Designing Mid-Air TV Gestures for Blind People Using User- and Choice-Based Elicitation Approaches

Nem Khan Dim[‡], Chaklam Silpasuwanchai[*], Sayan Sarcar[‡], Xiangshi Ren[*‡]
nemkdim@gmail.com, {chaklam, sayan.sarcar, ren.xiangshi}@kochi-tech.ac.jp
[*]Center for Human-Engaged Computing, Kochi University of Technology, Kochi, Japan
[‡]School of Information, Kochi University of Technology, Kochi, Japan

ABSTRACT
Mid-air gestures enable intuitive and natural interactions. However, few studies have investigated the use of mid-air gestures for blind people. TV interactions are one promising use of mid-air gestures for blind people, as "listening" to TV is one of their most common activities. Thus, we investigated mid-air TV gestures for blind people through two studies. Study 1 used a user-elicitation approach where blind people were asked to define gestures given a set of commands. Then, we present a classification of gesture types and the frequency of body parts usage. Nevertheless, our participants had difficulty imagining gestures for some commands. Thus, we conducted Study 2 that used a choice-based elicitation approach where the participants selected their favorite gesture from a predefined list of choices. We found that providing choices help guide users to discover suitable gestures for unfamiliar commands. We discuss concrete design guidelines for mid-air TV gestures for blind people.

Author Keywords
Blind People; Mid-Air Gesture; TV Interaction; Choice-Elicitation Approach; User-Elicitation Approach.

ACM Classification Keywords
H.5.2 [Information Interfaces and Presentation]: User Interfaces: Evaluation/Methodology.

INTRODUCTION
Mid-air gesture interfaces have become increasingly popular due to its intuitive and natural control. However, only a very few studies [10] have explored the use of mid-air gestures for blind people. Consequently, blind people have not enjoyed equal access to the benefits provided by such technologies. One promising use of mid-air gestures for blind people is their use for TV interactions because TV is one of the most common and essential activities of many blind people [2]. We have confirmed this fact through our preliminary interviews with blind people where we found that they used TV for more than 7 hours per week. Furthermore, the interviews informed us that blind people often have difficulty finding their remote controls or figuring out the button layout. Such problems suggest that employing mid-air gestures for interaction with TVs is a useful alternative method.

To design usable and accessible gestures, we determined that the first important step was to understand the mental mapping of blind users when performing mid-air gestures. Thus, we investigated the mid-air gesture preferences of blind people for TV interaction using a user-elicitation approach. With the user-elicitation approach, we were able to formulate a classification of gesture types and the frequency of various body parts usage. However, we found two primary limitations with the user-elicitation approach. First, our participants experienced much difficulty and stress trying to come up with their own gestures for some commands because they did not have adequate understanding of the design alternatives and possibilities. Consequently, there was a gesture consensus only for commands that were familiar to users. Second, for unfamiliar commands, our participants suggested many abstract/random gestures which seemed to be rather inefficient and unsuitable for gesture inputs.

To address these limitations, we conducted a second study which adopted a choice-based elicitation approach [17] as a complementary approach where the participants were given a predetermined range of possible gestures from which to choose. All participants reported that they appreciated the gesture choices as the choices helped them to discover more suitable gestures. We also found that the choice-based elicitation approach can help confirm the preferences of certain gestures, while guiding the participants to reach a consensus for unfamiliar commands. We conclude our work with some generalized implications for mid-air TV gesture interaction design, as well as some discussion regarding the use of the choice-based elicitation approach.

Our contributions include:

1. We analyzed 180 defined gestures from which we developed a classification of gesture types for blind people and determined the frequency of body parts usage.

2. We leveraged the choice-based elicitation elicitation approach to address commands that had little consensus among participants and about which blind participants had little understanding or awareness.

DIS 2016, June 4–8, 2016, Brisbane, QLD, Australia

DOI: http://dx.doi.org/10.1145/2901790.2901834.

3. We present a complete gesture set of mid-air TV gestures for blind people.
4. We present design guidelines for mid-air gestures for blind people by analyzing patterns of user gestures and their mental mappings.
5. We discuss the usefulness and general guidelines of the choice-based elicitation approach.

RELATED WORK

Here we present the most relevant works regarding gesture design based on user-elicitation, gesture interactions for blind people and the choice-based elicitation method.

User-elicitation Method

The user-elicitation approach is a recently developed approach for designing suitable gestures. This approach was originally based on the guessability technique [14, 7], where participants were asked to define gestures, given some visual effects of commands/actions.

In terms of effectiveness, Morris et al. [12] and Wobbrock et al. [8] suggested that a user-elicitation approach can produce more highly preferred gestures, compared to gestures designed by experts who have the tendency to develop more "physically and conceptually complex gestures". In terms of learnability, Nacenta et al. [13] found that user-defined gestures were easier to learn, remember and recall, compared to other predefined gestures.

Motivated by the effectiveness of this approach, many user-elicitation works have been conducted, e.g., tabletops [8], mobile phones [9] and TV [22]. Most works resulted in promising design practices and a consensus set of suitable gestures which were determined by the highest agreed gestures among participants.

Informed by promising findings from previous studies, we conducted a user-elicitation study to determine the gesture preferences of blind people for TV interactions.

Gesture Interactions for Blind People

Previous studies on defining suitable gestures for blind people mainly focused on touchscreen surfaces in terms of enhancing hardware to provide access to touch screen kiosks (e.g.,[3, 21]) and exploring interaction techniques for mobile touchscreens ([11, 20, 16, 18, 19]). Kane et al. [15] performed a touchscreen gesture elicitation study and compared the performance of sighted with the performance of blind people. They found that blind people have different gesture preferences, and their ways of performing gestures are significantly different (e.g., the gestures of blind people tend to be noticeably exaggerated).

A few studies also explored motion gestures to interact with mobile devices. Ruiz et al. [9] presented an elicitation study of motion gestures for mobile interaction. The 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 consider gestures of people with visual impairments. More recently, Dim and Ren [5] studied motion gestures for blind people in mobile phone interactions. They found that blind people mostly perform metaphorical gestures reflecting actions from their daily activities. Thus, such gestures had a high consensus for familiar activities related to daily activities such as *making phone call* and *voice search*, but they had a low consensus for unfamiliar commands such as *zooming*.

Although some studies have been done for touchscreen gestures and motion gestures in mobile phones for blind people, very few works have been done on suitable mid-air gestures. Thus, we conducted user studies to determine suitable mid-air gestures for blind people, especially for TV interactions.

Choice-based elicitation Method

To our knowledge, only a little study has employed choices for gesture design, with the exception being the work of Silpasuwanchai and Ren [17]. They employed choices to understand how gamers perform simultaneous gestures during intense gameplay. However, there was no deep discussion regarding the use of choices for gesture design.

Where users have little or no idea of the range of possibilities for suitable gestures, employing designer-selected choices can serve as one way to non-threateningly introduce designer ideas into the design process without making the users feel compelled. In addition, allowing users to select their preferred choices can reduce the chance of designers introducing overly-complicated gestures, as suggested by Morris et al. [12]. This also increases the chance of settling on a gesture that is suitable in both the view of designer and that of the users. Informed by this potential benefit, we adopted the choice-based elicitation approach for our study.

In summary, while designing suitable gestures for blind people is an ongoing challenge, our study leverages the benefits of both user-elicitation and choice-based elicitation to determine the most suitable mid-air gestures for blind people.

PRELIMINARY INTERVIEWS

Prior to the studies, we conducted a preliminary interview with 8 blind participants (ages ranged from 37 to 65 years). The purpose of the interview was to gain an understanding of the current problems in TV interactions of blind people and their experiences with mid-air gesture interfaces. Each participant was paid $10 for their participation.

TV Interactions

We observed that the use of TV is one of the most common and essential activities of our participants. All our participants regularly 'listen' TV for more than 7 hours per week. Although they can not see, they enjoyed the experience through audio. Their reasons for using the TV were listening to sports, news, drama, music and other programs. Our participants mentioned that they often had a hard time finding their remote controls as well as figuring out the button layout on the remote control. For example, one participant stated, "*I am always frustrated when someone changes the position of the remote control and I cannot find it.*" Another participant mentioned that it was difficult to learn the layout of remote controls as there was no standard layout of TV remote controls. For example, one of the participants mentioned, "*Many times, I cannot distinguish* *buttons for changing channels and changing the TV volume on the remote control, so I just use the buttons on the TV.*" When asked about the possibilities of using mid-air gestures, all participants agreed that it would be useful if they could sometimes use mid-air gestures in their interactions with TV.

Experiences in Mid-Air Gesture Interfaces

None of our participants had any knowledge of mid-air gesture interfaces but one had experience with touch gestures in smartphones. When asked about experiences in mid-air gesture interfaces, one participant stated, "*Generally, I understand that mid-air gesture is about moving your body parts, for example, waving the hand for saying goodbye. But I have no idea about how I can use it for communicating with any devices.*" When asked the participant who used a smartphone, that participant mentioned, "*With my smartphone, I only do flick gestures (to the left and the right) for browsing the next and previous contents. I do not know what mid-air gesture interfaces would look like.*" Even though blind people perform daily gestures, they are not necessarily able to transfer their knowledge to mid-air gesture design due to their lack of knowledge and experience with mid-air gesture interfaces and with technology in general.

Thus to design usable mid-air gesture interfaces, there is a need to better understand the preferences and mental mappings of blind people. To determine blind people' mid-air gesture preferences in interactions with TVs, we conducted a user-elicitation study.

STUDY 1: USER-ELICITATION APPROACH

The objective of this study was to determine blind people's mid-air gesture preferences for TV interactions. The participants were asked to define gestures for a set of TV interaction commands. Gesture classifications and frequency of body parts usage, agreement scores, gesture defining times and subjective assessments were analyzed.

Participants

Twelve blind participants (7 males, age range = 26 to 78 years, M = 53.9 years, SD = 12.64) were recruited. One of the participants could see light, two of them could see the presence of objects but not able to distinguish shapes. One could see objects. The other participants were totally blind. Ten of the participants were 'early blind' (age 0 to 3 years), and the rest two were 'late blind' (one at 6 years and the other at 17 years). All participants regularly 'listen' TV for more than 7 hours each week. None had experience using mid-air gestures for TV control or any other devices. Eleven participants were right-handed and one was ambidextrous. Each participant was paid $15 for the experiment.

Commands

We identified commonly used TV commands. We also selected commands that are specifically used by blind people, e.g., *Voice Guide* [1]. A total of 15 commands were selected for the experiment (see Table 1).

No.	Command	Description
1	Open	Power on TV
2	Close	Power off TV
3	Channel	Change channel
4	Favorite Channel	Go to saved channels
5	Next	Go to next channel/menu
6	Previous	Go to previous channel/menu
7	Volume Up	Increase volume
8	Volume Down	Decrease volume
9	TV Guide	Check channel list and time
10	Pause	Pause
11	Play	Play
12	Menu	Open menu
13	Yes	Answer Yes to system question
14	No	Answer No to system question
15	Voice Guide	Activate voice guidance for blind users

Table 1. TV commands used in Study 1 and 2.

Procedure

Our study design used a user-elicitation approach similar to that of [9, 8]. As with previous elicitation studies, we did not want participants to take recognizer issues into consideration, and to observe the users' best possible gestures.

At the start of the experiment, participants were asked to define gestures for 15 TV commands in randomized order. To identify the most preferred gesture and reduce any ordering effects, participants were allowed to reuse the same gesture for multiple commands.

Each command name was communicated by the experimenter, along with the description of the command's effect (e.g., "*Open TV*" followed by "*This command will power-on your TV*").

During gesture defining, participants were instructed to think-aloud while performing their gestures, confirming the start and end of their performed gesture and describing their corresponding rationale. To determine how well they imagined the gestures, we recorded gesture defining times, i.e., the time between command instruction and gesture execution.

After each group of commands was defined (each group contains approximately 5 commands), participants were asked to evaluate their gestures, using similar evaluation to that of [8]: "The gesture I performed is a *good match* for its purpose"; "The gesture I performed is *easy to perform*"; "The gesture I performed is *tiring*".

The experiment took around 1.5 hours, and it was audio and video recorded for qualitative data analyses.

Results

Our results include a gesture taxonomy, body parts used, agreement score, gesture defining time and subjective ratings.

Classification of Mid-Air Gestures

As noted in related work, gesture classifications have been developed for surface gestures [8] and mobile interactions [9]. However, no work has established a taxonomy of mid-air gestures for blind people. We classified gestures collected from the study along five dimensions: *nature, form, axes of motion, binding* and *flow*. Each dimension was further classified into multiple subcategories (see Table 2 and Figure

Dimension	Category	Description
Nature	*Symbolic*	Gesture depicts similarity of symbols.
	Metaphorical	Gesture indicates a metaphor.
	Deictic	Gesture refers to spatial information.
	Abstract	Gesture-command mapping is arbitrary.
Form	*Static pose*	Hand pose is held in one location.
	Dynamic pose	Hand pose changes in one location.
	Static pose with path	Hand pose is the same but the hand is moving.
Axes of motion	*Sagittal*	User moves the hand/leg forward and backward.
	Horizontal	User moves the hand/leg left and right.
	Longitudinal	User moves the hand upward/downward.
	Compound-axes	User hand movement includes more than one direction.
	Stationary	User performs the gesture keeping the hand or fingers at one location.
Binding	*Independent*	Gesture requires no information about object positions or body parts.
	Body-centric	User moves the hand/leg using spatial reference of the body.
	Body-referenced	User moves the hand with respect to other body parts.
Flow	*Discrete*	Response will occur after user performs the gesture.
	Continuous	Response will occur while user is performing the gesture.

Table 2. Taxonomy of mid-air gestures based on 180 gestures.

1a). Gesture taxonomies obtained in our study and those in [8] and [9] were different for two main reasons: (1) the gestures were performed for different interactions (i.e., surface gestures, motion gestures in mobile phones and mid-air gestures), and (2) they were performed by users with significantly different capabilities (i.e., sighted users and blind users).

The *Nature* dimension is a classification of gesture-meaning relationships. The other four dimensions are concerned with the characteristics of physical movements involved in performing the gestures (trajectories, movement plane, gesture flow, etc.). In the (1) *Nature* dimension, symbolic gestures are depictions of symbols, for example, drawing a letter 'V' for the *Voice Guide* command. As we expected, very few symbolic gestures were performed in our study, given that these gestures require visual references. (2) Metaphorical gestures linked to their meaning by mechanical determinism (i.e. linked to the logical meaning of daily actions or feelings, not linked to their visual similarities). For example, some of our participants put their hands near by the ears for the *Voice Guide* command indicating that "I want to hear". Metaphorical gestures reflecting common daily actions were the most frequently performed gestures by our participants. Among the metaphorical gestures, we found some conventional gestures. For example, four of our participants performed 'OK gesture' for the *Yes* command. (3) Deictic gestures refer spatial information of objects. For example, when asked to perform a gesture for the *Menu* command, one of our participants said, "I would put the *Menu* command at the upper right corner of the screen and point at it". It is important to note that, unlike sighted people who are able point at visual on-screen contents, our participants performed pointing gestures only relying on their proprioception. That is, they could only refer the content position to hand movements in distinct and specific directions which are related to their body (i.e. upper left, upper-right, etc.) rather than by accurate reference to the display or on-screen contents on the display. (4) Abstract gestures are linked to the meaning neither by the similarity nor by any other relationship that allows one to infer the meaning from the gesture. Arbitrary gestures were the second most performed gestures in our study. For example, when asked to perform a gesture for the *Voice Guide* command, one of our participants came up with the gesture which showed a finger stating no clear gesture-meaning relationship.

In the *Form* dimension, (1) static pose is where the hand pose is held in one location, for example, showing the palm for the *Stop* command or showing 'OK' for the *Yes* command. (2) Dynamic pose is where the hand pose changes within one location, for example, the *blink gesture* (close and open the palm) for the *Open TV* command or the *double tap* by the foot for the *Yes* command. (3) Static pose with path is where user moves the hand but hand pose is not changed, for example, moving the hand from left to right for the *Next* command or moving the hand upward for the *Volume Up* command. These gestures were the most frequently performed gestures. One interesting observation is that our participants mostly used distinctly exaggerated movements, that is, moving their entire forearm or arm instead of only the wrist and the hand when performing hand or finger gestures. When asked about the reasons, our participants commented that they were not sure of the boundary of interactions, thus making large movements seemed to reassure them and give them confidence that they had performed the actions in a manner that made them clearly recognizable. While gestures with *static pose with path* were the most frequently performed in our study, it is noticeable that there can be cases where the user moves the hand position and also changes the hand pose (i.e. *dynamic pose with path*). However, none of our participants performed such gestures in our study.

The *Axes of motion* dimension shows the direction of user movements of the hand, foot or fingers. (1) Sagittal gestures are those where users move the hand forward and backward. Examples include moving the hand backward to represent 'undoing' for the *Pause* command and pushing the finger forward as 'pushing a button' for *Open TV*. (2) Horizontal gestures are those where users move the hand or foot left and right, for example, moving the hand left and right for *Next/Previous*. (3) Longitudinal gestures are those where users move the hand upward and downward, for example, moving the hand upward for *Volume Up*. (4) Compound-axes gestures are those where user movements include more than one spatial dimensions. For example, some of our participants drew a circle in the air for *TV Guide*. Another example is moving the hand in the upper-right direction where our participant stated, *"I would put the Menu command at the upper-right corner of the screen and point at that direction."* (5) Stationary gestures are those where users perform their gestures keeping their hands or fingers in one location. With respect to the *form* dimension, stationary gestures are *static* and *dynamic gestures* while sagittal, horizontal, longitudinal and compound-axes gestures are *static gestures with path.*

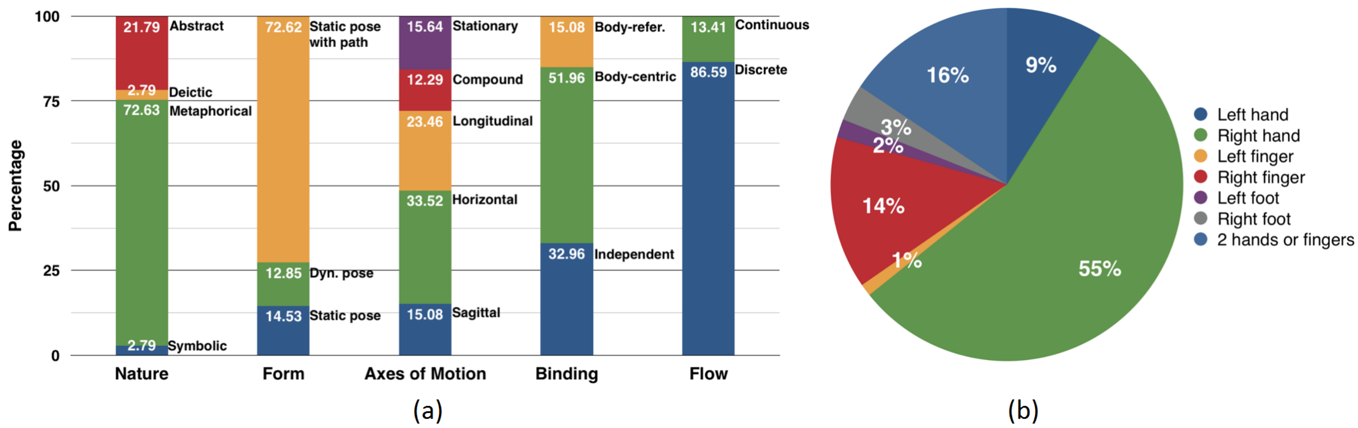

Figure 1. Percentage of gesture types and body parts used.

In the *Binding* dimension, (1) independent gestures require no information about the world or the body. Examples are showing 'OK' sign for the *Yes* command and showing the palm for *Stop*. These gestures can be performed anywhere regardless of their relation to the user's body or objects on the screen. (2) Body-centric gestures are performed by users by referring spatial information to the body, for example, moving the hand to the left of the body or by moving the foot forward. These gestures were the most performed gestures by our participants. While it is understandable that *body-centric gestures* are specifically selected and performed for certain commands, such as *Next/Previous* commands (moving the hand left and right), we found that our participants performed *body-centric* gestures in many other cases as well. Examples are waving the hand left and right for the *Play/Pause* commands, drawing the hand backward to represent undoing for the *No* command and moving the hand in an upper-right direction to represent skipping for the *Favorite Channel* command. (3) Body-referenced gestures were performed by referring other body parts, for example, putting the hands near by the ears for *Voice Guide*. Another example was that our participants performed *body-referenced gesture* by opening the hands (clasp two hands together and then separate them) for commands such as the *Open TV, Play* and the *Voice Guide* commands.

In the *Flow* dimension, (1) discrete gestures are performed and responded to as one event, for example, the *blink gesture* for the *Open TV* command. (2) Continuous gestures require ongoing recognition. One example is *raising the hand* for the *Volume Up* command. *Discrete* or *continuous gestures* were performed depending on the users' perceptions of the system responses to their gestures for the commands. For the *Open TV* command, it was expected that the system would respond by opening the TV instantly after the gesture; the participants performed *discrete gestures* for that command. On the other hand, for the *Volume Up* command, the participants continuously moved the hand upward as if progressively increasing the volume to certain desirable level. Similarly, for the *TV Channel* command, the participants rotated the hand steadily as if to continue until they found the channel they wanted.

Use of Body Parts

Figure 1b shows the use of body parts. One-handed gestures with the right hand were the most preferred, followed by one finger gestures, two hand/finger gestures, and foot gestures.

Hand gestures were commonly performed in the form of movements, such as waving the hand left and/or right for *Next* and *Previous*, waving the hand left and/or right for *Select Channel*, and moving the hand up and down for*Volume Up/Down*. Finger gestures commonly featured the index finger to draw a circle to activate *TV Guide* or to pointing in order to *Open/Close* the TV. Fingers were also used to perform 'OK' gesture for the *Yes* and the *crossed fingers* gesture for *No*. Foot gestures were used by some participants when they wanted to do the same gesture they had already performed using the hands or fingers. For example, one of our participants who moved the right foot to right for the *Channel* command mentioned, "I would like to use my leg because I have done the same gesture for *Next* using the hand." Although the participants were allowed to suggest the same gesture for different commands, they often used the feet reasoning that they wanted to use different gestures from the previous gestures they had already performed. We observed that, our participants performed their foot gestures mostly for abstract gestures where they could not suggest any gesture-meaning relationship. For example, one participant moved the right foot to the right for the *Menu* command stating, "When I have no idea what gesture to use, I just change the body part".

Agreement Score

To determine the gesture consensus among participants, we calculated the agreement score using a formula from [8]. The calculation of agreement score is as follows:

$$A_r = \sum_{P_i} \left(\left| \frac{P_i}{P_r} \right| \right)^2 \tag{1}$$

where P_r represents all gestures performed for event r and P_i is a subset of identical gestures from P_r. A_r ranges 0 to 1. Gestures were considered to be identical when they had similar trajectories and poses.

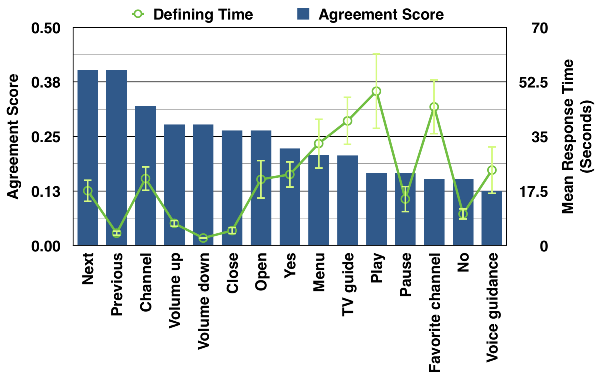

Figure 2. Agreement scores and gesture defining times of the user-elicitation approach.

Figure 2 shows the agreement score of each command. We found that user-defined gestures in our study had a slightly lower agreement score (0.24) compared to gestures of sighted people for TV control (0.42 in [22]) and other user-defined studies (0.28 and 0.32 in [8]). There were some commands that had relatively high agreement scores such as *Next* and *Previous*. However, for more abstract commands such as *Voice Guide* and *Favorite Channel*, participants reported that they had a lot of difficulty imagining the gestures. As a consequence, most participants came up with arbitrary gestures about which they had no clear rationale as to why they defined the gesture in that particular way.

Defining Time

Gesture defining time is the time between command instruction and gesture definition. It can measure the speed and ease of defining of commands which can help inform how well participants can imagine their gestures. The mean defining time of all commands was 22.92 seconds ($SD = 12.02$). Figure 2 shows the mean defining time for each command. Some commands showed consistency between agreement scores and defining times. For example, *Previous* and *Volume Down* both had high agreement scores and low defining times. Similarly, there were commands such as *Play* and *Favorite Channel* that participants had difficulty imagining and consequently there was low consensus among the participants. On the other hand, there were also some commands with low agreement scores but also low defining times, e.g., *Pause* and *No*, suggesting that there were certain commands where, although each participant had some clear idea of the gestures, they shared little commonality between participants' preferred gestures.

Subjective Ratings

Unlike past works, we found that the highest agreed gestures in our user-elicitation study had no significantly higher good match rating compared to other gestures. This indicates that, in our case, popularity did not necessarily identify better gestures over the worse ones. This was largely due to the fact that our participants had a lot of difficulty imagining suitable gestures and partly due to their lack of experiences with gesture interactions.

Discussion

The classification of gestures revealed that most gestures performed by our participants came from their daily experiences, for example, raising the hand for the *Open TV* command which mimics switching on the light and pushing the hand forward for the same command which mimics pushing buttons on their household appliances and devices. From the *binding* dimension, we observed that our participants preferred *body-centric* gestures, that is, moving hands or fingers to directions with respect to the body (left, right, forward, backward, etc.). Given that our participants did not perform gestures combining dynamic poses with path gestures, we observed that they preferred gestures with simple hand movements. When performing those gestures, user movements could include any directions (of left/right, forward/backward and upward/downward) and most of the gestures occurred at the *frontal plane* of users. Also, for the gesture recognition, it is important to note the inclusion of compound gestures, for example, moving the hand to draw circle in the air (i.e. the hand moves to more than one spatial dimension in a single gesture). Regarding the physical movements when performing gestures, gesture recognition should be developed to accommodate exaggerated hand movements which were preferred by our participants. Gesture recognition should also be developed with attention to user expectations regarding the system's response. For example, *continuous gestures* such as gestures for *Volume, Select Channel* would require on-going (sustained) recognition of gestures.

Regarding the use of body parts, we observed that right and left hands, right fingers, and 2 hands and fingers were mostly used for the *metaphorical, deictic and symbolic gestures*. The left finger, left foot and right foot were not recommended for gesture design, given that these body parts were not commonly used and we observed that they were mostly used for abstract commands.

Some commands in the user-elicitation study achieved low-agreement scores, given our participants had no clear idea of mid-air gesture interfaces and design. Also, our participants were less influenced by seeing gestures of other people. Defining gestures for the given TV commands was difficult for our participants especially for abstract and unfamiliar commands such as *Voice Guide, No, Favorite Channel, Play, Pause, TV Guide* and *Menu* commands. The higher defining time revealed the fact that our participants had much higher difficulty imagining suitable gestures in the user-elicitation approach. Thus, to address these gaps, we recruited the same participants and conducted a choice-based elicitation study where participants were asked to select their preferred gestures from a predefined list. To minimize any possible effect of the user-elicitation approach on the choice-based elicitation approach, we set sufficient time gap (around 1 month) between Study 1 and 2.

STUDY 2: CHOICE-BASED ELICITATION APPROACH

Study 1 allowed us to develop an understanding of user gestures and their mental mappings through the detailed analyses and classification of gestures. However, we found that our participants had difficulty suggesting suitable gestures for certain TV commands, for example, *Voice Guide, No, Favorite Channel, Play, Pause, TV Guide* and *Menu*. The objective of Study 2 was to adopt the choice-based elicitation approach to better understand the preferences of gestures in commands where users had less idea of gesture design. In choice-based elicitation, participants were asked to select their preferred gestures from a list of designer-selected choices. The choice list contained two data columns, one for gesture choices, and another for commands, with the relationship as many-to-one, respectively.

Selecting Choice Candidates

We regard the process of selecting which choices to include in the choice list and how choices are introduced to participants to be key to the effectiveness of the choice-based elicitation approach. Four measures were considered.

First, there is abundant evidence to show that too many or too few choices may decrease user engagement and can possibly confuse users [6]. However, there exist no simple rules to decide the suitable number of choices. To decide the number of choices, we adhered to two grounds - (1) modern working memory studies show that humans can only hold about 4 chunks at a time in their memories [4]; (2) consistent with the memory capacity theory, our initial pilot study with two blind people shows that they preferred a range of 2 to 4 choices. Based on these two grounds, we kept our number of choices to no more than 4.

Second, in order to ensure that each gesture choice is reasonable, there is a need to select gesture choices that are considerably intuitive. We recognized that there are many different ways to construct the choice list and that this depends largely on designer preferences. To demonstrate one possible way, we derived our choices from three places (Study 1, related work, and designers) - (1) we selected the highest-agreed gestures for each command from Study 1 (26.67% of gestures in the choice list). We also selected any gesture from Study 1 that we intuited to be potentially intuitive and suitable (28.33% of gestures in the choice list); (2) we derived some gesture choices from existing works (e.g., [22]); (3) we designed some gesture choices based on our own experience with blind people[1]. Also, gesture classification derived from Study 1 also helped us to design the most suitable gestures for Study 2, for example, selecting the body parts, avoiding symbolic gestures wherever possible. The final list of gesture choices achieved a high interrater reliability among authors and one independent rater (Kappa=0.919, $p < 0.001$).

Third, to ensure that each choice was equally exposed and judged, experimenters asked participants to perform each choice, instead of only passively listening to the different choices. This also enabled the participants to weigh each gesture based on how comfortable it was to use.

Last, to prevent any possible ordering effect of choices (i.e., due to the recency or primacy effect on choices), we decided that all choices should be presented to participants in randomized order.

[1]One of the authors has been working consistently with blind people for approximately three years.

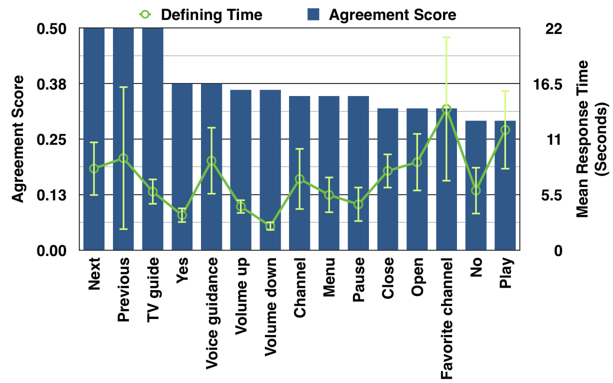

Figure 3. Agreement scores and response times of the choice-based elicitation approach.

Experimental Design

To understand how choices can help us better understand users, the same 12 blind participants in Study 1 were recruited. We used the same set of commands as in Study 1. The procedure was also conducted in a similar manner to Study 1, except that the participants were provided with a list of choices and were asked to select their most preferred choices. Each command name was communicated by the experimenter, along with the description of the command effect. The experimenter then provided the first gesture choice, i.e., the experimenter explained how to perform the gesture, followed by asking the participants to perform the gesture. For each command, the experimenter went through all choices in a similar manner. Gesture choices were presented in a randomized order. The experimental session took around 1.5 hours.

Results

Agreement score, gesture defining time,subjective ratings and the mid-air gesture set were presented.

Agreement Score

The choice-based elicitation approach achieved a noticeably higher agreement score (0.37) (Figure 3). This result was initially expected because choices reduced the possible variations between participants. Nevertheless, the choice-based elicitation approach was found to be useful, particularly for commands where participants have no clear ideas of their own, e.g., *Favorite Channel*, *Voice Guide*, *TV Guide*. Participants reported that the supplied gesture choices enabled them to discover more suitable gestures, particularly for difficult commands.

Defining Time

To understand how long it took the participants to select their choice, we measured the time between the choice presentation time and decision time. The mean defining time of all commands was 7.16 seconds (SD = 3.13). Figure 3 shows the mean defining time for each command. We found that the gesture choices can guide users to discover more suitable gestures.

Subjective Ratings

We found that the highest agreed gestures had significantly higher good match rating (Z = -2.612 , $p < 0.01$) compared to other gestures. This indicates that, in the choice-based elicitation study, popularity does identify better gestures over the worse ones. We were thus confirmed that the choice-based elicitation approach can help confirm the preferences of certain gestures.

Mid-Air Gesture Set

Based on the highest agreed gestures, we developed the mid-air gesture set for TV commands (see Figure 4). Nine of the 15 commands shared similar highest-agreed gestures in both user- and choice-based elicitation approaches. These nine commands includes Open/Close, Volume Up/Down, TV guide, Yes, Previous/Next, and Voice Guide. Another six commands were mainly based on the choice-based elicitation approach which include Play/Pause, Change Channel, Favorite Channel, Menu, and No. Using the choice approach, we found a good average consensus on these six commands (0.32).

Discussion

Gesture agreement scores for all commands increased in the choice-based elicitation study. Higher agreement scores were expected given that the choice-based elicitation approach has fewer choices than the user-elicitation approach. However, we observed the effectiveness of the choice-based elicitation approach especially for abstract or unfamiliar commands which had quite low agreement scores in the user-elicitation study. The agreement scores for those commands noticeably increased in the choice-based elicitation study: *Voice Guide* (increased by 192.31%), *TV Guide* (138%), *Favorite* (113.33%), *Pause* (105.88%), *No* (93.33%), *Yes* (72.73%), *Play* (70.59%) and *Menu* (66.67%) respectively. Gestures for commands that had the highest agreement scores in the user-elicitation study (*Next, Previous, Channel, Volume Up, Volume Down, Close* and *Open*) had the highest scores in the choice-based elicitation study and in the same order .

The choice-based elicitation approach also helped confirm that there was indeed some commonality in the participants' mental mappings to the gestures, and this commonality could be more likely observed when all participants were equally informed about the different possibilities within the design space.

GENERAL DISCUSSION

We have explored the use of the user- and choice-based elicitation approach, and our studies showed promising results. This section discusses (1) mental model observations, and (2) user-elicitation and choice-based elicitation.

Mental Model Observations

Our quantitative data, qualitative data and video recordings enable us to capture user mental models that will benefit the design of mid-air gestures for blind people. Guidelines are summarized in Table 3.

Symbolic vs metaphors

We observed that the mid-air gestures of blind people were quite different from those of sighted people in a way that sighted people performed a lot of *symbolic gestures* [22] and our participants performed very few (fewer than 3%) of those

Guidelines	Criteria
Metaphors	Gestures should be designed mimicking metaphor in users' daily lives. Avoid symbolic gestures such as drawing letters (e.g., 'M', 'X'). Avoid gesture that needs visual references (e.g., sizes, distances). For measurements, system should recognize the fine-grained self-referenced movements of users as measures of degree.
Logical Mappings	When mapping gestures to daily metaphors, *actions* (e.g., gesture that represents opening a menu book for the *Menu* command) should be used. Mimicking *visual similarities* (e.g., gesture that represents the *Menu* as a list of items) should be avoided.
Body-centric	Movements in performing the gestures should be easily relatable to user body (e.g., moving left and right, backward and forward). If pointing are required to trigger a menu, users should be able to rely on reference only to their own bodies (e.g., moving the hand upward, upper left). The need of accurate reference to the display or on-screen contents on the display should be avoided.
Commonly Used Device	Gestures should be designed to mimic the use of users' older devices whereas possible. Attention should be paid to differences of commonly used devices of blind and sighted people (e.g., smartphone vs feature phone).
Reversible gestures	Reverse gestures should be designed for dichotomous commands. Reverse or related gestures should also be designed for similar commands (e.g., *Pause, Close TV*) where possible.
Exaggerated user movements	Gesture recognition should allow freedom of movements for performing user gestures. Expect slow pace due to big movements in gestures performed by blind users. Avoid time-based gesture recognition.

Table 3. Design Guidelines and Criteria.

gestures. Instead, *metaphoric gestures* were mostly performed by our participants. For example, while the gesture for *Menu* was drawing letter 'M' in [22], the gesture for the same command in our study was the "*Open Book*" gesture (Figure 4). Most of our participants chose that gesture stating that, "*This gesture is like opening a menu book at a restaurant*". Thus, it is important for designers to understand which daily actions of blind people can be exploited as potential gestures in interactions with digital devices.

Also, for the *Volume Up/Down* commands, gestures of sighted people included having the non-dominant hand as a visual reference of measurement while moving the dominant hand upward/downward increasingly with reference to it [22]. However, none of our participants performed gestures that referenced measurement. Thus, for commands like *VolumeUp/Down*, gesture recognition should be developed so that users can rely only on their non-visual skill via proprioception, that is, the system should recognize the fine-grained self referenced movements of users as measures of degree (increase/decrease).

Visual resemblances vs logical mapping

It is important to understand how blind people perceive the meaning of certain commands. For example, the highest agreed gesture in user-elicitation study for *TV Guide* was

Figure 4. Highest-agreed gestures in choice-based elicitation approach.

drawing a circle in the air stating, "*the meaning of a circle is 'all' and a TV Guide also shows all TV channels.*" This gesture was also the most agreed gesture for *TV Guide* in the choice-based elicitation. During the choice-based elicitation study, although the participants were given other gesture choices for the same command such as moving the hand from up to down (as if to indicate a channel list), none of our participants chose that gesture. This informed us that the imagination of our participants for certain commands can be very different from that of sighted people (i.e. visual-based imagination of the *TV Guide* as a channel list vs logically mapping as *showing the whole*). Thus, designing gestures that are related to visual resemblances should be avoided as these gestures may not be intuitive to blind users.

Body-centric

As discussed in earlier sections, our participants mostly performed *body-centric* gestures (i.e. moving the hands or fingers to left/right, forward/backward, and so on). From our semi-structured interviews, we also observed that spatial references relating to their body was a great aid in many situations (for example, saying *"coffee cup at the left and mobile phone at the right"*). Thus, it is clear that *body-centric gestures with simple movements* are desirable for blind users because they could apply their proprioceptions.

Also, both sighted and blind users may simply want to select contents on the screen when they have no gesture ideas for triggering the commands. In our study, *deictic gestures* were performed by our participants stating that they would like to place the menu items and select them. However, finding and selecting the on-screen contents will be different between sighted and blind users, given that blind users have no visual cues to help them find the content on the screen. For such cases, gesture recognition should allow users to rely on non-visual skills such as proprioception, for example, users can select the content just by moving their hand in certain directions (upper-left, right, etc.) with reference only to their own bodies.

I want to use old things that I have learned

During the study, our participants often came up with gestures that they commonly used to interact with appliances and devices in their daily lives. For example, the highest agreed gesture in both studies for the *Select Channel* command was rotating the hand left and right which mimicked rotating the channel dial on old televisions. Similarly, the participants selected a "*Push button* gesture in mid-air for the *Open/Close TV* command which mimicked pushing buttons in their mobile phones. Thus it is important for designers to understand the commonly-used actions used by blind people when interacting with their commonly used devices and appliances. Although the influence of daily device usage may also be applicable for sighted users, the devices used and the interaction experiences can be quite differently perceived between blind and sighted people (e.g., smartphone vs. feature phones). Thus, designers should pay close attention to those differences when designing gestures that mimic the use of older devices.

Reversible gestures are more memorable

We observed that our participants used consistent mapping of gestures for many commands, for example, *Next/Previous, Volume Up/Down, Play/Pause*, and so on. Notably, reversible gestures were very important for blind users for gesture learnability and memorability because mapping similar and opposite things was an efficient learning aid for them in many cases. During Study 1, one of the participants stated, *"I do an open hands gesture (i.e. claps the two hands then separate them) one time for Play, so I would do the same gesture 2 times for*

Open TV because they are similar commands." This implies that even for commands that are not directly dichotomous, our participants preferred mapping similar or opposite gestures where possible. Thus, it is a good idea to design reversible gestures wherever possible.

Big movements are more recognizable

All our participants used noticeably larger movements when they performed gestures. This can be partially due to the lack of feedback during gesture performance in our study. We speculated that by exaggerating their movements our participants compensated for the lack of feedback intending to eliminate uncertainty regarding the perception of their gestures and ensuring that their movements were clear enough to be recognized by the system. Thus, gesture recognition should be developed to allow freedom of movement. Also, time-based gesture recognition should be avoided because blind people may perform gestures with big movements at a different pace to sighted people.

User-Elicitation and Choice-Based Elicitation

We observed that the choice-based elicitation approach can work well when the user has no clear idea of gesture possibilities. Nevertheless, we cannot disregard the results from user-elicitation. In particular, we found that the results from the user-elicitation proved to be very useful. They enabled us to understand the gestures that users were familiar with and the commands that they had little idea about and those that they had difficulty imagining gesture(s) for. Conducting only the choice-based approach may not allow us to obtain such insights. By combining the information from both the user-elicitation and the choice-based elicitation approaches, we gained useful insights for designers, as demonstrated in previous section. We thus suggest that it would be beneficial to use both approaches during gesture design processes.

Participants commented on the benefits of the choice-based elicitation approach in three ways. First, it helped users understand what is possible within the design space. By learning through the choices, participants commented that they understood better about the different possibilities in the design space. This understanding was reflected by one participant who stated that, "*For some commands, it was just impossible to imagine, I really appreciated the choices, as I could learn through examples.*" Second, the choice approach made participants to think actively about all the possibilities presented to them. Participants commented that the process of selecting their preferred choices caused them to think actively, i.e., to carefully weigh the benefits and disadvantages of each gesture. One of the participants commented that it was fun as they felt like they were 'wearing the hat of a designer'. Third, by giving choices, our participants felt more confident when they discovered that the gestures they had preferred were also in the choice list. All blind participants reported that they appreciated the gesture choices as they helped guided them to decide on their preferred gestures. One participant stated, "*The choice list helps me to express what I really want.*", another commented, "*This is like a game and it's fun!*".

For designer convenience, we summarized our experience of using the choice-based elicitation approach in the form of five generalized guidelines.

1. We should consider the number of choices based on the memory and attentional capacity pertaining to different demographics. For example, old people may prefer a fewer choices, compared to young people.

2. It is essential for users to actually perform each gesture choice instead of just passively to look at or listen to a description of a gesture. By actually performing the choice, users can better estimate the suitability of the gesture.

3. The criteria for selecting candidate gestures for the choice list should be specified as objectively as possible. These criteria may vary depending on the priorities set by the designers, e.g., high-comfort, high-speed and/or low-error rate, etc.

4. It is important to carefully consider the issue of learnability in the choice-based elicitation approach, as the selected gesture may not have originated from the user's mind. It is thus essential to pay special attention to gesture learnability in the choice-based elicitation approach.

5. Although we found that the choice-based approach is effective for our blind participants, it is also worthwhile to conduct a user-defined approach to better understand users. In other words, user-defined and choice-based approaches have their own respective advantages and they can well complement one another limitations.

CONCLUSION AND FUTURE WORK

We conducted two user studies: a user-elicitation study and a choice elicitation study to determine the mid-air gesture preferences of blind people in TV interactions. From Study 1, we developed and propose a classification of gesture types and the frequency of body parts usage. From Study 2, we presented a gesture set for blind people in TV interactions. Study 2 also confirmed that providing choices can help reconfirm the preferences of gestures for unfamiliar commands. We conclude by discussing concrete design guidelines for mid-air TV interaction gestures for blind people.

In our study, the number of participants and diversity in their demographic background were limited. Thus, we will extend our work to include more diverse blind participants (with diverse level of sight, age, the age of onset of blindness). Also, we will further explore the use of user- and choice-based elicitation approaches and more theoretical grounding of choice candidate gestures.

ACKNOWLEDGMENTS

We thank the members of the Center for Human-Engaged Computing in Kochi University of Technology for their support. We also thank anonymous reviewers for their thoughtful suggestions. This study has been partially supported by the Grant-in-Aid for Scientific Research of Japan under Grant No. 25330241.

REFERENCES

1. Fraser Hamilton Alex Carmichael, Helen Petrie and Jonathan Freeman. 2003. The Vista Project: Broadening Access To Digital TV Electronic Programme Guides. *PsychNology Journal* 1, 3 (2003), 229–241.
2. American Foundation for the Blind. 1997. Blindness statistics. (1997). Available: http://www.afb.org/info/blindness-statistics/statistical-snapshots-archive/technology-related-statistics-archive/technology-video-description/2345, Accessed on June 30, 2015.
3. Gregg Vanderheiden Chris Law. 2000. The development of a Simple, Low Cost set of Universal Access Features for Electronic Devices. In *Proc. CUU 2000*. ACM, 118. DOI:http://dx.doi.org/10.1145/355460.355545
4. Nelson Cowan. 2010. The Magical Mystery Four: How Is Working Memory Capacity Limited, and Why? *Current Directions in Psychological Science* 19, 1 (2010), 51–57. DOI:http://dx.doi.org/10.1177/0963721409359277
5. Nem Khan Dim and Xiangshi Ren. 2014. Designing Motion Gesture Interfaces in Mobile Phones for Blind People. *Journal of Computer Science and technology* 29, 5 (2014), 812–824. DOI: http://dx.doi.org/10.1007/s11390-014-1470-5
6. Sheena Iyengar. 2011. *The Art of Choosing*. Twelve.
7. Brandon Rothrock Jacob O. Wobbrock, Htet Htet Aung and Brad A. Myers. 2005. Maximizing the guessability of symbolic input. In *Proc. CHI EA 2005*. ACM, 1869–1872. http://doi.acm.org/10.1145/1056808.1057043
8. Meredith Ringel Morris Jacob O. Wobbrock and Andrew D. Wilson. 2009. User-defined gestures for surface computing. In *Proc. CHI 2009*. ACM, 1083–1092. http://doi.acm.org/10.1145/1518701.1518866
9. Yang Li Jaime Ruiz and Edward Lank. 2011. User-defined Motion Gestures for Mobile Interaction. In *Proc. CHI 2011*. ACM, 197–206. DOI: http://dx.doi.org/10.1145/1978942.1978971
10. Kibum Kim, Xiangshi Ren, Seungmoon Choi, and Hong Z. Tan. 2016. Assisting people with visual impairments in aiming at a target on a large wall-mounted display. *International Journal of Human-Computer Studies* 86 (2016), 109 – 120. DOI: http://dx.doi.org/10.1016/j.ijhcs.2015.10.002
11. Gregory Abowd W. Keith Edwards Matthew Bonner, Jeremy Brudvik. 2010. No-Look Notes: Accessible Eyes-free Multi-touch Text Entry. In *Proc. Pervasive 2010*. Springer, 409–427. http://dx.doi.org/10.1007/978-3-642-12654-3_24
12. Andrew D. Wilson Meredith Ringel Morris, Jacob O. Wobbrock. 2010. Understanding users' preferences for surface gestures. In *Proc. GI 2010*. CIPS, 261–268. http://dl.acm.org/citation.cfm?id=1839214.1839260
13. Miguel A Nacenta, Yemliha Kamber, Yizhou Qiang, and Per Ola Kristensson. 2013. Memorability of Pre-designed and User-defined Gesture Sets. In *Proc. CHI 2013*. ACM, 1099–1108. http://doi.acm.org/10.1145/2470654.2466142
14. Michael Nielsen, Moritz Störing, Thomas B. Moeslund, and Erik Granum. 2004. A Procedure for Developing Intuitive and Ergonomic Gesture Interfaces for HCI. In *Gesture-Based Communication in Human-Computer Interaction*, Vol. 2915. Springer, 409–420. DOI: http://dx.doi.org/10.1007/978-3-540-24598-8_38
15. Jacob O. Wobbrock Shaun K. Kane and Richard E. Ladner. 2011. Usable Gestures for Blind People: Understanding Preference and Performance. In *Proc. CHI 2011*. ACM, 413–422. DOI: http://dx.doi.org/10.1145/1978942.1979001
16. Jeffrey P. Bigham Shaun K. Kane and Jacob O. Wobbrock. 2008. Slide Rule: Making Mobile Touch Screens Accessible to Blind People using Multi-touch Interaction Techniques. In *Proc. ASSETS 2008*. ACM, 73–80. DOI:http://dx.doi.org/10.1145/1414471.1414487
17. Chaklam Silpasuwanchai and Xiangshi Ren. 2015. Designing concurrent full-body gestures for intense gameplay. *International Journal of Human-Computer Studies* 80 (2015), 1 – 13. DOI: http://dx.doi.org/10.1016/j.ijhcs.2015.02.010
18. Jaime Snchez and Fernando Aguayo. 2006. Mobile Messenger for the Blind. In *Proc. ECRIM 2006*. Springer, 369–385. http://dx.doi.org/10.1007/978-3-540-71025-7_24
19. Jaime Snchez and Eduardo Maureira. 2006. Subway Mobility Assistance Tools for Blind Users. In *Proc. ECRIM 2006*. Springer, 386–404. http://dx.doi.org/10.1007/978-3-540-71025-7_25
20. Hugo Nicolau Daniel Gonalves Joaquim A. Jorge Tiago Guerreiro, Paulo Lago. 2008. From Tapping to Touching: Making Touch Screens Accessible to Blind Users. *IEEE Multimedia* 15, 4 (2008), 48–50. DOI: http://dx.doi.org/doi:10.1109/MMUL.2008.88
21. Gregg Vanderheiden. 1996. Use of Audio-haptic Interface Techniques to Allow Nonvisual Access to Touchscreen Appliances. In *Proc. HFES 40*. Springer, 1266. DOI: http://dx.doi.org/10.1177/154193129604002430
22. Radu-Daniel Vatavu. 2012. User-defined Gestures for Free-hand TV Control. In *Proc. EuroiTV 2012*. ACM, 45–48. http://doi.acm.org/10.1145/2325616.2325626