AN EMPIRICALLY-DERIVED MODEL FOR PREDICTING COMPLETION TIME OF CURSOR POSITIONING TASKS IN DUAL-TASK ENVIRONMENTS

Sc.D. Dissertation Defense

Martin J. Schedlbauer

April 4, 2006

Chair: *Dr. Jesse M. Heines, University of Massachusetts Lowell*

Committee: *Dr. Giampiero Pecelli, University of Massachusetts Lowell*

Dr. Holly Yanco, University of Massachusetts Lowell

Dr. Robert Pastel, Michigan Technological University

Outline of Presentation

- Motivation of Research
- Statement of Research Problem
- Foundation of Model
- Statement of Research Questions
- Experimental Methodology
- Results and Discussion
- Application of Results
- Summary and Future Work

Motivation: Navigation System

- Maptech Navigator
- Touch screen
- GPS navigation

- But: frustrating to use…
- Why?
- What guidelines can be provided to designers to make mobile systems easier and safer to use?
- What is the definition of safety?

The Problem

- Is the interaction safe?
- In other words, is:

$$T_{Completion} + T_{Switch} \leq T_{Neglect}$$

- The research questions then become:
  - How can completion time be defined?
  - How can completion time be calculated?
  - What factors influence completion time and how?

Task Completion Time Model

- Task Completion Time (*TT*) is defined as the time it takes to perform a series of sub-tasks that constitute an interactive transaction with a system:

$$TT = LT + \sum_{i=1}^{n} (RT_i + MT_i)$$

  where *LT* is learning time, *RT* is the choice reaction (or recognition) time, and *MT* is the mean movement time

- Completion time increases in dual-task situations due to context switch and "task scheduling" (Hoffman & Lim, 1997; Shin & Rosenbaum, 2002).

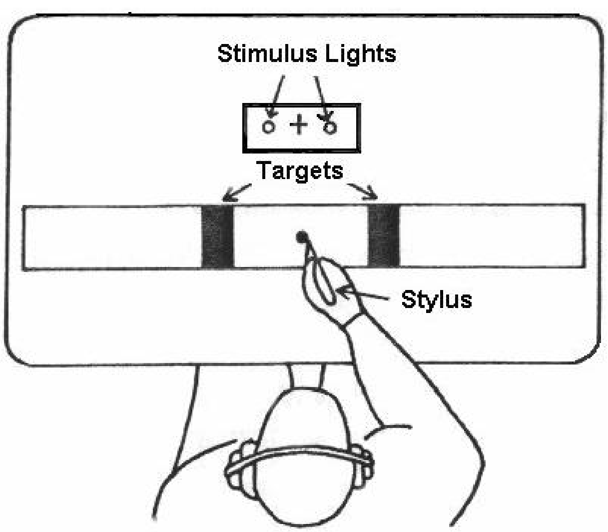
HCI Engineering Models

- Movement Time (*MT*) Predictions
  - Fitts' Law (Fitts, 1954) and recent variations by Meyer *et al*. (1988), Kvålseth (1980), and Oel *et al*. (2001)
- Reaction Time (*RT*) Predictions
  - Hick-Hyman Law (Hick, 1952; Hyman, 1953)
  - Kvålseth's Decision Model (Kvålseth, 1996)
- Learning Time (*LT*) Predictions
  - Power Law of Practice
  - Exponential Law of Practice (Heathcote *et al*., 2001)

Fitts' Law

$$MT = a + k\log_2\left(\frac{D}{W_e} + 1\right)$$

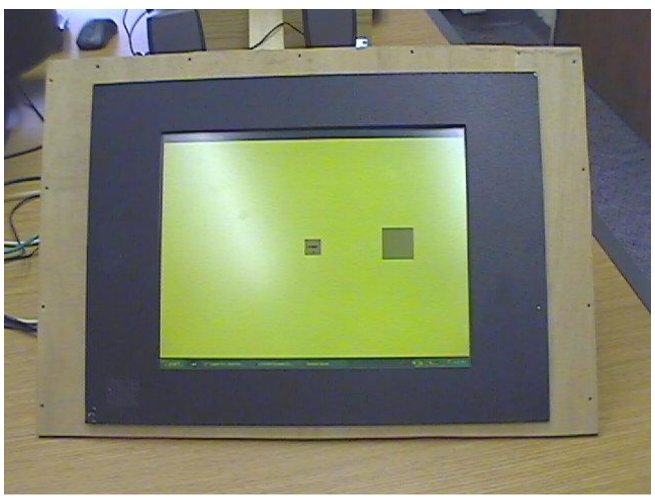
where *a* and *k* are empirically derived coefficients, *D* is distance to target (amplitude of movement), and W_e is effective width of the target.



- Derived from Shannon-Hartley Theorem 17
- Describes spatially constrained rapid aiming (pointing) tasks
- Applies to many different input devices

Input Methods & Devices

- Based on survey of currently deployed systems, these input devices were tested:
  - touch input with finger and stylus
  - trackball
  - isometric joystick



Research Questions

- Does Fitts' Law hold when changing posture from sitting to standing?
- Does movement time increase in dual-task situations and how does it affect Fitts' Law?
- Does Fitts' Law hold when targets are not stationary?
- How can task completion time of multi-step interactions be predicted?
- Can specific heuristics for the design of user interfaces of mobile computing be derived?

Experimental Validation

- The research questions were investigated experimentally through five IRB-approved laboratory tests with human subjects
- The experiments were carried out using a specially developed Java software platform (*MTE*)
- The data was then analyzed in "*R*" using appropriate statistical tests (Wilcoxon, *t*-test, ANOVA, Pearson Moment)
- Statistical significance was set at the generally accepted level of 5% ($p < .05$)

Experiments

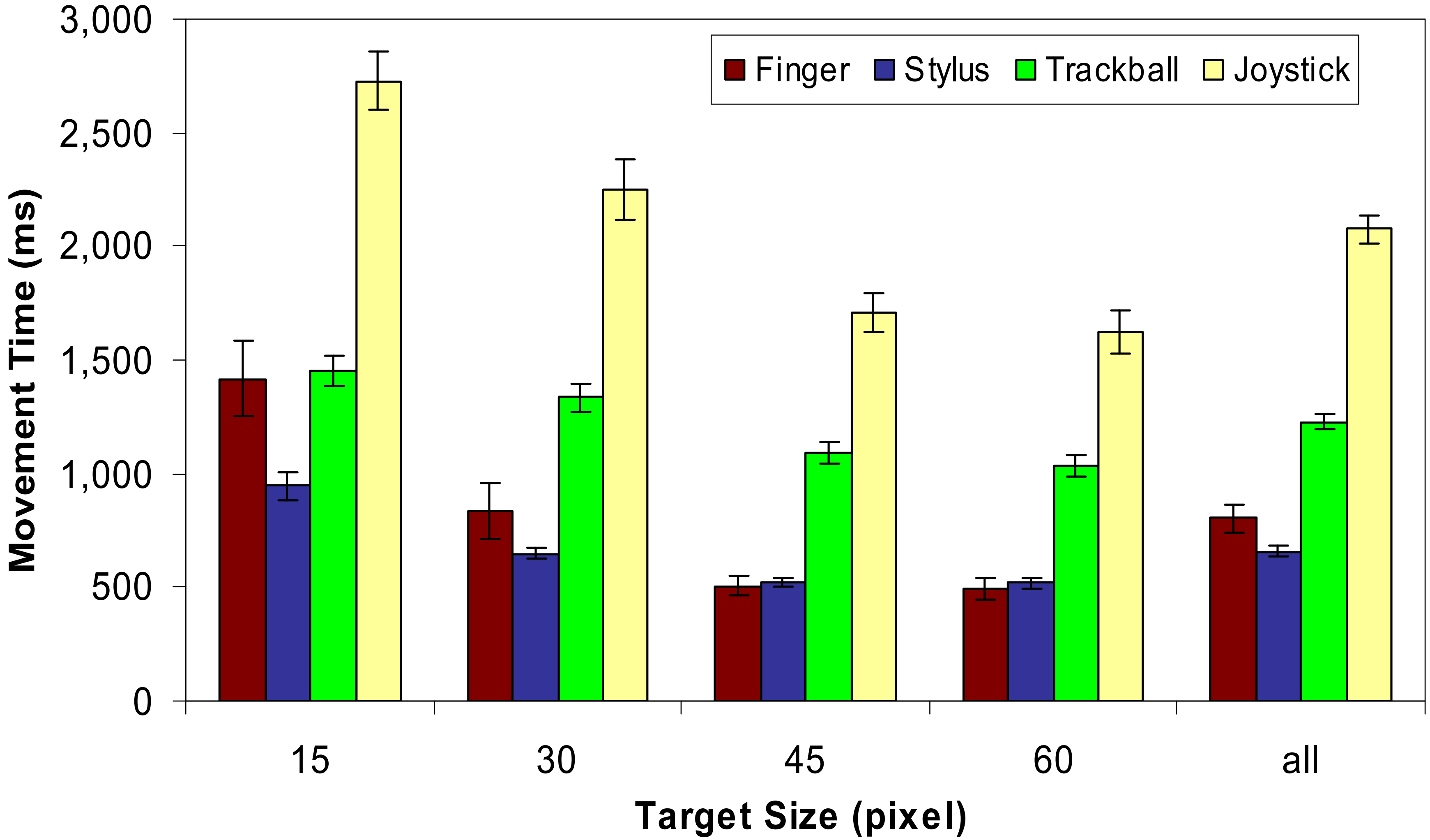
- Experiment I: Validation
  - validate experimental apparatus and compare against published results; investigate potential negative effects of posture
- Experiment II: Dual-Task Situations
  - determine if the presence of a second cognitive and manual task has a negative effect on aiming performance
- Experiment III: Effect of Walking
  - investigate the effect of walking on aiming performance
- Experiment IV: Non-Stationary Targets
  - determine if Fitts' Law holds for the selection of vibrating targets
- Experiment V: Multi-Step String Entry Task
  - validate task completion time model for a multi-step numeric string entry task on a soft keypad

Experiment I: Validation

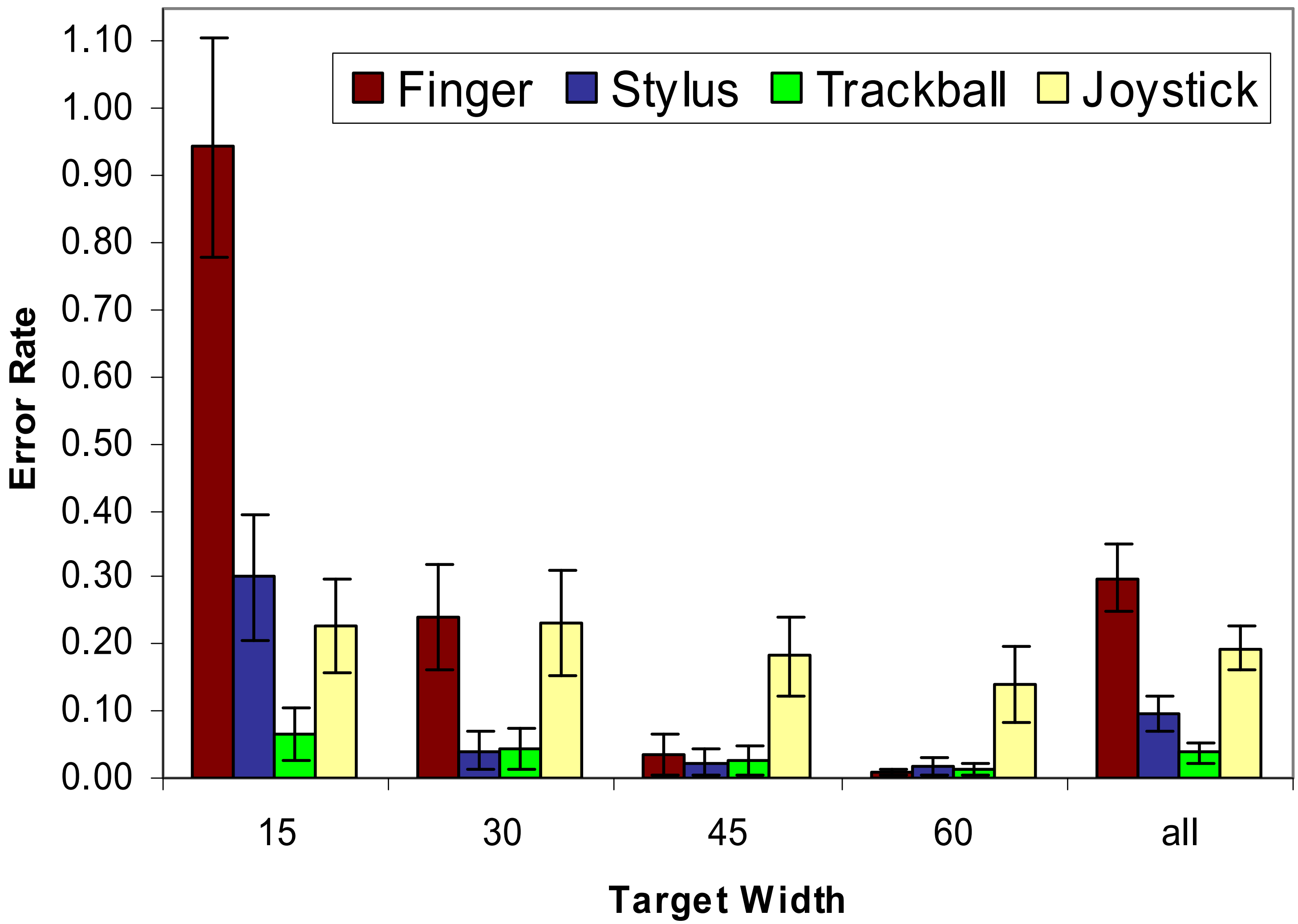
Experiment I: Method

- Subjects
  - 11 volunteers (9 men, 2 women); all right handed
- Apparatus
  - Finger, stylus, trackball, isometric joystick
  - Gateway Tablet PC with 15" Elo touch screen
- Experiment Design
  - Four blocks of 20 randomly positioned circular targets
  - Each block varied target size: 15, 30, 45, 60 pixels
  - Repeated while sitting and standing

Device Performance: Movement Time



Device Performance: Error Rate



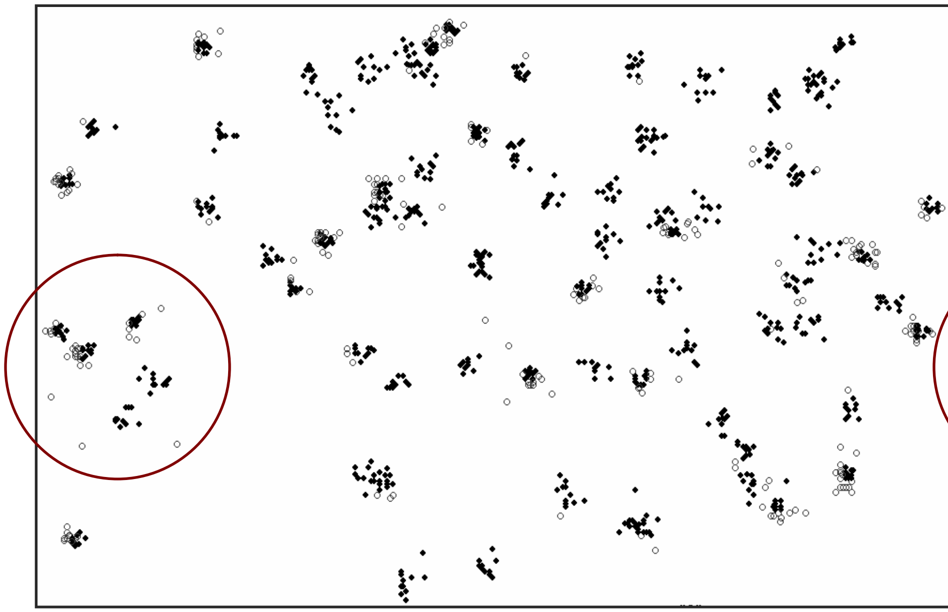
Overall Device Performance

$$TP = \frac{\overline{ID}}{\overline{MT}}$$

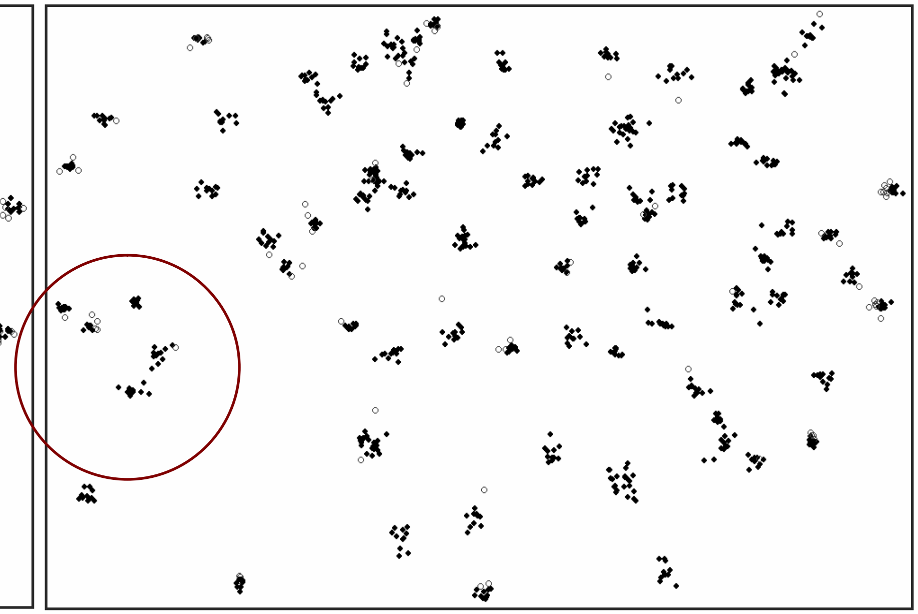
Device	Mean MT	Mean ER	TP	User Rank
Finger Touch	803	29.8%	3.37	1
Stylus	659	9.5%	4.11	2
Trackball	1228	3.8%	2.21	3
Joystick	2075	19.4%	1.32	4

Unit for movement time (MT) is milliseconds (*ms*), unit for throughput (TP) is bits/second (*bps*)

Pointing Accuracy



Finger Touch



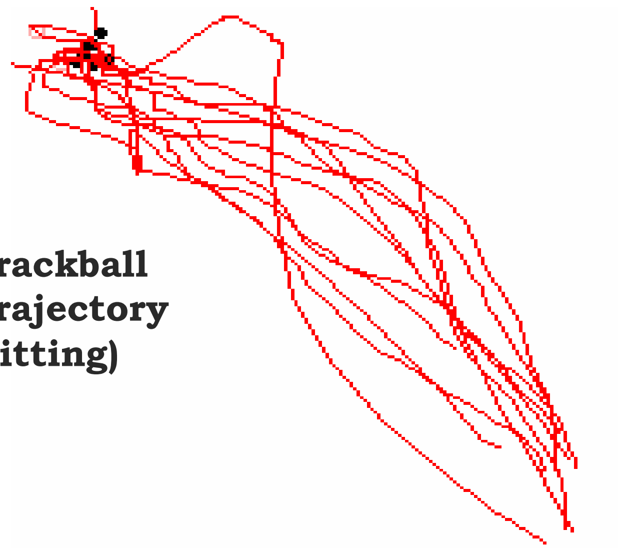
Stylus Touch

- Finger touch shows wider spatial variability of selection end points, explaining lower accuracy and higher error rate
- Variability is more pronounced for targets less than 15*mm*
- Differences are statistically significant (ANOVA, $p < .001$)

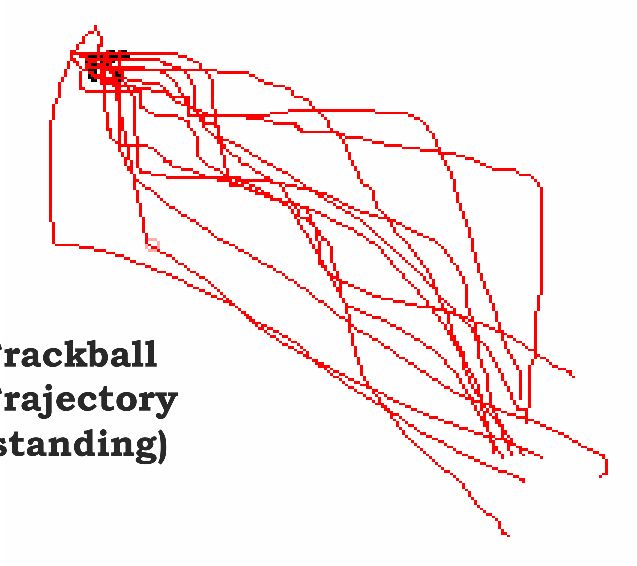
Effect of Posture: Sitting *vs.* Standing

- Effect on finger touch:
  - no significant difference in performance and accuracy
- Effect on stylus touch:
  - no significant difference in performance and accuracy
- Effect on trackball:
  - performance unchanged, but 102% increase in error rate to 7.7%
- Effect on isometric joystick:
  - performance unchanged, but a increase in error rate to 24.4%
- Spatial variability of selection end points more pronounced
- Two-way ANOVA shows a significant interaction between error rate and width and posture as factors

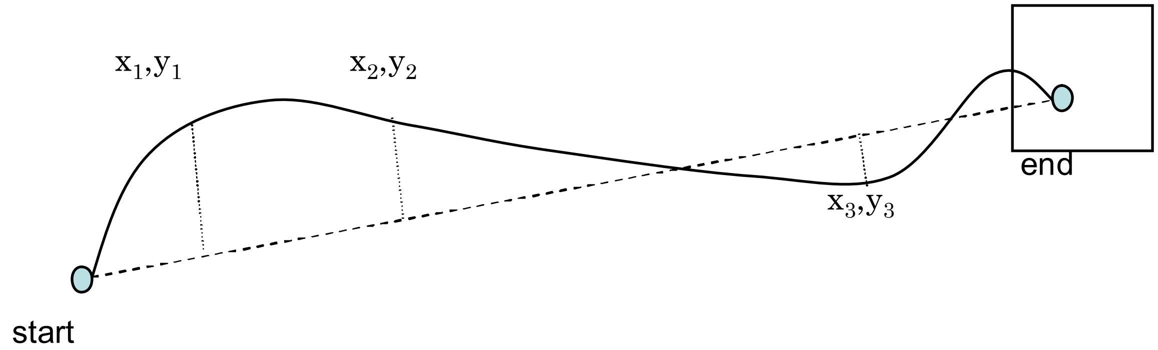
Movement Variability



Trackball Trajectory (sitting)



Trackball Trajectory (standing)



- Movement variability is more pronounced for trackball ($p = .04$) when standing, but not for the isometric joystick
- Between the devices, the variability is significant ($p < .0001$)
- Higher variability results in more targeting errors and longer movement time ($R = .30$, $p < .001$)

Conclusions: Experiment I

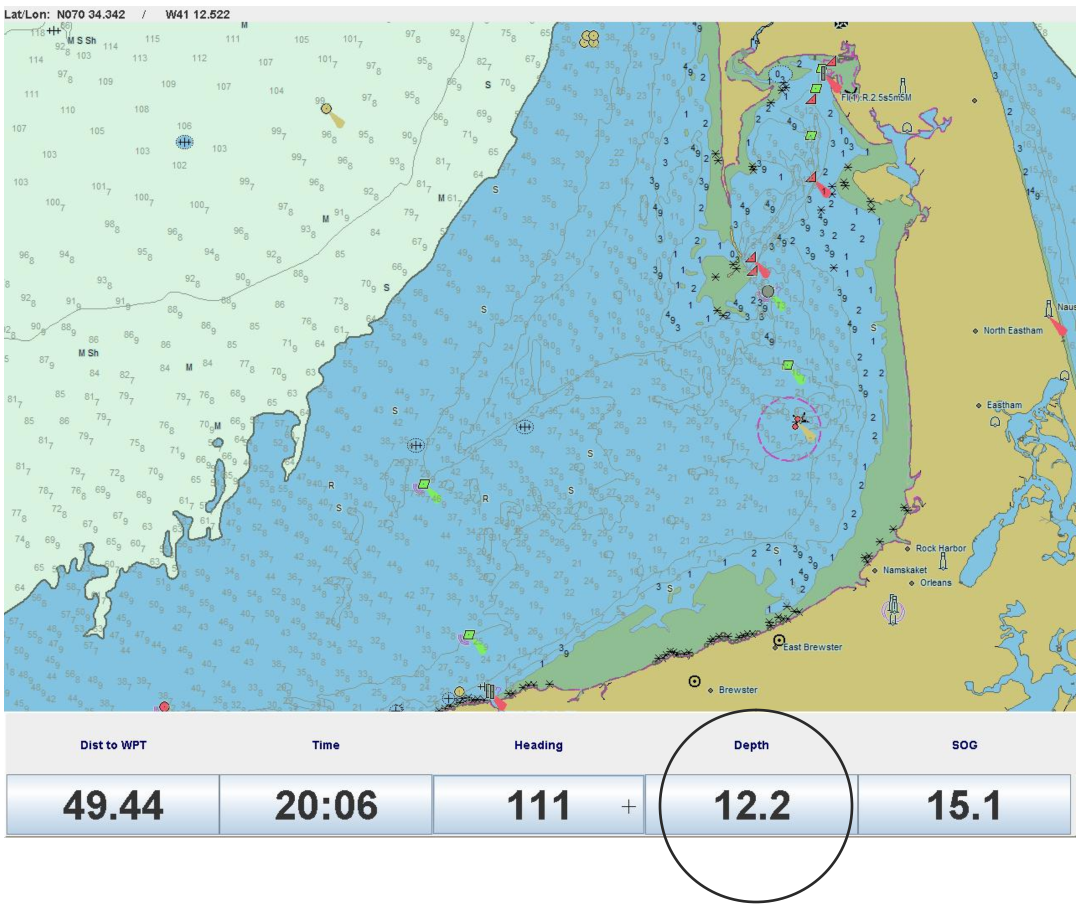
- The experiment shows that the *MTE* tool is suitable for exploration and research
- The predictive strength of Fitts' law remains unchanged when standing
- Touch input is least affected by posture
- Indirect input methods show a significant decrease in accuracy when standing
- Touch input is the best overall choice for ubiquitous applications

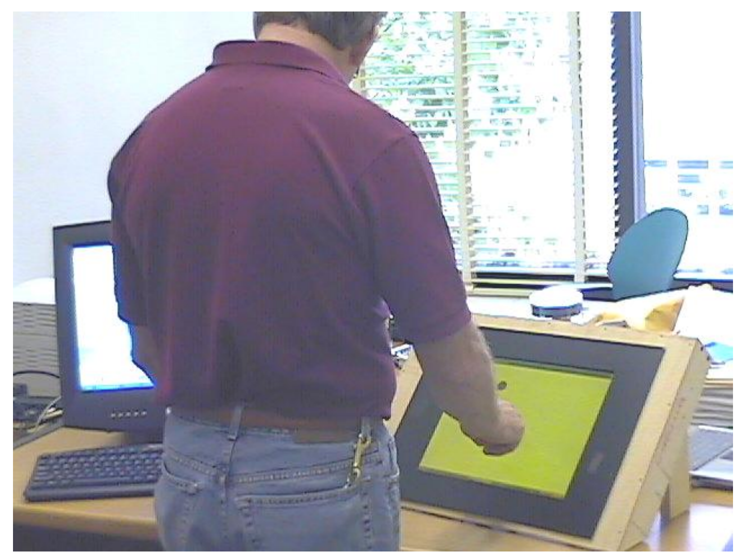
Experiment II: Dual-Task Situations

Experiment II: Method

- Subjects
  - 8 volunteers (7 men, 1 woman); all right handed
- Apparatus
  - Finger, stylus, and trackball
  - Gateway Tablet PC with 15" Elo touch screen; second PC with 17" monitor
- Experiment Design
  - Four blocks of 20 randomly positioned circular targets
  - Each block varied target size: 15, 30, 45, 60 pixels
  - Standing only with secondary cognitive/manual task

Apparatus for Experiment II





Hypothesis

- Based on research by Hoffman & Lim (1997) and Shin & Rosenbaum (2002), aiming performance should decrease as reaction time becomes a factor:

$$TT = a + bRT + c\log_2\left(\frac{A}{W} + 1\right)$$

- Choice reaction time among 2 targets:

$$TT = a + \underbrace{200\sqrt{2}} + c\log_2\left(\frac{A}{W} + 1\right)$$

increase in regression intercept expected

Conclusions: Experiment II

- An average increase of 334*ms* in the regression intercept is found, which is slightly more than the increase expected by the value for choice time
- Additional increase likely due to switch time
- Correlations for Fitts' law decrease slightly but are still strong (between .77 and .95, $p < .001$)
- Task completion time increases predictably and linearly in a dual-task situation
- Error rate remains the same for all but finger touch which increases significantly to 38.1% ($p < .001$)

Experiment III: Effect of Walking

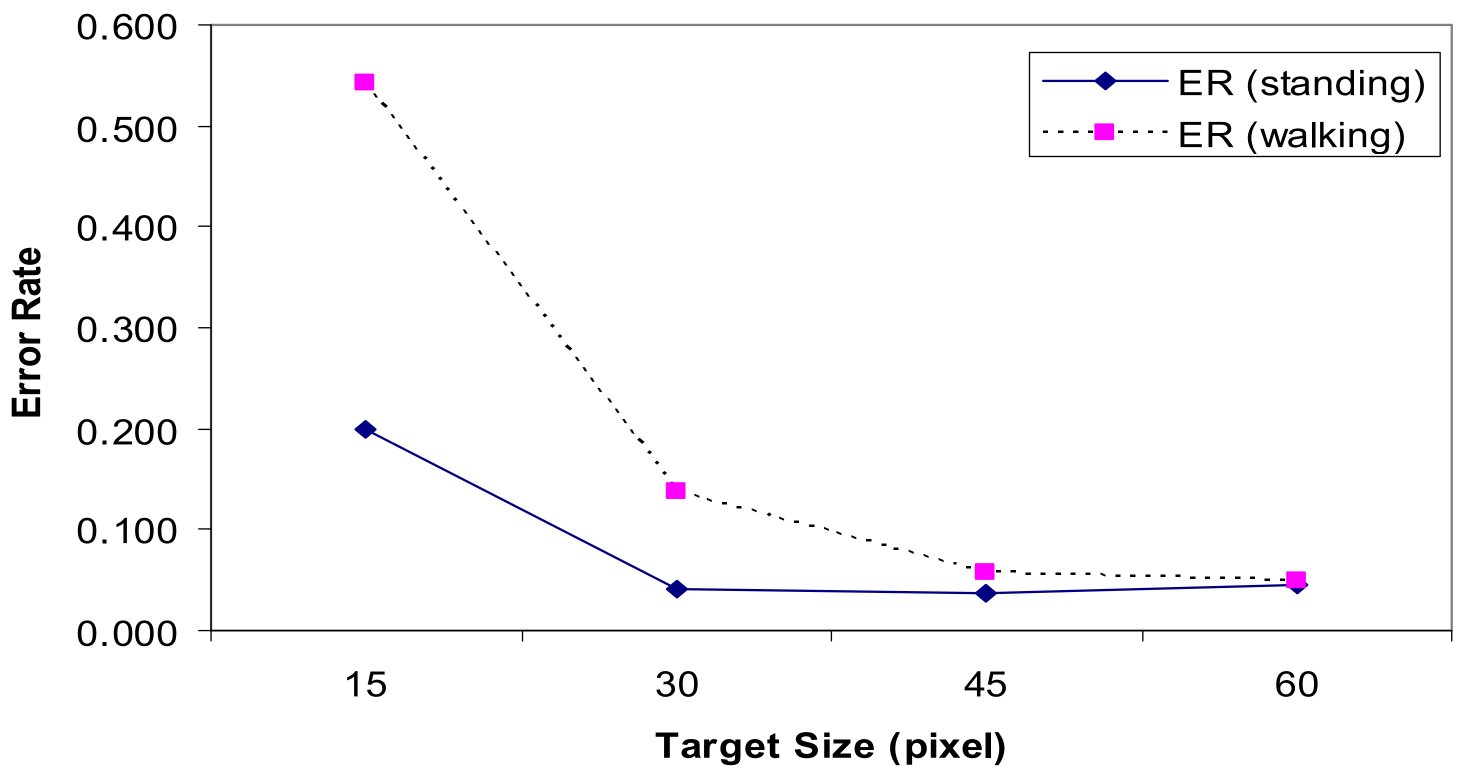
Experiment III: Method

- Subjects
  - 11 volunteers (9 men, 2 women); all right handed
- Apparatus
  - Stylus touch
  - Fujitsu Stylistic LT with 8.4" touch screen
- Experiment Design
  - Four blocks of 20 randomly positioned circular targets
  - Each block varied target size: 15, 30, 45, 60 pixels
  - Repeated while standing and walking

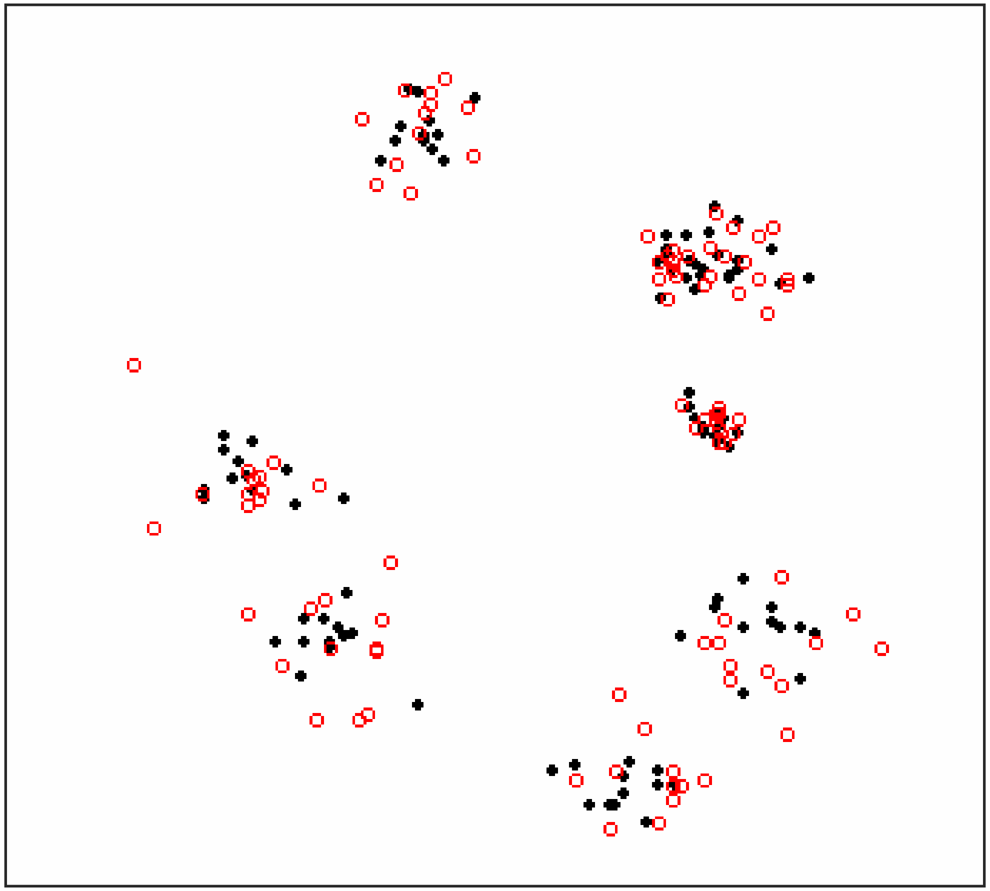
Apparatus

Effect of Walking

- No difference in the movement time while walking
- Fitts' law is valid ($R = .93$, $p < .001$)
- <u>But</u>: error rate increases from 8.1% to 19.6% ($p < .0001$)



Spread of Selection Endpoints



walking (open circles) vs. standing (filled circles)

- The higher error rate while walking is caused by less accurate selection of targets
- The standard deviations of the selection end points differ significantly ($p < .001$)
- Higher deviation requires more target tolerance for maximized task completion time
- Finger touch would likely lead to increase the error rate based on results of previous experiments

Conclusions: Experiment III

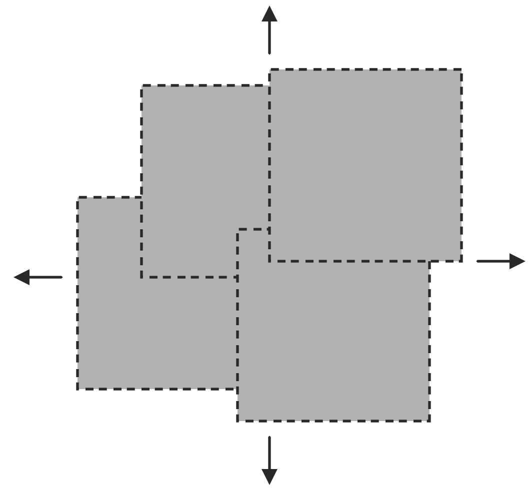
- Fitts' law is a robust predictor of aiming performance during walking, which implies that noise in the production of limb movements is accounted for by Fitts' law
- There is little increase in the movement time, but a significant increase in the error rate, particularly for small targets (< 7*mm* or 30*px*)
- While walking, movement time increases more rapidly with the difficulty of the movement
- Subjects slowed walking speed noticeably when selecting the smallest target (3.2*mm* or 15*px*) indicating a neglect of the walking task in favor of scheduling the aiming task

Experiment IV: Non-Stationary Targets

Experiment IV: Method

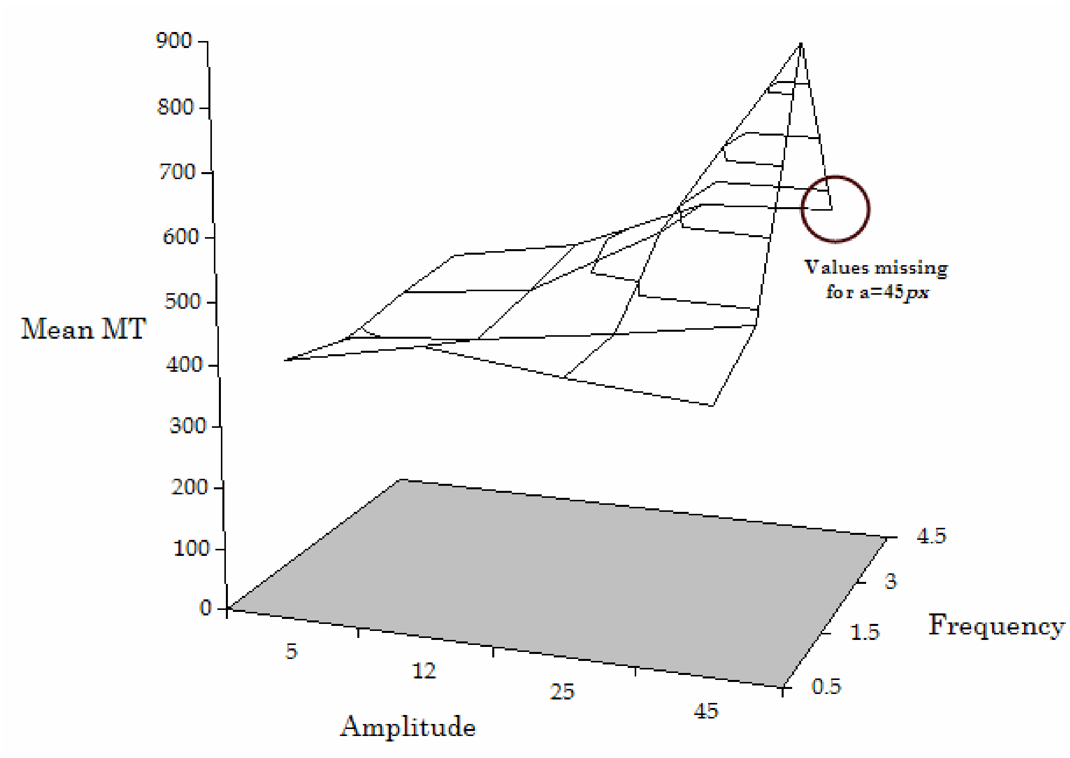
- Subjects
  - 11 volunteers (9 men, 2 women); all right handed
- Apparatus
  - Finger and stylus
  - Compaq iPaq PC with 15" Elo touch screen
- Experiment Design
  - 64 blocks of 10 randomly positioned square targets
  - Each block varied target size, frequency, and amplitude:
    - 25, 45, 60, and 80 pixel wide targets
    - 5, 12, 25, and 45 pixel amplitudes
    - 0.5, 1.5, 3.0, and 4.5 *Hz* frequencies
  - Data block (w=25*px*, a=45*px*, f=4.5*Hz*) omitted

Apparatus



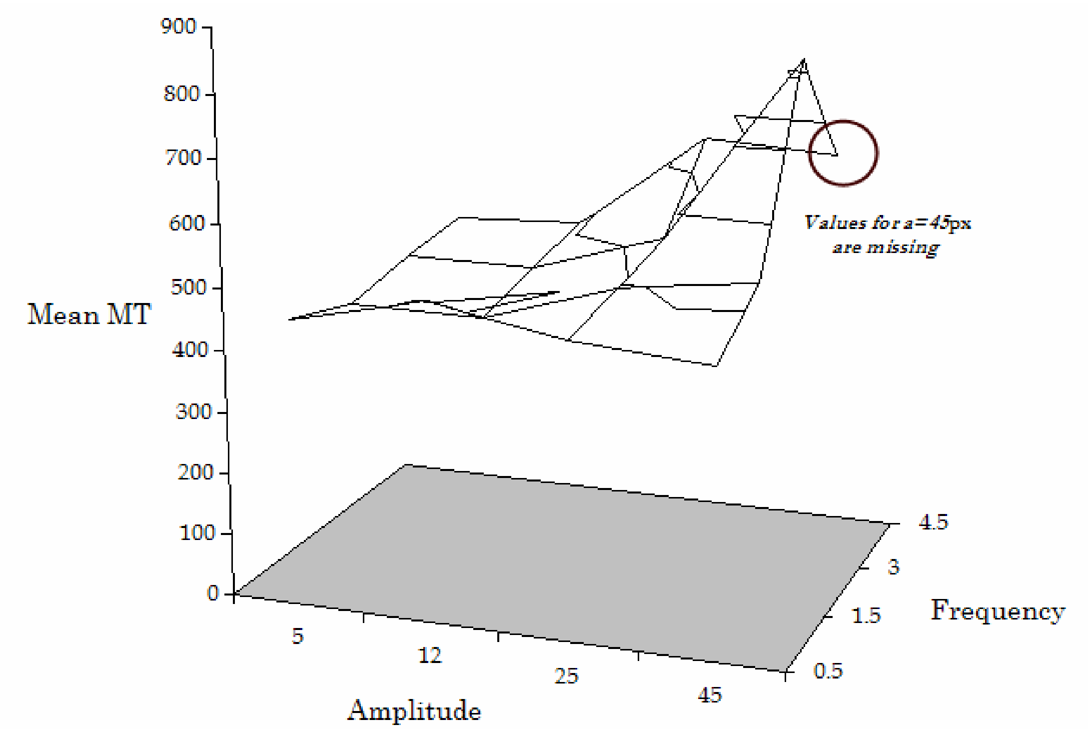
- Parameters:
  - *frequency (f)*
  - *amplitude (a)*
  - *target size (w)*

Movement Time for Finger Touch



- Capture time relatively unaffected by motion when $f < 2Hz$ and $a < 25px$
- Fitts' law predicts capture time quite well for small vibration (jitter)
- Error rates reach "catastophic" realm when $f > 3.0Hz$ and $a > 40px$
- Fitts' law fails when vibration increases above these threshold rates

Movement Time for Stylus Touch



- Similar observations as for finger touch
- Large targets do not have a significantly higher capture time when *f* and *a* are large
- Compensate for high velocity motion with larger target areas

Dual-Phase Model for Capture

- Analysis of the motion suggests that capturing is a two-step process:
  - rapid aiming motion toward vicinity of target
  - accurate homing phase in which probe is centered on the target
- Using the growth in *CT* as frequency (*f*) and amplitude (*a*) increase, yields:

$$CT = k + b\log_2\left(\frac{D}{W'}+1\right)^2 + c\left(\frac{a}{W+H}f\right)^2\log_2\left(\frac{P^2}{A}+1\right)$$

Fit of Data with Dual-Phase Model

- Finger touch: $R^2 = .75$ ($p < .001$)
- Stylus: $R^2 = .76$ ($p < .001$)
- Best fit is observed as amplitude and frequency increase and Fitts' law rapidly shows poor correlation with the data ($R^2 < .10$)
- Comparing Fitts' law and the Dual-Phase Model using an ANOVA shows that the Dual-Phase Model only becomes significantly better when $w < 30px$ or when $f > 2.5Hz$ and $a > 25px$

Conclusions: Experiment IV

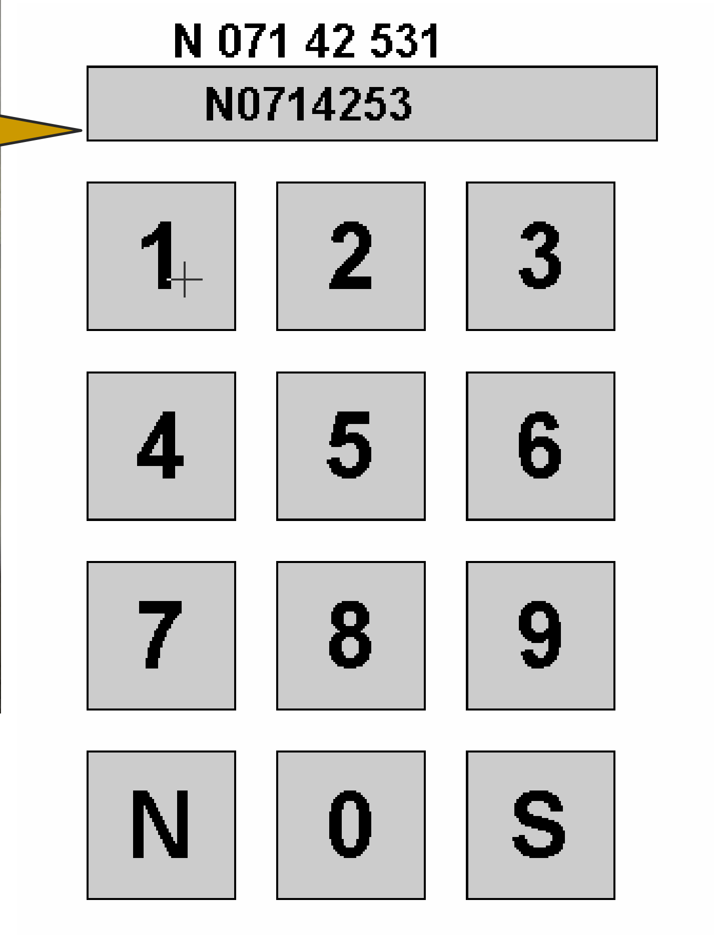
- Fitts' law remains an acceptable approximation for estimating capture time of vibrating targets as long as the frequency and amplitude are small
- The *Dual-Phase Model* is a significantly better predictor of capture time as the speed of the movement increases
- The *Dual-Phase Model* can also be applied to predicting aiming performance for jittering images used in web page advertising or computer games

Experiment V: Multi-Step String Entry Task

Experiment V: Method

- Subjects
  - 8 volunteers (8 men); all right handed
- Apparatus
  - Finger, stylus, and trackball
  - Compaq PC with 15" Elo touch screen
- Experiment Design
  - Four blocks of 8 numeric strings containing 9 characters
  - Each block varied
    - Key size: 35 and 50 pixels (10.5 and 15*mm*)
    - Edge-to-edge gap: 5 and 15 pixels (1.5 and 4.5*mm*)

Apparatus



Predictive Model

- According to Jax *et al.* (2003), four main issues are at the core of motor behavior:
  - task ordering (choice reaction)
  - learning
  - perceptual-motor integration
  - movement
- Therefore, task completion time of an interaction with *n* subtasks can be described by:

$$TT \approx LT + \sum_{i=1}^{n} \left(RT_i + MT_i \right)$$

The Model Details

$$RT_{HH} = 150\log_2(12) = 498ms$$

$$RT_K = 150\sqrt{12} = 519ms$$

$$\Rightarrow TT \approx LT + \sum_{i=1}^{n}(508 + MT_i)$$

reaction time described by Hick-Hyman or Kvålseth

assume expert user, so $LT = 0$

$$TT \approx LT + \sum_{i=1}^{n}(RT_i + MT_i)$$

$$MT_{ij} = a + k\log_2\left(\frac{D_{ij} + W_j}{W_j} + 1\right)$$

key-to-key movement time described by Fitts' law

regression coefficients from previous experiments

Proposed Predictive Model

- The Task Completion Time Model (*TCTM*) for numeric string entry on a soft keypad is:

$$TT \approx a + b\sum_{i=1}^{n}\left(508 + MT_i\right)$$

Experiment Results: Finger Touch

- Mean task completion time increases from 5.9*s* to 6.3*s* when the key size decreases from 15*mm* to 10.5*mm* (ANOVA, $p < .001$)
- Gap size does not have a significant effect on mean task completion time and error rate
- Mean error rate increases from 1.9% to 5.1% when the key size decreases from 15*mm* to 10.5*mm* (ANOVA, $p < .001$)
- Fit of prediction model is $R = .44$ ($p < .01$) with Fitts' law describing movement time

Centering Model

- Fitts' law is not well suited for aiming tasks with small amplitudes in dual-task situations
- Task is more focused and centering the probe over the target is more deliberate
- Therefore, Fitts' law must take probe size and target area into account:

$$MT = a + b\log_2\left(\frac{D \times P}{A} + 1\right)$$

- Fit improves to $R = .72$ ($p < .001$)

Experiment Results: Stylus Touch

- Mean task completion time increases significantly from 5.4*s* to 5.8*s* when the key size decreases from 15*mm* to 10.5*mm* (ANOVA, $p = .05$)
- Mean error rate increases from 1.5% to 3.4% when the key size decreases from 15*mm* to 10.5*mm* ($p < .001$)
- Compared to finger touch, the mean task completion time and error rate are significantly lower ($p < .001$)
- Fit of prediction model is $R = .58$ ($p < .01$) with Fitts' law, but $R = .67$ ($p < .001$) with Centering Model

Experiment Results: Trackball

- Mean task completion time increases significantly from 8.8*s* to 9.6*s* when the key size decreases from 15*mm* to 10.5*mm* (ANOVA, $p < .0001$)
- Mean error rate (2.2%) does not show a significant difference between the 10.5*mm* and 15*mm* key sizes
- Trackball has the highest mean task completion time, but the lowest error rate
- Difference in error rates between trackball and stylus is not significant, but is significant compared to finger touch
- Fit of prediction model is $R = .50$ ($p < .01$) with Fitts' law, but $R = .80$ ($p < .001$) with Centering Model

Conclusions: Experiment V

- Centering Model is a better model in dual-task situations and should be used for touch input instead of Fitts' law
- Touch input is convenient and simple for mobile devices, but utilizing a stylus is preferable for high accuracy tasks in which task completion time and accuracy are critical
- Buttons should be at least 15*mm* in size and present a maximized target area with an edge-to-edge gap of at least 5*mm*
- Model of human motor-control is dependent on the task type and is subject to task interference

Looking at the Navigator Again

- Buttons do not presenting maximum area
- Buttons are too small (10×12*mm*)
- Gap size is too small (3*mm*)
- No stylus (stylus is faster and has higher accuracy); no trackball (most accurate)
- Lower accuracy during motion requires fast error correction
  - virtual cursor pad not sufficient
  - task completion time > neglect time
  - potentially unsafe to use
  - high error rate frustrates user

Mobile Phone Dialing

- Informal experiment shows that *TCTM* has the potential for predicting dialing time on mobile phones
- New phones have recessed keys flush with surface similar to touch input
- Mean dialing time of a ten-digit phone number was found to be 5.4*s*
- Model predicts dialing time between 4.7*s* and 6.0*s* depending on key movement prediction model
- Goodness-of-fit of $R^2 = .86$ ($p = .02$)

Summary of Contributions

- Investigation of Fitts' law for multiple input devices in alternate postures (standing and walking)
- Investigation of Fitts' law in dual-task situations
- Formulation of Centering Model for touch input in dual-task situations that better models motor-control behavior
- Validation of Dual-Phase Model for acquisition of vibrating targets
- Derivation and validation of Task Completion Time Model for keypad entry tasks
- Clarification of design heuristics of touch interfaces that maximize performance and accuracy
- Construction of an open and extensible research tool for conducting evaluations of movement time models

Future Work

- Extend the task completion model to steering tasks and extend Accot & Zhai's (1997) Steering Law
- Verify the model *in situ* aboard a vessel
  - *MTE* has already been extended to take input from an accelerometer
- Expand the kinematic analysis capabilities of *MTE*
- Conduct experiments to determine if Laws of Decision hold in dual-task situations
- Verify applicability of Dual-Phase Model to web advertisements that use jittering images
- Conduct usability studies of navigation systems and compare against the proposed guidelines

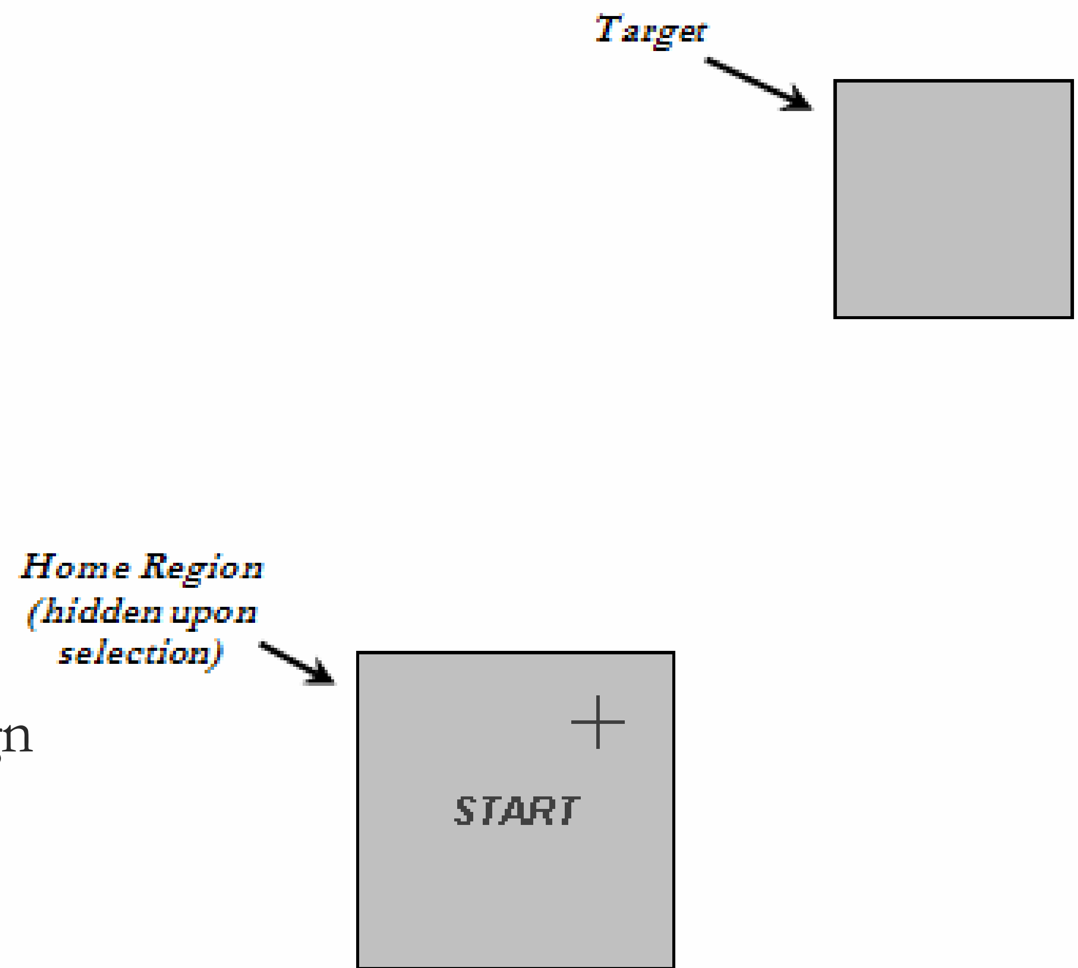
Thank you!

Questions or Comments?

Research Tools

- Most of the tools used in this research area are *ad hoc*, with the possible exception of:
  - *WinFitts* (Douglas *et al.*, 1999)
    - C++ with MFC
    - Configurable only through code changes
  - *Generalized Fitts Law Model Builder* (MacKenzie *et al.*, 1993)
    - Macintosh and DOS
- Interactive exploration is difficult with these tools
- *MTE* was developed as an open source, extensible, and configurable experimental platform for rapid exploration of movement time models

Capabilities of *MTE*

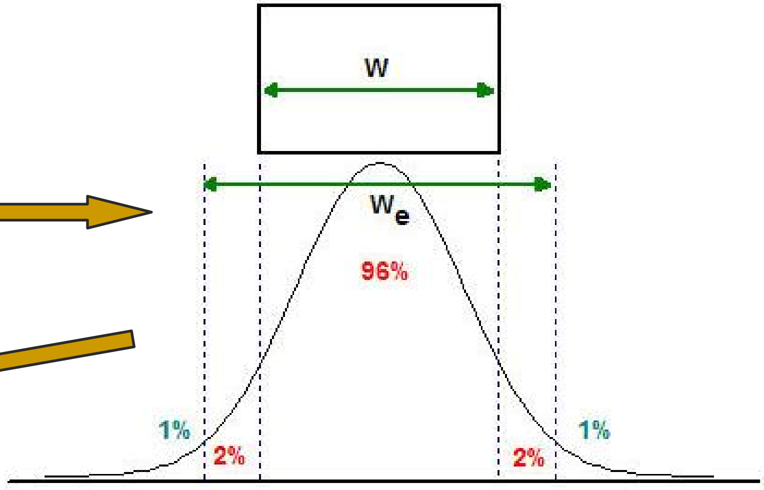


- Experiment configuration
- Comprehensive GUI
- Data set manipulation
- Statistical analysis
- Model evaluation
- Remote execution
- Files in XML and CSV
- Extensibility through design patterns
- Exports to "R"
- Sharing of results and experimental configurations for between-study analysis

Effective Width

$$H = \log_2\left(\sigma\sqrt{2\pi e}\right) = \log_2(4.133\sigma)$$

$$z \pm 2.066$$



$$W_e = 4.133 \times \sigma$$

deviation of selection end points from mean

- 4% error rate is considered to be normal
- Normalize all error rates into this range so that comparisons across studies and devices are meaningful

Alternative Models

- Alternative models provide more accurate predictions in some cases:
  - Meyer's Law

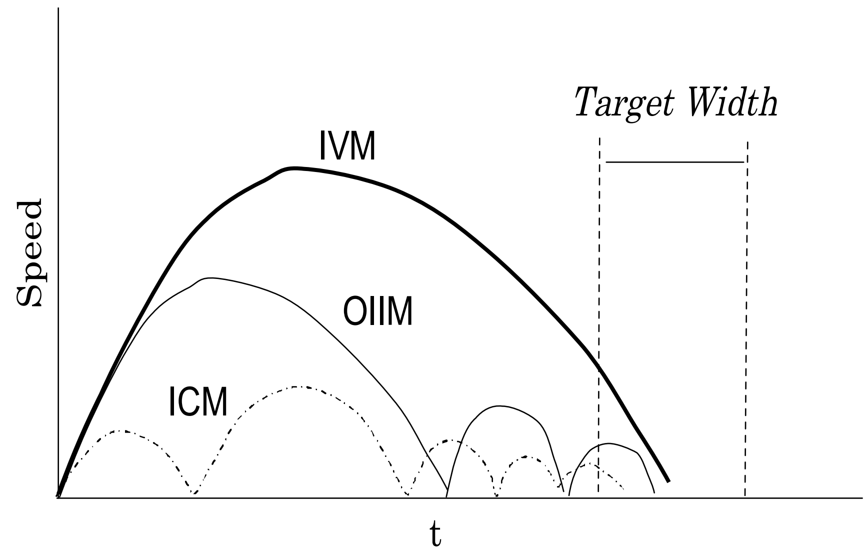
$$T = a + k\sqrt{\frac{D}{W}}$$

  - Kvålseth's Law

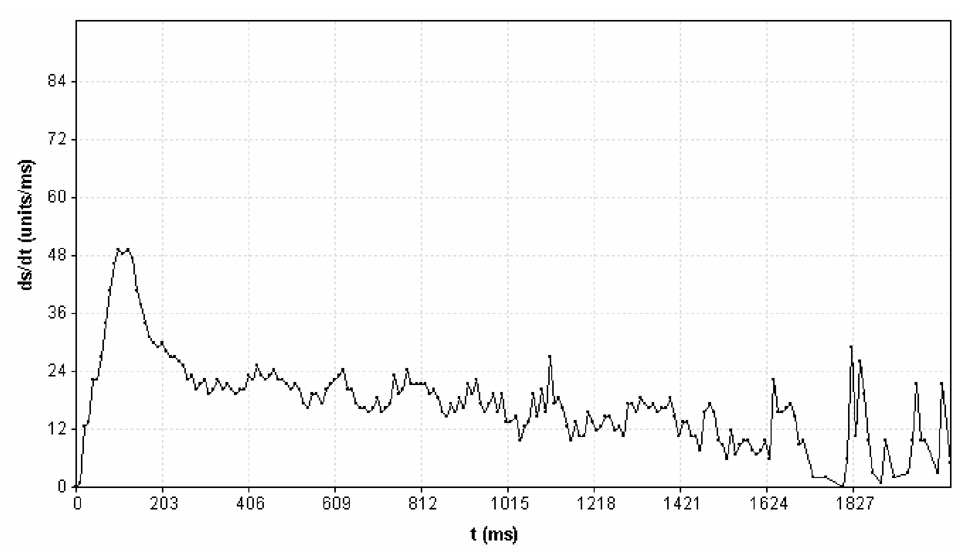
$$T = a + b\left(\frac{D}{W}\right)^c$$

Kinesthetic Models

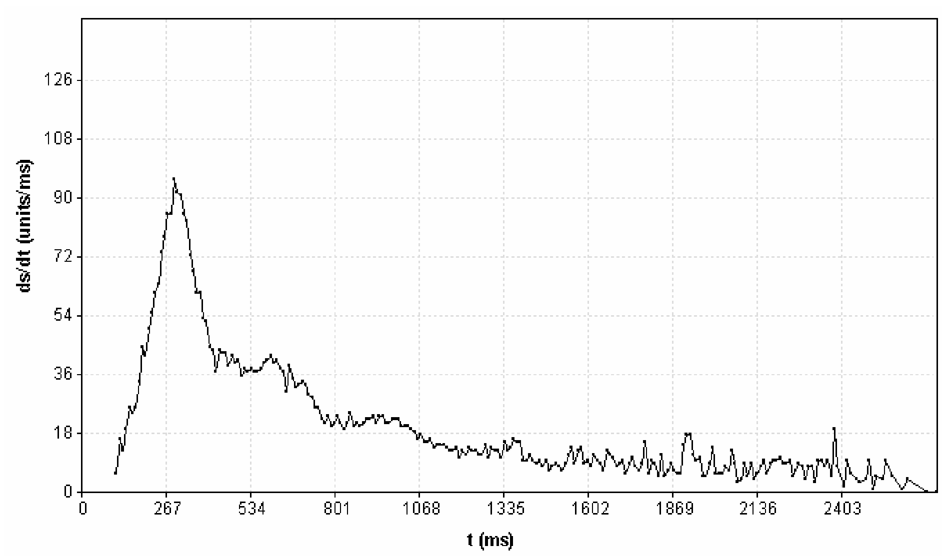
- Different theories of human motor control:
  - iterative corrections model (*ICM*)
    - series of discrete submovements
  - impulse variability model (*IVM*)
    - initial muscle impulse then gliding
  - optimized initial impulse model (*OIIM*)
    - initial impulse, some corrective submovements
- *ICM* and *OIIM* presume a closed-loop with visual and proprioceptive feedback
- Fitts' Law as approximation



Conformance to Motor Control Models



small amplitude movements in dual-task situations



standard Fitts aiming tasks

- Results suggest that multiple motor control models are at work
- OIIM governs rapid aiming movements with large amplitudes
- ICM governs focused aiming movements with short amplitudes in dual-task situations
- Explains poor fit with Fitts' law

Fit of Prediction Model

- Correlation of *Movement Time* and *Index of Difficulty* ranges from .87 to .97 ($p < 0.001$)
- Results agree with other researchers' studies
- Fitts' law does not accurately predict performance for finger touch of targets less than about 15*mm*
- Relationship for small targets is in fact quadratic:

$$MT_{touch} = 269 + 63ID^2$$

Hick-Hyman Law

- Model for predicting choice reaction time when making a decision among *n* choices:

$$RT_n = RT_1 + kH$$

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

where *k* is empirically derived, p_i is the probability of choice *i*.

- Alternative model by Kvålseth (1996):

$$RT_n = 200\sqrt{n}$$

Law of Practice

- The time it takes to perform a task on the n^{th} try can be stated as:

$$T(n) = T_1 \frac{1}{n^a}$$

where a is an empirically derived constant and T_1 is the time of first try.

- Alternative model by Heathcote *et al.*:

$$T(n) = T_1 e^{-\alpha n}$$

Change of ID with D and W

d/w	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
25	1.81	1.17	0.87	0.70	0.58	0.50	0.44	0.39	0.35	0.32
50	2.58	1.81	1.42	1.17	1.00	0.87	0.78	0.70	0.64	0.58
75	3.09	2.25	1.81	1.52	1.32	1.17	1.05	0.95	0.87	0.81
100	3.46	2.58	2.12	1.81	1.58	1.42	1.28	1.17	1.08	1.00
125	3.75	2.86	2.37	2.04	1.81	1.62	1.48	1.36	1.26	1.17
150	4.00	3.09	2.58	2.25	2.00	1.81	1.65	1.52	1.42	1.32
175	4.21	3.29	2.77	2.43	2.17	1.97	1.81	1.67	1.56	1.46
200	4.39	3.46	2.94	2.58	2.32	2.12	1.95	1.81	1.69	1.58
225	4.55	3.61	3.09	2.73	2.46	2.25	2.08	1.93	1.81	1.70
250	4.70	3.75	3.22	2.86	2.58	2.37	2.19	2.04	1.92	1.81
275	4.83	3.88	3.35	2.98	2.70	2.48	2.30	2.15	2.02	1.91
300	4.95	4.00	3.46	3.09	2.81	2.58	2.40	2.25	2.12	2.00
325	5.07	4.11	3.56	3.19	2.91	2.68	2.50	2.34	2.21	2.09
350	5.17	4.21	3.66	3.29	3.00	2.77	2.58	2.43	2.29	2.17
375	5.27	4.30	3.75	3.38	3.09	2.86	2.67	2.51	2.37	2.25
400	5.36	4.39	3.84	3.46	3.17	2.94	2.75	2.58	2.44	2.32