

Predictive Model for Group Selection Performance on Touch Devices

Per Bjerre¹, Allan Christensen¹, Andreas K. Pedersen¹,
Simon A. Pedersen¹, Wolfgang Stuerzlinger², and Rasmus Stenholt¹

¹ Aalborg University, Denmark,

² School of Interactive Arts + Technology (SIAT), Simon Fraser University, Canada

Abstract. Users spend hours making selections with ineffective tools, we therefore examine selection methods for efficiency in various touch trials. In a preliminary study three alternative selection methods were identified, we compared these to a smart selection tool. The study showed that a single selection method was the fastest; however, when the amount of targets increased, a multiple selection tool became more efficient. A secondary study with more targets revealed similar results. We therefore examined the single selection method against traditional selection methods in a user study. The results reveal a model of the average action and time cost for all methods within the parameters of mental preparation and target addition. The study revealed that the most favored selection methods were a lasso and brush selection tool. The study provides evidence towards a predictive model of selection performance for multiple target selection trials.

1 Introduction

Selecting multiple objects can be tedious when interacting with systems. These selection operations tend to require high user accuracy and intense visual attention. Users have a large amount of tools available to perform selection tasks, and these tools can have different purposes. Novice users have difficulties identifying the most optimal selection tool for a given scenario. Stenholt showed that users prefer using simple tools several times rather than a more complex one fewer times [1]. This indicates that selection tools are not always used as intended.

We examine Lazy Selection, a stroke based tool designed for efficient target selection, with the ability to select multiple targets with a single stroke action. The stroke crosses the elements of interest and can be of any length. The system interprets the users stroke and predicts the users intent, requiring less user accuracy compared to normal selection techniques [2]. One issue with such methods is that they sometimes select the wrong targets in a drawing (denoted prediction errors). Prediction errors occur when there are ambiguities between the users strokes and the drawing.

For the first user study, we examine how to adjust prediction errors on a touch device, by expanding the tools available for handling these errors with user defined gestures. We investigate which of these tools perform best in terms

of time usage and amount of errors. In the second user study we increased the amount of targets to examine the effect. For the third user study, we investigate the most efficient tool from the previous studies against traditional selection tools in various scenarios to determine, which of these tools perform best in terms of actions, time usage, and amount of errors when the target amount is increased. The study results in a linear regression model of the most efficient selection methods for multiple selection trials.

2 Related Work

Touch Interaction Direct touch manipulation involves the user interacting directly with content on a touch-sensitive display. The advantage of direct touch was highlighted by Forlines et al., who found that the interaction felt more natural when interacting through direct touch compared to a mouse [3]. The results showed that it was faster to interact directly. Yet, when target size decreased, the selection time increased. Regarding hand preference with direct touch, previous studies showed that participants prefer unimanual gesture interaction (one handed interaction) and that unimanual gestures used less time than bimanual ones (two handed interaction) [4].

Wills examined different types of selection and identified three criteria that interaction techniques should adhere to: simple, powerful, and forgiving [5]. Users also prefer to use simpler interactions repeatedly over complex ones that require fewer actions to complete a task [1, 6].

Traditional selection tools for image manipulation Image manipulation tools, such as Gimp [7], provides users with a variety of selection tools for selecting desired content in an image. James et al. mentioned that selection tools are crucial for detailed image manipulation to accurately apply effects to individual parts of an image [8]. Selection tools highlight an area in an image, where the user can then apply effects. In GIMP one of the common tools for region selection is Rectangle Select. With this tool the user draws a rectangle and selects everything in the associated area. This tool is also used in file managers for file and folder selection [5]. Another common selection tool is Lasso. This tool selects all targets inside a freely defined (closed) path that the user is drawing [8, 5]. According to James et al., on a touch based device, users find the Lasso tool more precise compared to mouse users on a desktop [8, 5]. This is also supported by Kin et al., as their study identified a higher selection performance for one finger touch compared to mouse-based interaction [9]. In a similar study, Stenholt tested the Rectangle, Brush, Lasso, and Magic Wand selection methods in various 3D selection trials. The results showed that Brush was faster than Lasso. In the experimental trials, Rectangle used significantly fewer actions than the other tools, but was also the slowest [1].

Gestalt Principles To identify the best alternative for multi-touch interaction, we explored the Gestalt principles. These principles were first examined in 1923 [10]. Whenever points (or previously formed groups) have one or several characteristics in common, they get perceptually grouped and form a new,

larger visual target, known as a Gestalt, which is often explained as shape or form [11]. The Gestalt principles apply to the visual, auditory, tactile, and haptic modalities [12–14]. Here we focus only on the visual modality.

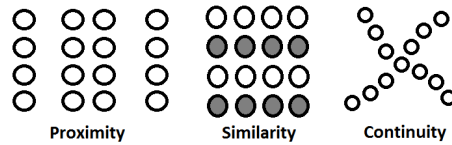


Fig. 1. Illustration of some Gestalt principles. Proximity refers to the distance between targets. Similarity groups targets that are similar in appearance. Continuity refers to groups that appear when targets lie on a line or curve.

There is no definitive list of Gestalt principles, but some of the commonly used ones are proximity, similarity, closure, continuity, and symmetry, e.g., [15, 14]. Some of these can be seen in Figure 1. The proximity principle assumes that if a number of targets lie close together, they will naturally be considered to be grouped together. If a number of targets lie scattered, but have similar visual features, then they can also form a group [16]. Co-linear targets also form groups. Similarity implies that targets that look similar will be perceived as a group, which can be due to a variety of cues. The similarity principle can be divided into categories, such as form, size, color, brightness, orientation, and texture. All these categories can work towards making targets look similar and therefore enable the targets to form Gestalts [16].

Perceptual Grouping Methods To further investigate the Gestalt principles, we explore relevant work that applied the principles for selection methods and to design experiments.

Thofissson created an algorithm to find perceptual groupings with the Gestalt principles of proximity and similarity [16]. Desolneux et al. stated that their participants perceived groups, when targets appeared inside a closed curved line or when targets aligned symmetrically across a straight line. Lastly, they identified that the continuity occurs when targets align in a line [11] Dehmeshki and Stuerzlinger created PerSel, which utilizes the Gestalt principles to predict the group of targets to select based on a flick gesture on a single target as input. They found that PerSel outperforms standard selection methods such as the Lasso and Rectangle selection techniques, when users have to select targets that are perceptually grouped [17].

Surveying the work discussed above, we can state that the individual targets to be selected in our experiment should have the same size, to ensure an equal baseline. Several studies used four traditional selection tools: brush, rectangle, lasso, and tap. We decided to use tapping and Lazy Selection as our baseline to compare against. Lastly, we also use stimuli based on the Gestalt principles, also because they help users quickly understand the task.

3 Methods and Materials

A pilot study focused on user interactions and accuracy of Lazy Selection established that the predictions are frequently inaccurate. This was also pointed out by the authors of Lazy Selection [2]. In our pilot study, users colored sketches on a tablet with the original Lazy Selection software. We asked them to indicate when and where a selection error occurred and to correct the selection with the tools available in Lazy Selection. We observed that inaccuracies occurred when targets were positioned in dense clusters or when selections involved many targets. The fat finger problem also increased the chance for errors, as it makes it hard to select specific targets without selecting other surrounding targets [18, 19]. The results also showed that users would redo a selection, rather than utilize the provided adjustment tools. If the initial prediction was undesirable, subsequent predictions were typically also not desired.

We then conducted an elicitation pilot by prompting users to use gestures intuitive to them to adjust such predictions. Our observations identified that people preferred direct interaction on the targets, rather than more abstract gestures to perform this task. These results are consistent with the findings of Forlines et al. [3]. A distinct pattern observed throughout the test was that users preferred to swipe and flick on targets as an interaction technique for selecting or de-selecting targets. Other common interaction patterns included using movements similar to how an eraser would be used on paper, as well as using symbols on top of targets such as a line or a cross, typically to remove targets from the selection. However, the described interaction technique of flicking or swiping on a target to “remove” the selection (more precisely the highlighting associated with selection) was predominant.

Based on the findings from our pilots and previous research, we chose three gestures to adjust the predictions of Lazy Selection. These gestures consist of two gestures based on the pin and flick concept and a click (Tap) gesture. Figure 2 illustrates all four gestures investigated in our work. Tap was selected due to being the simplest possible gesture for toggling a selection. The two pin and flick gestures were included because of their similarity to the delete gesture suggested by Wobbrock et al. [4]. Moreover, several users suggested this gesture during the elicitation pilot. With the Lazy Selection tool, users would first mark an area where they wanted to change the selection. When they lifted their finger, the algorithm would then toggle the selection inside the marked area, as determined by the Lazy Selection algorithm. Using Tap, users could then select individual targets by tapping said targets. For the uni- and bimanual pin and flick gestures, users would first pin the target(s) they wanted to act upon and then perform a flick gesture. In the unimanual method, the fingers that initially performed the pin on the undesired target(s) would also perform the flick gesture, which meant that the interaction could be performed with a single hand. For the bimanual method, the algorithm first waits for a pin gesture to occur on one (or more) target(s) and then looks for a flick at a separate location. This will then act on the pinned target(s). This operation can be applied with one or two hands, as desired. However, users were encouraged to use the method with both hands.

The flick serves both as a confirmation of the pin gesture and also simulates that the selection is “flicked away” from the target(s).

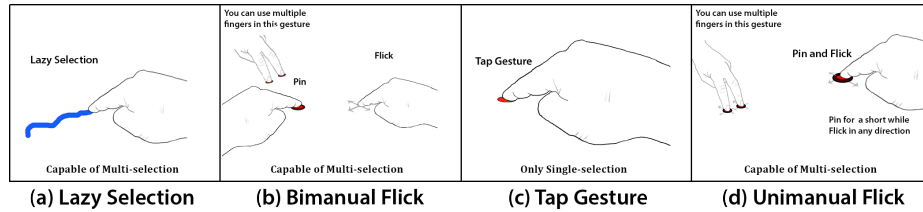


Fig. 2. Illustration of the gestures used during the study. (a) shows interaction with the original Lazy Selection technique. The user draws a stroke over targets to select them. (b) illustrates interaction with the Bimanual Pin and Flick gesture. The user pins one (or more) targets with one hand, and flicks away with the other hand to indicate the selection of the pinned targets. (c) shows an interaction with a tap gesture. The user taps and releases his/her finger on each target that needs to be selected. (d) illustrates an interaction with the Unimanual Pin and Flick gesture. The user first pins targets with one or more fingers and then toggles the selection by flicking with these fingers.

3.1 First User Study

The first user study compared the four mentioned interaction methods, with the original implementation of Lazy Selection as the control. The purpose of this study was to determine, which touch-based gestures are efficient to correct the predictions made by the Lazy Selection algorithm.

Hypotheses The study had four hypotheses:

1. The **amount of actions** used is significantly different between the methods.
2. The **completion time** is significantly different between the methods.
3. The **amount of user-generated errors** between the methods is significantly different.
4. The **user preference** between the methods is significantly different.

Within the stated hypotheses we expect Tap to have the lowest completion time and be the most efficient method for small sets of targets, as it is the simplest method [20]. For larger sets, we expect Tap to decrease in efficiency, as it needs one action per target. We expect that users will prefer the unimanual pin and flick gesture, as it seems to be very intuitive to use, and was also suggested by users during initial testing. According to our observations in the pilots, we do not expect a significant difference in completion time between the individual methods, as all methods seem to be equally fast at correcting predictions. Lastly, the bimanual pin and flick gesture is expected to be the least preferred gesture, as research and our initial pilot indicate that users prefer to use one hand when interacting with a tablet [21, 4].

Response variables We identified three main response variables for our user study: number of actions, completion time, and number of user-generated errors. We also investigated the participants preferred interaction technique through questionnaires and an interview. This yielded both qualitative and quantitative measurements. The quantitative measurements consisted of five point Likert scales to rate the methods. The qualitative data provided us with the participants thoughts on the best interaction technique.

The system tracked the number of actions, by counting the number of times a gesture was completed. Timing began when participant touched the surface and ended when a trial was completed. The last response variable was the user-generated errors. For this we recorded video of the participants during the trials. An error occurred whenever the participants selected something other than the desired targets and had to correct the prediction.

Experimental Procedure and Equipment We ran the user study on an Android tablet that interacted with a desktop-based implementation of Lazy Selection over Splashtop (www.splashtop.com) via a sufficiently fast wireless network. This kept our test comparable with previous results from Lazy Selection, while also investigating its potential for touch interaction. The setup also makes future comparisons to mouse-based interactions easily accessible.

Thirty participants (6 females) participated in the experiment. Ages were between 20 and 27, with an average of 23.27. All participants were students and were either at least familiar with touch interfaces and painting applications.

The user study was designed as a factorial experiment, with interaction methods and complexity as factors. The interaction method factor had four levels: Lazy Selection, Tap, unimanual pin and flick, and bimanual pin and flick. The complexity factor had three levels: easy, medium, and hard.

For the user study, a prediction scenario was created, where users had to deselect specific targets in the trials. In terms of user actions, the prediction scenario functions exactly the same as a selection scenario and thus all results are comparable with selection tasks. Each interaction method was evaluated with three trials involving drawings of different complexity. Each participant was first introduced to the study. Then a questionnaire with demographic questions was filled, followed by an explanation of how the trials would work. The participant was then placed in front of the tablet. For each interaction method a piece of paper with an illustration of the gesture mechanics was placed in front of the user. The illustrations can be seen in Figure 2. Subsequently and for each trial, the facilitator presented the participant with a drawing with a predefined set of erroneously selected targets (indicating a prediction error). Such errors were colored in red and needed to be removed from the trial for successful completion. After a trial had been completed, the next one was presented. The three complexity levels used during the experiment are illustrated in Figure 3. Selection complexity was defined by the proximity of targets, as well as the amount of overlapping targets. This makes it progressively harder to adjust the selection to the correct result. The order of the trials always increased in complexity. After

the participant had completed all three trials for an interaction method, a short training session for the next interaction method was used to familiarize subjects. Once they felt comfortable enough, they resumed the trials.

A within subject design was used, where each user went through all combinations of interaction methods and complexity levels. Therefore each participant went through a total of 12 trials (4 input methods times 3 complexity levels). The order of presentation for each interaction method was determined via a Latin Square to reduce potential learning effects.

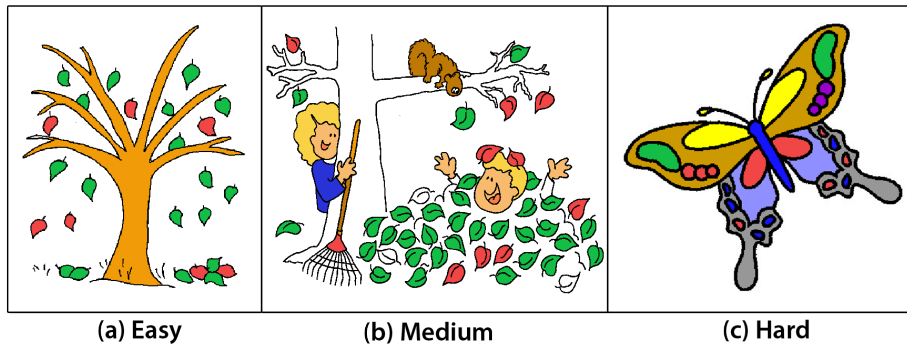


Fig. 3. Illustration of the drawings used for experimental trials. In each drawing the participant had to de-select all red targets. The complexity of the drawings increases throughout the experiment. (A) easy task, with nine red-colored targets in the image. (B) medium task, with ten targets. (C) hard task, with seven targets to de-select.

Results The analysis of the quantitative data was done with the R statistical software package with $\alpha = 0.05$. The analyzed data is based on the response variables: actions, time, and errors. Repeated measures ANOVA was used for all analysis. In the cases where the normality and homoscedasticity assumptions could not be upheld, the non-parametric Friedman test was used instead. The qualitative measurements gathered through the experiment, i.e., the favored interaction method, and easiest method to use.

Hypothesis 1 The data for number of actions was not normally distributed. A Friedman Ranked Sum test revealed that H1 is supported, because the Lazy Selection method used significantly fewer actions per exercise in relation to the difficulty of the exercises compared to the other designs ($\chi_{11}^2 = 165.87, p = 0$).

Hypothesis 2 The data for time usage was not normally distributed. A Friedman Ranked Sum test revealed that H2 is supported ($\chi_3^2 = 48.4, p \ll 0.001$). The test identified that Tap was significantly faster than the other interaction techniques, with Lazy Selection also being faster than both uni- and bimanual pin and flick. A

logarithmic plot of the distribution of time values averaged across all complexity levels can be seen in Figure 4.

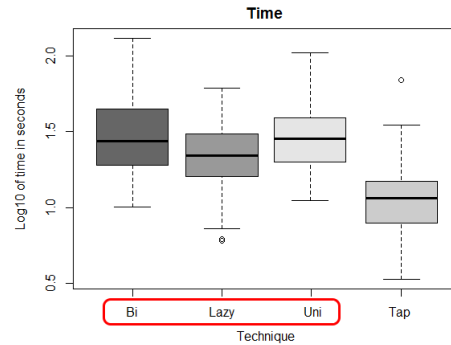


Fig. 4. Logarithm of time in the first user study. The figure indicates that the tap gesture needed significantly less time across all complexity levels. The red rounded rectangle indicates no significant difference between the contained methods.

Hypothesis 3 This hypothesis deals with the number of errors for each interaction method. Users generally made few selection errors during the experiment. The data was not normally distributed. Yet, the Friedman Ranked Sum test revealed that there is a significant difference between Tap and Lazy Selection with Tap being superior ($\chi_{11}^2 = 34.22$, $p = 0.0003$). Due to the overall low selection error count we could not detect other differences.

Hypothesis 4 For the ratings of the interaction methods, we used a Friedman Ranked Sum test to identify significant differences in the preference between the different interaction methods ($\chi_3^2 = 52.82$, $p \ll 0.001$). According to this test, Tap was ranked highest, followed by Lazy Selection, then unimanual pin and flick, and finally bimanual pin and flick.

Qualitative Data Table 1 summarizes the collected qualitative data. The participants who preferred Tap stated intuitiveness and simplicity as the primary reasons for their preference. However, most users mentioned that for more complex images with more targets, they would likely prefer the other techniques due to being able to select multiple targets with one action. The participants who preferred Lazy Selection mentioned the ability to select multiple targets at once as a positive feature. They also highlighted that the method still supported tapping individual targets. A large majority of the participants, 27 (90%), preferred to use one hand during the trials. Participants mentioned that for the bimanual pin and flick gesture one could mark targets with one hand, and then lift the hand before doing the flick with the other, thus being able to see which targets would be affected. This could partially counteract the fat finger problem [18, 19].

Method	User preference	Ease of use
Tap	22 (73.33%)	20 (66.66%)
Lazy Selection	7 (23.33%)	6 (20%)
Bimanual	0	1 (3.33%)
Unimanual	0	0
Both Lazy Selection & Tap	1 (3.33%)	3 (10%)
Total	30 (100%)	30 (100%)

Table 1. Table showing user preference and ease of use ratings for the four techniques.

3.2 Second User Study

The results from the first user study indicated that for small sets of targets (up to 10), Tap performs the best. Consequently, we decided to run a second user study to investigate performance in more complex scenarios, to identify if the results change with higher target counts.

The hypotheses and the experimental setup of this study are identical to the first one (Section 3.1 & 3.1). However, the selection trials were based on the visual Gestalt principles [14]. Examples of target arrangements for the second user study can be seen in Figure 5. The Gestalt groupings were focused primarily on the principles of proximity, similarity, and continuity. Test participants again had to remove all red colored targets.

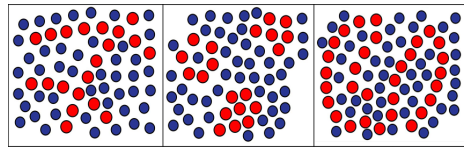


Fig. 5. Examples of drawings used for trials during the second user study. In each drawing the participant had to de-select all the red targets. There was a minimum of 17 and a maximum of 43 targets.

Twelve participants (3 females) participated in this study, between 23 and 30 years old with an average of 25.16. The purpose of the test was to examine if and how some of the selection methods benefit from Gestalt-based target groups.

Results For the results of the second user study, we examine the same response variables for trials involving larger target groups. We expect the results to be similar to the first user study, but with the multiple target selection methods having an advantage in terms of all response variables, because more targets had to be selected in the trials. The data from the second user study did not follow a normal distribution.

Hypothesis 1 The Friedman Ranked Sum test revealed that there is a significant difference for the amount of actions used between all interaction methods ($\chi_3^2 = 33.91$, $p \ll 0.001$). The test reveals that Lazy Selection used the fewest actions, followed by Tap, then the unimanual gesture, and the bimanual method with the most actions. The distribution of logarithmic action values averaged across all complexity levels can be seen in Figure 6. A logarithmic transformation is used to better illustrate the differences between the techniques.

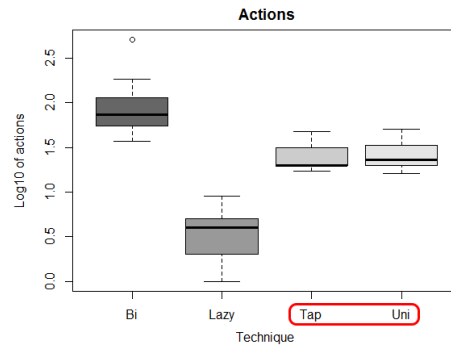


Fig. 6. Logarithm of number of actions in the second study. The figure indicates that Lazy Selection needed on average a significantly lower amount of actions across all trials, whereas the bimanual pin and flick gesture required the most actions. The red rounded rectangle indicates no significant difference between the contained methods.

Hypothesis 2 For the amount of time used to de-select each group, the test reveals that there is a significant difference between all methods, ($\chi_3^2 = 33.3$, $p \ll 0.001$). The Friedman Ranked Sum test identified that the fastest interaction method was Tap, followed by Lazy Selection, which is faster than unimanual pin and flick. Bimanual pin and flick is the slowest.

Hypothesis 3 The Friedman Ranked Sum test for the amount of user-generated errors during the study shows the same result as the previous one, namely that there is no significant difference between the interaction techniques ($\chi_3^2 = 1.57$, $p = 0.67$). Again, users only made few errors during the test.

Hypothesis 4 Based on a Friedman Ranked Sum test, we identified a significant difference between the user ratings of the gestures during the study ($\chi_3^2 = 27.52$, $p \ll 0.001$). Lazy Selection and Tap were rated highest, with no significant difference between the two, significantly higher than uni- and bimanual pin and flick, with no significant difference between these two.

Qualitative In the qualitative questions, 3 out of 12 subjects stated that they preferred Tap, whereas 9 out of 12 stated a preference towards Lazy Selection. A

Test of Equal Proportions test shows that there is a significant difference between these ratings ($\chi_1^2 = 4.17, p = 0.04$). When asked about the easiest design to use, 6 out of 12 participants stated that Tap was the easiest to use, whereas the other 6 stated that Lazy Selection was the easiest one. Further comments on this question shows that the participants who found Lazy Selection to be the best tool, but also stated that Tap was easier to use, clarified that they needed to examine the path they wanted to take before using Lazy Selection, whereas they could use Tap instantly. All participants preferred to interact with a single hand during the trials, and mentioned some of the same potential improvements to some of the gestures, such as being able to lift their hands to see the selection before completing the flick. Furthermore, several participants mentioned a dislike for the predictions of the Lazy Selection algorithm.

The results from the second user study shows that the efficiency of Tap is substantially reduced when the amount of targets increases. In the study Lazy Selection needed the least time and number of actions. Tap is still a good and simple selection method, but does not scale well, as every additional target requires a selection. To further investigate the efficiency and scalability of Tap in a selection environment, we chose to conduct another user study, where we investigate Tap in different selections scenarios. This enables us to identify *when* the effectiveness of Tap is less than other existing selection methods.

The results from the second user study shows that the efficiency of Tap is substantially reduced when the amount of targets increases. In the study Lazy Selection needed the least time and number of actions. Tap is still a good and simple selection method, but does not scale well, as every additional target requires a selection. To further investigate the efficiency and scalability of Tap in a selection environment, we chose to conduct another user study, where we investigate Tap in different selections scenarios. This enables us to identify *when* the effectiveness of Tap is less than other existing selection methods.

3.3 Final User Study

The main purpose of this user study is to create a predictive model, from which we can determine when one selection method is superior to another. The instability caused by the Lazy Selection algorithm led to a general dislike of the prediction choices. Instead of Lazy Selection we decided to investigate a simple brush, as used in several image manipulation software. To generalize the model a bit further, we chose to test lasso and rectangle selection methods. Figure 8 shows the methods used in the study. The goal of the predictive model is to provide us with information as to when one selection method is superior in both time and number of actions compared to the other methods.

Trial Design To reduce the bias for any individual selection method in the stimuli, the study consists of four sets of trials, where each set is targeted at being most advantageous for a single specific method, while still being possible to achieve with all methods. This ensures that one method does not have an

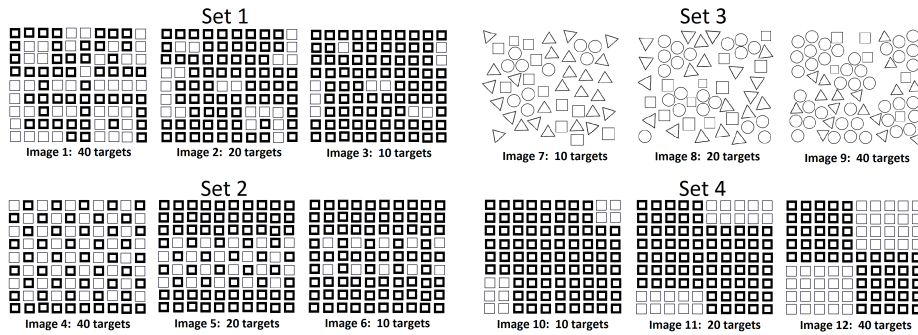


Fig. 7. Illustration of the stimuli used for the final user study trials. The trials consisted of four different images, with three complexity levels in each set (10, 20, and 40 targets). For the images in set 3, the users had to select all the circles. In the rest of the images, the users had to select all the targets with a thin black line and avoid selecting the incorrect targets (targets with a thick black line). Each set was created so that one selection method would have an advantage. Set 1 was targeted towards Brush, Set 2 towards Tap, Set 3 towards Lasso, and Set 4 was targeted towards Rectangle.

advantage throughout the entire study. For each set, three images with a similar number of targets were created. The amount of targets was set to 10, 20 and 40, to explore the scalability of all the methods. The four sets of trials can be seen in Figure 7, where the users had to select all the correct targets in the images and avoid selecting the incorrect targets. In trial set 1, 2, and 4 the correct targets are the squares with a thin black line and the incorrect targets are the ones with a thick black line. In set 3, the users had to select all the (groups of) circles and avoid all the other targets. This set was targeted to give an advantage to Lasso selection. The design of the trials is again based on the Gestalt principles of similarity and proximity.

Hypotheses For the model, we look at the same response variables as in the earlier studies (Section 3.1), i.e., actions used, time spent, and user errors (incorrect target selections). These are tracked in the same way as in the previous

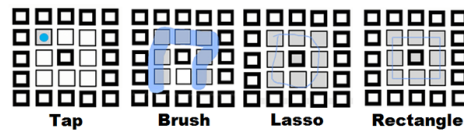


Fig. 8. Illustration of the selection methods used in the final user study. Tap and Brush select the targets that they directly interact with, whereas Rectangle and Lasso select all targets inside the selected region (including the center target). All methods can be utilized as a tap, when selecting single targets.

studies. The user ratings and comments given throughout the study are examined, to see if the qualitative and quantitative results are homogenous.

Within the stated hypotheses we expect Tap to use the highest amount of actions on average, as it is a single target selection method and therefore requires the user to select all targets individually. This is also supported by the results of the second user study. We expect that the negative effect of this increases as the number of targets grows. We also expect that Tap will be among the fastest for images with 10 targets. We expect Brush to be the fastest tool for images with a high amount of targets, due to the direct interaction. This reduces time, compared to the more abstract selection tools (i.e., Lasso and Rectangle) [4]. For selection errors, we expect Brush to make the highest amount of errors, as pilot testing shows that Brush is more prone to errors due to the thickness of the brush and the direct interaction, which can cause users to easily select incorrect targets. We do however expect that the amount of errors during the experiment is low in general, as in the previous studies. Lastly, we expect users to prefer Brush as it seems to be a fast and simple tool for the trials, whereas Rectangle seems to have a disadvantage from being inflexible in most trials.

Response variables The response variables used in the final study are the same as those in the two previous ones, namely: number of actions, completion time, and number of user-generated errors. These are also tracked in the same way as in the previous studies.

Experimental Procedure and Equipment The final user study used the same equipment and software as the previous studies. Sixteen participants (2 female) took part in the study. Ages were between 21 and 27, with an average of 24.19. All participants reported to be students and daily usage of touch devices.

The user study was designed as a factorial experiment, with interaction methods and image complexities as factors. The interaction method had four levels: Tap, Lasso, Rectangle, and Brush.

The images used for the trials had three levels of complexity, which in this case was defined as the amount of targets that needed to be selected. The complexity of the images was set to 10, 20, and 40 target selections, and the users had to make selections on four different trials, with three levels of complexity each, which meant 12 trials in total. A within subjects design was used, where each user would try all interaction methods on all the different complexity levels, for each trial. The order of presentation was determined via Latin Squares to reduce potential learning. Each method was furthermore evaluated through two conditions, where one was a timed condition and the other was a perfect completion condition. In the timed condition users had 25 seconds to select as many targets as possible and the amount of errors was tracked for later analysis. In the perfect completion condition, users had to select all targets in the image and correct all selection errors. A trial was completed when no selection errors were present, and all the correct targets had been selected. The four sets of three complexity levels used during the experiment are illustrated in Figure 7.

In the user study, each participant was first introduced to the project. Then a demographic questionnaire was filled, followed by an explanation of how the timed and perfect conditions would work. The participant was placed in front of the tablet and presented with the four selection methods, to ensure that s/he understood the interactions required to complete the trials. This was done through a short training session, where the users could practice the individual interaction methods. After the training session, the user would begin with the timed or perfect condition. This was counter-balanced so that half of the participants started with timed, and the other half with perfect. For each condition, the users would complete all trials, followed by the trials for the second condition. After a participant had completed both conditions with all interaction methods, we administered the final questionnaire and a semi-structured interview.

Results We performed a two-way ANOVA on both the actions and time data as dependent variables, with the selection methods and the two conditions as independent variables. We then used linear regression to examine the differences between the individual selection methods as a function of the amount of targets in the trials. For each response variable, we further analyzed the two conditions individually, to see the differences in results between the time condition and the perfect completion condition. For each of the hypotheses we modeled the time and number of actions with a linear equation, $y = TA * x + MP$, depending on the number of targets x . The constant term corresponds roughly to the mental preparation (abbreviated as MP) and the linear factor to the time/actions required to add a target (abbreviated as TA). For the qualitative measurements, we asked also how users went about the task.

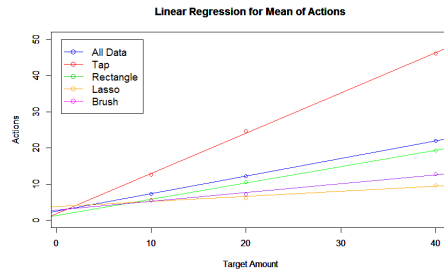


Fig. 9. Figure illustrating the mean distribution of the action data for the timed condition. Each line represents a selection method. The R^2 value for all data is $R^2 = 99.99\%$, for Tap: $R^2 = 99.82\%$, Rectangle: $R^2 = 99.89\%$, Lasso: $R^2 = 91.73\%$, and Brush: $R^2 = 97.21\%$. The graph for time looks similar, except that the conditions are not spread out as much.

Hypothesis 1 Hypothesis H1 states that there would be a significant difference in the number of actions used between the selection methods. The results for

both conditions are shown in Table 2. The actions data did not comply with the assumptions of the ANOVA. Instead ART (Aligned Rank Transform for non-parametric Factorial Analysis) was used [22]. The results show that there was a significant difference in the amount of actions used between the selection methods ($F_{3,45} = 20980.22$, $p \ll 0.001$). The Tukey-Kramer post-hoc test showed that there was a significant difference between all selection methods with Tap requiring the most actions followed by Rectangle, Brush, and Lasso in the given order. There was no significant difference in the actions used between the two conditions. Using linear regression we examine the differences between the selection methods for the individual conditions. For the timed condition, the test revealed that there was a significant difference between the methods for all data ($r^2 = 0.19$, $p \ll 0.001$). Tap requires the least actions for MP, but the most actions per TA. Conversely, Lasso required the most actions for MP, but the least actions per TA. The distribution of the action data for the time condition can also be seen in Figure 9, where each colored line represents a selection method. For the perfect condition, the test revealed that there was a significant difference between the methods ($r^2 = 0.23$, $p \ll 0.001$). The results show that Rectangle requires the fewest actions for MP, and Tap again has the most actions per TA. Lasso again requires the most actions for MP, but has the fewest actions per TA.

Methods	MP (timed)	TA (timed)	MP (perfect)	TA (perfect)
All Methods	2.44*	0.49*	1.97*	0.55*
Tap	1.84	1.11*	1.63	1.10*
Rectangle	1.30	0.45*	0.34	0.55*
Lasso	3.85*	0.14*	3.48*	0.22*
Brush	2.76*	0.25*	2.41	0.32*

Table 2. Table showing the coefficients ($y = TA * x + MP$) describing the average amount of actions. Both the timed and perfect condition are shown. * indicates significance, $p < 0.05$.

Hypothesis 2 The hypothesis H2 states that there would be a significant difference in the time usage between the selection methods. Table 3 shows the times. A two-way ANOVA was performed between the two conditions and the selection methods, after applying a logarithmic transformation to the data. The results show that there was a significant difference in the amount of time used between the selection methods ($F_{3,1531} = 17.70$, $p \ll 0.001$ and $1 - \beta \cong 1$, $\eta^2 = 0.28$). A Tukey-Kramer post-hoc test showed that Tap used significantly more time than the other methods. There was also a significant difference in time between the two conditions ($F_{1,1531} = 45.61$, $p \ll 0.001$ and $1 - \beta \cong 1$, $\eta^2 = 0.25$). A Tukey-Kramer post-hoc test showed that significantly more time was spent on the perfect condition. Using linear regression we examine the differences between the selection methods for the individual conditions. The results from the timed condition revealed that all methods had a significant difference in completion

time for the trials ($r^2 = 0.42$, $p \ll 0.001$). The results showed that Tap had the highest time per TA in a trial. Brush had the lowest MP time. Lasso had lowest time per TA, but it also the highest MP of the methods. For the perfect condition, the test revealed that there was a significant difference between the methods ($r^2 = 0.34$, $p \ll 0.001$). The results show that Tap had the highest time per TA. Lasso had the lowest time per TA, but also had the highest MP of all the methods. Rectangle had the lowest MP time.

Methods	MP (timed)	TA (timed)	MP (perfect)	TA (perfect)
All Methods	3.09*	0.34*	2.15*	0.51*
Tap	3.05*	0.48*	3.00*	0.56*
Rectangle	3.25*	0.34*	1.16	0.55*
Lasso	3.53*	0.26*	3.07*	0.43*
Brush	2.52*	0.28*	1.36	0.49*

Table 3. Table showing the coefficients ($y = TA * x + MP$) describing the average time in seconds. Both the timed and the perfect condition are shown. * indicates significance, $p < 0.05$.

Hypothesis 3 Hypothesis H3 deals with the number of selection errors that users made during the study. Users generally made few selection errors, with a few exceptions. To analyze the selection errors for the study we compute the sensitivity and specificity for each method. This was only done for the timed condition, as there were no errors in the perfect condition. Sensitivity is calculated by dividing the amount of correct selections with the amount of possible correct selections. Dividing the number of remaining incorrect targets with the possible incorrect targets yields specificity. We used a Test of Equal Proportions to calculate the sensitivity and specificity. The computed data can be seen in Table 4.

Method	Sensitivity (percent)	Specificity (percent)
Tap	97.88%	99.92%
Rectangle	98.73%	95.66%
Lasso	99.89%	95.39%
Brush	99.51%	96.33%

Table 4. Table illustrating the sensitivity and specificity data for the timed condition. The sensitivity and specificity is shown in percent.

Sensitivity is significant $\chi^2_3 = 107.7297$, $p \ll 0.001$. The data reveals that Tap was the worst and Lasso selection the best method in terms of sensitivity. For specificity the test showed a significantly different amount of selection errors that users made ($\chi^2_3 = 476.4897$, $p \ll 0.001$). Users made the most incorrect target selections with Lasso, and the fewest incorrect target selections with Tap.

Hypothesis 4 Hypothesis H4 states that there would be a significant difference in user preference between the methods. Based on a Friedman Ranked Sum test, we identified a significant difference between the user ratings of the selection methods during the study ($\chi^2_3 = 15.42$, $P < 0.001$). Brush and Lasso were rated highest, with no significant difference between the two. Rectangle and Tap received the lowest rating, with no significant difference between the two.

Qualitative Data In the questionnaire, participants were asked as to how they solved the trials. Nine participants (56.25%) stated that their goal was to complete the selections as fast as possible. Five participants (31.25%) had a combination between as fast as possible and with as few errors as possible. The last two participants (12.5%) tried to solve the trials with as few errors as possible without considering the time spent. From observations and comments from the participants we found that most participants wanted to be able to adjust the size of the tools during the completion of the trials. This was especially the case for Brush selection, as the implemented width was, according to several participants, too wide for the selections. Furthermore, some participants stated that Tap was annoying for the trials that required many target selections, but that it was a good selection tool for trials with fewer targets.

4 Discussion

In the first two user studies we initially expected that a lower number of actions would be necessary for the interaction techniques that supported selection of multiple targets with a single action (Lazy Selection and both pin and flick gestures). The results from both user studies confirm that Lazy Selection required fewer actions compared to the other methods.

In both user studies Tap was found to be the fastest interaction method, even when the amount of target selections was increased for the second user study. One potential explanation of this is that the Lazy Selection algorithm requires a bit more computation time than Tap and thus has a slight delay. More importantly, users often had to redo the interaction when selecting multiple targets with one action due to issues with gesture accuracy and/or failures of the Lazy Selection algorithm. These results further motivate the choice of using a simple Brush selection for the final user study. Our implementation of Brush was indistinguishable to Tap in terms of computation time.

In the final study we expected that a higher amount of actions would be necessary for Tap, as the other methods (Brush, Lasso, and Rectangle) were able to select multiple targets at the same time. Most users clearly understood that the tools were able to select more targets at once if applicable. The results from the final study confirm the expectation that for each added target (TA), Tap on average needed at least twice as many actions as the other methods. Yet, the constant, mental preparation (MP), cost of the task was lower for Tap than most of the other methods. These results were conclusive for both the timed and perfect completion conditions. This also corresponds to the comments from

the second user study, where some participants mentioned that it required more time to examine the path they wanted to take with the multiple selection tools, compared to Tap.

The results further show that in our creation of a model for the performance of the selection methods, Lasso was the most efficient tool in the amount of actions needed for a higher amount of targets. An example of this is for the timed condition, Rectangle is the most efficient tool in the amount of actions used until about seven targets, then Brush is the best for eight and nine targets, and Lasso is more efficient for trials from ten and above selections. However, given that these values are so close together, we can only state with authority that Lasso and Brush require the fewest actions for the highest number of targets. The Friedman Ranked Sum test for 40 targets confirmed these findings as Tap was ranked lowest, followed by Rectangle, and Lasso, and Brush. There was no significance between the latter ($\chi^2_3 = 33.525$, $p \ll 0.001$).

For the completion time, we initially expected Tap to be the fastest selection method for trials with few targets (as in our other studies), and that Brush would be the fastest for higher amounts of targets. However, these expectations were not met, as Tap was never the fastest selection method for both the timed and perfect conditions. For the timed condition, Brush was the fastest selection tool, which corresponds to the results found by Stenholt [1]. Based on some exploratory extrapolations, we believe that Lasso might eclipse Brush beyond 50 targets, but this is far from conclusive.

For errors, we calculated the sensitivity and specificity of the individual methods for the timed condition. While the amount of incorrect selections is low, there is an overall difference between the methods. The data shows that while users made the most correct selections with Lasso (within the given time frame), they also made the most incorrect selections with this method, more than with Brush. These results contradict the qualitative results, as users stated that the width of Brush was too wide which made it easier to inadvertently select targets compared to Lasso. For Tap the opposite was the case, as users had the fewest correct selections within the given time, and also the fewest incorrect selections. Observations during the user study confirm this, as most users tended to have issues selecting all targets within the given time frame for Tap, but made few selection errors. These results indicate that even though Tap might be a slow selection method (especially for trials with many targets), the direct interaction of the method makes it easier to be precise.

The expectation in terms of user preference was that, due to its simple nature, users would prefer Brush, as it might be the best method for most situations. However, this was only partially true, as Lasso was rated the highest, but with no significant difference between the two. For the final study, Tap was rated lowest, but with no significant difference between it and Rectangle. Inspecting the qualitative data reveals that some users tended to prefer Lasso, as it could be used as a thin version of Brush. Enabling users to adjust the width of Brush might be beneficial, but the error results indicate that this is not the case. The

qualitative results also revealed that most users chose to complete the trials as fast as possible without committing too many selection errors at the same time.

The implementation of the selection methods allowed them to be utilized as a tap for selecting single targets. The observations during the user studies showed that most users tended to use all the methods as Tap in the trials where the targets were not clustered (mainly seen in set 2 of the final user study). This further shows that Tap is simply unbeatable for a small amount of targets or where the targets are spread out over the entire image. Some users also stated this during the study.

5 Conclusion and Future Work

This work initially discussed several alternative approaches for adjusting a selection for the smart selection algorithm, Lazy Selection. The main goal was to identify if the adjustment of the set of selected targets could be improved through touch-based interactions. Three interaction techniques were designed for this task; a single touch tap (Tap), and a unimanual and bimanual pin and flick gesture. The results of two user studies showed that only Tap had an advantage compared to Lazy Selection in terms of time, while Lazy Selection required fewer actions on average. The advantage of Tap was especially relevant in trials with a low amount of target selections. This motivated a final user study to examine at which level other techniques become superior in efficiency.

The main purpose of the final user study was to build a predictive model around different existing selection methods. In the study Lazy Selection was replaced with simple brush selection and compared against lasso, rectangle and tapping. We investigated the number of actions and time needed to select targets in a timed, and perfect condition. The contribution of the final user study:

- A predictive model for the amount of targets and time needed to select a set of targets.
- Users preferred the lasso and brush selection methods.
- Tap has poor scalability in regards to the amount of targets.
- The lasso and brush selection methods improves in efficiency as the target amount increases.
- Participants prefer Tap for trials with a low amount of targets.

6 Acknowledgments

Many thanks to the authors of Lazy Selection for making their implementation available to us. Thanks to all those who participated in the various studies in the project. Some of the work was performed at York University. We acknowledge funding from NSERC.

References

1. Stenholt, R.: Efficient Selection of Multiple Objects on a Large Scale. *VRST'12* (2012) 105–112
2. Xu, P., Fu, H., Au, O.C., Tai, C.L.: Lazy Selection: A Scribble-based Tool for Smart Shape Elements Selection. *SIGGRAPH'12* **31**(6) (2012) 51–60
3. Forlines, C., Wigdor, D., Shen, C., Balakrishnan, R.: Direct-Touch vs. Mouse Input for Tabletop Displays. *CHI'07* (2007) 647–656
4. Wobbrock, J.O., Morris, M.R., Wilson, A.D.: User-Defined Gestures for Surface Computing. *CHI'09* (2009) 1083–1092
5. Wills, G.J.: 524.288 Ways to Say “This is Interesting”. *Information Visualization'96* (1996) 54–60
6. Stenholt, R., Madsen, C.B.: Poster: Brush, Lasso, or Magic Wand? Picking the Right Tool for Large-Scale Multiple Object Selection Tasks. *3DUI'12* **1** (2012) 163–164
7. Kimball, S., Mattis, P.: *GNU Image Manipulation Program*. (2014)
8. James, D.: *Crafting Digital Media - Audacity, Blender, Drupal, GIMP, Scribus, and other Open Source Tools*. Springer (2009)
9. Kin, K., Agrawala, M., DeRose, T.: Determining the Benefits of Direct-Touch, Bimanual, and Multifinger Input on a Multitouch Workstation. In *Proc. Graphics Interface* (2009) 119–124
10. Wertheimer, M.: Untersuchungen zur Lehre von der Gestalt. *Psychologische Forschung* **4** (1923)
11. Desolneux, A., Moisan, L., Morel, J.M.: *From Gestalt Theory to Image Analysis: A Probabilistic Approach*. Volume 34. Springer (2008)
12. Chang, D., Nesbitt, K.V., Wilkins, K.: The Gestalt Principles of Similarity and Proximity Apply to Both the Haptic and Visual Grouping Elements. *AUIC'07* **64** (2007) 79–86
13. Gallace, A., Spence, C.: To What Extent Do Gestalt Grouping Principles Influence Tactile Perception? *Psychological Bulletin* **137**(4) (2011) 538–561
14. Todorovic, D.: Gestalt Principles. *Cognitive Science Conference94* **3**(12) (2008) 5345
15. Rome, E.: Simulating Perceptual Clustering by Gestalt Principles. *25th Workshop of the Austrian Association for Pattern Recognition, OAGM/AAPR* (2001) 191–198
16. Thórisson, K.R.: Simulated Perceptual Grouping: An Application to Human-Computer Interaction. *Proceedings of the Sixteenth Annual Conference of the Sognitive Science Society* (1994) 876–881
17. Dehmeshki, H., Stuerzlinger, W.: Design and Evaluating of a Perceptual-Based Object Group Selection Technique. *BCS'10* (2010) 365–373
18. Au, O.K.C., Tai, C.L., Fu, H.: Multitouch Gestures for Constrained Transformation of 3D Objects. *Computer Graphics Forum* **31**(2) (2012)
19. Wigdor, D., Wixon, D.: *Brave NUI World: Designing Natural User Interfaces for Touch and Gesture*. Morgan Kaufmann (2011)
20. Mizobuchi, S., Yasumura, M.: Tapping vs. Circling Selections on Pen-Based Devices: Evidence for Different Performance-Shaping Factors. *CHI'04* (2004) 607–614
21. Wagner, J., Hout, S., Mackay, W.E.: BiTouch and BiPad: Designing Bimanual Interaction for Hand-held Tablets. *CHI'12* **1** (2012) 2317–2326
22. Wobbrock, J.O., Findlater, L., Gergle, D., Higgins, J.J.: The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only ANOVA Procedures. *CHI'11* (2011) 143–146