with whether the gesture is a simple single gesture or a compound gesture that is composed of more than one single gesture. Fig.5 illustrates the percentages of gestures in each category of dimension and complexity. The gestures tend to include more translations and rotations than simple single motions.

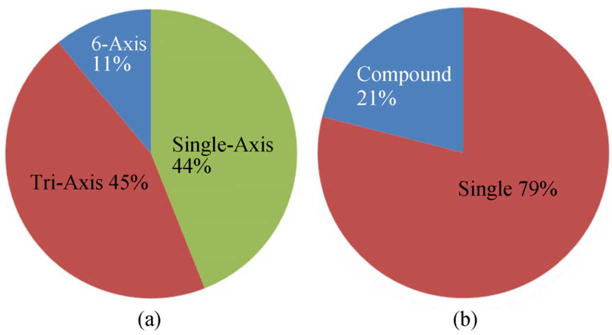

Fig.5. Physical characteristics of gestures. (a) Dimension. (b) Complexity.

3.4.5 Subjective Responses

After each session of the experiment, the participants rated the gestures they performed in terms of goodness, easiness and frequency of use. Overall, the participants gave the gestures average scores of 6.14 (SD = 1.069) for goodness, 6.26 (SD = 0.41) for easiness and 6.10 (SD = 0.81) for frequency. All the ratings were found to be equally and relatively high. This convinces us that our participants were very receptive to motion gesture interactions.

3.4.6 Open-Ended Use-Cases

After the experiment, we asked the participants to suggest possible tasks where motion gesture interactions can be effective for them. We received considerable suggestions from some of the participants. One of them suggested "Application Pause" mentioning "Sometimes I want to pause the screen reader of my phone on the way. It is very annoying when I cannot easily find the button to do that". The same participant also suggested considering a motion gesture for "Power Off" stating the same reason, i.e., the difficulty in finding buttons. Another considerable use case was proposed by one participant who mentioned "It would be very great if I can define my preferred gestures for calling frequently used numbers. Sometimes, finding only 3 numbers for an emergency call is still difficult for me". Similarly, another participant mentioned "I use my phone more as a music player than making calls. It would be great if I can control the player using motion gestures."

Once we obtained tasks suggested by some participants, we asked them to perform gestures for the tasks they suggested. We then treated suggested use cases as additions for the next sessions asking the participants to define gestures in the think-aloud method, in order to find design heuristics for possible tasks. After performing a gesture for "Application Pause", one participant asked "If I can pause the audio player, motion gesture will be useful to resume the player". The same participant commented "Activating the screen reader (i.e., Act on Selection) is like pressing the OK button or the Enter key. So, having Cancel and Close Program functions would also be useful."

As mentioned before, most of the use cases suggested came from relating similar or opposed functions, for example, "Pause" and "Resume". Through the interviews for open-ended use cases, we noted that our blind participants were very alert to the possibility of motion gestures for more accessible mobile interaction. This awareness accordingly brings new potential and challenges to designers of smartphone applications and vendors of smartphone devices.

3.4.7 Differences from Gestures of Sighted People

Ruiz *et al.*[7] performed a user-defined study for motion gesture interaction that was applied in sighted people. Since blind people and sighted people have different visual experiences, we hypothesized that they would have differences in the gesture generation process and in the gestures themselves. Because we could not have full access to the dataset of the previous study by Ruiz *et al.*, it is difficult to do a detailed comparison of the gestures between blind people and sighted people. However, it is worth highlighting some significant differences through the user-defined gesture set and gesture taxonomy.

Gesture Taxonomy. Ruiz *et al.*[7] presented 4-fold taxonomic themes for motion gestures in their study with sighted people: real-world metaphor, physical (direct manipulation), symbolic (visually depicting a symbol) and abstract gestures. Indeed, symbolic gestures and physical gestures primarily rely on visual capability. Our blind participants did not perform symbolic or physical gestures, and we could not include them in the taxonomic themes of gestures. Instead, we added natural gestures and consistently mapped gestures in the taxonomic themes of our study.

Gesture Generation Process. In Ruiz *et al.*'s study[7], for "hanging up the call", the user-defined gesture of sighted people was found to remove the phone from ear and rotate the screen like hanging up the phone receiver on a telephone. None of our participants mentioned rotating the screen for hanging up the call because they rarely consider whether the screen is facing back or front. Rather, the majority agreement was just "removing from the ear and putting somewhere". This implies that blind people may not be aware of some visually demanding actions that sighted people do. Also, sighted people's gestures were found to mimic the use of a magnifying glass for "Zoom In/ Zoom Out" tasks, and shaking the phone to "Return to Home Screen" with the reason of clearing current contents. On the other hand, gestures performed in our study were linked to the meaning by mechanism determinism, for example, drawing circles for zooming and flipping the phone front and back for returning to home screen. Again, Ruiz *et al.* reported that most of the gestures performed by their participants were slight flips because the participants concerned about the visual feedback on the screen. On the other hand, the physical characteristics of gestures in our study were large movements including more rotations and translations. This implies that the differences in visual capability, daily experiences, daily device uses and expected feedback make differences in the gestures performed by blind and sighted people. These differences are worth taking into consideration when designing motion gesture interfaces.

3.5 Discussion and Implications

In this subsection, we discuss some broader implications of our results for motion gesture interactions and mobile interactions.

For blind users, one of the primary reasons for accessibility problems in smartphones is the difficulty they have in learning the interfaces on an invisible screen. For the sake of the learnability and memorability of motion gestures for blind people, designers should consider consistent mapping where related or reverse gestures are available for similar or opposed tasks. For example, designing relative gestures for "Application Pause", "Program Close" and "Power Off" can make the interfaces more learnable because they are more logical to the blind.

As noted, gesture generation by blind people was primarily influenced by metaphors from daily life, thus we argue that the most usable and memorable motion gestures are those designed to best reflect real world metaphors from the users' daily lives. This design implication also indicates the need of the participation of representive blind users in design processes. Also, designers should pay specific care not to include gestures that are unexpected by these users. As discussed earlier, blind people are sometimes not aware of the visual-based actions that sighted people perform (for example, a gesture for hanging up the call). Designers should avoid gestures including these kinds of actions. Wherever possible, symbolic gestures and the direct depiction of visual objects should also be avoided.

Regarding the physical characteristics of gestures, we found that our participants used large movements to generate gestures so that their gestures were undoubtedly recognizable enough. With respect to the physical characteristics of gestures performed by blind people, the demand for motion accuracy should be reduced. This means that the gesture recognition and the supporting sensors should allow flexible freedom of movement to perform gestures for interactions.

The user-defined gesture set of our study also informed us that the most ergonomic interactions need combinations of different modals optimized for a specific task or context. Thus, various sensors integrated with today's smartphones should be used to support multimodal inputs and outputs for the most intuitive and natural interactions.

Recalling suggestions in open-ended interviews, gesture customization is a very acceptable and beneficial interface option for blind users. More customizable motion gestures should be available for simple tasks on smartphones.

4 Study 2: Motion Gesture Interface for Making Calls

In the second study, we investigated the usability of motion gesture interfaces implemented by gestures suggested in the first study. In this study, the participants used both a smartphone with a motion gesture interface and a feature phone with a button interface, to browse contacts and make calls. The study was motivated by three research questions: 1) Do motion gesture interfaces provide more efficient use of smartphones compared to traditional feature phones? 2) What do blind users actually experience when using motion gesture interfaces? 3) What design implications can be learned for smartphone assistive interfaces?

4.1 Motion Gesture Interfaces

Besides the user-defined motion gestures set, the first exploratory study also provides some design insights to consider when developing interfaces for the second study. We identified three common guidelines for designing the motion gesture interfaces for smartphones. First, feedback was provided to every gesture the participant made. In the first study, gestures were found to be created by using large movements. It was questionable whether the lack of feedback affected the gestures performed and if the participants still use large movements for gestures when feedback is provided. The interfaces in the second study were thus designed to provide vibration feedback or speech feedback to each gesture input. Second, motion gestures were designed to minimize the need for motion accuracy. This is related to the first consideration. The physical characteristics of gestures performed in the first study suggested that we should allow users to have more freedom of movement for doing gestures. Third, motion gestures were designed, following consistent mapping. Understanding the users' reliance on consistent mapping in the first study, we designed the motion gestures to consistently map wherever possible.

Following the design guidelines above, we developed a set of interactions that allow users to browse a contact list and make a call nonvisually. The task selection for this study was based on the fact that browsing a phone book and making a call are two of the most common and fundamental functions available on a smartphone. Although there are advanced functions available on smartphones, even the most basic functions such as receiving and making a call still impose limitations on visually impaired people. These limitations cause other smartphone affordances to be out of reach for visually impaired users. We were motivated to investigate how motion gesture interactions enable this user group to access the most basic smartphone functions.

4.2 Participants and Apparatus

Eight participants (2 females, 6 males) were recruited. The ages ranged from 22 to 49 years (*mean* = 28.37 years). Three of the participants were blind and five of them were blind-folded during the study. All of the blind participants had participated in the first study. Two of them are totally blind and one can see light. All participants are right-handed. Each was paid $10 for their time and effort for the experiment. Fig.6 shows the participant using the motion gesture interfaces.

Fig.6. Participant using motion gesture interface which provides vibration and speech feedback.

For the study, we used a Samsung Galaxy Nexus smartphone and a Panasonic EZ180 feature phone. For the motion gesture interface in the smartphone, we used custom software to recognize the participants' gesture and to perform corresponding functions. The custom software read the value of accelerometer sensor on the smartphone and the sensor values were mapped to the smartphone functions. The system was developed in Java Eclipse IDE (Integrated Development Environment). All the experimental sessions were video-recorded.

4.3 Procedures

Each experiment started with a practice session where the two systems in the smartphone and the feature phone were explained and demonstrated. Fig.7 describes the interactions used, and Table 2 describes feedback provided to each step of the experimental task.

In each trial, the participant started by browsing the contact list. Contact lists were set up in the same order

Table 2. Feedback Provided to Each Step of the Experimental Task

Step	Feedback
Browse phone book	Vibration to let the participants know that the gesture is recognized
Next contact	Speech — reading out the contact name
Previous contact	Speech — reading out the contact name
Make call	Vibration
Hang up call/home screen	Vibration

in the smartphone and the feature phone. The participants selected the contact name instructed by the experimenter by going next and previous. When the contact was selected, the participant made a call and held the phone for 5 seconds. They hung up the call after 5 seconds. The procedures occurred with the same steps and in the same order in both systems.

The participants were allowed to practice on the systems until they could successfully perform the tasks. At the end of the practice session, the participant was handed the smartphone and was asked to perform the experimental task. After completing the experimental task with the smartphone system, the participant was handed the feature phone and was asked to perform the same task. The participant performed two trials for each system.

During the study, task completion time and errors were recorded for each trial. Task completion time was the time elapsed from the moment the participant started browsing the contact list until they hung up the call. After all trials, participants were asked to complete a questionnaire about the two systems. Participants indicated their agreement with three statements about each system using 7-point Likert scale (1 for strongly disagree, 7 for strongly agree). The statements used in the questionnaire were:

- The system is easy to learn.
- The system is easy to use.
- Using the system is not tiring.

(a) (b) (c) (d) (e)

Fig.7. Motion gesture interfaces of making a call. (a) A flick gesture to browse phone book. (b) Flip motion to right is used to browse next contact. (c) Flip motion to left is used to browse previous contact. (d) Flip backward to select the contact and make a call. (e) Flip forward to hang up the call.

The experiment ended with the experimenter collecting all subjective comments and suggestions from the participants.

4.4 Result and Analysis

4.4.1 Task Completion Time and Errors

Task completion time and errors were recorded during the study. The participants completed the tasks faster with the motion gesture interfaces than with the button-based interfaces ($F_{(1,7)} = 8.761$, $p < 0.050$). The mean time for task completion by using the motion gesture interface was 16.56 seconds ($SD = 2.35$) while the mean time for button interface was 24.75 seconds ($SD = 9.17$). During the study, half of the participants started using the smartphone system while the other half started using feature phone. However, there was no learning effect or significant effect of the systems order on the performance of the participants. Fig.8 shows task completion time by using each system.

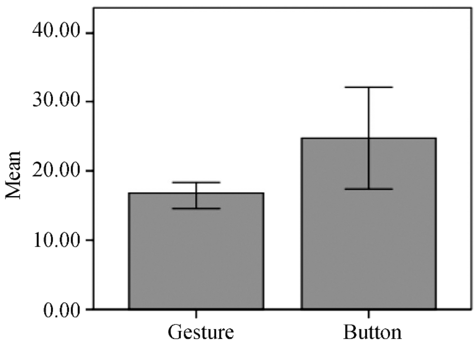

Fig.8. Task completion time for the two systems.

Overall, very few errors were made by using either system. The difference in error rates between the two systems was not statistically significant. The mean values of errors were 0.25 for the motion gesture interface ($SD = 0.462$) and 0.50 for the button interface ($SD = 0.925$) respectively. Blind-folded participants made more errors with button interface because they had less experience in using the feature phone. But blind participants were able to use both systems without any errors.

4.4.2 Subjective Responses

After the experiment, the participants completed questionnaires about the two systems. Among the three questions, "ease of use" was found to have statistical significance ($F_{(1,7)} = 6.818$, $p < 0.050$). Participants indicated that using the motion gesture interface was easier than using the button interface. The questions for "ease of learning" and "less fatigue" revealed no significant differences between the two systems. Fig.9 depicts the subjective responses of participants to the two systems.

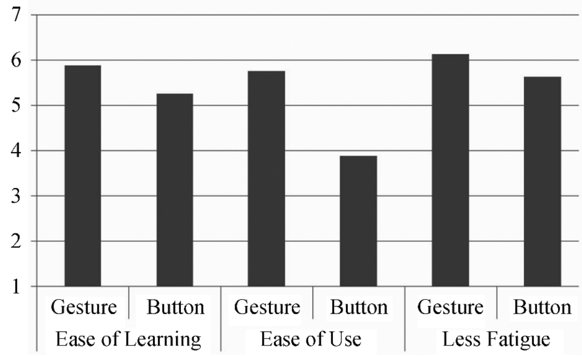

Fig.9. Subjective responses of participants to the two systems.

4.4.3 Qualitative Results

Following all the experimental sessions, participants answered the questions about their experiences using the two systems and any suggestions.

Motion Gesture Interfaces. Participants commented positively on the motion gesture interface. They mentioned that the motion gesture interface was more natural, and easier to remember and to use. As expected, blind-folded participants who are smartphone users preferred the motion gesture interface to the button interface. Two of the three blind participants preferred the motion gesture interface. One of them was neutral or positive about both systems, because she was familiar with the feature phone. One negative comment about the motion gesture interface was that it would become tiring while browsing many contacts. Furthermore, one of the blind participants commented that grouping contacts may be useful to reduce browsing time. He also commented that the touch-fling gesture for opening the contact book was very comfortable. One of the blind participants also commented that he was never confident when using a touch screen device, but motion gestures would be useful for blind users.

Button Interface. Most of the participants agreed that the button interface delivered slower response compared to the smartphone system, and they had difficulty in finding the buttons. Blind participants who are feature phone users mentioned that many buttons were arranged on a small space and it was difficult to distinguish the buttons.

Feedback. One noticeable comment from all participants is that vibration feedback was very understandable and comfortable. In the study, vibration feedback was provided to every action that had no speech output. The participants mentioned that vibration feedback made them sure that the gesture they performed was recognized by the system.

4.5 Discussion

The study results convinced us that motion gesture interfaces could offer successful mobile interactions that enable non-visual interaction for blind users. The positive comments about the motion gesture interface primarily focused on its simplicity and quick access. However, according to the analysis of the qualitative feedback, we must declare that motion gesture interfaces are effective mostly for discrete tasks such as "Make Call", "Hang Up Call", "Enter/Act on Selection", "Cancel", "Resume". Therefore, for continuous tasks such as navigating and scrolling, motion gesture interfaces should be combined with other interaction techniques that are efficient for supporting continuous tasks.

Positive comments about vibration feedback also provided us with insights for non-visual output in mobile contexts. The use of vibrations can be extended to convey rich non-visual information to blind users. Various kinds of information can be encoded in different patterns of vibration. However, today smartphones are usually equipped with only one vibration motor. We hope that smartphone vendors will equip more vibration motors that provide richer interactions in the future.

Regarding the issue of feedback and gesture size from the first study, we paid attention to the participants' gestures when provided with vibration feedback and speech feedback. For this purpose, we performed close analysis of each participant profile using the videos recorded during the study. We found that most of the participants performed small gestures when speech output was provided (i.e., browse next or previous contact). On the other hand, most participants (six out of eight) performed large gestures when only vibration feedback was provided (e.g., when making and hanging up a call). Therefore, designers should pay attention to the feedback type when designing motion gestures.

Finally, we learned that a successful assistive interface is not only about usability and performance. The users' confidence and motivation to use the system, and the users' experiences while using the system make the interfaces most effective. In our study, we found that the participants were very pleased with the simplicity and performance of the interfaces we proposed. During the training session of experiment, one of the participants stated "It is very interesting, I will use a smartphone if I can control it this way!" Although motion gestures are not always the best in every situation, our study met a good match between task and interaction techniques that suit the users' abilities. This encouraged the users to use the system. Therefore, instead of building the interfaces only focusing on knowledge about the users' disabilities, it is important to identify the users' abilities and to find the best interaction technique for a given situation.

5 Future Work

The results of this study suggest that motion gesture interfaces are potentially useful and worthy of further exploration. However, the limitations of this study remain in the number and diversity of blind participants. Therefore, to further investigate more implications for designing motion gestures for blind people, we will expand the study with more participants. Also, motion gestures are neither the only way nor always the best way in every situation for every user. It is very questionable if motion gestures are equally effective for aged users who have less arm-hand steadiness to perform gestures. We will thus extend the study with more diverse user groups, more possible tasks and interaction techniques such as touch gesture and 3D hover gestures.

6 Conclusions

In this paper, we conducted two studies regarding motion gesture interaction in a mobile phone for blind people. From the first study, we presented usable motion gestures for blind people for mobile interactions. This study also brings us closer understanding of motion gesture creation process and gesture taxonomy of blind people. Then we provided implications for mobile interactions and motion gesture interfaces. Using the user-defined gestures and design considerations suggested from the first study, we developed motion gesture interfaces with vibration and speech feedback. In the second study, we investigated the usability and qualitative feedback of motion gesture interfaces. Findings from the second study indicated that motion gesture interfaces are more efficient than traditional button interfaces. Also, motion gesture interface gained more positive comments from the participants. Through the study results, we provided implications for designing smartphone interfaces.

Acknowledgement The authors are grateful for the work and the support of all members of the Center for Human Computer Interaction and Ren Lab in Kochi University of Technology.

References

[1] Kane S K, Jayant C, Wobbrock J O, Ladner R E. Freedom to roam: A study of mobile device adoption and accessibility for people with visual and motor disabilities. In *Proc. the 11th ASSETS*, October 2009, pp.115-122.

[2] Poggi I. From a typology of gestures to a procedure for gesture production. In *Proc. Int. Gesture Workshop*, April 2001, pp.158-168.

[3] Poggi I, Pelachaud C, Caldognetto E M. Gestural mind makers in ECAs. In *Proc. the 2nd AAMAS*, July 2003, pp.1098-1099.

[4] Iverson J M, Goldin-Meadow S. Why people gesture when they speak? *Nature*, 1998, 396(6708): 228.

[5] Karam M, Schraefel M C. A taxonomy of gestures in human computer interaction. Technical Report, ECSTR-IAM05-009, Electronics and Computer Science, University of Southampton, 2005.

[6] Wobbrock J O, Morris M R, Wilson A D. User-defined gestures for surface computing. In *Proc. the 27th CHI*, April 2009, pp.1083-1092.

[7] Ruiz J, Li Y, Lank E. User-defined motion gestures for mobile interaction. In *Proc. CHI,* May 2011, pp.197-206.

[8] Rekimoto J. Tilting operations for small screen interfaces. In *Proc. the 9th UIST*, November 1996, pp.167-168.

[9] Metzger C, Anderson M, Starner T. FreeDigiter: A contact-free device for gesture control. In *Proc. the 8th ISWC*, Oct. 31-Nov. 3, 2004, pp.18-21.

[10] Jones E, Alexander J, Andreou A, Irani P, Subramanian S. GestText: Accelerometer-based gestural text entry system. In *Proc. CHI*, April 2010, pp.2173-2182.

[11] Wigdor D, Balakrishnan R. TiltText: Using tilt for text input to mobile phones. In *Proc. the 16th UIST*, Nov. 2003, pp.81-90.

[12] Weberg L, Brange T, Hansson Å W. A piece of butter on the PDA display. In *Proc. CHI*, March 2001, pp.435-436.

[13] Liu J, Zhong L, Wickramasuriya J *et al.* User evaluation of lightweight user authentication with a single tri-axis accelerometer. In *Proc. the 11th MobileHCI*, Sept. 2009, Article No. 15.

[14] Kray C, Nesbitt D, Dawson J, Rohs M. User-defined gestures for connecting mobile phones, public displays and tabletops. In *Proc. the 12th MobileHCI*, September 2010, pp.239-248.

[15] Kurdyukova E, Redlin M, André E. Studying user-defined iPad gestures for interaction in multi-display environment. In *Proc. IUI*, February 2012, pp.93-96.

[16] Liang H, Williams C, Semegen M, Stuerzlinger W, Irani P. User-defined surface+motion gestures for 3D manipulation of objects at distance through a mobile device. In *Proc. the 10th APCHI*, August 2012, pp.299-308.

[17] Obaid M, Häring M, Kistler F, Bühling R, André E. User-defined body gestures for navigational control of a humanoid robot. In *Proc. the 4th ICSR*, October 2012, pp.367-377.

[18] Vatavu R. User-defined gestures for free-hand TV control. In *Proc. the 10th EuroITV*, July 2012, pp.45-48.

[19] Lee S, Kim S, Jin B *et al.* How users manipulate deformable displays as input devices. In *Proc. CHI*, April 2010, pp.1647-1656.

[20] Piumsomboon T, Billinghurst M, Clark A, Cockburn A. User-defined gestures for augmented reality. In *Proc. CHI*, April 27-May 2, 2013, pp.955-960.

[21] Jung H, Qin S. User-defined gesture sets using a mobile device for people with communication difficulties. In *Proc. the 17th ICAC*, Sept. 2011, pp.34-39.

[22] Kane S K, Wobbrock J O, Ladner R E. Usable gestures for blind people, understanding preference and performance. In *Proc. CHI*, May 2011, pp.413-422.

[23] Kane S K, Bigham J P, Wobbrock J O. Slide rule: Making mobile touch screens accessible to blind people using multi-touch interaction techniques. In *Proc. the 10th ASSETS*, Oct. 2008, pp.73-80.

[24] Landau S, Wells L. Merging tactile sensory input and audio data by means of the talking tactile tablet. In *Proc. EuroHaptics*, July 2003, pp.414-418.

[25] Pirhonen A, Brewster S, Holguin C. Gestural and audio metaphors as a means of control for mobile devices. In *Proc. CHI*, April 2002, pp.291-298.

[26] O'Neill E, Kaenampornpan M, Kostakos V, Warr A, Woodgate D. Can we do without GUIs? Gesture and speech interaction with a patient information system. *Personal and Ubiquitous Computing,* 2006, 10(5): 269-283.

[27] Zhao S, Dragicevic P, Chignell M *et al.* Earpod: Eyes-free menu selection using touch input and reactive audio feedback. In *Proc. CHI*, April 2007, pp.1395-1404.

[28] Sánchez J, Aguayo F. Mobile messenger for the blind. In *Universal Access in Ambient Intelligence Environments*, Stephanidis C, Pieper M (eds.), Springer, pp.369-385.

[29] Yfantidis G, Evreinov G. Adaptive blind interaction technique for touch screens. *Universal Access in the Information Society*, 2006, 4(4): 328-337.

[30] Azenkot S, Wobbrock J O, Prasain S, Ladner R E. Input finger detection for non-visual touch screen text entry in Perkinput. In *Proc. Graphics Interface*, May 2012, pp.121-129.

[31] Buzzi M C, Buzzi M, Donini F *et al.* Haptic reference cues to support the exploration of touchscreen mobile devices by blind users. In *Proc. CHIItaly*, Sept. 2013, Article No. 28.

Nem Khan Dim is a Ph.D. candidate of the School of Information, Kochi University of Technology, Japan. She received her B.S. degree in computer science from Dagon University, Yangon, in 2008, and M.S. degree in computer science from University of Yangon, in 2011. Her research interests are assistive interfaces, natural user interface and haptic interfaces.

Xiangshi Ren is a professor in the School of Information and the director of the Center for Human-Computer Interaction at Kochi University of Technology, Japan. He is the founding president of the International Chinese Association of Computer Human Interaction (ICACHI). He was a visiting faculty researcher at IBM Research (Almaden), and visiting professor at the University of Toronto. His research interests include all aspects of human-computer interaction, particularly multimodal interactions and user interface design. Prof. Ren received a B.E. degree in electrical and communication engineering, and M.E. and Ph.D. degrees in information and communication engineering from Tokyo Denki University in Japan in 1991, 1993, and 1996, respectively. He is a senior member of ACM and IEEE, and a member of IPSJ, IEICE and Human Interface Society.