Hindawi
Advances in Human-Computer Interaction
Volume 2022, Article ID 9362916, 13 pages
https://doi.org/10.1155/2022/9362916

Research Article

Midair Gestural Techniques for Translation Tasks in Large-Display Interaction

Vera Remizova, Yulia Gizatdinova, and Veikko Surakka

Faculty of Information Technology and Communication Sciences, Computing Sciences Unit, TAUCHI, Tampere University, Kalevantie 4, 33100 Tampere, Finland

Correspondence should be addressed to Vera Remizova; vera.remizova@tuni.fi

Received 26 August 2021; Revised 20 May 2022; Accepted 1 June 2022; Published 9 July 2022

Academic Editor: Marco Porta

Midair gestural interaction has gained a lot of attention over the past decades, with numerous attempts to apply midair gestural interfaces with large displays (and TVs), interactive walls, and smart meeting rooms. These attempts, reviewed in numerous studies, utilized differing gestural techniques for the same action making them inherently incomparable, which further makes it difficult to summarize recommendations for the development of midair gestural interaction applications. Therefore, the aim was to take a closer look at one common action, translation, that is defined as dragging (or moving) an entity to a predefined target position while retaining the entity's size and rotation. We compared performance and subjective experiences (*participants* = 30) of four midair gestural techniques (i.e., by fist, palm, pinch, and sideways) in the repetitive translation of 2D objects to short and long distances with a large display. The results showed statistically significant differences in movement time and error rate favoring translation by palm over pinch and sideways at both distances. Further, fist and sideways gestural techniques showed good performances, especially at short and long distances correspondingly. We summarize the implications of the results for the design of midair gestural interfaces, which would be useful for interaction designers and gesture recognition researchers.

1. Introduction

Midair gestural interaction applications have been gaining popularity for large public displays [1, 2] and augmented reality [3] and in smart spaces [4]. One of the most frequently used actions in midair interaction with large screens is the one that *translates* (drags or moves) objects (or cursor) to an indicated target position while retaining their object's size and rotation [5, 6]. In conventional mouse interaction, this corresponds to, for example, scrolling or browsing actions. Based on earlier findings, *translation* could be a promising alternative for a traditional pointing interaction [7]. A variety of gestural techniques have been envisioned for translation tasks in large-display interaction. Most of the research work so far in this area was elicitation studies and focused on the evaluation of user preferences. Rarely has the actual recognition technology been used to automatically detect the dynamic hand gestures and collect quantitative characteristics of the interaction. Even rarer were alternative gestural techniques directly compared to see whether some gestural techniques are more appropriate than others for translation manipulation. However, both quantitative and qualitative measurements are needed to inform interface developers how different gestures would impact the performance and user experience of the interface [5, 8].

The contribution of the present work is twofold. Firstly, we looked at which gestural techniques have been used for translation tasks. We reviewed what kinds of user studies have been conducted and what metrics have been used in the studies. We referred to user studies across different application domains that concentrated on the design of gesture sets for translation tasks or associated interface schemes. We chose to concentrate only on the most common interaction techniques favored by the researchers. To date, comparisons of such techniques remain rare. Of particular interest in designing gestural techniques is in support of the execution of translations accurately and conveniently in large-display settings. We also focused on relatively simple and physically easy techniques, those that had achieved the best performance, and those that were favored by the users. As a result, we unified the names of selected interaction techniques. The set of translation gesture candidates that emerged from previous works included interaction by *palm*, *fist*, *pinch*, and *sideways* (as described in section *Design of the Selected Techniques*). Secondly, the four translation techniques were evaluated in a controlled user study, in which participants performed the repetitive translation of 2D objects to short and long distances. The aim was to quantify the effects of the translation distance on user performance and subjective experiences and systematically investigate differences between the techniques.

2. Midair Translation

The goal of the literature analysis was to review the translation gestures empirically evaluated in user studies. Then, based on the results of such evaluations, select candidates for controlled and systematic evaluation in our user study. We considered gestures suitable for translation when they satisfied the translation definition presented above. The following inclusion criteria were met. We considered translation only on the horizontal plane because there is a tendency for users' bodies to physically move in a horizontal rather than a vertical direction [9]. Further, only one-handed gestures were considered that have been shown to overcome a bimanual interaction with large displays in terms of the naturalism for real-life public situations [10]. The examination included general surveys of touchless gestural interactions from [5, 8, 11, 12]. We also referred to user studies that concentrated on the design of gesture sets to help users navigate using translation or associated interface schemes across different application domains. Patents and technical papers without user studies were excluded. In total, we evaluated 64 publications.

2.1. Interaction Techniques Suitable for Translation Task. From the review, we identified 11 interaction techniques in 19 publications. These differed by nature of actions and hand posture (see Table 1). We unified the names of gesture techniques into five categories by the primary gesture posture or action used.

As Table 1 shows, nine reviewed studies investigated the *swipe* gesture as the most often studied technique. *Swipe* generally follows a fluid manipulative gesture mechanic [15, 18] when a user makes a quick move to the left or right by the hand or finger. It was used with translation commands like next/previous for remote TV control [13, 14, 17–19, 21, 29], scroll tasks [15, 16], and drag and drop tasks [20]. Among these, Carreira et al. [15] reported that the *swipe* gesture provided better precision and control than *fist* gestures.

The second most popular interaction technique was translation by *palm* applied to tasks as scrolling [13, 24], browsing [23], and dragging items [22]. There is little specific research on the evaluation of the interaction performance of *palm* gestures for translation tasks. However, this technique was well investigated and applied often to pointing and selection tasks [30].

The third most popular technique for translation was the *fist* interaction for scrolling and browsing 2D and 3D objects [2, 25, 26]. The *fist* interaction was found as "faster, more fun, and intuitive" than the *palm* interaction [2].

The fourth most popular interaction technique was *pinch*, used for both scrolling [3, 7] and drag and drop [27] tasks, which are key interactions for more advanced applications. Farhadi-Niaki et al. [27] showed that the *pinch* gesture demonstrated better performance than hand *circling* for a drag and drop task in a WIMP interface.

The fifth technique mentioned in the literature for translation tasks was *sideways*, which was applied for image scrolling [28]. Koutsabasis and Domouzis [28] found out that the *sideways* gestures were more usable than both *swipe* or *circling*.

The literature analysis showed that translation by *palm*, *fist*, and *pinch* could be implemented as a clutch and release sequence. That is a relative pointing with a clutching gesture at the beginning and a release gesture at the end as demonstrated in [7]. The clutching gesture enables the engagement of an entity and the start of manipulation. The entity then follows hand movement while the clutching gesture continues; this could be *palm* [22–24], *fist* [2, 25], or *pinch* [7]. The release gesture allows one to exit from the clutching state and disengages the system from the interaction [7, 25].

Sideways gesturing belongs to autoscrolling mechanics [31] in which translation is carried out while the hand is held in a specific position. It describes all types of techniques that utilize some anchor point relative to the screen or user's body. For example, when the user's head serves as the anchor point, keeping the hand on the left (or right) from the head would translate the object to the left (or right, respectively). These include autoscrolling gestures when the user places the arm in relation to the anchor point, and the system starts the automatic translation of the object in a given direction as long as the user keeps the arm or hand in a dedicated place.

As the literature review revealed, there are generally five different gestural techniques used for translation so far: *swipe*, *fist*, *palm*, *pinch*, and *sideways*. The results of the studies are mixed and somewhat contradictory. In various studies, different gestures have been favored by the participants. For instance, the *swipe* that was preferred by participants in the elicitation studies for image browsing was not validated by the usability test [28]. While these results should be taken with caution (due to different gesture recognition technology and conditions used in the studies), they indicate at least a need to compare the performance of the interaction techniques against each other.

2.2. Design of the Selected Techniques. The goal of the literature review was to identify potential gestural techniques for accurate and fatigue-free continuous translation in the different distances. Most studies in which the *swipe* gesture was favored by the participants were elicitation studies (i.e., did not use the actual recognition devices and algorithms).

Table 1: The comparison of user studies designing interaction techniques for translation tasks.

Gestural interaction technique	Referenced literature	Application domain *(recognition technology or study approach)*	User interface task	Comparative study *(comparable gestures)*	Performance metrics
Swipe	[13]	TV control *(elicitation study)*	Next/previous	No	No
	[14]*	TV control *(Microsoft Kinect)*	Next/previous	No	Yes
	[15]*	General *(Microsoft Kinect)*	Scroll	Yes *(swipe versus fist)*	Yes
	[16]*	General *(elicitation study)*	Scroll	No	No
	[17]*	TV control *(elicitation study and Microsoft Kinect)*	Next/previous	No	No
	[18]	Video communication *(Microsoft Kinect)*	Next/previous	No	No
	[19]	TV control *(elicitation study)*	Next/previous	No	No
	[20]*	Learning application *(Microsoft Kinect)*	Drag and drop	No	No
	[21]*	TV control *(elicitation study)*	Next/previous	No	No
Palm	[22]*	Tiled display *(Microsoft Kinect)*	Drag items	No	No
	[23]	Small display *(Microsoft Kinect)*	Browsing	No	No
	[13]	Smart TV *(elicitation study)*	Scroll	No	No
	[24]*	General study *(Microsoft Kinect)*	Cursor control	Yes *(dwelling versus moving)*	No
Fist	[25]*	General *(Intel RealSense)*	Translation	No	No
	[26]	General *(elicitation study)*	Move	No	No
	[2]*	Public information display *(Microsoft Kinect)*	Browsing	Yes *(palm versus fist)*	No
Pinch	[7]*	General *(Leap Motion)*	Scroll	No	Yes
	[27]	General *(Microsoft Kinect)*	Drag and drop	Yes *(pinch, circling versus mouse drag)*	Yes
	[3]	AR menu environment *(Microsoft's HoloLens)*	Scroll	No	No
Sideways	[28]	Image browsing *(Microsoft Kinect sensor)*	Horizontal scrolling	Yes *(sideways, wheel versus swipe)*	Yes

Publications where a detailed gesture interaction description was provided are marked with*.

These studies did not shed light on the actual performance of the *swipe* in real-world interaction scenarios, nor did they test the performance of the users in continuous swiping to long distances. For translation to long distances, the user would need to swipe multiple times (10 or even more), which would lead to severe hand fatigue in continuous interaction as noted by [32]. Our piloting tests (*participants* = 5) confirmed that the swiping gesture in long-distance interaction was challenging for users due to physical tiredness. We also observed that because the users tried to "optimize" the repetitive swiping, they tended to decrease the amplitude and speed of the swipe that negatively affected the recognition performance of the gesture. We could not include swipe gestures (which are mainly preferred for selecting next/previous items) and were left with the four other interaction techniques: *fist*, *palm*, *pinch*, and *sideways*. A complete gesture set design is shown in Figure 1 for the two different gesture mechanics: clutch and release (*fist*, *palm*, and *pinch*) and autoscrolling (*sideways*).

The clutch and release mechanics of gesturing by *fist*, *palm*, or *pinch* included three key states: (1) clutching to grab an object, (2) translation manipulation, and (3) object release. The clutching state implied that a user first shows an open palm to the sensor to engage the system and then makes a clutching gesture to grab an object. A user needs to clench their fist (*fist* technique), remain with an open palm (*palm* technique), or make a pinching gesture (*pinch* technique). In the translation state, a hand with a clutching gesture is transferred in horizontal space towards the targeted place (interaction by *fist*, *palm*, and *pinch*). In the release state, a designated gesture stops translation and disengages the system from the interaction. The choice of a release gesture for *fist* and *pinch* techniques was based on the literature review. Thus, the release gesture for *fist* and *pinch* techniques was an open palm as a natural movement [7, 15] and a fist for the *palm* technique, as in [33].

For the *sideways* techniques, the translation starts with the hand moving towards the right or left edge of the screen and continues by keeping the hand stationary on the edge of the screen. Removing the hand from the screen edge stops the translation.

3. Methods

3.1. Prototype. To keep the task generic, a visual display for translation was a horizontal slider with 20 bars (see Figure 2(a)). Two distance conditions were implemented for translation tasks: short (varied from 1 to 10 bars) and long (varied from 11 to 20 bars). A depth image from the gesture recognition application was shown in the right-bottom

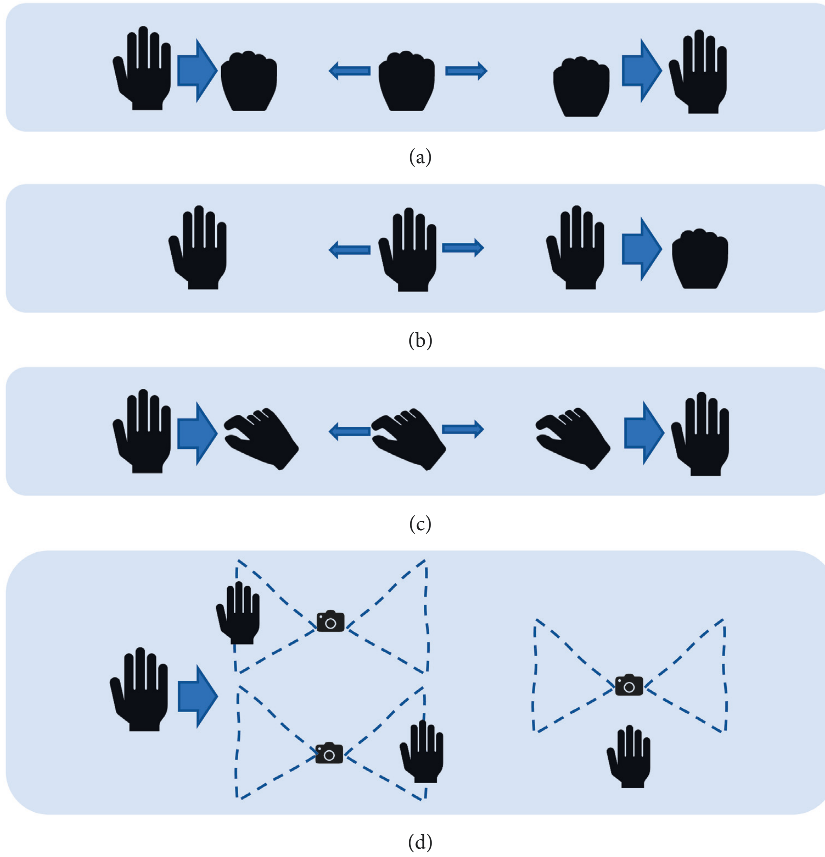


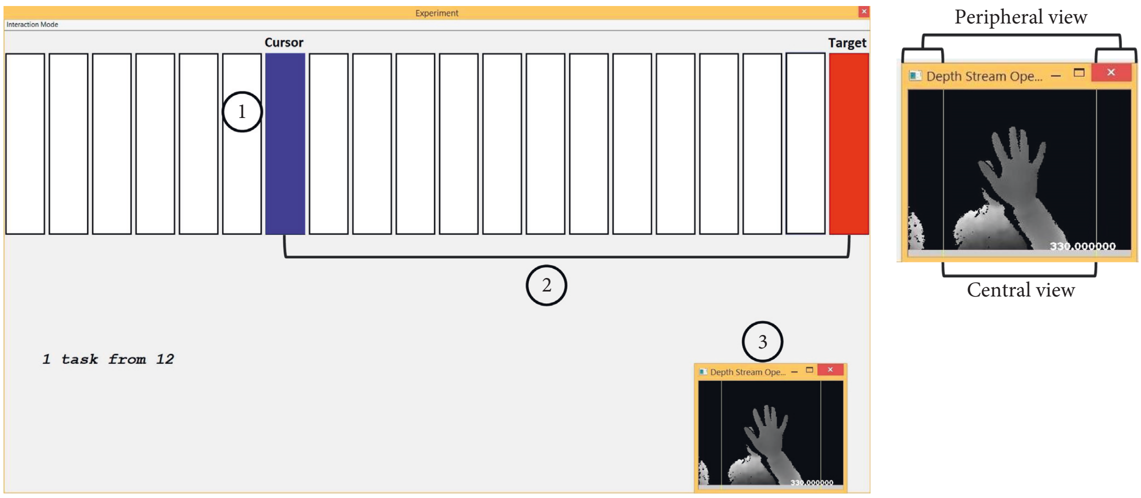
FIGURE 1: Interaction techniques design: (a) fist, (b) palm, (c) pinch, and (d) sideways.

corner of the screen, as Figure 2(b) illustrates. To guide the correct performance of the gestures, specific visual and audio feedback for translation states were defined. The corresponding locations of the item to be translated and its designated destination place were marked on the screen by the words "Cursor" and "Target" correspondingly. In each task, the item to be translated was initially colored gray and turned blue after the participant engaged the system with a clutching gesture (see Figure 2(a)). After the item was translated to its designated place and the participant performed the release gesture, a white screen appeared with a notification that the task was completed, and a "beep" signal was played.

3.2. Apparatus. The experiment was conducted in a large space without direct sunlight. The projection screen size was 2.8×1.6 meters with a resolution of 1920×1800 pixels. The camera (Intel® RealSense™ F200 with a color resolution of 640×480 pixels and a depth resolution of 640×480 pixels) was installed on a tripod and fixed depending on the height of the participant (see Figure 3). The distance between the large screen and the participant was 4 meters. The distance between the sensor device and the participant's hand was 50–70 cm. The computer used for the experiment was a Dell laptop with a 64-bit Windows 8 Professional operating system, Intel Core i7-4610M CPU, 3 GHz, and 8 GB RAM. The keyboard used was a Dell SK-8125. The projector used was a VPL-CH370 5,000 lumens WUXGA 3LCD Basic Installation.

3.3. Software Implementation. The experimental software prototype was created in VS 2013 with the help of Intel® RealSense™ SDK 7.0.23.8048 using Intel® RealSense™ camera F200 as a sensor device. A relative pointing was implemented for translation that has no direct correspondence between physical (location of the hand) and virtual (a graphical representation of an object on the screen) systems of coordinates. Translation movement can start at any point of the physical space as long as the beginning and the end of the translation gesture can be reliably defined [34]. This eliminates the need to bring the hand to an exact point in space to "reach" a certain object for interaction by *fist*, *palm*, and *pinch*. 2D coordinates of the hand's center of mass were used to determine the hand position.

To make a translation, the difference between previous and current hand positions was calculated for *fist*, *palm*, and *pinch* techniques. The translation of one step (the cursor moves one bar to the left or right on the slider view) was performed when the difference between hand positions in the horizontal plane was more than 15 pixels. The translation by *sideways* was executed when the hand was on the border of the camera view during a fixed number of frames. The item was translated by one step if the hand was on the border of the camera view for more than 10 frames.



(a) (b)

Figure 2: (a) Screenshot of the experiment software prototype: (1) engaged bar changed its color from gray to blue; (2) translation distance; and (3) depth image. (b) Visualization of the depth view with its three parts: central view in the middle and peripheral views on the sides.

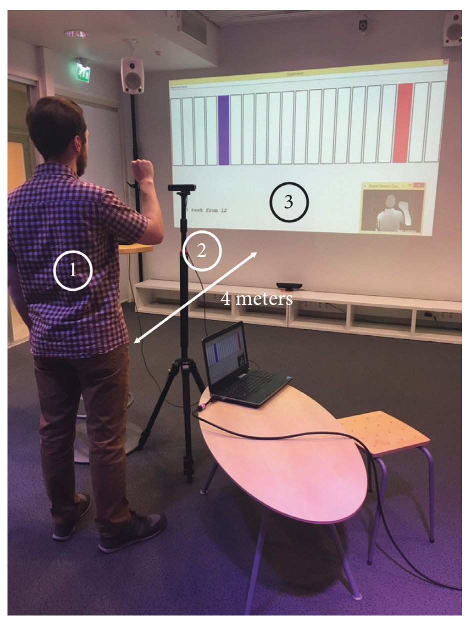


Figure 3: Study setup. (1) User standing in front of the gesture sensor; (2) gesture sensor; (3) large screen.

The gestures currently recognized as a part of the Intel® RealSense™ SDK were used as state elements of the interaction techniques. The predefined gesture types and system alerts of the PXCHandData interface were used to fire and handle hand postures and motions in real-life interaction: FIST for fist, SPREADFINGERS for palm, FULL_PINCH for pinch, and for sideways alerts ALERT_HAND_OUT_OF_LEFT_BORDER and ALERT_HAND_OUT_OF_RIGHT_BORDER. Recognition of gestures for translation by fist, palm, and pinch was on the central camera view, and the sideways translation was on the peripheral view (see Figure 2(b)).

We conducted a pilot study to confirm candidate techniques for translation tasks and review their design. We run the pilot study with five participants. The study was performed with the software prototype and equipment presented above. Firstly, we asked the participants to translate a cursor and select a specified item using five interaction techniques: *swipe*, *fist*, *palm*, *pinch*, and *sideways*. As a result, the translation by *swipe* had a significantly low interaction speed compared to other gestural techniques. Users reported that the *swipe* was challenging, and the hand got tired after a few gesture repetitions. These results were in line with the literature review and confirmed the rejection *swipe* from our study. Secondly, we determined the optimal difference between hand positions for *fist*, *palm*, and *pinch* techniques and the number of frame intervals for *sideways*. For each interaction technique, we asked the user to repeat the translation three times with different parameter values and then evaluated and analyzed the users' movement time and subjective opinions. Thus, we defined the optimal difference between hand positions, such as allowing to go through 20 bars in one clutch inside of the camera view. And the number of frame intervals for sideways interaction was chosen as enough to remove the hand from the border in time.

3.4. Participants. Thirty able-bodied university staff members and students (13 females and 17 males with mean age $M = 27$, $SD = 5.98$, range 19–45) participated in the experiment. The participation was voluntary, and there was no monetary compensation. All participants were novices at touchless gestural interaction. Among them, 13 had no previous experiences with gestural control devices, and 17 had only a few experiences of interacting with body movements using Microsoft Kinect, Nintendo Wii, or HTC Vive for entertainment purposes. Twenty-eight participants were right-handed, and all had a normal or corrected-to-normal vision.

3.5. Procedure. Upon arrival at the laboratory, participants were informed about the study's aims, equipment, and the test room. They were also asked to fill out a consent form and a background questionnaire. They were further introduced to the graphical interface (the slider), the depth view, and the task. The camera placement was adjusted depending on their height. It was also explained that the technology had certain limitations and that the system's best performance could be achieved while keeping the hand at a distance of 50–70 cm from the camera. A general recommendation was that slower movements allowed a larger number of bars to be passed with one translation gesture. It was also noted that the hand should be facing the sensor device, but the participant could freely move in the camera field view. Participants were instructed to use their dominant hand throughout the experiment.

After this, the experiment started with the first block, where the first translation technique was tested. The use of the technique was explained to the participant, including both clutching and release gestures. First, the experimenter showed examples of how to use the technique, after which the participant had a short practice trial in translating items using the technique. The participant also had a chance to ask questions.

As soon as the participant learned to use the technique, the first experimental trial started. The task was to acquire the target as quickly and accurately as possible. After the translation was performed, the participant pressed the SPACE bar on the physical keyboard to start another task. If the translation task was not successful after one minute, the system automatically moved to the next task. The four interaction techniques were presented in a counterbalanced order. Movement direction (left and right) was varied to ensure that the tasks covered a range of conditions for translation in large-display interactions. Thus, for each movement direction on each distance condition, three targets were presented on the slider. The cursor and target presentations were randomized. The participant could take a rest if needed between the experimental blocks.

After each block, the participants rated their experiences with each technique. The ratings were given with eleven nine-point bipolar adjective scales based and supplemented on NASA Task Load Index [35]. The scales varied from negative (−4) to positive (4) experience with the center point representing a neutral point, for example, neither unpleasant nor pleasant. The eleven bipolar adjective pairs used were *general evaluation* (poor/good), *pleasantness* (unpleasant/pleasant), *quickness* (slow/quick), *accuracy* (inaccurate/accurate), *efficiency* (inefficient/efficient), *physical demand* (difficult/easy), *mental demand* (difficult/easy), *temporal demand* (high/low), *frustration* (high/low), *distractibility* (difficult/easy), and *usability* (unusable/usable). The scales were explained to the participant. Each metric consisted of an explanatory question, as presented in Table 2.

After the experiment, a final rating scale and a free-form questionnaire about using the gestural techniques for the translation tasks were provided. The participants ranked the four gestural techniques in order of preferences from the most (1) to least (4) preferable. The whole experiment took 35–50 minutes per participant to complete.

3.6. Design. The experiment was a two-way (four interaction techniques × two translation distances) within-subjects factorial design.

The dependent variables were movement time (*MT*), error rate (*ER*), and target reentries (*TRE*). The *MT* was defined as the time measured from the first recognized clutching gesture to the acquisition of the target by release gesture. The *ER* was the ratio of unsuccessful acquisition attempts on one distance condition for both movement directions per a sequence of 6 trials. The *TRE* was a ratio of a target overshooting and then coming back per a sequence of 6 trials.

Each participant completed 4 blocks of 12 trials with one block per combination of movement distance and direction. With 30 participants, the total number of trials was $30 \times 4 \times 2 \times 2 \times 3 = 1440$.

4. Results

4.1. Data Analysis and Preprocessing. The data analyses were performed with data that included only successful trials. The trials were considered successful if their execution time did not exceed one minute. Notably, nearly all participants succeeded in the translation tasks with all four techniques. This happened in 100% of cases for the translation by *palm*, 99.31% of cases for the translation by *fist* and *pinch*, and 98.6% for the translation by *sideways*. Therefore, the total number of deleted trials was $2 + 2 + 4 = 8$, which was 0.7% of all the original trials. As an outlier detection, mean values in each block for each performance measure were first calculated individually for each participant. The individual mean block values exceeding ± 2SD from the sample block

Table 2: User satisfaction evaluation metrics and corresponding questions.

Satisfaction metrics	*Question (rated on 9-point Likert scale)*
General evaluation	How can you evaluate the technique in general?
Pleasantness	How pleasant was it to use the technique?
Quickness	How fast was the system reacting to your actions?
Accuracy	How accurate was the system reacting to your actions?
Efficiency	How successful were you in accomplishing what you were asked to do?
Physical	Demand how physically demanding was the task with the technique?
Mental demand	How mentally demanding was the task with the technique?
Temporal demand	How hurried or rushed was the pace of the task?
Frustration	How insecure, discouraged, irritated, stressed, and annoyed were you?
Distractibility	How does the software distract you from the performance of tasks?
Usability	How usable can you find the technique?

mean were then excluded (3.5% of the blocks) and finally replaced by a recalculated sample block mean.

A 4 (interaction technique) × 2 (distance) within-subject analysis of variances (ANOVA) was performed on the data. Greenhouse-Geisser corrected degrees of freedom were used in case of violation of sphericity. Bonferroni corrected pairwise t-tests were used for post hoc pairwise comparisons. One-way repeated measures ANOVAs were performed on the data to break down the interaction effect. A nonparametric Friedman test was applied to compare the subjective ratings. For pairwise comparisons of subjective ratings, the Wilcoxon signed-rank test with a Bonferroni correction was conducted.

4.2. Movement Time. The grand mean movement time MT was 8.68 s. The longest MT was 27 s for translation by *pinch* at the long distance. The shortest MT was 2.22 s for a *sideways* translation on the short distance (see Figure 4(a)). The ANOVA showed statistically significant main effects of the interaction technique ($F_{3,63} = 10.01$; $p < 0.001$) and the distance ($F_{1,21} = 80.11$; $p < 0.001$) on the MT. There was no statistically significant interaction of the main effects ($F_{3,63} = 2.01$; $p < 0.76$). Table 3 shows the results of pairwise post hoc tests.

4.3. Error Rate. The grand mean of error rate ER was 1.09 per sequence of 6 trials. The highest ER was 3.8 for translation by *pinch* at the long distance. The lowest ER was 0.2 for translation by *palm* at the short distance (see Figure 4(b)). Two-way ANOVA showed statistically significant main effects of interaction technique ($F_{1.4,28.1} = 43.19$; $p < 0.001$) and distance ($F_{1,20} = 12.05$; $p < 0.002$) on ER. The interaction of the main effects was also significant ($F_{1.8,35.6} = 4.11$; $p < 0.011$). The one-way repeated measures ANOVAs showed a significant effect of interaction technique for both the long ($F_{1.3,31.7} = 22.34$; $p < 0.001$) and the short distance ($F_{1.8,40.2} = 17.6$; $p < 0.001$). Table 4 shows the results of pairwise post hoc tests.

4.4. Target Reentries. The grand mean of target reentries TRE was 2.88 per sequence of 6 trials. The highest mean TRE was 21.7 for translation by *pinch* at the long distance. The lowest mean TRE was 0 for *sideways* translation at long and short distances (see Figure 4(c)). Two-way ANOVA showed statistically significant main effects of interaction technique ($F_{1.4,30.2} = 19.62$; $p < 0.001$) and distance ($F_{1,21} = 28.9$; $p < 0.001$) on TRE. The interaction of the main effects was also significant ($F_{1.9,41.1} = 14.5$; $p < 0.001$). The one-way repeated measures ANOVAs showed a significant effect of interaction technique for both the long ($F_{1.5,37.5} = 27.7$; $p < 0.001$) and the short distance ($F_{1.6,35.6} = 11.74$; $p < 0.001$). Table 5 shows the results of pairwise post hoc tests.

4.5. Subjective Ratings. Participants' preference ratings are shown in Figure 5. Friedman tests showed a statistically significant effect for participants' preferences of interaction techniques for translation ($\chi^2(3) = 37.6$; $p < 0.001$). Wilcoxon signed-rank test showed that there was a statistically significant preference in translation by *fist* versus *pinch* ($Z = 4.0$; $p < 0.001$), *fist* versus *sideways* ($Z = 3.7$; $p < 0.001$), *palm* versus *pinch* ($Z = 4.3$; $p < 0.001$), and *palm* versus *sideways* ($Z = 3.7$; $p < 0.001$).

Figure 6 summarizes the satisfaction results in boxplots separately for each interaction technique. Post hoc pairwise comparisons with the Wilcoxon signed-rank test showed a statistically significant increase in participants' subjective rating for interaction techniques on *general evaluation* ($\chi^2(3) = 32.32$; $p < 0.001$), *pleasantness* ($\chi^2(3) = 38.20$; $p < 0.001$), *quickness* ($\chi^2(3) = 32.37$; $p < 0.001$), *accuracy* ($\chi^2(3) = 31.49$; $p < 0.001$), *efficiency* ($\chi^2(3) = 26.30$; $p < 0.001$), *physical demand* ($\chi^2(3) = 25.20$; $p < 0.001$), *mental demand* ($\chi^2(3) = 20.94$; $p < 0.001$), *frustration* ($\chi^2(3) = 12.78$; $p < 0.005$), *distractibility* ($\chi^2(3) = 22.11$; $p < 0.001$), and *usability* ($\chi^2(3) = 26.60$; $p < 0.001 p < 0.001$), but not on *temporal demand*. Post hoc pairwise comparisons with the Wilcoxon signed-rank test showed a statistically significant increase in participants' subjective rating for interaction techniques (see Table 6). As we see from Table 6, interactions by *fist* and *palm* were more preferable by participants almost in all scales. Table 7 lists some participants' comments about their experience of translation by interaction techniques.

5. Discussion

Our results showed that the interaction by *palm* was significantly faster than that by any other techniques. The

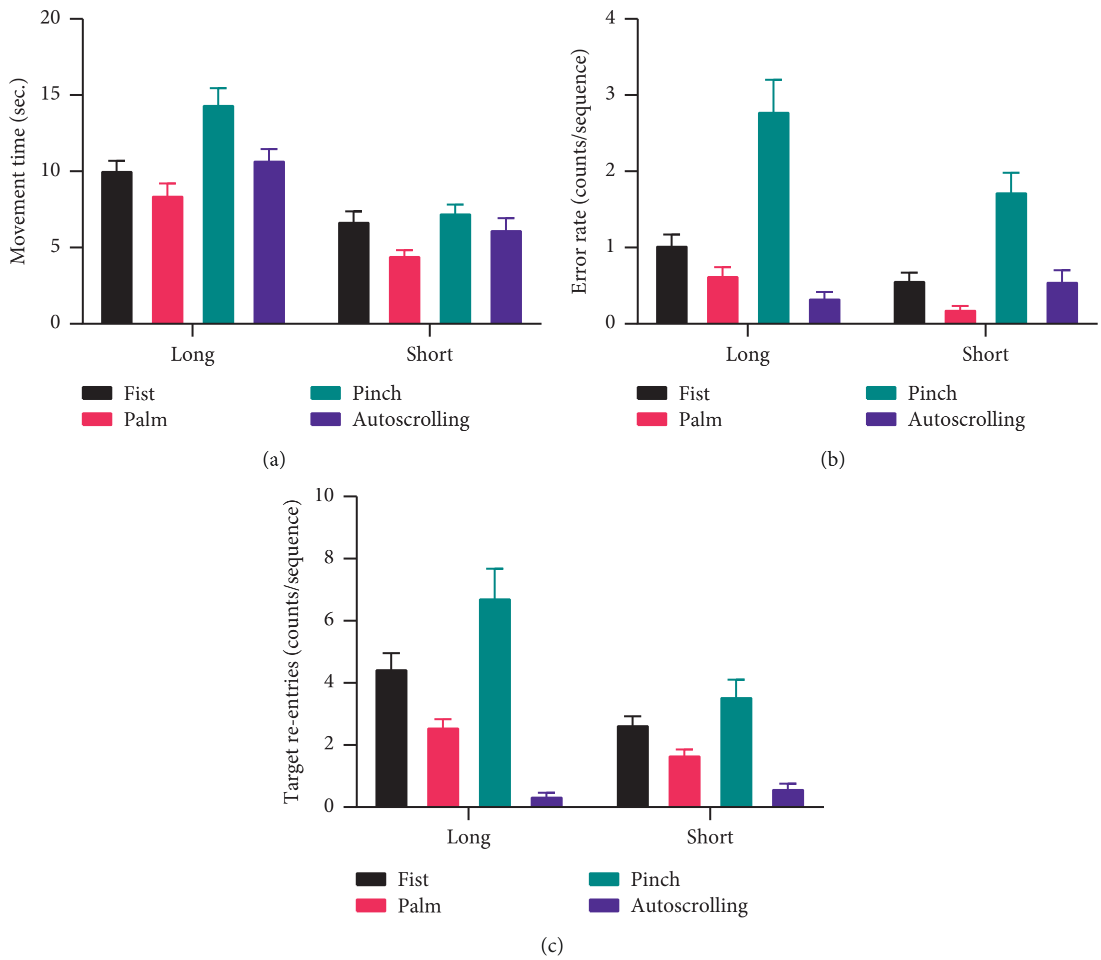


Figure 4: Plot of mean values for (a) MT; (b) ER; (c) TRE. Error bars represent ±1 SEM.

Table 3: Bonferroni corrected post hoc pairwise comparisons of *MT (sec)* for all interaction techniques.

Interaction technique		Mean difference	Significance
Slower	Faster	(*MD*)	(*p*)
Long distance			
Pinch	Fist	5.123	$p = 0.001$
Pinch	Palm	6.871	$p < 0.001$
Pinch	Sideways	5.589	$p = 0.001$
Short distance			
Fist	Palm	2.57	$p = 0.004$
Sideways	Fist	2.107	$p = 0.006$
Pinch	Palm	2.659	$p < 0.001$
Pinch	Sideways	2.589	$p = 0.001$

As we can see from the table, the pinch was the slowest technique at both distances.

findings from movement time data supported the assumption that the longer the translation distance, the more the movement time required for all interaction techniques. The *pinch* technique was significantly slower than the other techniques.

The results further revealed that the translation distance influenced interaction techniques in different ways, both in terms of accuracy and efficiency. So, the *sideways* technique was significantly more accurate than *fist* at long distances (see Table 3). Furthermore, the error rate analyses showed that significantly more errors were made with the *pinch* technique than with other techniques at all distances. The

Table 4: Bonferroni corrected post hoc pairwise comparisons of *ER* for all interaction techniques.

Interaction technique		Mean difference	Significance (p)
Less accurate	More accurate	(*MD*)	
Long distance			
Pinch	Fist	1.771	$p < 0.001$
Pinch	Palm	2.084	$p < 0.001$
Fist	Sideways	0.687	$p < 0.001$
Pinch	Sideways	2.423	$p < 0.001$
Short distance			
Pinch	Fist	1.098	$p < 0.001$
Pinch	Palm	1.417	$p < 0.001$
Fist	Palm	0.421	$p < 0.001$
Pinch	Sideways	1.206	$p < 0.001$

As we can see from the table, the pinch was the worst in accuracy at both distances, but the fist was less accurate than the sideways at the long distance and the palm at the short distance.

TABLE 5: Bonferroni corrected post hoc pairwise comparisons of *TRE* for all interaction techniques.

Interaction technique: Less efficient	Interaction technique: More efficient	Mean Difference (*MD*)	Significance (*p*)
Long distance			
Fist	Palm	1.709	$p = 0.002$
Pinch	Fist	2.309	$p < 0.004$
Fist	Sideways	4.103	$p < 0.001$
Pinch	Palm	4.006	$p < 0.001$
Palm	Sideways	2.211	$p < 0.001$
Pinch	Sideways	6.544	$p < 0.001$
Short distance			
Fist	Palm	0.922	$p = 0.003$
Fist	Sideways	2.121	$p < 0.001$
Pinch	Palm	0.640	$p = 0.004$
Palm	Sideways	1.476	$p < 0.001$
Pinch	Sideways	4.226	$p < 0.001$

As we can see from the table, the sideways was the most efficient at both distances, and the pinch was the least efficient than other techniques at the long distance.

trends suggested that increasing distance leads to an increasing number of errors only for interaction by *pinch*. *Sideways* interaction was significantly more efficient at both distances (see Table 4). However, interaction by *fist* was significantly more efficient than that by *pinch* only at long distances. Thus, translation by *pinch* was significantly more likely to overshoot than the other three techniques at long distances.

Putting together the statistical analyses of qualitative and quantitative evaluation, it seems that the best technique was interaction by *palm*. This might be because this technique was rated as low in overall difficulty, accurate, and easy to detect. Previously, it had been found that dwell-time-based interaction by *palm* gesture was the best gesture for item selection because of being the most intuitive [30, 33, 36] and more accurate [2] than *fist* techniques. Thus, our findings seem to be in line with the earlier studies investigating midair gestural interaction.

However, the Midas touch problem occurred during the translation by *palm*. Midas touch is one of the biggest problems for midair gestural interaction and relates to the detection of unintended gestures [7]. In our study, the Midas touch could appear as accidental clutching when an open palm was detected in the camera view. Accidental palm detection was also noticed in previous studies [23]. The solution for this kind of problem could be the conditions that help to avoid an accidental system engagement, for example, to dwell with an open palm or to place a hand in the center of the camera view.

For all interaction distances, translation by *pinch* had significantly worse results. This was also reflected in the ratings of participants' preferences. In line with this, the *pinch* technique received also significantly more negative user feedback than *palm* and *fist*. There are several possible explanations for this finding. Firstly, the low quality of recognizing the *pinch* gesture at hand rotation leads to an increase in errors. Secondly, users complained that targeting with this technique required extra attention, which distracted them from the main tasks. Similar findings have been noted previously [27]. In addition, hand fatigue was felt with prolonged interaction by *pinch*.

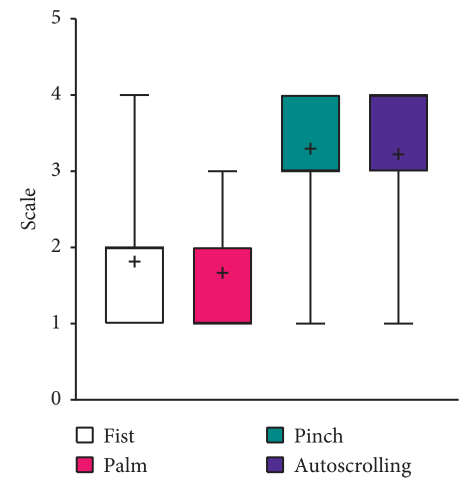


FIGURE 5: Plot of mean values for participants' preferences.

It is noteworthy that *sideways* translation showed a good result in quantitative evaluation measures such as error rate and target reentries. Thus, overshoots occurred less frequently in *sideways* interaction than in others at both distances. However, interaction by *fist* overshot significantly fewer times than by *pinch* only at short distances. Further, at long distances, the *sideways* translation resulted in significantly fewer errors than *fist* and *pinch*. However, the *sideways* technique received a low evaluation across all subjective rating scales (see Figures 5 and 6). This likely happened because it took more time for the participants to understand how the *sideways* interaction works. It seems that, when using this technique, participants had to make more body movements when compared to other techniques. For example, participants were required to take a step to the left or right for better gesture recognition, which was done reluctantly. This finding also reflects the argument made by Jakobsen et al. [37] that users prefer not to move during the midair interaction, even if the tasks are hard to select at a distance. It is possible that the use of both hands interchangeably could improve the user evaluation of this interaction technique as was done by Koutsabasis and Domouzis [28]. However, it was noted previously by Walter et al. [10] that a user tends to use one hand only. Although the participants needed to choose the hand for interaction in the present study, our findings seem to be in line with Koutsabasis and Domouzis [28].

Finally, based on the findings from the end questionnaire (see Figures 5 and 6), translation by *fist* received positive user evaluation across all rating scales. This was also reflected in the movement time and error rate. Thus, translation by *fist* was significantly faster and more accurate than by *pinch*. However, the statistical analysis of measurable parameters showed that the *fist* technique was significantly slower and less accurate than *palm* at short distances. Additionally, we found that the participants evaluated the interaction by *fist* as significantly more efficient than by *palm*. In earlier studies, interaction by *fist* was found fast, intuitive, and

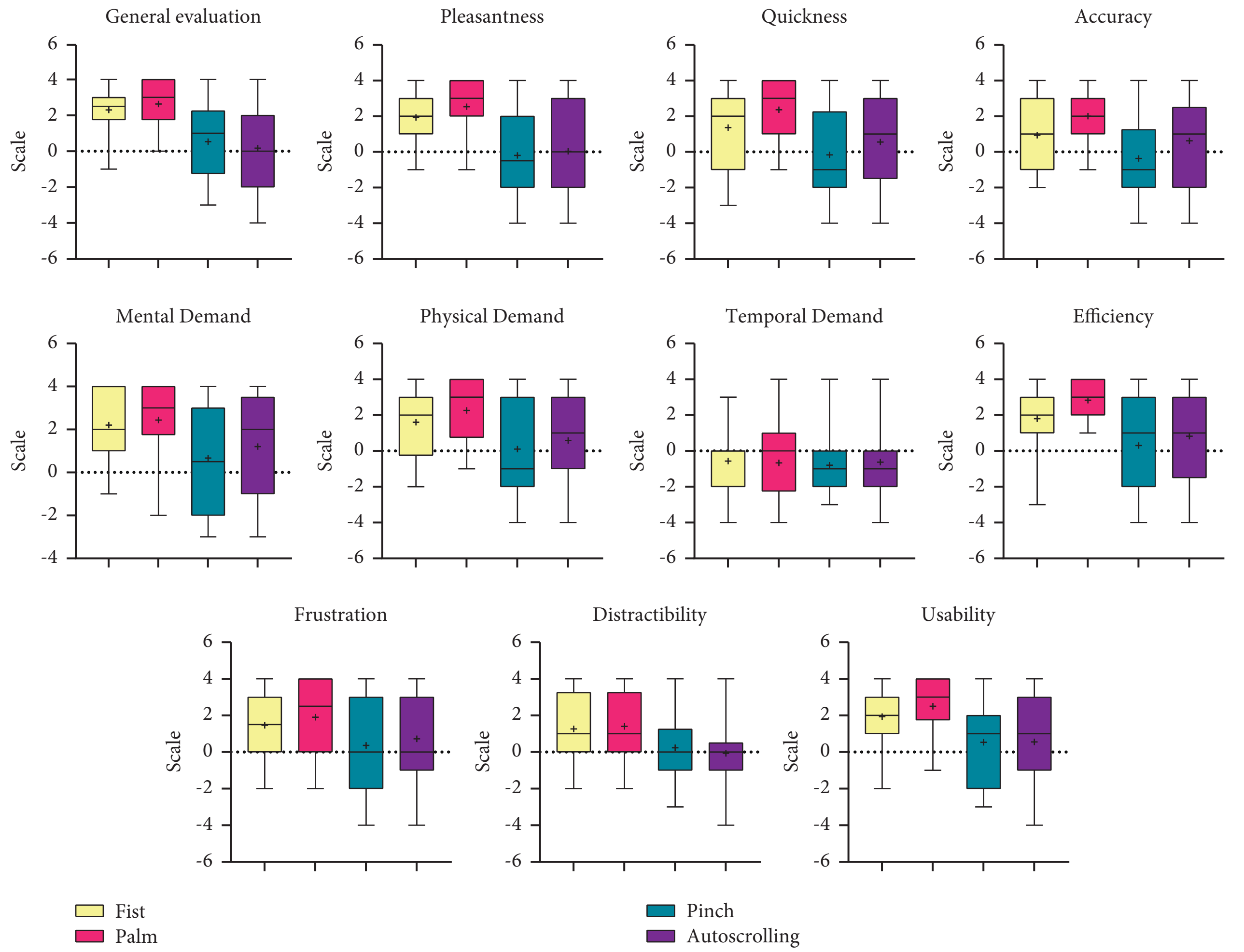


Figure 6: User satisfaction results across different interaction techniques.

required less physical demand [2, 38]. Thus, our findings regarding the *fist* technique are in line with previous findings.

In general, for translation tasks, both mechanics, relative pointing (with clutching and release gestures) and autoscrolling, have the potential to be utilized in real-time large-display interactions. The current findings are in line with the previous results of Gupta et al. [7] in that this interaction type is independent of target size but dependent on translation distance. In particular, the use of translation offers promising alternatives for the ray-casting pointing and selection tasks. However, while the metrics of utilizing interaction techniques for the translation tasks in certain scenarios are clear, they do not apply to real-life applications. Hence, the question of how to integrate them into applications with multiple control types and more complex interaction scenarios needs to be further explored. In real-life applications, users could also, for example, handle text entry, change the sizes of the objects, or rotate them. This implies different interaction techniques with optimal performance and could require compromises to integrate with other techniques.

To summarize our findings for future research, the following notions can be made. Gestures that may look simple and have good performance in static interaction, such as *pinch*, may prove challenging in translation. Such gestures demand more concentration from users and distract them from the primary task. This, in turn, led to fatigue and a reduction in performance. The use of simple gestures (such as *palm*) has several benefits, including being easy to learn, preventing fatigue in continuous movement, and enhancing recognition accuracy. However, the possibility of accidental detection of unintended midair gestures should be minimized. Furthermore, translation distance affects interaction techniques differently. Thus, interaction by *fist* had a significantly better performance at short distances compared with interaction by *sideways* and by *pinch* in terms of movement time and error rates. In contrast, *sideways* interaction seems to be a good alternative for translation tasks at long distances; a user just needs to extend his/her hand to the border of the screen in the required direction of the object's movement. *Sideways* interaction offers promising performance in terms of error rate and target reentries. In this study, interacting *sideways* required some participants to step to the left or right for better gesture recognition. Therefore, developing midair gestural interaction techniques should consider that the gestural interaction technique involves only hand movements and not full-body movements.

Table 6: Statistically significant reduction of participants' subjective rating based on post hoc pairwise comparisons with Wilcoxon signed-rank test.

Interaction technique preference			
More	Less	*Z*	*Significance* (*p*)
General evaluation			
Fist	Pinch	3.509	$p < 0.001$
Fist	Sideways	3.481	$p < 0.001$
Palm	Pinch	4.119	$p < 0.001$
Palm	Sideways	3.904	$p < 0.001$
Pleasantness			
Fist	Pinch	3.783	$p < 0.001$
Fist	Sideways	3.320	$p = 0.001$
Palm	Pinch	4.464	$p < 0.001$
Palm	Sideways	3.742	$p < 0.001$
Quickness			
Fist	Pinch	2.694	$p = 0.007$
Palm	Pinch	4.298	$p < 0.001$
Palm	Sideways	2.865	$p = 0.004$
Accuracy			
Fist	Palm	3.088	$p = 0.002$
Fist	Pinch	2.768	$p = 0.006$
Palm	Pinch	4.379	$p < 0.001$
Efficiency			
Fist	Palm	3.402	$p = 0.001$
Fist	Pinch	3.517	$p < 0.001$
Palm	Pinch	4.122	$p < 0.001$
Palm	Sideways	3.386	$p = 0.001$
Physical demand			
Fist	Pinch	3.055	$p = 0.002$
Palm	Pinch	3.816	$p < 0.001$
Palm	Sideways	2.990	$p = 0.003$
Mental demand			
Fist	Pinch	3.222	$p = 0.001$
Palm	Pinch	3.812	$p < 0.001$
Frustration			
Palm	Pinch	2.979	$p = 0.003$
Distractibility			
Fist	Pinch	3.211	$p = 0.001$
Palm	Pinch	3.442	$p = 0.001$
Efficiency			
Fist	Pinch	2.956	$p = 0.003$
Fist	Sideways	3.238	$p = 0.001$
Palm	Pinch	4.060	$p < 0.001$
Palm	Sideways	3.287	$p = 0.001$

Table 7: Participants' comments about the use of interaction techniques.

Interaction technique	*Original comments from participants*
Palm	"Was easy and quick"; "felt very fluid and natural"; "the most pleasant technique, it was organized in such a way that I did not get distracted by technical problems."
Fist	"Was the easiest to hit, but it did not recognise my palm so well"; "was so natural to do, for example, when I grab something, and that is why it is so pleasant."
Pinch	"Getting the pinch was often really hard"; "it was more difficult to perform the task, because it was hard to pose my hand in a way that is visible for the camera"; "it seemed that the camera kept losing track of my hand that was frustrating"; "I did not like this because of the fatigue it caused many times."
Sideways	"Was slow, and moving left was difficult, I also needed to move a lot and it was hard to hit the target"; "touching edges was hard to learn and hard to use on the left"; "was easy since I did not have to move my hand just take it down when I needed to stop."

6. Conclusion

We studied midair gestural techniques for translation effect in large-display interaction. The choice of interaction techniques for the studied translation tasks was justified by the literature review. The results showed statistically significant differences in movement time, error rate, and target reentries. Translation by palm was favored over pinch and sideways at both distances. The *sideways*, on the other hand, provided the most efficient translation at both distances. Most participants found that interaction by *palm* was the easiest and preferable, as opposed to *pinch* and *sideways*. Based on both quantitative data and subjective opinion, it is clear that translation by *pinch* was the worst interaction technique, and this interaction technique for translation tasks may not be recommended. Moreover, our findings also showed that interaction techniques differed by performance depending on the target distance. This justifies in some way why the results of the literature review were mixed and somewhat contradictory. Thus, the interaction by *fist* displayed a good performance and potential for usage at short distances and *sideways* interaction at long distances.

In future work, we plan to investigate how these gestural techniques for translation tasks combine with other midair interface tasks and work in real-world applications.

Data Availability

The underlying experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was completed as a part of doctoral studies funded by the Faculty of Information Technology and Communication Sciences at the Tampere University, Cognirem project, financed by the Academy of Finland (Grant no. 326430) and CIMO Fellowship under Grants TM-14-9500 and TM-16-9980. The authors thank all the volunteers and all publications support and staff who wrote and provided helpful comments on previous versions of this paper.

References

[1] C. Ackad, A. Clayphan, M. Tomitsch, and J. Kay, "An in-the-wild study of learning mid-air gestures to browse hierarchical information at a large interactive public display," in *Proceedings Of the 2015 ACM International Joint Conference On Pervasive And Ubiquitous Computing*, Osaka, Japan, 2015.

[2] S. Yoo, C. Parker, J. Kay, and M. Tomitsch, "To dwell or not to dwell: an evaluation of mid-air gestures for large information displays," in *Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction*, Parkville, VIC, Australia, 2015.

[3] L. Van den Bogaert and D. Geerts, "User-defined mid-air haptic sensations for interacting with an AR menu environment," *Haptics: Science & Technologie Alimentaire. International Conference on Human Haptic Sensing and Touch Enabled Computer Applications*, Springer, Berlin, Germany, pp. 25–32, 2020.

[4] V. Mäkelä, H. Korhonen, J. Ojala et al., "Investigating mid-air gestures and handhelds in motion tracked environments," in *PerDis '16: Proceedings of the 5th ACM International Symposium on Pervasive Displays*, New York, NY, USA, 2016.

[5] R. Aigner, D. Wigdor, H. Benko et al., "understanding mid air hand gestures: a study of human preferences in usage of gesture types for hci, s.l: microsoft research," 2012, http://www.microsoft.com/en-us/research/publication/understanding-mid-air-hand-gestures-a-study-of-human-preferences-in-usage-of-gesture-types-for-hci/.

[6] D. Mendes, F. Relvas, A. Ferreira, and J. Jorge, "The benefits of DOF separation in mid-air 3D object manipulation," in *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*, pp. 241–268, Munchen, Germany, 2016.

[7] A. Gupta, T. Pietrzak, C. Yau, N. Roussel, and R. Balakrishnan, "Summon and select: rapid interaction with interface controls in mid-air," in *Proceedings of the ACM Conference on Interactive Surfaces and Spaces*, New York, NY, USA, 2017.

[8] P. Vogiatzidakis and P. Koutsabasis, "Gesture elicitation studies for mid-air interaction: a review," *Multimodal Technologies and Interaction*, vol. 2, no. 4, 2018.

[9] R. Ball, C. North, and D. A. Bowman, "Move to improve: promoting physical navigation to increase user performance with large displays," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, San Jose, California, USA, 2007.

[10] R. Walter, G. Bailly, and J. Muller, "StrikeAPose: revealing mid-air gestures on public displays," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Paris, France, 2013.

[11] C. Groenewald, C. Anslow, J. Islam, C. Rooney, P. J. Passmore, and B. L. Wong, "Understanding 3d mid-air hand gestures with interactive surfaces and displays: a systematic literature review," in *Proceedings of the BCS Human Computer Interaction Conference 2016*, Poole, UK, 2016.

[12] P. Koutsabasis and P. Vogiatzidakis, "Empirical research in mid-air interaction: a systematic review," *International Journal of Human-Computer Interaction*, vol. 35, no. 18, pp. 1747–1768, 2019.

[13] J. Ahn and K. Kim, "Investigating smart TV gesture interaction based on gesture types and styles," *Journal of the Ergonomics Society of Korea*, vol. 36, no. 2, pp. 109–121, 2017.

[14] A. S. Arif, W. Stuerzlinger, E. J. d. M. Filho, and A. Gordynski, "Error behaviours in an unreliable in-air gesture recognizer," in *CHI '14 Extended Abstracts On Human Factors In Computing Systems*, pp. 1603–1608, Association for Computing Machinery, New York, NY, USA, 2014.

[15] M. Carreira, K. L. H. Ting, P. Csobanka, and D. Gonçalves, "Evaluation of in-air hand gestures interaction for older people," *Universal Access in the Information Society*, vol. 16, no. 3, pp. 561–580, 2017.

[16] E. Chan, T. Seyed, W. Stuerzlinger, X.-D. Yang, and F. Maurer, "User elicitation on single-hand microgestures," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, San Jose, California, USA, 2016.

[17] H. Dong, A. Danesh, N. Figueroa, and A. El Saddik, "An elicitation study on gesture preferences and memorability toward a practical hand-gesture vocabulary for smart televisions," *IEEE Access*, vol. 3, pp. 543–555, 2015.

[18] J. Tang, R. Xiao, A. Hoff, G. Venolia, P. Therien, and A. Roseway, "HomeProxy: exploring a physical proxy for video communication in the home," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Paris, France, 2013.

[19] R.-D. Vatavu and I.-A. Zaii, "Leap gestures for TV: insights from an elicitation study (2014)," in *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online*, Newcastle Upon Tyne, United Kingdom, 2014.

[20] B. Vogel, O. Pettersson, A. Kurti, and A. Huck, "Utilizing gesture based interaction for supporting collaborative explorations of visualizations in tel," in *Proccedings of the Seventh IEEE International Conference on Wireless, Mobile and Ubiquitous Technology in Education*, Takamatsu, Japan, 2012.

[21] H. Wu, J. Wang, and X. Zhang, "User-centered gesture development in TV viewing environment," *Multimedia Tools and Applications*, vol. 75, no. 2, pp. 733–760, 2016.

[22] A. Viveros and E. Rubio, "Kinect ©, as interaction device with a tiled display. Human-computer interaction interaction modalities and techniques. HCI 2013," in *Proceedings of the Lecture Notes in Computer Science*, Las Vegas, NV, USA, 2013.

[23] F. Garzotto and M. Valoriani, "Touchless gestural interaction with small displays: a case study," in *Proceedings of the Biannual Conference of the Italian Chapter of SIGCHI*, Trento, Italy, 2013.

[24] F. Tian, F. Lyu, X. L. Zhang, X. Ren, and H. Wang, "An empirical study on the interaction capability of arm stretching," *International Journal of Human-Computer Interaction*, vol. 33, no. 7, pp. 565–575, 2017.

[25] L.-C. Chen, Y.-M. Cheng, P.-Y. Chu, and F. E. Sandnes, "Identifying the usability factors of mid-air hand gestures for 3D virtual model manipulation," in *Universal Access In Human–Computer Interaction. Designing Novel Interactions. UAHCI 2017. Lecture Notes In Computer Science, 10278*, M. Antona and C. Stephanidis, Eds., Springer, Berlin, Germany, 2017.

[26] A. Pereira, J. P. Wachs, K. Park, and D. Rempel, "A user-developed 3-D hand gesture set for human-computer interaction," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 57, no. 4, pp. 607–621, 2015.

[27] F. Farhadi-Niaki, S. Etemad, and A. Arya, "Design and usability analysis of gesture-based control for common desktop tasks," in *Human-Computer Interaction. Interaction Modalities And Techniques. 15th International Conference, HCI International 2013, Part IV, 215-224). Las Vegas, NV, USA*, M. Kurosu, Ed., Springer, Berlin, Germany, 2013.

[28] P. Koutsabasis and C. K. Domouzis, "Mid-air browsing and selection in image collections," in *Proceedings of the International Working Conference on Advanced Visual Interfaces*, Bari, Italy, 2016.

[29] H. Wu and J. Wang, "User-defined body gestures for tv-based applications," in *Proceedings of the Fourth International Conference on Digital Home*, Guangzhou, China, 2012.

[30] O. Erazo, A. Vicuña, R. Pico, and B. Oviedo, "Analyzing mid-air hand gestures to confirm selections on displays," *Communications in Computer and Information Science*, vol. 895, pp. 341–352, 2019.

[31] K. Käki, P. Majaranta, O. Špakov, and J. Kangas, "Effects of haptic feedback on gaze based auto scrolling," in *NordiCHI '14: Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast*, Helsinki, Finland, 2014.

[32] S. Lenman, L. Bretzner, and B. Thuresson, "Using marking menus to develop command sets for computer vision based hand gesture interfaces," in *NordiCHI '02: Proceedings of the Second Nordic Conference on Human-Computer Interaction*, pp. 239–242, Århus, Denmark, 2002.

[33] L. Hespanhol, M. Tomitsch, K. Grace, A. Collins, and J. Kay, "Investigating intuitiveness and effectiveness of gestures for free spatial interaction with large displays," in *Proccedings of the International Symposium on Pervasive Displays*, New York, NY, USA, 2012.

[34] E. Velloso, J. Turner, J. Alexander, A. Bulling, and H. Gellersen, "An em pirical investigation of gaze selection in mid-air gestural 3D manipulation," in *Proceedings of the Human-Computer Interaction – INTERACT 2015. INTERACT 2015. Lecture Notes in Computer Science*, Bamberg, Germany, 2015.

[35] S. Hart, "NASA-task Load Index (NASA-TLX); 20 Years later," in *Proccedings of the Human Factors and Ergonomics Society 50th Annual Meeting, 904-908*, Santa Monica, California, USA, 2006.

[36] R. Walter, G. Bailly, N. Valkanova, and J. Müller, "Cuenesics: using mid-air gestures to select items on interactive public displays," *P*, in *roceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices & Services*, New York, NY, USA, 2014.

[37] M. R. Jakobsen, Y. Jansen, S. Boring, and K. Hornbæk, "Should i stay or should i go? selecting between touch and mid-air gestures for large-display interaction," in *Proccedings of the 15th Human Computer Interaction (INTERACT)*, Bamberg, Germany, 2015.

[38] Y. T. Hsieh, A. Jylhä, V. Orso, L. Gamberini, and G. Jacucci, "Designing a willing-to-use-in-public hand gestural interaction technique for smart glasses," in *Proccedings of the Conference on Human Factors in Computing Systems*, 2016.