How Can Adding a Movement Improve Target Acquisition Efficacy?

Alexander R. Payne, Beryl Plimmer, Andrew Mcdaid, Andrew Luxton-Reilly, T. Claire Davies

How can Adding a Movement Improve Target Acquisition Efficacy?

Alexander R. Payne[1*], Beryl Plimmer[1], Andrew McDaid[1], Andrew Luxton-Reilly[1], T. Claire Davies[2]

[1]University of Auckland, Auckland, New Zealand
apay876@aucklanduni.ac.nz, {b.plimmer, andrew.mcdaid, a.luxton-reilly}@auckland.ac.nz

[2]Queen's University, Kingston, Canada
claire.davies@queensu.ca

Abstract. People with motor impairments, such as cerebral palsy (CP), have difficulty acquiring small targets with a mouse. To improve upon this many assistive technologies enlarge targets, generally by introducing an extra movement. Often this improves accuracy however there appears to be a time penalty of performing all aspects of a movement twice. We investigate if it is possible for an extra movement to improve efficacy, and if this time penalty can be counterbalanced by a reduction of errors and corrective movements.

We measure overall interaction efficacy in a controlled 1D experiment. Participants acquire targets under three conditions: a single movement, a double movement, and using an example assistive tool. We anticipate that a double movement may only increase efficacy when the single movement target is so small that corrective movements are disproportionately time consuming. Therefore we investigated the effects of task scale, and of motor control. The results show that it is possible for two movements to be more efficient than one. However, this appears to be an edge case that only occurs at a very small scale. We suggest that tool designers must focus on how and why their tool is going to be attractive to users, since in real world situations it is unlikely to improve pointing efficacy. Users may choose to use a tool because it improves accuracy, or requires less effort, but they are unlikely to use it because it is faster.

Keywords: Cerebral palsy · Motor impairment · Breadth · Depth · Fitts's Law · Mouse pointing · Target size

1 Introduction

Motor impairments make acquiring small targets with a mouse more difficult. For all users, with or without motor impairments, acquiring targets that are too small for their motor ability is slower than Fitts's Law predicts due to motor and visual difficulty [1]. So, motor impairments such as cerebral palsy (CP), or those resulting from aging, can result in slow target acquisition that is not well modelled by Fitts's Law [2-4].

Accessibility tools often make target acquisition easier by replacing a single input movement with two or more inputs. For example: Motor-Magnifier, Visual-Motor-Magnifier, Click-and-Cross, Cross-and-Cross, Ballistic Square, Scanning Area Cursor (all [5]), Expansion Cursor [6], Adaptive Click-and-Cross [7], and Pointing Magnifier [8]. In all these tools, the purpose of the extra movement is to allow for increased target sizes thus avoiding the requirement for fine pointing. The difference between these tools and simply making all targets larger is they do not rely on unused screen space being available. The cost of these tools is that the user has to make extra inputs.

Despite generally improving accuracy, these tools often result in slower interaction. This paper asks what is the potential for two movements to be more efficient than one when a fair comparison is made? When or how can more movements be faster, and is there potential to improve throughput based on improved accuracy?

The benefit of two movements is the reduced pointing difficulty for each movement, so we investigate the potential improvements from a purely pointing perspective. Therefore, our experiment (Fig. 1) aims to avoid other effects such as interaction techniques, aesthetics, and navigation.
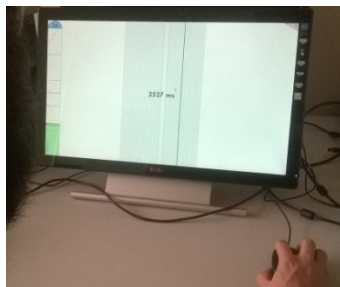

Fig. 1. Task Equipment. The task focuses on pointing and not other aspects of a realistic interface, such as aesthetics and navigation.

As a side note, previous work has suggested that menu navigation is slower with increased depth of menus [9-11]. So, even if it is possible for pointing to be faster with an assistive technology that adds a movement, navigation is a further design obstacle that must be considered.

Our experiment aims to uncover the fundamental potential of assistive pointing techniques that replace a single movement with two or more. Hence one and two movements are compared in as fair a scenario as possible. This does not aim to be realistic (Fig. 2 shows screen shots of the task), but focuses on *if* and *how* pointing can be more efficient with additional movements. As an example assistive technique, an Expansion Cursor will be compared to single and double movements (Fig. 3).

Of the example assistive tools mentioned, only two resulted in faster interaction and that was for a very specific situation. Both the Visual-Motor-Magnifier and the Click-and-Cross [5] were quicker than a standard cursor for people with motor impairments pointing at 4 pixel targets. For 16 pixel targets they were both slower, and for people without motor impairments they were slower for all target sizes.
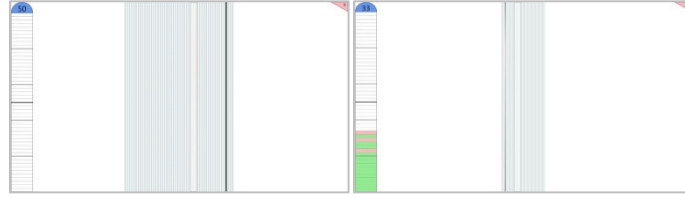
Fig. 2. Screen layout for our 1D clicking task. Left shows *Larger* scale artificial screen, right shows *Smaller* scale artificial screen. This is the *Single* movement task.
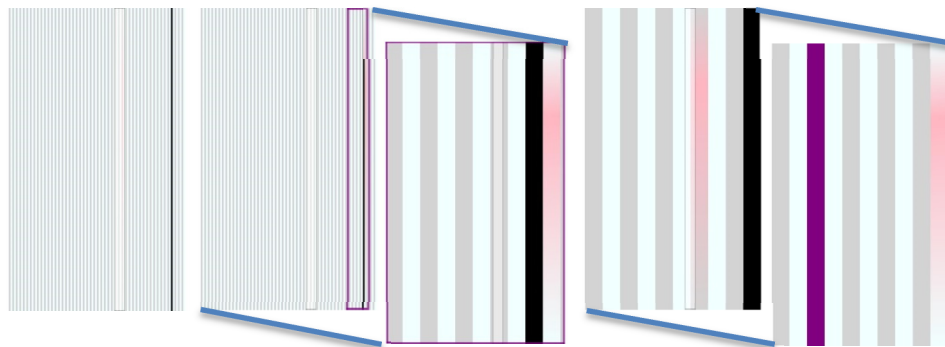
Fig. 3. Left: *Single* movement trials. Centre: *Expansion* trials where columns inside the purple box are expanded when the mouse button is held down. Right: *Double* movement trials. For all trials the black column is the target and the semi-transparent column is the starts zone. In *Double* trials the purple column is the second target. The pink highlight shows cursor location.

We hypothesize that adding a movement has the most potential to improve efficacy when corrective movements are problematic. In the case where Visual-Motor-Magnifier and Click-and-Cross [5] were faster than standard pointing, participants made significantly fewer sub-movements using the tools despite an addition input movement being required. Previous work also suggests small and difficult targets take disproportionately long to acquire according to Fitts's Law [1]. Adding a movement may avoid the problem of corrective movements since the targets can be larger.

We examine the effects of reduced motor ability and small scale movements since they are likely to result in increased time spent making corrective movements.

To investigate motor ability the experiment involved participants with and without CP. We hypothesize that increasing target size via an extra movement will be more useful to participants with CP since CP has been shown to result in increased number of sub-movements [12-14]. It seems reasonable that the increased time spent making corrective movements should diminish for larger targets.

Considering task scale is a useful way of highlighting the issue of small targets and corrective movements. According to Fitts's Law, scale should not affect acquisition times, therefore any changes observed are precisely due to the nature of the target. Since it is not clear if participants with CP would adhere to Fitts's Law anyway, we only investigate the scale effect for participants without CP. We hypothesize that adding a movement may be more beneficial for smaller scale trials.

2 Background Theory & Task Rationale

We explore the potential of two movements to communicate more efficiently than one movement from a pointing perspective. This should highlight the potential of assistive tools to improve interaction efficacy. As a third condition, we also included an assistive tool that makes use of adding a movement. For this we used a modified version of the Expansion Cursor [6]. (Fig. 3 shows trial types).

Our experiment aims to *fairly* compare the three trial types (*Single*, *Expansion, Double*). We measure the time it takes to communicate a set amount of information, using a set amount of screen space. Also in the interests of fairness, we aim to generalize for all possible target locations, and all possible cursor start locations. The fairness of our experiment set-up is based on two key theories: information theory [15], and Fitts's Law [16].

2.1 Information Theory and Fitts's Law

Information theory suggests the amount of information being communicated, H, is dependent on the number of possible outcomes, n, and the probability of each outcome occurring, p_i. This is defined by the Shannon-Weiner [15] measure of information (Eq. 1). Information is maximized if the probability of each outcome occurring is equal [15], and the equation reduces to Eq. 2.

$$H = \sum_{i=1}^{n} p_i \log_2 \left(\frac{1}{p_i} \right) \tag{1}$$

$$H = \log_2(n) \tag{2}$$

Hence, the number of possible alternatives directly informs the amount of information being communicated. In terms of the experiment task, the number of possible outcomes is the number of potential targets. To calculate the efficacy a screen layout, we must know the amount of information being communicated. Therefore the probability of selecting each target location should be equal.

Fitts's Law [16] relates movement time, MT, to movement amplitude, A, and target width, W. The key implication is that smaller movements to larger targets should be the fastest. A current iteration of this equation is Eq. 3 [17]. The MT also depends on the participant via constants a and b.

$$MT = a + b \times \log_2 \left(\frac{A}{W} + 1 \right) \tag{3}$$

Fitts's original experiment was based on a tapping task between 1D targets, and has been widely applied to HCI and 2D targets [18]. Similar to Chapuis & Dragicevic's experiment [1], our experiment reflects Fitts's task [16] by using 1D targets. For reasons further explained in 2.3, this reduces the complexity of controlling for possible cursor start locations and maintaining equal probability for each outcome.

Our experiment is concerned with communication rate; the time required (Fitts's Law) to communicate a set amount of information (information theory). The task

screen layout (Fig. 1, Fig. 2, Fig. 3) is intended to be as efficient as possible for a set screen size. It is based on the assumption that communication rate is maximized for a screen full of targets (number of possible outcomes is maximized and target width is maximized) that are equally likely to be selected (so all target positions are tested).

2.2 Scale, Motor Skills, and Magnification Level

These theories also informed our experiment set-up in terms of expected factors (scale, motor skills, and magnification level) and making fair comparisons.

Scale relates to the size of the screen and the targets. If the screen is larger, targets will become larger as well as the distance between them. Fitts's law suggests scale is not important as long as the amplitude to width ratio remains constant. However, this does not necessarily apply when small targets are involved. There is a small target effect where acquiring smaller targets takes longer than larger targets for the same A/W ratio [1]. We expect this effect to be important since single movements to smaller targets may encounter it, whereas double movements potentially avoid it.

A person's motor skills may also affect the potential benefits of adding a movement. The small target effect is due to motor effects and the visibility of targets [1]. Therefore, a person's motor skills are likely to interact with this effect. We investigate this by comparing participants with and without CP. Cerebral palsy is a common physical disability resulting from damage to the brain at, or around, birth [19]. (See [20] for a modern definition). We expect adding a movement to be more beneficial for people with motor impairments since fine pointing can be avoided.

Aside from scale and motor skills, magnification level may affect the potential of adding a movement. We define magnification level as the ratio of target sizes between *Double* and *Single* conditions. If target width is important for speed, it seems plausible that increasing the magnification level could further benefit an additional movement. Previous research of accessibility tools used magnification levels of 3 [6] and 4 [5, 7, 21]. For this experiment we chose a fixed magnification level that was as high as practically possible, at a value of 10. The rationale being that a higher magnification level should allow the best chance of larger targets (for *Double* trials) to avoid the small target effect, whilst the smaller targets (for *Single* trials) may still suffer from it.

2.3 Controlling for Target Locations and Cursor Start Position

To generalize the efficacy of a screen layout, we must assume that each target could be selected, and that the cursor could start from any position. Practically it is not possible to perform trials to every possible target location for every possible cursor location. Hence, we must control for target position and cursor start location.

The easiest way to control for target location is to perform trials to every target, but this requires performing too many trials. For a 1D target layout, position can be defined by distance from the center of the screen. Thus, we can halve the number of trials by performing trials to every second target. This also maintains equal probability of each target position being selected so we can calculate the amount of information being communicated.

We aim to generalize the results for the cursor starting from any location on the screen. Since this affects the movement amplitude, it must be controlled for. Therefore, each target location requires an actual cursor start location that can account for all possible start locations. For each target position we calculate an expected average movement time and then infer a representative movement amplitude. These calculations can be found in the supplementary material. Note that these calculations do not depend on a and b values, and are greatly simplified by making the task 1D instead of 2D. For 2D targets, direction would have an effect on target width [22].

3 Methodology

The experiment involved 2 participant groups (W/o CP, CP), 2 scales (*Larger*, *Smaller*), and 3 trial types (*Single*, *Expansion, Double*). Our key measure reflecting communication rate was time per information (Time/Info). Other performance measures were task completion time, accuracy, reaction time, traversal time, corrective time, and click time. We also considered subjective measures of user preference and workload to examine how user perceptions relate to the primary performance measures.

3.1 Participants

There were 15 participants; 10 without cerebral palsy (group name W/o CP) and 5 with CP (group name CP). Ethical approval was granted by our university and all participants (and parents for the two participants under 16) signed a consent form. These populations were recruited to establish whether motor skills interact with the potential benefits of adding a movement. Previous work suggests the latter group should take longer to complete pointing movements [2], whilst still being regular mouse users. All participants had normal or corrected to normal vision. The entire group W/o CP (37.1 years ± 11.2) used their preferred right hand.

Table 1. Participant information for participants with CP. MACS: Manual ability classification system [23].

Participant	Hand Used	Age	MACS Level
1	Right	34	II
2	Left	64	I
3	Right	13	I
4	Right	15	I
5	Left	38	I

Part of the definition of CP is that it causes activity limitation [20]. The manual ability classification system (MACS) is a validated method of qualifying the effects of CP on people's hand activities [23]. We sought participants who regularly use a computer mouse, and were therefore from levels I & II. Further (self-reported) demographics of the group with CP can be seen in Table 1.

3.2 Test Conditions

Participants sat at a desk and performed a target acquisition task. A Dell S2240T 21.5" Monitor (resolution 1920x1080) plugged into a Lenovo T520 laptop computer (Windows 7 Professional SP1) was used with a Dell optical mouse (Fig. 1). The pointer speed was the second fastest option, enhance pointer precision was enabled.

All tasks involved acquiring 1D targets. The screen layout (Fig. 2) had an artificial screen containing all the target columns, and a progress bar showing how many trials had been completed. The artificial screen width was 500 or 200 pixels depending on the scale being tested (*Larger* or *Smaller*). Both participant groups performed the *Larger* scale trials. Only the group W/o CP performed the *Smaller* scale trials because the two pixel targets are too small for participants with CP to reliably acquire. In fact they were chosen to be small for people W/o CP to engender the small target effect between *Single* and *Double* trials.

The artificial screen area differed from an actual screen since the cursor could leave the area during trials; edges did not constrain movement. The corresponding target widths are shown in Table 2. The scales were also chosen since the Expansion Cursor [6] study suggested most participants with CP will prefer the *Double* movement compared to 5 pixel targets (*Larger* scale), and most participants W/o CP will prefer an extra movement for 2 pixel targets (*Smaller* scale).

Table 2. Target widths (pixels) for different conditions. For *Expansion* trials, users could enlarge targets from the *Single* width (5 or 2) to the *Double* width (50 or 20) with a zoom.

Scale	Single	Expansion	Double	Groups
Larger	5	5 => 50	50	W/o CP, CP
Smaller	2	2 => 20	20	W/o CP

The initial movement of every trial was from a semi-transparent start zone (30 pixels wide) to a target column. The location of the start zone varied depending on the target location and the desired movement amplitude. For each target, the center-to-center amplitude of the movement was determined based on the target width and location. Using Fitts's Law, a theoretical average movement time to a specific target was calculated for the cursor starting from all possible locations (within the artificial screen). This average time was used to calculate a representative movement amplitude which was then used in the actual experiment. See supplementary material for the maths used. As mentioned, this does not depend on a and b values from Fitts's Law.

This experiment was intended to test communication via motor input, and visual acuity was not supposed to be a limiting factor. Target columns did not have a border, but were alternating colors in order to make them visually distinguishable. Furthermore, a semi-transparent pink highlight was applied to the column of the current cursor location. This ensured a high contrast change occurred when the pink highlight was over the black target column. It was easy to see if the cursor was over the target.

For all trial types, targets were acquired by releasing the mouse button over them. Videos of trials being performed can be viewed in the supplementary material.

Single.

For *Single* movement trials, the artificial screen was divided into 100 target columns (Fig. 3, left). As such, the target width was either 5 or 2 pixels depending on the scale (see Table 2). The current target column was black.

Double.

For *Double* movement trials, the artificial screen was divided into 10 target columns so the target width was either 50 or 20 pixels (Fig. 3, right). The initial target column was black. Once it was acquired, it would return to its original color and a second target column would appear purple. These columns had to be distinguishable since in some instances, the second target column would be the same column as the initial target. Once the second target was acquired, the trial would finish.

Expansion.

For the *Expansion* trials, there were 100 target columns, the same as for *Single* trials. However, when the mouse button was depressed the columns near the cursor would expand to be 10 times larger (Fig. 3, center). The expanded targets were the size of the *Double* movement targets. As with the other trials, the target was acquired via releasing the mouse button. This meant that participants could choose to zoom in on the target and correct their movement to make sure they were acquiring the right target. Depression of the mouse button initiated the zoom with a focal point on the column in which the mouse was positioned. Participants could also choose to ignore the zoom, and just perform *Expansion* trials in exactly the same way as *Single* trials.

To make it clearer which target columns would be expanded, a purple box followed the cursor (Fig. 3, center). All the columns inside the purple box were expanded once the zoom was activated. When the zoom was active, the zoom area could not be moved. If the target was not inside the purple box after zooming, it could not be selected and the trial counted as a miss.

3.3 Procedure

In a test session, participants performed blocks of 25 trials twice for the three trial types at one scale. Therefore each test session involved 150 trials (25×2×3). The order of trial types was counterbalanced within a test session in the order ABCCBA. Ordering of trial types was also counterbalanced between participants. This should reduce confounding effects of learning. Also, at the start of each block of 25 trials participants had the opportunity to practice that trial type until ready to start.

For each trial type, 50 of the 100 possible target location/combinations were used across a test session. For the *Single* and *Expansion* trials, every second target location was used so targets appeared at every possible distance from the screen center. For *Double* trials, equivalent target combinations were used (i.e. every initial target position, and every second final target position). The direction of movements was also controlled: there were equal numbers of left and right movements. To discourage anticipatory movements, the movement direction appeared random to participants.

Each test session was conducted at a fixed scale. The test sessions for the group with CP only examined the *Larger* (500 pixel) scale. The group W/o CP performed the two scales, each one in a separate test session. Between the participants W/o CP, ordering of test session scales was counterbalanced.

Participants initiated individual trials by clicking the start zone (30 pixels wide). After a randomized wait period (1–2s at 0.1s intervals) the target column turned black. Experimental design ensured that the timing and location of the target column being designated was not predictable. To prevent anticipation and to control movement amplitudes, the target would not appear if the cursor was outside the start zone. If the cursor left the start zone for an extended period (0.5s) the trial was aborted and the participant had to reinitiate the trial by clicking the start zone again. Participants were instructed to "select the target as quickly and accurately as you can."

Once the target was acquired, participants would receive speed and accuracy feedback. After a successful trial, the target column and a box surrounding the artificial screen would turn green. This would be recorded in the progress bar (see Fig. 2). If the trial was inaccurate, the feedback was red. After a trial, the participant's time was also displayed. This allowed a sense of speed without being judgmental. Also, participants were unaware of the performance of other participants.

Once a trial type was completed, participants reported workload measures via a NASA TLX based survey [24]. The six Likert scale questions asked about three task properties (Mental Demand, Physical Demand, Time Demand) and three personal experience properties (Your Performance, Effort, Frustration). As per NASA TLX methodology, participants gave a score out of 20. Once a test session was completed, participants were also asked for their most and least preferred trial types.

3.4 Data Processing

The data of the 3,750 trials was recorded in Excel workbooks. Trials were discarded when the participant was distracted (29), confused about the task (10), activated the zoom near the start zone (20), or when they did not notice the second target in the *Double* task (10). The distracted and confused trials were, for example, when the participant was looking away from the screen, or when they did not know where the target was. These were discarded since they did not reflect the user's pointing performance. Trials were discarded based on observation during experimentation in conjunction with viewing the recorded data in Excel.

For each of accuracy, time per information, preference, and workload, one value was calculated for each trial type of a test session (i.e. one value for 50 trials). Accuracy was defined as the percentage of successful trials.

For time per information values, time is the total time of all trials, successful or not (but excluding discarded trials). Information was defined as the total number of successful trials multiplied by inherent information, $\log_2(100)$, in a single trial. For *Expansion* trials, zooming in on the wrong location pushed the target out of the screen unexpectedly. In this case, the time recorded was not an accurate reflection of time spent pointing. In calculating time per information, the recorded time for any 'push

out' trial was replaced with an average time for that test block. However, if the push out occurred because of clicking near the start zone (i.e. not near target at all), then the trial was discarded. Also, if the push out click was quick (i.e. within the normal range of durations of *Single* trial clicks) and the zoom was not used, the time was not replaced since it did reflect time spent pointing.

Trial types were assigned a preference rank from first (1) to third (3) based on participants' responses to their most preferred and least preferred interactions.

The time measures (completion times, reaction times, traversal times, correction times, and click times) were calculated for each individual trial. Completion times were the time from when the target appeared to when the trial was completed. The four other times were components of the completion time. The reaction time was from when the target appeared to when the cursor had moved 10% of the direct distance to its final location. The traversal time was from when the cursor had moved 10% of the distance to 90% of the distance. The correction times were from 90% of the distance to when the mouse button was depressed. Lastly, the click time was from when the button was depressed to when it was released and the target was selected. For *Double* trials, these component times were the cumulative values of the two movements. For the *Expansion* trials, the component times could not be calculated since the distinction between correction time and click time is blurred.

3.5   Statistical Analysis

Statistical analysis was performed using IBM's SPSS Statistics 22. The only independent variable directly analyzed was *Trial Type*.

For accuracy, time per information, preference, and workload measures, non-parametric Friedman tests were used to test for overall significance. For pairwise comparisons, Wilcoxon signed-rank tests were used along with Bonferroni correction. Since there were 3 comparison pairings, significance was set at $p<0.017$. Reported p-values are unadjusted.

A repeated measures mixed model was used for time values. *Trial Type* and *Trial#* were used as repeated variables. *Trial Type* was the only fixed factor, whilst *Participant* was a random variable. The data set were natural log transformed to improve kurtosis and skewness except for the traversal times. Outliers that were 3.29 standard deviations outside the mean were deleted.

4   Results

The following sections discuss the results. Mean completion times and standard error values are shown in Table 3 and Fig. 4, along with medians and quartiles for Time/Info and accuracy. Fig. 5 shows means and standard error values of times for the different components of the task. Note that in the terms *Smaller* and *Larger* refer to test sessions by the group W/o CP. The group with CP only used larger targets.

4.1 Time per Information (Time/Info)

Trial Type had a significant effect on Time/Info for the group W/o CP for both *Larger* ($\chi^2(2) = 9.600$, p=0.008) and *Smaller* scales ($\chi^2(2) = 16.800$, p<0.001), but it did not reach significance for the group with CP ($\chi^2(2) = 5.200$, p=0.074).

In pairwise comparisons for *Larger* scale trials (the group W/o CP), the difference between *Single* and *Expansion* trials was closest to significance (Z = -2.191, p = 0.028). For 9 of the 10 participants, *Single* trials elicited a lower Time/Info value. The difference between *Single* and *Double* trials did not reach significance (Z = -1.886, p = 0.059). The medians were almost identical, although for 9 of the 10 participants *Single* trials elicited a lower Time/Info value than *Double* trials.

For *Smaller* scale, *Double* trials had significantly less Time/Info than both *Single* (Z = -2.803, p = 0.005) and *Expansion* (Z = -2.803, p = 0.005). In both cases for all 10 participants, the *Double* trials elicited lower Time/Info. *Single* trials did not result in significantly less efficacy compared to *Expansion* trials ($Z = -2.191$, $p = 0.028$).

4.2 Completion Times & Components

For all groups, *Trial Type* significantly affected completion times (W/o CP *Larger*: F(2,978.261)=54.782, p<0.001; W/o CP *Smaller*: F(2,982.607)=91.763, p<0.001; CP: (F(2,481.980)=18.311, p<0.001). For W/o CP *Larger*, *Single* trials were performed significantly faster than both the *Double* trials (p<0.001) and the *Expansion* trials (p<0.001). Contrary to the *Larger* scale results, for the *Smaller* scale the *Single* trials were performed significantly slower than both the *Double* trials (p<0.001) and the *Expansion* trials (p<0.001). *Double* trials were also significantly faster than *Expansion* trials (p<0.001) for this scale.

For the group with CP, the *Single* trials were performed significantly faster than both the *Double* trials (p=0.012) and the *Expansion* trials (p<0.001). This is similar to

Table 3. Median values (with quartiles) for Time/Info and accuracy, and mean completion times (with standard error) across different groups and trial types. * denotes that *Trial Type* was a significant main effect of the relevant measure for that participant group and scale.

	Single	**Expansion**	**Double**
W/o CP – Larger			
Time/Info (ms/bit)*	269 (234, 298)	308 (257, 330)	268 (257, 313)
Completion Time (ms)*	1649 (56)	1845 (63)	1833 (62)
Accuracy (%)	96.0 (93.5, 98.5)	97.0 (92.2, 100)	100 (96.9, 100)
W/o CP – Smaller			
Time/Info (ms/bit)*	422 (393, 500)	346 (317, 425)	292 (269, 333)
Completion Time (ms)*	2448 (94)	2164 (83)	1911 (74)
Accuracy (%)*	93.9 (87.9, 94.0)	96.0 (87.7, 98.5)	98.0 (96.0, 100)
CP			
Time/Info (ms/bit)	420 (356, 667)	448 (338, 935)	364 (313, 545)
Completion Time (ms)*	2283 (429)	2656 (499)	2453 (461)
Accuracy (%)*	77.6 (69.5, 89.9)	83.3 (82.6, 92.9)	100 (86.0, 100)
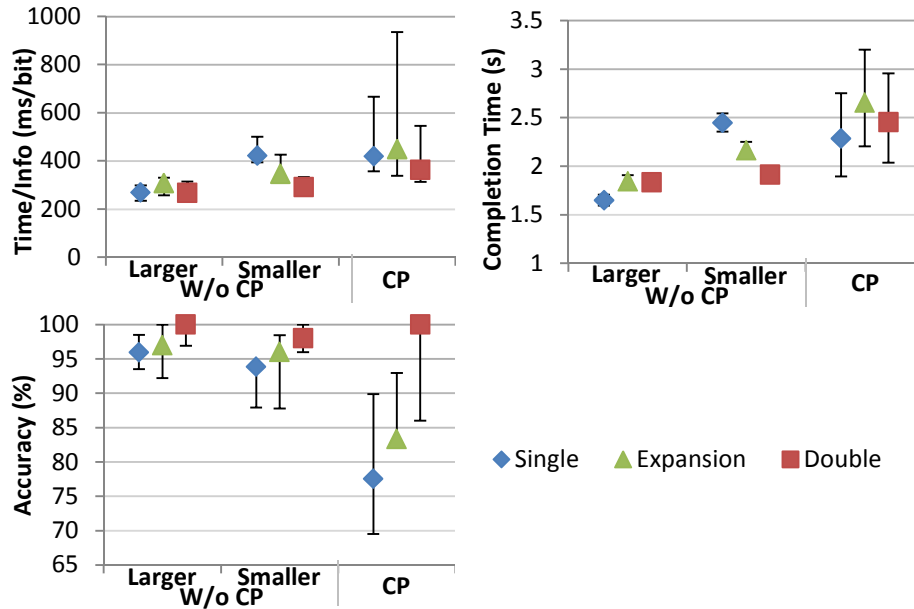
Fig. 4. Time/Info, task completion times, and accuracy across different groups and trial types. Mean completion times are shown with error bars for standard error (since this is based on analysis of variance). Other measures show medians with error bars for the 25th and 75th percentiles (since analysis was based on non-parametric Friedman tests).

the *Larger* scale results for the group W/o CP. However, *Double* trials were significantly faster than *Expansion* trials (p=0.005), akin to the *Smaller* scale results.

For each test session type (W/o CP *Larger*, W/o CP *Smaller*, CP) all component times were significantly different (all p<0.001) between *Single* and *Double* trials. The durations of reaction times, traversal times, and click times were all larger for *Double* trials. Conversely, corrective times were always shorter for *Double* trials even though two targets had to be acquired. One apparent interaction is that the difference in correction times between *Single* and *Double* trials is the largest for the *Smaller* scale. This is also the only case where Time/Info significantly reduced for *Double* trials.
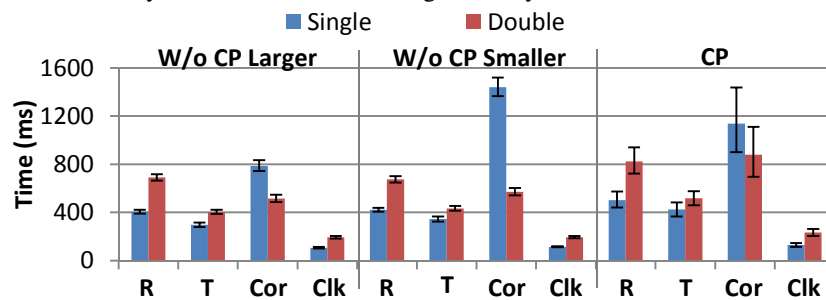
Fig. 5. Graphs of mean component times (ms) with standard error bars of *Single* and *Double* trials. R: Reaction time, T: Traversal time, Cor: Correction time, Clk: Click time.

4.3 Accuracy

Trial Type had a significant effect on accuracy for the W/o CP *Smaller* trials ($\chi^2(2) = 6.821$, p = 0.033) and the group with CP ($\chi^2(2) = 6.400$, p = 0.041).

For W/o CP *Smaller*, there was a significant difference between *Double* and *Single* trials (Z = -2.550, p = 0.011). During *Double* trials 9 of the 10 participants were more accurate than during *Single* trials.

For the group with CP, the difference between *Double* and *Single* trials was closest to significance (Z = -2.023, p = 0.043). Despite not reaching significance, all 5 participants were more accurate for *Double* trials than *Single* trials.

4.4 Preferences

For W/o CP *Smaller* scale, *Trial Type* had a significant effect on preference rank-ranking, ($\chi^2(2) = 6.200$, $p = 0.045$). *Double* trials were significantly preferred over *Single* ones (Z = -2.495, $p = 0.013$). *Double* was ranked higher than *Single* by 9 participants, and lower by the other 1. The actual preference responses can be seen in Table 4.

As an observation for both *Larger* and *Smaller* scale trials, half the participants W/o CP preferred *Double* trials the most and almost none disliked it. When the scale got smaller, it seems votes went from *Single* towards *Expansion*. The data of participants with CP does not reveal any preferences in particular.

Table 4. Votes for most preferred and least preferred trial types. * denotes that *Trial Type* was a significant main effect that participant group and scale.

|	Preferred (+Most/-Least)		
	Single	**Expansion**	**Double**
W/o CP – Larger	+4/-4	+1/-5	+5/-1
W/o CP – Smaller*	+1/-7	+4/-3	+5/-0
CP	+1/-2	+2/-1	+2/-2

4.5 Workload Measures (NASA TLX)

None of the task properties (Mental Demand, Physical Demand, Time Demand) var-varied significantly with *Trial Type* for any test session. The main workload measure that varied with *Trial Type* was Frustration (Fig. 6). This was significant for the group W/o CP at both *Larger* ($\chi^2(2) = 7.784$, $p = 0.020$) and *Smaller* scales ($\chi^2(2) = 9.556$, $p = 0.008$), but not for the group with CP ($\chi^2(2) = 5.333$, $p = 0.069$).

For the *Larger* scale trials, there was a significant difference between *Double* and *Expansion* trials (Z = -2.699, p = 0.007). For 9 of the 10 participants, *Double* trials were less frustrating than *Expansion* trials. For the other participant, they were tied. The median ratings were 1.0 and 3.5 respectively.

For *Smaller* scale trials, *Double* trials were significantly less frustrating than *Single* trials (Z = -2.524, $p = 0.012$), but not significantly less than the *Expansion* trials (Z = -

2.243, *p* = 0.025). For *Expansion* trials, 7 participants rated frustration as higher, two as equal, and one as lower than *Double* trials. For *Single* trials, 8 participants rated
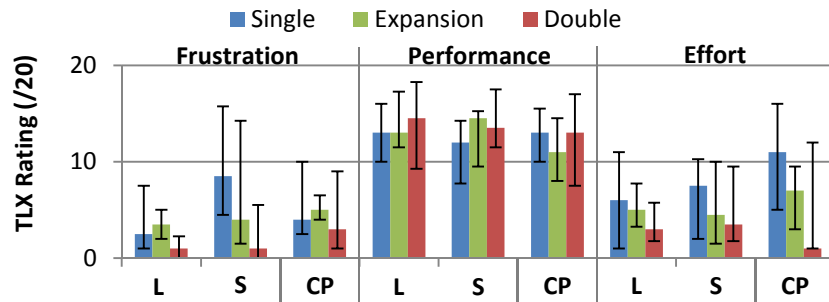
Fig. 6. Median TLX workload measures for Frustration, Performance and Effort. Error bars show the 25th and 75th percentiles. L: W/o CP *Larger*, S: W/o CP *Smaller*, CP: CP.

frustration as higher and two as equal to *Double* trials. The median values for *Single*, *Expansion*, and *Double* trials were 8.5, 4.0 and 1.0 respectively.

Further examination of the group with CP revealed Frustration was most likely to be different between *Double* and *Single* trials (Z = -2.121, p = 0.034). This seems counterintuitive since the medians suggest there should be a greater difference between *Double* and *Expansion* trials. Yet for all 5 participants, *Double* trials were less frustrating than *Single* trials. Regardless, the Friedman test suggested there was no main effect of *Trial Type* (recall, p=0.069).

Performance ratings for the group W/o CP changed significantly with *Trial Type* for *Smaller* scale trials (χ^2(2) = 6.324, p = 0.042). No pairwise comparisons were significant, but in general participants rated their performance as being worse for the *Single* trials compared to both the *Expansion* trials (Z = -2.222, p = 0.026) and the *Double* trials (Z = -2.200, p = 0.028). In both cases, 8 of the 10 participants rated their performance as worse for *Single* trials. The median value for *Single* was 12.0, whereas the median for *Expansion* was 14.5 and the median for *Double* was 13.5.

Also for *Smaller* scale, changes in Effort were close to significance (χ^2(2) = 5.852, p = 0.054). If we permit ourselves to perform pairwise comparisons on the Effort data, participants came closest to rating *Single* trials as requiring more effort than *Double* trials (Z = -2.207, p = 0.027), and to a lesser extent *Expansion* trials (Z = -1.897, p = 0.058). Compared to *Double* trials, 6 of the 10 participants rated effort as being higher for *Single* trials, the other 4 participants rated it as being equal. Compared to *Expansion* trials, 5 participants rated effort higher for *Single* trials, 4 rated them equal, and 1 rated the *Single* trials as requiring less effort. The medians were 7.5 for *Single* trials, 4.5 for *Expansion* trials, and 3.5 for *Double* trials.

5　　Discussion

We investigated the potential of improving pointing efficacy by replacing one movement to a smaller target with two movements to larger targets. Since pointing was the primary interest we sought to minimize the role navigational and cognitive effects such as searching for targets and deciding which one to acquire. However, the sample assistive tool tested did alter functionality and involved decision making since there was access to an elective zoom lens.

We sought conditions where two movements to larger targets decreases the time it takes to communicate a set amount of information. We expected this would only be possible when corrective movements were problematic. So we expected scale (i.e. the small target effect) and motor skills to influence the likelihood of an extra movement increasing efficacy.

5.1　　Scale

Scale was investigated for participants W/o CP and had a clear effect on the potential of adding a movement. For the *Larger* scale, adding the extra movement resulted in significantly slower completion times, and potentially less effective communication. In stark contrast, adding the extra movement resulted in significantly faster and more effective communication for *Smaller* scale trials. This importance of scale is likely due to the small target effect [1], whereby reducing the scale results in reduced conformity to Fitts's Law for smaller *Single* targets, but not for larger *Double* targets.

This is also reflected in the component times. When the scale is reduced there is a large increase in the correction times of *Single* trials. By contrast, all of the other component times do not appear to vary greatly across the different scales. Furthermore, correction times are the only component where *Double* trials are faster than *Single* trials. All of this suggests that corrective movements are a problem for the *Smaller* scale *Single* trials, and as a result the *Double* trials increase efficacy.

5.2　　Motor Impairment

The effects of users' motor skills were more difficult to interpret. The results of both the *Smaller* scale trials and the group with CP's trials suggested they were subjectively more difficult than the *Larger* scale trials since they were generally slower. However, the results of the group with CP did not replicate the results of the *Smaller* scale trials.

In contrast to the *Smaller* scale trials, it is difficult to determine if the small target effect is present in the *Single* versus *Double* results for the group with CP. For that group, there was a significant difference in accuracy (and almost frustration) between the two trial types. These differences suggest the *Single* targets were sufficiently small to be 'too small', whilst the *Double* targets were not. Despite this, the *Single* trials were still significantly faster than the *Double* ones.

If we compare the component times for the group with CP, it is clear that the difference in correction times between *Single* and *Double* trials bears more similarity

to the *Larger* scale trials than the *Smaller* scale ones. We hypothesized that motor impairments would increase the benefit of an additional movement since motor impairments tend to result in an increased number of sub-movements [12-14]. Whilst *Single* trials did on average involve over a second of correction time for the group with CP, the significant reduction in correction time for *Double* trials was not enough to make them faster. This pattern is similar to the *Larger* scale trials even though the completion times suggest the difficulty is more like the *Smaller* scale trials.

Another clear distinction is that there is much greater variability in the results of participants with CP. Whilst this trend is common in pointing tasks [2, 25-27], it also suggests a larger sample size would benefit future research.

5.3   Expansion Cursor

Since single movements are likely to be faster than double movements for most target sizes larger than 5 pixels, it seems any assistive tool should only add a movement when the user finds the target to be very small. The Expansion Cursor accommodates this with a discretionary zoom. However, there appears to be a cost to this functionality. In every performance measure the Expansion Cursor was never the best. It seems that depending on the situation, a single movement or a double movement may be optimal, but not the Expansion Cursor. The point is that any assistive tool that aims to switch between one or two movements based on context will not be 100% efficient. In the case of the Expansion Cursor there is probably a time cost associated with the decision to use it or not. Likewise there is an added dexterity requirement compared to *Double* trials: the user must hold the mouse button down whilst they make corrective movements. This possibly has an effect on accuracy.

These issues may appear specific to the Expansion Cursor. However any assistive tool that adds a movement implicitly has to have some mechanism to determine when to add that movement. That decision, whether performed automatically by the computer or the user, will not be correct 100% of the time. Hence there will be a cost relative to the optimal scenario of a single or double movement.

It should be noted that most tools in the literature do not involve a mechanism for users to make a single movement when the target is easy to acquire (e.g. all of the enhanced area cursors in [5]). Avoiding this mechanism does not solve the problem of costs associated with it since the assistive tool is likely to become a hindrance for larger targets. If the tool is still useful for larger targets, it suggests the user would benefit from targets being universally bigger in motor space. This universal enlargement can always happen by reducing the mouse speed (i.e. the C-D gain), hence a tool that improves acquisition of larger targets appears to be redundant.

5.4   User Experience

So far, this discussion has neglected to mention what the participants thought about the experiment. From a user-perspective, there seems to be a larger risk to making targets too small, than there is to introducing an extra movement layer.

When comparing the smallest targets (*Smaller* scale *Single* trials, 2 pixels) to larger ones (*Smaller* scale *Double* trials, 20 pixels), participants reported increased frustration and effort that resulted in worse performances. This was reflected in participants' comments. For *Single* trials, participants mentioned that you "find yourself going over and back trying to get it (the target)", you had to be more precise, it was a matter of getting it right, and that it was frustrating trying to make such a small adjustment when you were 1 pixel off the target.

By contrast, half of the participants preferred the *Double* movement for *Larger* scale trials, even though it was significantly slower than *Single* trials. They mentioned that it was easier, the targets were bigger and less stressful, and that you didn't have to think too much about it. Only one participant rated *Double* as their least preferred for *Larger* trials. Furthermore, all 5 of the participants in the group with CP rated *Double* trials as less frustrating than *Single* trials. Again, this is despite *Double* trials being significantly slower.

5.5   Summary of Potential to Improve Efficacy

There are two components to the potential for a double movement to improve efficacy (Time/Info). The first is accuracy. *Double* trials seemed likely to improve accuracy although that was only significant for *Smaller* scale trials. Improved accuracy effectively increases the amount of information communicated per target acquired.

The second aspect of improving efficacy is time. *Double* trials inherently increased reaction time, traversal time, and click time, presumably since they all had to be performed twice. However an additional input that enlarges targets can potentially reduce correction times enough to reduce completion times overall.

Both accuracy and speed gains appear to relate to the difficulty of acquiring very small targets. *Double* movements can avoid this difficulty since the targets can be larger. However, the benefit of larger targets seems to diminish when the *Single* movement target is already large enough. This is apparent in the different results between 2 and 5 pixel wide *Single* targets. If the *Single* targets were a more reasonable size (i.e. even bigger than 5 pixels), the potential for an extra movement to benefit target acquisition efficacy presumably disappears.

How motor skills interact with all of this is less clear. Previous work has demonstrated that CP can result in an increased number of sub-movements for pointing tasks [12-14]. The results agree with this since the group with CP appear to spend more time making corrections than the group W/o CP for the same task scale. However, this increase in corrections does not seem to result in added potential for two movements to be faster than one. *Double* movements were significantly slower than *Single* movements for both participant groups when performing the same task scale.

5.6   Limitations

The limitations of this experiment relate to it being a controlled lab experiment. First, it did not aim to accurately represent real-life interaction. The targets were 1D instead

of 2D, and we aimed to minimize the interactions of navigation and other practicali-practicalities. Likewise, the participants were only aiming to click a black target column; they had no other intrinsic task goal informing their experience. This test scenario avoided confounds and allowed pointing to be explicitly investigated. Also, if this trade-off is to be more thoroughly understood, future work should consider a broader range of magnification levels and scales.

6 Conclusion

In this experiment we have demonstrated that it is possible for two movements to result in more effective pointing than one movement. However, it seems that this is dependent on the single movement being to a target that is frustratingly small and requires a disproportionate amount of time spent making corrective movements. In this experiment, even a 5-pixel wide target was not sufficiently small for an extra movement to improve efficacy for people with and without motor impairments. Only a 2-pixel wide target was small enough. Outside of correction time, *Double* movements resulted in significant increases in all other time components. This presumably stems from all aspects of target acquisition having to be performed twice. Regardless of speed, *Double* movements tended to improve accuracy for all of the conditions tested. Again, it appears this potential improvement depends on the *Single* trial target being small enough to impair accuracy.

These results help define the potential of assistive tools that replace one movement to a small target with two movements to larger targets. Efficacy benefits are possible for very small targets, but practically unlikely. The design of tools that add a movement should focus on the ways they can potentially improve user experience, such as improving accuracy, reducing frustration and attentional demands. Improving speed and efficacy is unlikely to be a potential benefit. Furthermore, assistive tools need a mechanism of aiding small target acquisition, but not interfering with larger target acquisition.

The Expansion Cursor takes this distinction into account and allows users to choose when they wish to use a zoom lens. However, the results demonstrated that there is a cost to this mechanism. We propose that any mechanism that makes a distinction between easy and difficult targets is likely to entail performance costs relative to the ideal scenario of either a single or double movement. This is design issue common to assistive tools. Regardless, the main accessibility benefit of such assistive tools is making accurate target acquisition easier, but not faster.

References

1. Chapuis O., Dragicevic P.: Effects of motor scale, visual scale, and quantization on small target acquisition difficulty. 18, 13 (2011).
2. Davies T. C., AlManji A., Stott N.S.: A cross-sectional study examining computer task completion by adolescents with cerebral palsy across the manual ability classification system levels. 56, 1180-1186 (2014).

3. Almanji A., Payne A., Amor R., Davies C.: A nonlinear model for mouse pointing task movement time analysis based on both system and human effects. (2014).
4. Bakaev M.: Fitts' law for older adults: Considering a factor of age. , 260-263 (2008).
5. Findlater L., Jansen A., Shinohara K., Dixon M., Kamb P., Rakita J., Wobbrock J.O.: Enhanced area cursors: Reducing fine pointing demands for people with motor impairments. , 153-162 (2010).
6. Payne A. R., Plimmer B., McDaid A., Luxton-Reilly A., Davies T.C.: Expansion cursor: A zoom lens that can be voluntarily activated by the user at every individual click. , 81-90 (2016).
7. Li L., Gajos K.Z.: Adaptive click-and-cross: Adapting to both abilities and task improves performance of users with impaired dexterity. , 299-304 (2014).
8. Jansen A., Findlater L., Wobbrock J.O.: From the lab to the world: Lessons from extending a pointing technique for real-world use. , 1867-1872 (2011).
9. Kiger J. I.: The depth/breadth trade-off in the design of menu-driven user interfaces. 20, 201-213 (1984).
10. Landauer T. K., Nachbar D.: Selection from alphabetic and numeric menu trees using a touch screen: Breadth, depth, and width. 16, 73-78 (1985).
11. Zaphiris P., Shneiderman B., Norman K.L.: Expandable indexes vs. sequential menus for searching hierarchies on the world wide web. 21, 201-207 (2002).
12. Hwang F., Keates S., Langdon P., Clarkson J.: Mouse movements of motion-impaired users: A submovement analysis. , 102-109 (2004).
13. Hwang F., Keates S., Langdon P., Clarkson J.: A submovement analysis of cursor trajectories. 24, 205-217 (2005).
14. Saavedra S., Joshi A., Woollacott M., Van Donkelaar P.: Eye hand coordination in children with cerebral palsy. 192, 155-165 (2009).
15. Seow S. C.: Information theoretic models of HCI: A comparison of the hick-hyman law and fitts' law. Hum.-Comput.Interact. 20, 315-352 (2005).
16. Fitts P. M.: The information capacity of the human motor system in controlling the amplitude of movement. J.Exp.Psychol. 47, 381 (1954).
17. MacKenzie I. S.: Fitts' law as a research and design tool in human-computer interaction. Hum.-Comput.Interact. 7, 91-139 (1992).
18. Card S. K., English W.K., Burr B.J.: Evaluation of mouse, rate-controlled isometric joystick, step keys, and text keys for text selection on a CRT. Ergonomics 21, 601-613 (1978).
19. Bax M. C.: Terminology and classification of cerebral palsy. 6, 295-297 (1964).
20. Rosenbaum P., Paneth N., Leviton A., Goldstein M., Bax M., Damiano D., Dan B., Jacobsson B.: A report: The definition and classification of cerebral palsy april 2006. Dev.Med.Child Neurol.Suppl. 109, 8-14 (2007).
21. Ramos G., Cockburn A., Balakrishnan R., Beaudouin-Lafon M.: Pointing lenses: Facilitating stylus input through visual-and motor-space magnification. , 757-766 (2007).
22. Zhang X., Zha H., Feng W.: Extending fitts' law to account for the effects of movement direction on 2d pointing. , 3185-3194 (2012).
23. Eliasson A., Krumlinde-Sundholm L., Rösblad B., Beckung E., Arner M., Öhrvall A., Rosenbaum P.: The manual ability classification system (MACS) for children with cerebral palsy: Scale development and evidence of validity and reliability. 48, 549-554 (2006).

24. NASA: NASA TLX: Task load index. http://humansystems.arc.nasa.gov/groups/TLX/ (2016). Accessed July 2016.

25. Smits-Engelsman B. C. M., Rameckers E.A.A., Duysens J.: Children with congenital spastic hemiplegia obey fitts' law in a visually guided tapping task. 177, 431-439 (2007).
26. Payne A. R., Plimmer B., Davies T.C.: Repeatability of eye-hand movement onset asynchrony measurements and cerebral palsy: A case study. , 31-38 (2015).
27. Al Manji A., Davies C., Amor R.: Examining dynamic control-display gain adjustments to assist mouse-based pointing for youths with cerebral palsy. 3 (2015).**P**