

Comparing Free Hand Menu Techniques for Distant Displays using Linear, Marking and Finger-Count Menus

Gilles Bailly, Robert Walter, J.-P. Müller, Tongyan Ning, Eric Lecolinet

▶ **To cite this version:**

Gilles Bailly, Robert Walter, J.-P. Müller, Tongyan Ning, Eric Lecolinet. Comparing Free Hand Menu Techniques for Distant Displays using Linear, Marking and Finger-Count Menus. 13th International Conference on Human-Computer Interaction (INTERACT), Sep 2011, Lisbon, Portugal. pp.248-262, 10.1007/978-3-642-23771-3_19 . hal-00826676

HAL Id: hal-00826676

https://imt.hal.science/hal-00826676

Submitted on 28 May 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Comparing Free Hand Menu Techniques for Distant Displays using Linear, Marking and Finger-Count Menus

Gilles Bailly[1,2], Robert Walter[1], Jörg Müller[1], Tongyan Ning[1], Eric Lecolinet[2]
[1]Deutsche Telekom Laboratories, TU Berlin
[2]Telecom Paristech – CNRS LTCI
{gilles.bailly, robert.walter, hans-joerg.mueller, tongyan.ning}@telekom.de
eric.lecolinet@telecom-paristech.fr

Abstract. Distant displays such as interactive public displays (IPD) or interactive television (ITV) require new interaction techniques as traditional input devices may be limited or missing in these contexts. Free hand interaction, as sensed with computer vision techniques, presents a promising interaction technique. This paper presents the adaptation of three menu techniques for free hand interaction: Linear menu, Marking menu and Finger-Count menu. The first study based on a Wizard-of-Oz protocol focuses on Finger-Counting postures in front of interactive television and public displays. It reveals that participants do not choose the most efficient gestures neither before nor after the experiment. Results are used to develop a Finger-Count recognizer. The second experiment shows that all techniques achieve satisfactory accuracy. It also shows that Finger-Count requires more mental demand than other techniques.

Keywords: Finger-Counting, Depth-Camera, Public display, ITV, Menus

1 Introduction

Interaction with distant displays strongly differs from personal computers and interactive surfaces. Users generally do not have a mouse and a keyboard, nor can they reach a touchable surface. So, even common classical operations such as pointing [25], text entry [14] or command selection may be challenging.

In this article, we focus on command selection on distant displays for interactive television (ITV) or interactive public displays (IPD). In both cases, the number of available commands and services continuously increases, this requiring new menu techniques for presenting and organizing commands and for making it easy for users to select them. In the case of public displays, free hand menu techniques are especially relevant because they do not require the user to touch the screen, which can be dirty. They also avoid users having to make a detour to touch the screen, a useful property for passing-by scenarios (e.g., in subway

stations). Finally, free hand interaction does not require users to use their mobile phone, which may take time (the user must find the phone and wait for the network connection). Moreover, in the case of interactive television, free hand gestures are convenient for selecting frequent commands as they avoid the need to reach (or search) for the remote control [8].

Many menu techniques have been specifically designed for personal computers [1, 15, 26, 29], mobile devices [12, 24] or tabletops [2, 19], but menu techniques for distant displays have received limited attention and we are not aware of any comparative studies.

We propose three menu techniques for free hand interaction: Linear menu, Marking menu and Finger-Count menu. The Linear and Marking menu have similar behavior as their respective versions on personal computers or interactive surfaces: the cursor on the distant display is controlled by moving the hand and the command is selected by closing the hand. The Finger-Count menu is an adaptation of the technique proposed in [2] for multi-touch surfaces. It is a two-handed and multi-finger interaction technique that allows the user to perform commands by expressing a pair of digits with fingers "in the air": the left hand is used for choosing a category while the right hand serves to select an item in the corresponding category.

We performed two experimental evaluations. The first one, which is based on the Wizard-of-Oz protocol, was aimed to to understand how users would perform Finger-Count gestures in our free hand scenarios and to develop our Finger-Counting recognizer. The second study compared the three free hand menu techniques in an interactive TV scenario. Results show the accuracy rate of all techniques is above to 93% and that Finger-Count menu requires more mental demand. Our findings are relevant for the design of free hand menu selection for public displays and interactive television.



Fig. 1 Marking menus (Left), and Finger-Count menus (Right) on ITV.

2 Related work

Menu Techniques. Linear Menus are widely used for exploring and selecting commands in interactive applications. Several alternatives have been proposed for desktops [1, 7, 15, 16, 26 29], mobile devices [24] and interactive surfaces [2, 19]. Marking menus are certainly one of the most famous menu techniques. They combine Pie menus [7] and gestural interaction. In novice mode, the user selects commands in a circular menu. In expert mode, the menu does not appear and the user leaves a trail that is recognized and interpreted by the system. Marking menus are efficient as they favor the transition from novice to expert usage: users perform the same gesture in both modes [5]. Multi-Stroke menus [29] consist of an improvement of hierarchical Marking menus [16]: users perform a series of simple marks rather than a complex mark. This strategy improves the accuracy and reduces the total of amount of screen space [29].

Distant Displays. Studies on distant displays can be split into two main categories depending if users can use physical remote controls or not. Interactive TV is a typical case where users manipulate a physical remote control. However with the increasing number of services and multimedia data, users are forced to navigate in deep hierarchies or to manipulate remote controls overcrowded with buttons [8]. Performing free hand gestures can serve as a complementary modality for selecting frequent or favorite actions [13, 17]. For instance, a prototype of Marking menus based on computer vision-based hand gestures has been proposed in [17] to control frequent actions. However, this technique requires two specific registration poses, which are not appropriate for novice users. Moreover, this technique has not been experimentally evaluated. Finally, Microsoft recently introduced the Kinect, a combination of a RGB and a low-cost depth camera that enables users to play video games by performing body gestures in front of their TV.

While some public displays are multi-touch (e.g. CityWall [23]), this solution is not always appropriate as it forces users to stop walking for interacting. Moreover, some users may refuse to touch the display as it can be dirty. For these reasons, some projects have investigated computer vision to enable interaction with public displays [5, 21]. These projects however do not focus on menu selection.

Hand Gesture Interaction. Several interaction techniques based on hand gestures have been proposed [4, 27] especially in the context of virtual environments [10]. However, they generally use expensive and inconvenient hardware such as gloves that are not compatible with practical use. Some studies focus on computer vision based gesture recognition applications in HCI [18, 22]. However they mainly focus on recognition algorithms [17] and only few interaction techniques have been implemented for pointing [6, 25], manipulating data [20, 3] and even less for command selection.

Multi-Touch Interaction. Multi-touch interaction and free hand interaction share several similarities as users can use both hands and several fingers [28]. Several interaction techniques exploit multi-touch capabilities [2, 19]. For instance, the Multi-Touch Marking menu [19] is a technique that combines a Marking menu and chording gestures for selecting commands in a 2-level menu. The Finger-Count menu [2] is a two-handed and multi-finger technique that only counts the number of finger on the surface. This technique proved efficient on multi-touch surfaces [2] can be considered as a good candidate for free hand interaction, and it does not require distinguishing fingers.

While a variety of menu techniques have been investigated for conventional interfaces and interactive surfaces, and a few free hand techniques have been implemented, we are not aware of any comparative studies comparing free hand menu selection techniques such as Linear menu, Marking menu and Finger-Count menu.

Table 1: Summary of the main properties of Linear, Marking and Finger-Count menus.

	Linear	**Marking**	**Finger-Count**
Preview	Yes	No	Yes
Expert mode & Eyes-free selection	No	Yes	Yes
Fluid transition	No	Yes	Yes
Direct access	No	No	Yes
Gestures	Dynamic	Dynamic	Static
Number of items	About 8x8	About 8x8	5x5=25

3 Menu Techniques

We now present the three menu techniques that we designed for distant displays: the Linear menu, Marking menu and Finger-Count menu. Their main properties are summarized in Table 1.

3.1 Linear Menu

Users activate the menu by opening their hand, the palm in direction of the distant display. Items are organized vertically. The root menu is displayed on the left side while the submenu is displayed on the right side when a category (parent item) is selected. Users control a cursor by moving their hand "in the air". As soon as the cursor is over a category the corresponding submenu appears on the right side. To execute the desired command, the user has to "grab" the corresponding item. Users can perform this metaphorical gesture just by closing their hand when the cursor is over the item. We choose this end gesture delimiter rather than a delay to let the user control the system. Besides, delays are often perceived as too fast by novice users and too slow by expert users [9].

Linear menu properties. One important feature of Linear menus which is often underestimated is that they make it possible to previsualize submenus [1]. Users can quickly scan the content of submenus just by performing a vertical gesture over categories. As the Linear menu is based on pointing, it requires visual feedback that is not compatible with eyes-free selection.

3.2 Marking Menu

Our implementation of a Marking menu [15] is based on the Multi-Stroke menu [29] (Fig. 1, left): The root menu and the submenu are superimposed to avoid that submenus will be displayed outside of the screen [29]. Moreover, users perform two simple strokes rather than a compound stroke to maintain a high level of accuracy [17, 29]. In novice mode, the menu is always displayed in the center of the screen. The cursor is automatically located in the center in the menu when the user opens the hand with the palm in direction of the TV set. So, users select a category in the root menu just by performing a first stroke in direction of the corresponding item and then by "grabbing" it. The

corresponding submenu appears at the same location as the root menu and users execute the same mechanism to select the desired item. In expert mode, the menu does not appear and users only perform two straight strokes.

Marking menu properties. Marking menus have several advantages. First, they reduce the mean distance for selecting items thanks to their circular layout. Second, they make it possible to perform eyes-free selection as they are not based on positioning. Third, they favor the fluid transition from novice to expert mode as users perform the same gestures in both modes. Moreover, gestures are easy to learn thanks to spatial memory [2]. However, users cannot preview the submenus as menus are superimposed [29].

3.3 Finger-Count menu

The Finger-Count menu (Fig. 1, right) is an adaptation of Finger-Count shortcuts [2] from multi-touch surfaces to distant displays. The graphical layout is similar to the Linear menu except that the corresponding number of fingers to extend is displayed close to the item. In novice mode, the user selects a category in the root menu by exhibiting the corresponding number of fingers of the left hand and then selects the desired item in the submenu in the same way but with the right hand. The command is executed when the user closes both hands simultaneously. In expert mode, the user performs the same gestures except that the menu does not appear.

Finger-Count properties. Finger-Count menus also have several advantages. First, these gestures are natural: users interact with the system like basketball referees communicate with administration for signaling the number of the player called for foul: just by exhibiting fingers on each hand. Second, users can scan the different categories just by adding/removing fingers of the left hand. Third, users can perform eyes-free selection as they do not need the visual modality to show a given number of fingers. As for Marking menus, this technique favors the fluid transition from novice to expert usage as users perform the same gestures in both modes. Moreover, users have direct access to commands: experienced users can simultaneously exhibit fingers on both hands if they already know where the desired item is located. Contrary to Linear or Marking menus, users do not need to perform a dynamic gesture, a simple posture is sufficient to be recognized and

interpreted by the system. Finally, it is worth noting that this technique is little affected by counting behaviors: only the number of fingers is taken into account, not how digits are mapped to fingers. Moreover Asian people use different finger-counting methods than Western people except for digits from 1 to 5 (differences appear for 6 or more).

4 Study 1: Hand Posture for Finger-Count

The experiment is based on a "Wizard-of-Oz (WoZ)" prototype. Users were led to believe that they interacted with a fully implemented system, while the display was in fact controlled by a wizard in another room, observing the user through cameras. We used this methodology to observe and understand how users would naturally perform gestures using a distant display.

We did not study gestures for Linear and Marking menus because 1) pointing is quite common, 2) it is already implemented in commercial products and 3) implementing a Wizard-of-Oz protocol for a technique with direct feedback would be quite difficult. For these reasons, we only focused on finger-counting gestures, where the best hand posture is not known. The purpose of the first study was 1) to find which hand posture users would naturally choose and prefer 2) to identify the best position for the camera to recognize this posture and 3) to optimize our computer vision algorithm according to the resulting perspective.

We suspected that the context of use, e.g. interactive TV (users may lean back in a sofa) vs. public display (users stand in front of the device) can strongly impact how users perform gestures. Accordingly, we decided to compare Finger-Count hand postures for these two scenarios. Through an informal pre-study the following postures (Fig. 2) have been identified as promising and were tested during the experiment:

- Interactive TV scenario (sitting):
  - **Palm:** Palm facing the display
  - **Back:** Back of hand facing the display
  - **Fingertips:** Fingertips towards the display.
- Interactive public display scenario (standing):
  - **Palm:** Palm facing the display
  - **Back:** Back of hand facing the display
  - **Down:** Arms hanging loosely, back of hand facing the display.



Fig. 2 Palm (ITV), Back (ITV), Fingertips (ITV), Palm (IPD), Back (IPD), Down (IPD).

4.1 Experimental design

The stimulus consisted in displaying two digits on the display. Once the stimulus appeared, the participant could show the corresponding number of fingers of each hand. Feedback occurred as soon as the wizard recognized a valid posture.

For each scenario (ITV vs. IPD), the 3 different hand postures where assigned to 3 different blocks which were counterbalanced between participants with a Latin square design. We also introduced a first and last block where participants could choose a hand posture freely. The first block was used to observe which hand posture users would choose naturally without instruction. The last block was used to know which hand posture is preferred after the experiment. So in total, each participant performed 5 blocks of 5x5=25 selections (all finger combinations). 10 European participants (age 22-26, mean 25.3) were recruited from a HCI lecture and assigned to the ITV/IPD conditions randomly (5 for each condition). At the end of the experiment, participants filled out NASA TLX questionnaires to evaluate the mental and physical workload (100-point scale) for each technique, stated their preferred posture, and a short semi-structured interview was conducted. The display was a 52" display in landscape format with 1920x1080px resolution. A camera (Microsoft Kinect) was installed directly below the display. In the ITV condition the participant was sitting on a sofa at a distance of 1.6m from the display (recommended viewing distance[1]). In the public display condition, the user was standing at a distance of 1.6m from the display.

[1] http://www.sony.co.uk/hub/bravia-lcd-televisions/4/1

4.2 Results

Intuitively preferred hand posture. In the ITV condition, all five participants started (1st trial of the 1st block) with the palm posture. In the IPD condition, four participants started with the palm posture, while one participant started with the back posture. In ITV condition, for the 1st block all participants always chose the palm posture. In IPD condition, the palm posture was chosen the most often (67 times in total), but followed by the back posture (34 times) and mixed postures (different postures of left and right hand, 24 times).

Preferred hand posture after experience. In the 5th block in the ITV condition, the palm posture was chosen 53 times, followed by fingertips posture (45 times), mixed postures (19 times) and the back posture (8 times). In the IPD condition, the most popular posture was the back posture (54 times), followed by the down posture (38 times), the palm posture (19) and the mixed postures (14).

Workload. In the ITV condition, the mental workload of the back posture was significantly higher than for the palm posture (51 vs. 41, $p<0.05$). Similarly, the physical workload of the back posture was significantly higher than for the fingertips posture (62 vs. 26, $p<0.05$). In IPD condition, palm posture scored worst. The mental workload was significantly higher than both back and down postures (26 vs. 21 vs. 20, $p<0.05$). The physical and temporal workload of palm posture was higher than down posture (45 vs. 18 and 38 vs. 16, $p<0.05$). Frustration of palm posture was higher than for down posture (37 vs. 12, $p<0.05$).

Preference. For sitting, the fingertip posture was preferred, followed by the back and the palm posture. For standing, the back posture was preferred, followed by the down posture and the palm posture.

Observation. In general, we could observe four different strategies of using Finger-Count. Mostly, people took their hands down between trials, and opened the respecting number of fingers while lifting their hands (for palm and back gestures). Sometimes, they first opened the correct number of fingers and only then lifted their hands. Sometimes, they left their hands lifted and closed all fingers between trials, and a few participants sometimes only changed the number of fingers between trials. In most cases, participants showed the fingers of the left hand first. Sometimes, especially when the same number of fingers were shown on both hands, both hands were shown synchronously. Only one user sometimes showed his right hand first. Rarely,

participants had to correct the number of shown fingers, and in very few cases they actively looked at their hands. While for 1, 2, 4 and 5 fingers gestures were relatively consistent, users seemed to be unsure about which fingers to show when showing 3 fingers.

4.3 Discussion

There are two interesting observations from this study. First is, while almost all users initially used the palm posture, this is not the most efficient one. Second is, while for the ITV condition, the palm posture is chosen most often, for IPD condition, the back posture is chosen most often (after training). However, while for ITV the palm posture is chosen most often, it is also the least preferred technique, but requires less mental demand than e.g. the fingertip posture. While both the fingertips and down postures required less physical demand from the participants, they were not performed very often. We believe that this may be because participants prefer a certain expressivity of their gesture towards the display. Even for the down gesture, users did usually not let their arm hang loose as indicated in the instructions, but rather moved their arms slightly forward towards the display. Similarly, for the fingertips gesture, users often expressively lifted their arms from the legs and pointed towards the display. This may be either because users believe that their gestures are recognized better when hands are moved towards the display, or because they feel more comfortable when it is obvious for bystanders that their gestures are directed towards a display.
From the results of this study, we decided our implementation should recognize both palm and back gestures, so users could use both of them. Even if the palm posture is not the preferred nor the most efficient, it is chosen most often by participants both before and after training.

5 Implementation

We employed the recent and low-cost Microsoft Kinect[2] sensor which contains a depth camera providing 11 bit depth images in VGA resolution at a rate of 30 Hz. The server and client communicate via the

[2] http://www.xbox.com/en-US/kinect

widespread multi-touch protocol TUIO[3]. Our recognition software uses the OpenNI framework[4] including the PrimeSense NITE middleware[5] for hand tracking and OpenCV library[6] for computer vision algorithms.

Linear and Marking menu. The hand of the user is tracked in 3D space during the whole session. We use the point tracking capabilities of the NITE middleware, which is initialized by a *focus gesture*, which is a wave gesture in our experiment. The grabbing gesture is recognized by analyzing the contour of the user's hand (Fig. 3). To get the contour of the hand, we use the x,y position of the tracked hand point to segment the hand contour from the depth image by isolating the object that is within a depth range of 10 cm around the tracked point. We then determine the ratio between the area of the hand contour and the area of its convex hull. If the ratio exceeds 80%, we assume the hand is closed, otherwise it is opened. Motion blur artifacts in the depth image may lead to incorrectly recognized grabbing gestures if the hand is moving too quickly. So we used a minimal grabbing time of 500 ms increasing the overall time required for one selection but avoiding false positive detections.

Finger-Count menu. Contrary to Linear and Marking menus, counting fingers does not require hand tracking, hence no focus gesture is required. To count the fingers, we first isolate hands from the depth image by using a fixed threshold of 75 cm (see Fig. 3). This means that users must slightly extend their arms in the direction of the screen in order to interact with the system. In order to count the fingers, the contour of the hand is processed in the following steps (Fig. 3):

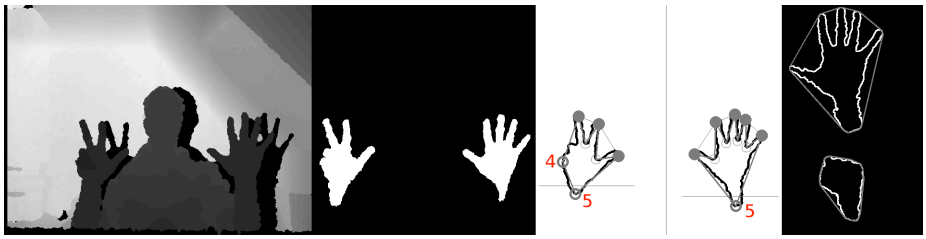


Fig. 3: left: depth map; middle: isolated hands; right: fingertip detection: hand contour (black), simplified hand contour (gray), fingertips (gray circles), vertices marked with 4 or 5 were removed from the list of fingertips in the corresponding processing step

[3] http://www.tuio.org/
[4] http://www.openni.org/
[5] http://www.primesense.com/?p=515
[6] http://opencv.willowgarage.com/wiki/

1. Approximation of the hand contour using the Douglas–Peucker algorithm
2. Determine the convex hull of the simplified contour
3. Consider all vertices of contour of step 1 that are also contained in the convex hull to be fingertips
4. Remove all vertices that hold large interior angles in the contour from step 1 from the list of fingertips (threshold of 57.3°)
5. Remove all vertices that are in the lower 10% of the hand contour from the list of fingertips.

The limited resolution of the depth camera and the amount of noise make the detection of fingertips a little bit unstable. Since false positives and false negatives on fingertip detection may occur randomly, we decided to apply a histogram based smoothing technique. For each frame (every 33ms), a new value of counted fingers is written into a 60 elements circular buffer. Then we use the buffer to identify stable states by finding the most frequent values over the last period of time. This filter adds a delay of 1 to 2 seconds before a new count can be recognized depending on the amount of image noise. Future improved depth sensors would enable faster recognition.

6 Study: Menu Techniques Comparison

The goal of this experiment was to compare the efficiency of three menu techniques for distant displays: the classical Linear menu, the Marking menu and the Finger-Count menu. Each menu was tested for novice and expert behavior. For this study we decided to focus on the interactive TV (i.e. sitting) condition.

6.1 Experimental design

We used a 46" full-HD LCD display in landscape orientation. The distance between the user and the display was 1.6m. For the Finger-Count condition, the distance between the Kinect and the user was 1m. In the other conditions, the distance was 1.5m.

Task, Stimulus and User's Behavior. The task consisted in selecting an item in a 5x5 menu hierarchy. We used the same labels as in [2], e.g., "Shape" for category and "Line" for item. The feedback was a

green/red square for correct/incorrect selection. We found interesting to evaluate menu efficiency for two types of user's behavior:

- *Novice behavior*. Users with novice behavior do not know the exact location of items: To simulate this behavior, the stimulus consists in showing the name of the target: users must navigate in the hierarchy to find and select the target.
- *Expert behavior*. Users know where the desired command is located and how to select it. To simulate this behavior, all necessary information is displayed: the target category and the target item are highlighted in blue to indicate the path. Moreover, the gestures for Marking menus and Finger-Count are displayed: 2 strokes for Marking menus and 2 digits for Finger-Count.

*Procedure and design***.** 12 European participants (aged 24-36, mean 28) were recruited from a subject pool with people of various professions and computer experience. They were briefed (~5 min) with written explanations and videos of the techniques. We also allowed them practicing in order to be sure they understood how to use the techniques. For each behavior, participants had to perform 2 blocks of 25 selections. Novice and expert behavior were evaluated in this order, familiarity increasing with blocks. After the experiment, participants filled out NASA TLX and SUS (System Usability Scale) questionnaires for each menu technique, stated their preferred technique, and a semi-structured interview was conducted.

We used three sets of contents to avoid learning effects between techniques. The order of techniques and sets were counterbalanced across participants with a Latin Square design. For each block, the order of appearance of items was randomized. The independent variables of the study were menus and behavior. Dependent variables were speed, accuracy, workload (NASA TLX) and usability (SUS). To sum up, the experiment involved: 14 participants x 3 menus x 2 behaviors x 2 blocks x 25 items = 4200 selections.

6.2 Results

*Accuracy***.** The accuracy was 94.2% for Linear menu, 95.3% for Marking and 93.4% for Finger-Count (Fig. 4). There were no significant differences in accuracy between techniques or user's behavior.

*Speed***.** Completion time was measured as the time from when the stimulus appeared to the time when the item was activated [1, 2, 29] (in

our case as soon as the system recognized a "grabbing gesture"). There was an interaction effect for menu technique and novice/expert mode (ANOVA, $F_{2,22}$=28 $p<0.01$) shown in Fig. 5. A post-hoc Tukey test revealed that in novice mode, Linear (6.6s) is significantly faster than Marking (7.2s), which is significantly faster than Finger-Count (8.5s). In expert mode, Linear (5.4s) is still significantly faster than Marking (5.8s) ($p<0.05$). The mean selection time for Finger-Count is 5.7s.

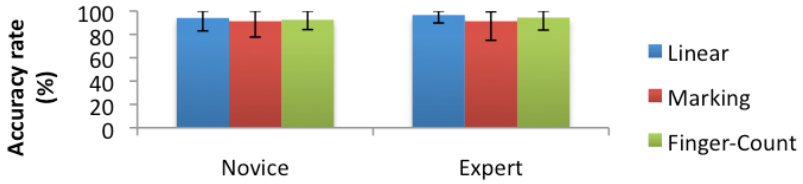


Fig. 4. Accuracy by technique and user's behavior (95% confidence interval indicated).

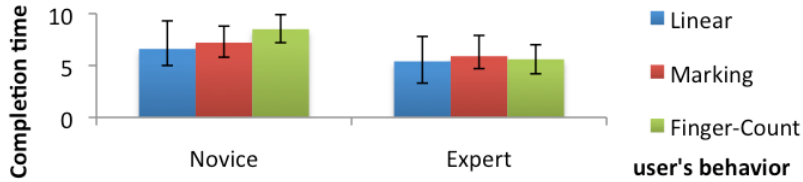


Fig. 5. Completion time by technique and user's behavior (95% confidence interval indicated).

Questionaires. A GLM repeated measures test on NASA TLX data revealed no significant effects on all questions except for mental demand ($F2,22 = 4,9$, $p< 0.03$). A post-hoc t-test revealed that Linear menu (24.4) required significantly less mental demand than Finger-Count menu (44.6). The mental demand of Marking menus was 28.1.

A GML repeated measures test revealed no significant effect on usability (SUS) for techniques. Participants stated they can learn techniques very quickly (Linear: 4.6/5; Marking: 4.3/5; Finger-Count:4,2/5) and found all techniques easy to use (Linear: 4.3/5; Marking: 3.8/5; Finger-Count:3.7/5). Finally, a Friedman test (ChiSquare=4.77, df=2, $p>0.05$) revealed no significant effect on techniques ranking. 7 Participants chose the Linear menu as their favorite technique, 3 the Finger-Count and 2 the Marking menu.

Observations. We observed that there were great differences in the ability of users to express Finger-Count gestures: while some users could easily express all gestures, others had surprising difficulties for moving their fingers. Furthermore, our recognizer required fingers to be

clearly separated while it can be bio-mechanically difficult to strongly separate middle and ring fingers. We also observed that some users seemed to be unsure about which fingers to show when showing 3 fingers. Finally, we observed that most of the users had difficulties to move their hand in a 2D-plane when interacting with Linear and Marking menu: they generally moved their hand in a hemi-spherical area and sometimes accidently left the threshold area of the recognizer.

7 Discussion

Accuracy. Results show that free hand gestures can be accurately performed during our experiment. Indeed, the three interaction techniques provide an accuracy rate superior to 93%. While the Kinect is already used for hand tracking in Xbox games, our experiment reveals that it can also be used to count fingers with our algorithm.

Speed. All techniques are much slower than their interactive surface pendants. Selection time for Linear menus was 5.3s compared to 2.0s in [2], Finger-Count 5.7s (1.8s in [2]), and Marking menu 5.8s (2.4s in [2]). There were small differences between techniques for expert users with a small advantage for Linear menus. This result can seem surprising as Marking and Finger-Count have been proved faster than Linear menus [2, 7]. A more detailed analysis of the implementation of Linear menu can partially explain this result. Contrary to common Linear menu implementations, all submenus share the same location to reduce the real-estate used to display the menu. So, for the third category, the submenu is placed at mid-height of the parent item. This improvement, which was proposed in [26], decreases the average distance for reaching items and thus decreases the mean selection time according to the Fitts' law [11]. A deeper analysis also reveals that the Marking and Finger-Count performances are underestimated. Indeed, the Marking menu uses two grabbing gestures requiring 0.5s while the Linear menu needs only one. Similarly, Finger-Count uses a filter which causes a delay of about 1s to compensate for the noise and the low resolution of the camera. Better camera hardware or algorithms can improve performance of these two techniques.

Finger-Count. Several participants mentioned that Finger-Count requires more mental demand than other techniques through the NASA TLX questionnaire or the open discussion. However, Finger-Count gestures did not require a high mental demand in the first study. One

reason may be that the task in the first study was easier as it did not imply menu selection. Moreover, in the first study, the recognizer was "perfect" and users never needed to make adjustments. Finally, free hand Finger-Counting seems to require more mental demand than the original version on multi-touch surfaces. In this latter case, users only need to touch the surface with the correct number of fingers and it does not matter which fingers are used. While this is also true for free hand counting, certain fingers must be extended and others folded. This makes the movement more difficult than just slightly moving the fingers to touch a surface. Due to cognitive and bio-mechanical constraints, certain finger postures may be difficult to perform.

Distant displays. In the context of ITV, free hand gestures cannot replace the physical remote control as tapping on buttons will remain easier and faster. However, several users mentioned that they would like to use free hand gestures as a complementary tool especially for selecting "favorite actions" or "switch lamps", for example. They also mentioned that it is beneficial if they do not need to look for the TV remote control nor to move to reach it on the coffee table. One participant also mentioned "only the guy with the physical remote control can interact with the TV; with free hand interaction everyone has the *power*". As no technique is significantly preferred, we recommend to let the choice to the user to configure the techniques that s/he wants to use.

In the context of public displays, people can walk, making pointing or directional gestures difficult to recognize by the system. The Finger-Count technique seems promising in this context as it is only based on posture and thus compatible with passing-by interaction. However, the high mental demand mentioned by participants is not compatible with immediate usability. So, we recommend to use Linear menus on public displays for novice users but to let the possibility for expert users to perform finger-count gestures as these two techniques are compatible.

4 Conclusion

In this paper, we presented and evaluated three menu techniques for interacting with distant displays using depth cameras: Linear, Marking, and Finger-Count menus. In a first study, we compared different hand postures for Finger-Count. While for an interactive TV scenario, palm posture seems suitable and intuitive, for a public display scenario back

posture seems much better suited. In a second study, we compared Linear, Marking and Finger-Count menus. While all three techniques achieve satisfactory accuracy, they seem to be much slower than their multi-touch counterparts or remote controls. There are relatively little differences between the techniques: Linear menus are slightly faster than Marking menus which are faster than Finger-Count in novice mode, while Finger-Count lies between Linear and Marking menus in expert mode. We believe that free hand menu techniques can be a valuable complement to touch and remote controls, and it may be best to leave users a choice for their individually preferred technique.
For future work, it would be important to decrease the delay introduced by our filtering techniques. This could mainly be achieved by using a depth camera with higher resolution and less noise, and improved recognition algorithms. Further, we have evaluated menu selection in an interactive TV scenario, such that evaluation in an (outdoor) public display scenario would be a next step. Finally, we would like to deeper investigate the impact of age (elderly people or children) on the acceptance of these techniques.

Acknowledgements

This work was supported by the Alexander von Humboldt Foundation and the QUAERO project. We thank H. Maktoufi, F. Alt and A. Roudaut.

References

1. Bailly, G., Lecolinet, E., and Nigay, L. Wave menus: improving the novice mode of hierarchical marking menus. In Proc. INTERACT'07. 475-488. Springer, Berlin (2007)
2. Bailly, G., Lecolinet, E. and Guiard, Y. Finger-count & radial-stroke shortcuts: 2 techniques for augmenting linear menus on multi-touch surfaces. In Proc. CHI '10. 591-594. ACM, NY (2010)
3. Benko H. and Wilson A. D. Pinch-the-sky dome: freehand multi-point interactions with immersive omni-directional data. In Proc. CHI EA '10. 3045-3050. ACM, NY (2010) .
4. Baudel, T. and Beaudouin-Lafon, M. Charade: remote control of objects using free-hand gestures. Commun. ACM 36, 7 (Jul. 1993), 28-35. ACM, New Yok (1993)
5. Beyer, G., Alt, F., Müller, J., Schmidt, A., Haulsen, I., Klose, S., Isakovic, K., and Schiewe, M.: Audience Behavior around Large Interactive Cylindrical Screens. In Proc. CHI '11. ACM, NY, (2011)
6. Bolt B. R. , “Put-that-there”: Voice and gesture at the graphics interface, ACM SIGGRAPH Computer Graphics, v.14 n.3, 262-270. ACM NY (1980)
7. Callahan, J., Hopkins, D. Weiser, M. and Shneiderman, B. An empirical comparison of pie vs. linear menus. In Proc. CHI '88. 95-100. ACM, NY (1988)

8. Cesar, P., and Chorianopoulos, K. The evolution of TV systems, content and users toward interactivity. NOW Foundation and Trends in HCI, vol.2: n°4. 373-95 (2009)
9. Cockburn, A., Gutwin, C., and Greenberg, S. 2007. A predictive model of menu performance. In Proc.of CHI '07. 627-636. ACM, NY (2007)
10. Dachselt, R., Hübner, A. Virtual Environments: Three-dimensional menus: A survey and taxonomy. IEEE Computers & Graphics. 31, 1, 53-65. (2007)
11. Fitts, P.M. The Information Capacity of The Human Mote; System in Controlling The Amplitude of Movement. Jeurnal of Experimental Psychology, 47, pp. 381-391(1954)
12. Francone, j., Bailly, G., Lecolinet, E. Mandran, N. and Nigay, L. Wavelet menus on handheld devices: stacking metaphor for novice mode and eyes-free selection for expert mode. In Proc. AVI'10. 173-180. ACM, NY (2010)
13. Freeman, W. T., and Weissman, C. 1995. Television control by hand gestures. Int. Workshop on Automatic Face- and Gesture- Recognition. IEEE, p. 179-183 (1995)
14. Jones, E., Alexander J., Andreas Andreou, A., Irani, P. and Subramanian, S. 2010. GesText: accelerometer-based gestural text-entry systems. In Proc. of CHI '10. 2173-82. ACM, NY (2010)
15. Kurtenbach, G. and Buxton, W. The limits of expert performance using hierarchic marking menus. In Proc. of INTERACT '93 and CHI '93. 482-487. ACM, NY (1993)
16. Kurtenbach, G. and Buxton, W. User learning and performance with marking menus. In Proc. ACM CHI '94, 258-264. (1994)
17. Lenman, S., Bretzner, L. and Thuresson, B. Using marking menus to develop command sets for computer vision based hand gesture interfaces. In Proc. Of NordiCHI '02. 239-242. ACM, NY (2002)
18. Lenman, S., Bretzner, L. & Thuresson, B. Computer Vision Based Recognition of Hand Gestures for Human-Computer Interaction. Technical report TRITANA-D0209 (2002)
19. Lepinski, g. J., Grossman T. and Fitzmaurice, G. The design and evaluation of multitouch marking menus. In Proc. Of CHI '10. 2233-2242. ACM, NY (2010)
20. Malik, S., Ranjan, A. and Balakrishnan. R. Interacting with large displays from a distance with vision-tracked multi-finger gestural input. In Proc. UIST '05. 43-52. ACM, NY (2005)
21. Michelis, D., and Müller, J. The Audience Funnel: Observations of Gesture Based Interaction with multiple large displays in a city center. Int. Journal of HCI, to appear.
22. Pavlovic, V., Sharma, R. and Huang, T. S. Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review. IEEE Trans. Pattern Anal. Mach. Intell. 19, 7 (July 1997), 677-695. IEE (1997)
23. Peltonen, P., Kurvinen, E. , Salovaara, A., Jacucci, G., Ilmonen, T., Evans, J., Oulasvirta, A., and Saarikko, P. It's Mine, Don't Touch!: interactions at a large multi-touch display in a city centre. In Proc. CHI '08. 1285-1294. ACM, NY (2008).
24. Roudaut A., Bailly G., Lecolinet E. and Nigay L. Leaf Menus: Linear Menus with Stroke Shortcuts for Small Handheld Devices. Conference INTERACT'09 (2009)
25. Schick, A., van de Camp, F., Ijsselmuiden, F. and Stiefelhagen R. Extending touch: towards interaction with large-scale surfaces. In Proc. of ITS '09. 117-124. ACM, NY (2009).
26. Tanvir, E., Cullen, J., Irani, P., and Cockburn, A. AAMU: adaptive activation area menus for improving selection in cascading pull-down menus. In Proc. of CHI '08. 1381-1384. ACM, NY (2008) .
27. Vogel D. and Balakrishnan. R. Interactive public ambient displays: transitioning from implicit to explicit, public to personal, interaction with multiple users. In Proc. of UIST '04. 137-146. ACM, NY (2004)
28. Wu, M. and Balakrishnan, R. Multi-finger and whole hand gestural interaction techniques for multi-user tabletop displays. In Proc. of UIST '03. 193-202. ACM, NY (2003)
29. Zhao, S. and Balakrishnan, R. Simple vs. compound mark hierarchical marking menus. In proc. UIST'04. 33-42. ACM, NY (2004)