EYES-FREE METHODS FOR ACCESSING LARGE AUDITORY MENUS

Raine Kajastila

Aalto University
Department of Media Technology
P.O. Box 15500, FI00076 Aalto, Finland
raine.kajastila@tkk.fi

Tapio Lokki

Aalto University
Department of Media Technology
P.O. Box 15500, FI00076 Aalto, Finland
tapio.lokki@tkk.fi

ABSTRACT

Two interaction methods for eyes-free control of a mobile phone or a media player are introduced. The methods include a gestural pointing interface and a touchscreen interface to a spherical auditory menu where feedback is provided using spatially reproduced speech. The methods could facilitate the eyes-free use of devices and also make them accessible for visually impaired users. The effectiveness of gestural and touchscreen interaction is compared to traditional visual interface when accessing large menus. Evaluation results prove that moderately fast and accurate selection of menu items is possible without visual feedback. Combining eyes-free interfaces, positions of menu items in 3D and a browsing method with a dynamically adjustable target size of the menu items allows the use of large menus with intuitive easy access.

1. INTRODUCTION

The design and use of audio-only and eyes-free interfaces has been emerging in recent years. They can bring better usability in situations where eyes-free operation is necessary [1]. Such cases include the competition of visual attention, absence or limitations of visual display, or reduction of battery life [2]. With proper design, an audio interface can be even more effective than its visual counterparts [2]. Although audio interfaces are not yet widely used in public, major companies have already realized their potential [3]. Furthermore, they are important as assistive technology for visually impaired users.

This paper presents two interaction methods that allow a reasonably fast and accurate way to navigate spherical auditory menus with large number of menu items. This work builds upon the author's previous work [4], which is presented in Figure 1. The user points or tilts the control device to different directions to browse an egocentric auditory menu. As the user browses the menu items, they are read out loud and the sound is reproduced from the correct 3D direction. Fast browsing is enabled with reactive interruptible audio design [2] and the accuracy in selection is enhanced by the dynamic movement of menu items and an expansion of the selection area. The initial study [4] proved that this type of gestural interaction with auditory menus is efficient and intuitive. The second novel interaction method uses touchscreen input with auditory menus. With good design, auditory menus can be combined to work seamlessly with visual menus [5], thus making eyes-free use of devices intuitive and easy. Touchscreens can also be a barrier for visually impaired users [6], but visual touch screen menus can be made easily accessible by using audio feedback.

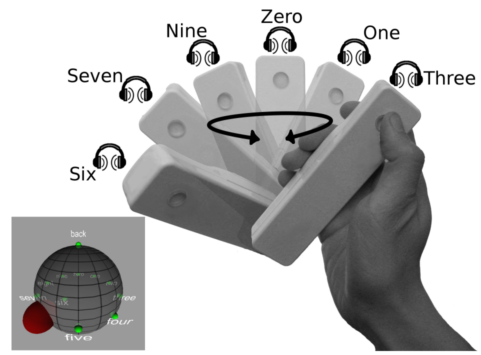

Figure 1: Gesture interface utilizing a mobile device mockup, as introduced in [4]. Auditory menu items can be accessed by pointing or tilting the device to desired direction. Browsing can be continued to the neighboring menu items by rotating the wrist.

The two menu configurations and the introduced interaction methods make it possible to browse auditory menus with large number of items (>100), for example, a contact list in a phone or a playlist in a music player. Furthermore, the interaction methods can be used to handle all basic controls of a modern mobile phone or a music player that contains either accelerometers or a touchscreen. It is also possible to construct a small multi-functional device consisting of only one button and an internal rotation sensing devices, for example, accelerometers or alternatively a touch surface device without a screen. Such robust devices without visual displays can be inexpensive and have low energy consumption but still offer the same functionalities as similar devices with a visual screen. A good example of a such a device is Apple's iPod shuffle [3], which gives feedback to users using synthesized speech. Our interaction and browsing techniques enable more sophisticated control of devices such as the iPod shuffle.

The novelty of the presented techniques lies in advanced auditory menus that can be used with two parallel interaction methods. In the gesture interaction, simple and intuitive wrist rotations are measured with three accelerometers. Touch interaction extends the circular touchpad implementation of earlier work [2] to be applied in touchscreen devices. Both methods can be combined with various browsing techniques and provide a way to access a high number of selectable content especially in 360-degree spherical auditory menus. Furthermore, menus enable the interoperability of the visual and auditory menus where the control logic remains the same in visual and auditory menus, as described in author's previous work [5].

2. RELATED WORK

The implemented interaction methods mainly build on auditory menus controlled with gestures or a touchscreen.

2.1. Auditory menus

In previous studies, different interaction methods with audio objects in menus and different spaces surrounding the user have been presented, and their input methods range from normal keypads and touch interfaces to gestural interfaces. Hiipakka and Lorho [7] used cursor keys for a spatial audio interface for music players. Their approach enabled a browsing system where spatialized sounds along the horizontal axis informed the user about context, menu structure and interaction possibilities.

Pirhonen et al. [8] tested a prototype eyes-free touch interface for a music player, in which finger sweeps on the screen controlled the playing of the music. The interface was provedn to be effective in eyes-free situations, and the results of the study pointed out that immediate audio feedback is crucial for user confidence.

Savidis et al. [9] presented a method of pointing interaction where a data glove, head tracker, voice recognition and headphones were used to produce a modifiable circular audio environment. They used the concept of auditory windows, where a subset of four sound objects was simultaneously played in a spatially larger area, while others were suppressed closer together.

An egocentric circular auditory menu, which is also applied in this paper, has been proposed many times earlier. Brewster et al. [1] used a directional head nodding interface to study four simultaneously playing menu items located around the user. They found egocentric menu design better than exo-centric and the selection method using the head-tracker was also successful in a mobile experiment. Brewster et al. [1] hypothesized that more than 8 simultaneously playing menu items would be difficult to handle with the system in their experiment. Circular auditory menu structures have also been applied in nomadic radio by Sawhney and Schmandt [10] and in diary application by Walker et al. [11].

Marentakis and Brewster [12] studied audio target acquisition in the horizontal plane with the aid of orientation trackers. The experiment focused more on how target width, distance and user mobility affects random target acquisition with a gestural pointing interface. They concluded that a pointing interaction with 3D audio is successful with mobile users. They also suggested that audio elements with feedback in egocentric audio displays could produce efficient design.

The user studies of circular auditory menus with touch input by Zhao et al. [2] showed that auditory menus can outperform typical visual menus used in iPod-like devices. The menu used by Zhao et al. is similar to the one presented in this paper. Their key elements include: 1) a touch interface similar to the iPod, where menu items are mapped to a circular touchpad; 2) direct reactivity to user touch input that gives control to the user without waiting periods; 3) interruptibility of the audio, where only one sound is played at a time, but its playing can be interrupted if user chooses to continue browsing; and 4) menu items that can be accessed directly without browsing through all items.

2.2. Gestural and touch control

Different types of tilting interfaces have been presented mainly for visual displays and writing applications. TiltType [13] has a writing interface where the tilt direction of the device can be used two-handedly to specify letters with the aid of 4 buttons. Wigdor and Balakrishnan [14] proposed a similar system in mobile phones, where tilting direction and numeric keypad press define the output character. Oakley and Park [15] presented a one-dimensional tilt menu system for mobile phones with tactile feedback. Tian et al. [16] studied a circular tilting menu using a pen, where the pen tilt direction was used to select visual menu items. One of the first tilting interactions was presented by Rekimoto [17]. He utilized a FASTRACK position and orientation sensor and applied tilting and a two-button-device to browse menus on a visual display.

Recently, the interaction with wrist rotations of horizontally held arm has been studied by Crossan et al. [18]. Their multipart mobile device consisted of a SHAKE sensor pack attached to the user's hand as a wristwatch and a Nokia N95 as visual feedback. They concluded that horizontal wrist rotation is a quite accurate and feasible control method, but simultaneous walking makes control more difficult.

The Wii Remote has been used in audio only music browsing by Stewart et al. [19] for moving in a large music collection located in surrounding 2D or 3D spaces. According to the authors, their Wii interface was not really usable. It only allowed the definition of general moving directions and was given negative evaluations by most of the users. The interaction methods presented in this paper would radically improve the usability of the music browser, because it would allow an exact definition of the direction of movement.

One important aspect of eyes-free auditory menus is their suitability as assistive technology for blind users. Guerreiro et al. [6] have implemented a gesture-based text entry method for touchscreen devices. In their NaviTouch interface, all the letters are accessed through vowels. The user first slides his finger vertically to find vowels that are read out loud. After hearing any of the vowels (e.g. A), the user can slide his finger vertically to find consonants that are after that particular vowel in the alphabets (e.g. B or C). The user makes one L-shaped gesture for each successful consonant selection. With this approach, the alphabets cannot be accessed directly, which slows down the writing process.

Another good example of touchscreen input is Slide Rule [20]. Kane et al. used the iPhone for eyes-free browsing of lists, selecting items, and browsing and changing music tracks. Kane et. al. also used similar L-shaped touch-gestures for browsing music tracks in Navitouch [6]. In the experiment, 10 album names were placed vertically in a list. Each item on the list could be listened to one at a time. The user first found a desired album with a vertical finger-swipe and continued the finger movement to the right to heari the track names read out loud. The multi-touch capability of the iPhone was utilized by tapping the screen with a second finger to select the desired track. Although the songs can be accessed by using only one continuous touch-gesture, this method does not solve the problem for music libraries holding hundreds of albums; queezing them into a vertical list would be difficult.

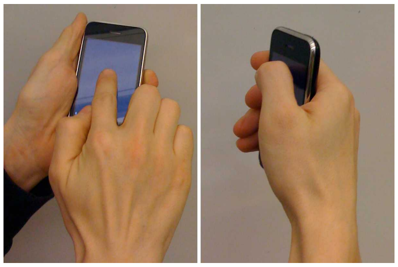

Figure 2: Two interaction methods was implemented as prototype applications for the iPhone. No visualizations of the auditory menu was shown in the device.

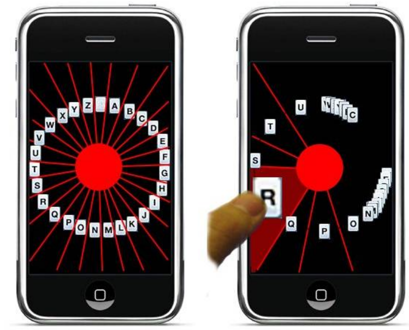

Figure 3: The menu items are defined as dynamically changing sectors on a screen. The visualization is exaggerated and artificially created. It is not needed in eyes-free auditory menu.

3. INTERACTION METHODS

In this work, two eyes-free interaction methods, wrist rotation gestures and touch-gestures of the finger are applied. The prototype implementation was aimed at the iPhone, but the interaction methods can be applied with other devices as well. Next, the novel control methods are introduced.

3.1. Gestural control method

Accelerometers along three axes can be used to sense the orientation of a device relative to the direction of earth's gravity. Problems can occur when the device is rotated horizontally along one axis, because accelerometers cannot sense that type of motion. This scenario can be avoided by holding the device as illustrated in Figure 2 and rotating it as show in Figure 1 [4]. The device can be used like a joystick by tilting the device slightly and rotating it 360 degrees with a gentle wrist gesture. To ease the targeting of menu items, the tilt angles of the device between 5 and 90 degrees are clamped to 90 degrees, i.e., zero elevation. This allows smaller wrist movements and prevents any tedious turning and twisting of the wrist. This control method is especially suitable for horizontal 360-degree menus, because all directions can be reached with equal effort and ease.

3.2. Touch-surface control method

A touch-surface (or a screen) can also be used to access a circular auditory menu, see Figure 2. Sectors extending from the center of the surface represent the menu items, as shown in Figure 3. The user can access any item directly by placing a finger on the surface and can continue browsing with a circular finger sweep. A selection is made by removing the finger from the surface. The center of the touch-surface is a safe area from where the finger can be lifted without making a selection. As explained later, special actions can be assigned to the center of the touch-surface.

4. SPHERICAL AUDITORY MENUS

The auditory menu described in this paper has similarities to the works of Brewster et al. [1] and Zhao et al. [2]. The key element is the use of interruptible audio and immediate reactivity to user input with an auditory display. The spoken menu items are played one by one while browsing a menu and the user has the ability to jump to the next item thus stopping the playback of the previous one. With slower motion, the user can hear all menu items one by one. When jumping to the other side of the menu, no sound is heard until the gesture or finger movement has slowed down. Thus, the user is in control and he/she can adjust browsing speed according to his/her own abilities.

When browsing faster, the user hears only the beginning of the sounds. Because the short sounds (or phonemes) represent the first letter(s) of the names, they help the user keep track of the position in a large menu. This feature was recently evaluated as beneficial and was suggested to be named "spindex" [21]. In the author's previous works [4][5] and in the auditory menu described in this paper, the spindexes are automatically generated when the user browses the menu. This is achieved by the auditory menu's instant reactivity to users' gestures by using prerecorded names or fast text to speech synthesis. By slowing down the browsing speed, the user can adjust the length of the spindex thus enabling an efficient search method for menu items starting with a same letter, letters or even word.

To enhance the selection accuracy, a dynamically adjusted target sector, where the item is active (played), is applied. As visualized in Figure 3 (left), if none of the items is active, the menu items have an even distribution of the target area. When a menu item is active, its target area expands in both directions reaching a 1.9 times larger target area. The value of 1.9 was chosen to leave a big enough target area for the neighboring menu items, because they shrink, allowing space for the expanding sector. This is done to facilitate easier browsing and selection by reducing undesired jumping between tightly packed menu items.

When a selection is made by removing the finger from the screen, it is important to give feedback to the user. There are many suggested feedback sounds, e.g., auditory icons [22], earcons [23] and spearcons [24]. The implementation presented in this paper, uses a fast replay of the selected menu item mixed with a short auditory icon. A short clink-sound is played immediately after the selection, followed by the fast replay of the selected menu item. The playback time of the sound is shortened considerably, but the user can still easily recognize the content. The clink-sound further clarifies that the selection was made. The changed pitch also indicates a feedback sound, not another menu item. In this way, the user gets immediate feedback and he/she can easily double-check whether a correct selection was made. A modified sound sample of a plucked guitar was used to inform the participants if

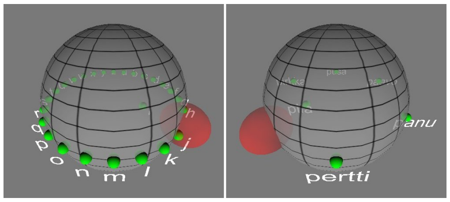

Figure 4: The two-layer menu. The alphabets are always found from the same position in the first menu layer. The second layer holds 6 names in alphabetical order and they are spread evenly.

they moved their finger to the center of the screen or pointed the device up. This was done to notify the participant in the case of an advanced one-layer menu that menu items have been spread evenly around the user.

The correlation between the touched screen location and the reproduced sound directions can help the user associate the sound to the specific menu item location [12]. Proper design can also improve the performance as the spatial menu item configuration becomes familiar to the user. Furthermore, each menu item is heard from a different spatial direction making it easier to distinguish them when browsing with increased speed. The binaural implementation for headphone reproduction applies head-related transfer functions (HRTFs) measured in house in an anechoic environment with a dummy head. In addition, the sounds are processed with a simple reverberation algorithm, which helps in externalization of auditory menu items. Informal listening tests suggest that spatial audio makes the use of the system easier. However, no user tests have been done regarding how much (if any) the mono sound reduces the usability and whether it is possible to use the system with only one headphone or loudspeaker.

The goal of this work is to study the selection speed and accuracy of auditory menus with large number (>100) of items. We designed two alternative auditory menu layouts containing a contact list of 156 (26 x 6) names. As explained below, one menu layout is a more traditional auditory menu with two layers and other uses a novel approach to fit all 156 names to one menu level. Earlier studies with different interaction methods have suggested that egocentric auditory menus could contain at maximum 5 [25], 8 [1], or 12 [2] menu items in usable scenarios. This is probably due to limitations in interaction devices, browsing methods or simultaneously played sounds. With our auditory menu layout the number of names displayed to the user can be dramatically increased compared to the number of names in Slide Rule [20].

4.1. Menu with two layers

The general layout of the two layer menu is visualized in Figure 4. The first menu level consists of 26 letters from A to Z, which are always found in the same locations and are placed in alphabetical order. When none of the menu items is selected, the target sector width in the first menu level is 13.85 degrees (360 / 26). After an item is activated by touching or pointing, its target sector expands to occupy 26.3 degrees (13.85 x 1.9). In the implemented prototype, the user can select a menu item by releasing the finger from iPhone's touchscreen. By selecting a letter, the user can advance to the second layer of the menu. The second layer holds six names starting with the chosen letter. They are evenly spread around 360 degrees and have a target area of 60 degrees. The target area of an active item is now 114 degrees. The names are listed in alphabetical order. For testing purposes, after a name selection, the menu automatically jumps to the first layer. In this prototype, there was no option for going back in the menu structure.

4.2. Menu with hundreds of items in one layer

The novel one-layer menu layout can handle very large number of items, as shown in Figure 5. The applied browsing method combines the benefit of a-priori known item positions in a static menu with large menus. In this approach, when none of the items is selected the menu items are in their absolute positions defined by alphabetical order. For example, all the names starting with the letter A are placed in alphabetical order to the sector that occupies the letter A in the menu shown in Figure 4 (left). Thus, the user can point to or touch desired position and hear one of the names starting with that letter. When none of the items is selected, the target sector width for each menu item is 2.31 degrees (360 /156).

When the user targets a particular item, its neighboring items are spread around evenly with a spacing of 40 degrees, and items farther away are grouped together (see Figure 5, right). If the desired menu item was not directly found, the user can continue browsing items with a rotating hand gesture or a circular finger sweep. The next item is always found 40 degrees forward and the previous one 40 degrees backwards, respectively. Now the target sector width is significantly larger - 76 degrees (40 x 1.9) for the active menu item. Such dynamic item placement greatly reduces undesired jumping between items. This spreading can also be seen as a different implementation of the fisheye distortion concept [26, 27]. In the preliminary tests, the spacing was set to 40 degrees, but further study is needed to find the optimal spacing for different interaction methods.

Initial testing suggested that it is possible to access the desired item without extensive training. This browsing method can significantly facilitate the browsing of alphabetically ordered large menus, such as a music playlist or an address book containing hundreds of items. In the visual domain, this is similar to the possibility of jumping directly to names starting with the same letter in an address book.

5. USER EXPERIMENT

A mobile user experiment was conducted to ascertain the performance of the introduced eyes-free menu browsing techniques. The test environment is depicted in Figure 6 (left). A participant navigated his way around four aligned chairs while choosing names from a large menu. The auditory menus were reproduced with headphones and the names to be chosen with different interaction methods were listed on the large projection screen.

5.1. Participants

Nine participants completed the experiment. All of them were males between 23 and 43 years old. The participants volunteered for the experiment, and they had no previous experience with the interaction methods and auditory menus used. Three participants owned iPhones. Additionally, three participants were left handed.

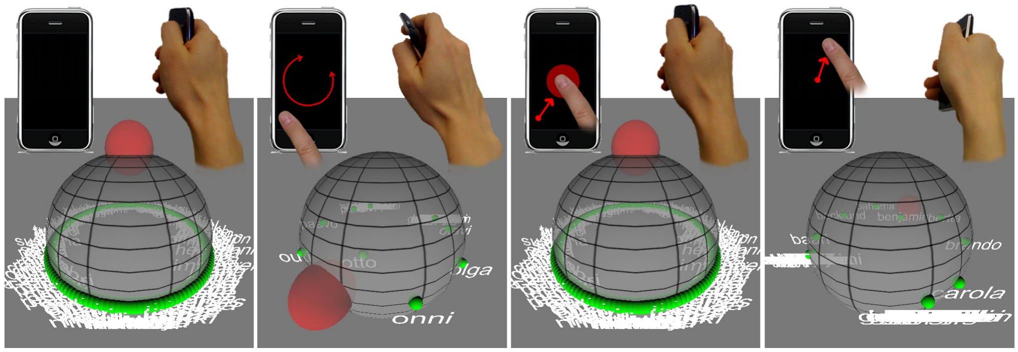

Figure 5: Method for browsing large eyes-free auditory menus. It combines qualities of absolute positioning of menu items and browsing easiness of smaller menus. The browsing method can handle hundreds of items that can still be accessed fast.

5.2. Apparatus

An iPhone was used as the test device. The iPhone was chosen because it had the desired features and only one device for all interaction methods was needed. The iPhone was connected to a Macbook Pro using a wireless network. The participants used Sennheiser HDR 120 wireless headphones whose transmitter was connected to the laptop's audio output. This setup allowed the participants to be truly mobile without the need for all the software running in the iPhone. The visual reference test was implemented with the iPhone's contact list application.

The connection between the iPhone and the laptop was implemented by using the Open Sound Control (OSC) protocol and a modified version of the free Mrmr software [28] installed on the iPhone. The auditory menu was implemented using Pure Data (PD) [9], which received the raw control information from the iPhone. High quality samples (fs = 44.1 kHz) for the menu items were pre-recorded by synthesizing them with text-to-speech software. All samples started immediately in the beginning of the sound file to ensure fast responses to user actions.

5.3. Procedure

After a brief tutorial of the two interaction methods, the participants were given headphones and an iPhone. The touch interface was introduced to the participants before the gesture interface. Both tutorials consisted of writing a test sentence with total of 25 letters by using the menu shown in Figure 4 (left). This took approximately 5 minutes. The visualization was shown on a laptop screen when writing the first word of the test sentence. Afterwards, the visualization was hidden and the participant practiced the rest of the sentence using audio cues only.

The actual experiment consisted of four tasks using auditory menus and one visual task giving reference time and accuracy. The gesture interface was used with one hand. In the touchscreen experiment, the device was held in one hand and, while the other hand's finger was used for the menu browsing. The two handed method was chosen, because it was expected that some participants (e.g., persons with small hands) might not feel comfortable browsing the menu with their thumb. However, many iPhone owners use their device one handedly. The participants were instructed to select the given names aiming at maximum speed with minimum errors. The names to be searched were shown in groups of 5 on a large projector screen, as shown in Figure 6 (left). Before every task the participants were allowed to rehearse the interaction method using a set of 5 names, which remained the same for all methods. The practice time was again restricted to 5 minutes. In the experiment, 11 slides containing 5 names were used. The next slide was revealed right after the last name of the previous slide was completed. The participants were instructed to carry on to the next name, even if they made a mistake. The names were Finnish first and last names.

The whole experiment including the training periods lasted about 1.5 hours. Throughout the main experiment the participants walked in a figures-of-eight around 4 chairs placed across the room, similar to the method used in [11, 1]. The spaces between chairs were 2, 1, and 2 meters, as seen in Figure 6 (left). The participants were instructed to keep a steady pace when walking. They were reminded to continue walking if they stopped for some reason during the experiment. We designed the experiment so that we would have exact reference time and accuracy from an existing visual application. The experiment did not contain a noisy environment or other unexpected events comparable to oncoming pedestrians in the real world. The aim was not to prove the general validity of the auditory menus, which have already been extensively studied. Instead, the mobile experiment was made under conditions that reveal if the presented interaction methods suffer any setbacks when the user is moving during which, for example, the hands could be shaking.

5.4. Design

The experiment was a simple factorial design, in which five different interaction methods were tested. The used auditory menus are explained in Section 4. The 5 methods were:

- *Reference* (Ref), the normal contact list of the iPhone without any auditory feedback.
- *Touchscreen one-layer* (T_1L), the eyes-free touchscreen input with 3D audio output. All names were directly accessible in one menu layer, see Figure 5.
- *Touchscreen two-layers* (T_2L), the eyes-free touchscreen input with 3D audio output. The subject selected first the first letter and then the name from a submenu, see Figure 4.

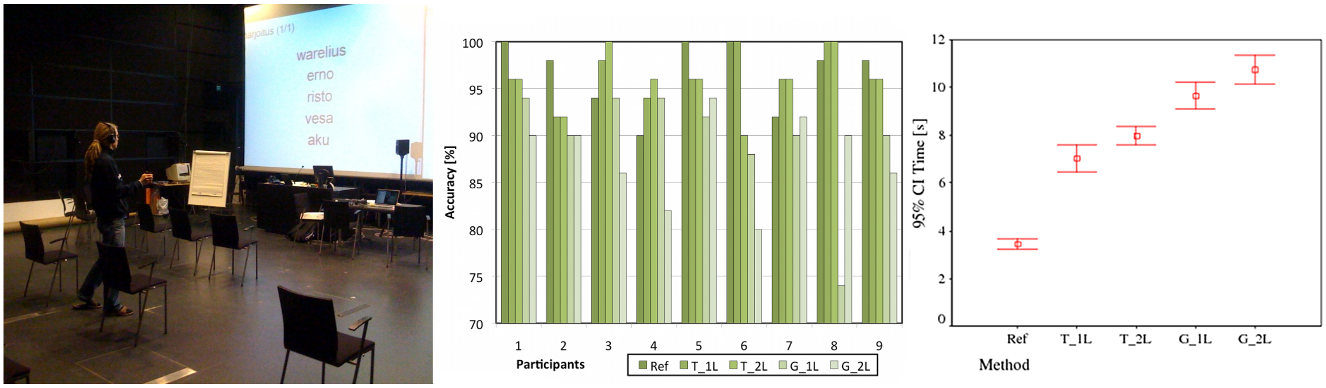

Figure 6: Left: The mobile user experiment. Center: Mean accuracies of individual participants with all tested interaction methods. The scale of the y-axis does not start from 0. Right: Means and 95% confidence intervals of selection times of all interaction methods.

- *Gesture-based one-layer* (G_1L), the eyes-free gesture input with 3D audio output. All names were directly accessible in one menu layer, see Figure 5.
- *Gesture-based two-layers* (G_2L), the eyes-free gesture input with 3D audio output. The subject selected initially the first letter of the name and then the name from a submenu, see Figure 4.

The order of the tested methods was randomized between the participants to ensure proper control groups. In all methods, the participants browsed the same contact list containing 156 names, 6 names for each of the 26 alphabets from a to z. The names to be searched were shown in groups of 5 using a big screen and the participants completed 11 lists using each interaction method. In order not to favor the audio interface, a visual presentation of stimuli was used for all five methods.

5.5. Data Collection and Analysis

In the audio-only methods, PD was used for recording user interactions to a log file with time stamped information about the selection process. The selection time was counted from the moment the finger touched the screen to the moment the finger was lifted and the selection was made. The time data was rounded to the nearest millisecond. The reference time was recorded using a simple program implemented with the iPhone's contact list application. The selection time was counted from the moment the participant touched a start button to the moment a selection was made. The collected data included the selected name and the time used to do the selection. This data enabled the derivation of the dependent variables, namely the time to find a name (time) and the correctness of selection (accuracy). The independent variables were the applied interaction method and the participant.

Each participant completed 11 lists of five names using all five methods. In total, 275 selections were made. The first list was considered as a training list, thus the data from the remaining 10 lists was were analyzed. The analysis method was similar to one used by Zhao et. al. [2]. The accuracy of each method was determined by checking the correctness of 50 selected names. Because the distribution of all raw selection times was positively skewed, only the median time of each 5-name list was used. Thus, from each participant 10 selection times per method were applied in a two-way analysis of the variance. The SPSS statistical software tool was used for the analysis.

6. RESULTS

The correctness of selected names is presented in Table 1, and the percentages of each participant are depicted in Figure 6 (center). The means of the median selection times are listed in Table 1. A two-way analysis of variances yielded a main effect for both the methods, F (4.405) = 154.77, $p < 0.001$, and the participants F (8.405) = 4.60, $p < 0.001$. Given the main effects observed, post hoc tests with Bonferroni confidence interval adjustments were performed to check for differences between the various input methods. All methods were found to differ significantly as seen also in Figure 6 (right). The interaction effect was also found significant, F (32,405) = 2.64, $p < 0.001$, indicating that some participants performed better with certain methods than others.

6.1. Result analysis

The eyes-free touchscreen input methods (T_1L and T_2L) were as accurate as the reference method. If compared with the earPod's [2] selection accuracy of 94.2% amongst 8 items, the accuracy 96.4% amongst 156 items can be considered very good in a mobile context. The gesture-based inputs (G_1L and G_2L) were slightly less accurate, but half of the participants made only a few errors (see Figure 6 (center)).

The mean selection time for the five interaction methods varied significantly. The touchscreen input methods were two times slower than the reference and the gesture-based interactions were almost three times slower than the references. However, the

Table 1: The results of the user experiment.

	Ref	T_1L	T_2L	G_1L	G_2L
Correct selections [%]	97.0	96.4	95.8	88.6	87.6
Selection times [s]	3.43	7.02	7.98	9.65	10.76

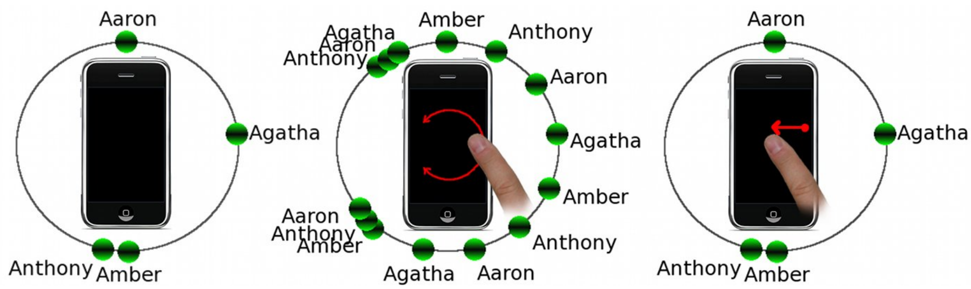

Figure 7: Advanced browsing method applied to small auditory menu with only few menu items.

novel methods were used without any visualizations, because the objective was an eyes-free usage scenario. The touchscreen input would certainly benefit from a visual representation of letters, as illustrated in Figure 3. It is also notable that one-layer menu structures were found to be faster than the two-layer structures with both touchscreen and gestures. With T_1L, the deviation of selection times is larger than with T_2L, indicating that sometimes the selection can be very fast, but in some cases finding the right name takes more time. With gesture-based interaction there are large individual differences in deviations.

7. DISCUSSION AND FUTURE WORK

The names and the alphabet can be browsed in both directions. Almost all participants commented about the difficulty in remembering the order of the alphabet while browsing the auditory menu. For audio-only browsing, the participants found it particularly hard to remember the reverse order of the alphabet. During the tasks, many participants paused browsing for a while and thought which way the desired letter or name would be found. Some participants even tried to find letters in the wrong place. During the test, the participants clearly learned the order of the alphabet and the positions where to find them. This might have increased the browsing speed towards the end of the experiment, although no significant learning effect was seen in the gathered data. Shorter search times can be expected after prolonged use of this type of auditory menus, because the alphabetical order (especially the reverse order) and the letter positions in the menu would become familiar to the user.

The accuracy of the gesture interface remained less (87.6 %) than observed in a previous study using large alphabet menus (95.1 %) [4]. This might be caused by the poor design of the menu item selection in the gestural interface. In the gestural interface, the selection was made by releasing the thumb from the touchscreen. This approach seemed good in the preliminary testing, but some participants did not find it comfortable. The participants complained that they could not grip the phone well enough without the thumb. Most errors were due to mistakenly selecting neighboring items instead of the intended one. It was observed that some participants released the thumb with great force and speed cousing the phone to jerk, which accidentally selected the next menu item. To overcome this problem, the buttons on the side of the iPhone can be used or, alternatively, an intuitive selection gesture can be assigned. Ergonomics should be one of the highest priorities in gestural interaction and buttons would make the use more comfortable and would ease the selection process.

The experiment also revealed a need for enhancing the touchscreen implementation. Some participants accidentally slid their fingers outside the touchscreen area. These actions were counted as finger lifts, and thus a wrong menu item was selected. These false finger lifts can be avoided either by discarding the finger lift actions that occur near the borders or attaching some kind of physical barrier to the borders.

The advanced menu item spreading in the one-layer menus (T_1L, G_1L) was proven effective in the experiment. There are many benefits in this approach. First of all, there is only one menu level and selection needs to be made only once when searching. This can reduce the possibility of an error and increase the selection speed. Also, the distance between menu items is always the same regardless of the number of items, which facilitates the browsing. Furthermore, this novel menu layout enables fast transition from one part of the list to another. This layout can be further enhanced to achieve faster and better usability. One improvement would be to define the start place to be always the first name in alphabetical order. When the user points the device or touches the screen, for example on the letter D, he/she would always know that names starting with D will be heard when browsing clockwise and with counter clockwise browsing the last name starting with C is found. The advanced spreading method could also be adapted to small menus with only a few menu items. An example of browsing a small contact list is depicted in Figure 7. Names can be positioned so that they are always found according to the second letter (see Figure 7, left). For example, the name Aaron is always positioned in front, and Amber behind, in the auditory menu. When the user points the device or touches the screen, the closest menu item becomes active and the rest is spread evenly around them. Items can be repeated when the user continues browsing the menu (see Figure 7, center). The user can always return to the absolute positioning of the menu items by moving the finger to the center of the screen (see Figure 7, right) or pointing the device up.

The usefulness of these improvements needs to be validated in further testing. Furthermore, after the user learns the absolute positioning and is able to directly access all of the letters of the alphabet, the menu layout could even be considered a powerful tool for eyes-free text-entry. Some of the improvements are being used and tested in the Funkypod (also known as Funkyplayer) software [5], which enables the browsing and playing of iPod's music library.

8. CONCLUSIONS

This work evaluated accurate and moderately fast gesture-based and touchscreen interaction methods with 3D auditory displays. The interaction method is well suited for eyes-free and mobile conditions. The control gestures are natural wrist movements, which are easy and intuitive to learn without extensive training. Synthesized speech samples were used with spatial sound reproduction to display menu items and give feedback about the selection to the user. The presented interaction methods and novel way of placing hundreds of menu items in one level of an auditory menu can be applied as an alternative control method for mobile devices, such as a simple mobile phone or a music player.

9. ACKNOWLEDGEMENTS

This research has received funding from the Academy of Finland, project no. [119092] and the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement no. [203636].

10. REFERENCES

[1] S. Brewster, J. Lumsden, M. Bell, M. Hall, and S. Tasker, "Multimodal 'eyes-free' interaction techniques for wearable devices," in *Proc. CHI*. ACM, 2003, pp. 463–480.

[2] S. Zhao, P. Dragicevic, M. Chignell, R. Balakrishnan, , and P. Baudisch, "earpod: Eyes-free menu selection using touch input and reactive audio feedback," in *Proc. CHI*. ACM, 2007, pp. 1395–1404.

[3] "iPod shuffle: portable music player," http://www.apple.com/, 2010, accessed 2.2.2010.

[4] R. Kajastila and T. Lokki, "A gesture-based and eyes-free control method for mobile devices." in *CHI'09 extended abstracts*, 2009.

[5] R. Kajastila, J. Vehviläinen, P. Palm, and T. Lokki, "Funkypod: Integrating auditory and visual menus." in *Proc. IWPASH*, 2009.

[6] T. Guerreiro, P. Lago, H. Nicolau, D. Gonalves, and J. Jorge, "From tapping to touching: making touch screens accessible to blind users," *IEEE Multimedia*, vol. 15, no. 4, pp. 48–50, 2008.

[7] J. Hiipakka and G. Lorho, "A spatial audio user interface for generating music playlists," in *Proc. ICAD*, 2003, pp. 267–270.

[8] A. Pirhonen, S. Brewster, and C. Holguin, "Gestural and audio metaphors as a means of control for mobile devices," in *Proc. CHI*. ACM, 2002, pp. 291–298.

[9] A. Savidis, C. Stephanidis, A. Korte, K. Crispien, and K. Fellbaum, "A generic direct-manipulation 3D-auditory environment for hierarchical navigation in non-visual interaction," in *Proc. ASSETS*. ACM, 1996, pp. 117–123.

[10] N. Sawhney and C. Schmandt, "Nomadic radio: Speech and audio interaction for contextual messaging in nomadic environments," *ACM Transactions on Computer-Human Interaction*, vol. 7, no. 3, pp. 353–383, 2000.

[11] A. Walker, S. Brewster, D. McGookin, and A. Ng, "A diary in the sky: A spatial audio display for a mobile calendar," in *Proc. BCS IHM-HCI*. Springer, 2001, pp. 531–540.

[12] G. Marentakis and S. Brewster, "Effects of feedback, mobility and index of difficulty on deictic spatial audio target acquisition in the horizontal plane," in *Proc. CHI*. ACM, 2006, pp. 359–368.

[13] K. Partridge, S. Chatterjee, V. Sazawal, G. Borriello, and R. Want, "Tilttype: accelerometer-supported text entry for very small devices," in *Proc. UIST*. ACM, 2002, pp. 201–204.

[14] D. Wigdor and R. Balakrishnan, "TiltText: using tilt for text input to mobile phones," in *Proc. UIST*. ACM, 2003, pp. 81–90.

[15] I. Oakley and J. Park, "A motion-based marking menu system," in *Proc. CHI*. ACM, 2007, pp. 2597–2602.

[16] F. Tian, L. Xu, H. Wang, X. Zhang, Y. Liu, V. Setlur, and G. Dai, "Tilt menu: Using the 3D orientation information of pen devices to extend the selection capability of pen-based user interfaces," in *Proc. CHI*. ACM, 2008, pp. 1371–1380.

[17] J. Rekimoto, "Tilting operations for small screen interfaces," in *Proc. UIST*. ACM, 1996, pp. 167–168.

[18] A. Crossan, J. Williamson, S. Brewster, and R. Murray-Smith, "Wrist rotation for interaction in mobile contexts," in *Proc. MobileHCI*. ACM, 2008, pp. 435–438.

[19] R. Stewart, M. Levy, and M. Sandler, "3D interactive environment for music collection navigation." in *Proc. DAFX*, 2008, pp. 13–17.

[20] S. K. Kane, J. P. Bigham, and J. O. Wobbrock, "Slide rule: making mobile touch screens accessible to blind people using multi-touch interaction techniques," in *Proc. ASSETS*. ACM, 2008, pp. 73–80.

[21] M. Jeon and B. N. Walker, "Spindex: Accelerated initial speech sounds improve navigation performance in auditory menus," in *Proc. HFES2009*, 2009.

[22] W. Gaver, "Auditory icons: Using sound in computer interfaces," *Human Computer Interaction*, vol. 2, no. 2, pp. 167–177, 1986.

[23] M. Blattner, D. Sumikawa, and R. Greenberg, "Earcons and icons: Their structure and common design principles," *Human Computer Interaction*, vol. 4, no. 1, pp. 11–44, 1989.

[24] B. Walker, A. Nance, and J. Lindsay, "Spearcons: speech-based earcons improve navigation performance in auditory menus," in *Proc. ICAD*, 2006, pp. 63–68.

[25] G. Lorho, J. Marila, and J. Hiipakka, "Feasibility of multiple non-speech sounds presentation using headphones," in *Proc. ICAD*, 2001, pp. 32–37.

[26] R. Spence and M. Apperley, "Data base navigation: an office environment for the professional," *Behaviour & Information Technology*, vol. 1, no. 1, pp. 43–54, 1982.

[27] G. Furnas, "Generalized fisheye views," in *Proc. CHI*. ACM, 1986, pp. 16–23.

[28] "Mrmr," http://poly.share.dj/projects/, 2010, accessed 2.2.2010.