

Understanding Multitasking Through Parallelized Strategy Exploration and Individualized Cognitive Modeling

Yunfeng Zhang and Anthony J. Hornof

Computer and Information Science

University of Oregon

Eugene, OR 97403 USA

{zywind, hornof}@cs.uoregon.edu

ABSTRACT

Human multitasking often involves complex task interactions and subtle tradeoffs which might be best understood through detailed computational cognitive modeling, yet traditional cognitive modeling approaches may not explore a sufficient range of task strategies to reveal the true complexity of multitasking behavior. This study proposes a systematic approach for exploring a large number of strategies using a computer-cluster-based parallelized modeling system. The paper demonstrates the efficacy of the approach for investigating and revealing the effects of different microstrategies on human performance, both within and across individuals, for a time-pressured multimodal dual task. The modeling results suggest that multitasking performance is not simply a matter of interleaving cognitive and sensorimotor processing but is instead heavily influenced by the selection of subtask microstrategies.

Author Keywords

Cognitive modeling; high performance computing; model comparison; multimodal; multitasking; task strategies.

ACM Classification Keywords

H.1.2 [User/Machine Systems]: Human information processing; I.6.4 [Model Validation and Analysis]; I.2.0 [General]: Cognitive simulation.

INTRODUCTION

A contemporary lack of understanding of the human ability to multitask—to pursue two goals in parallel, with each goal requiring and competing for the same human information processing components such as the high resolution vision at the point of gaze—puts humans in peril in human-computer interaction contexts [13]. As manufacturers race towards profit and consumers towards perceived social and lifestyle benefits, this lack of understanding of fundamental multitasking abilities paves the road for new interface designs that cause great harm [15]. As with seatbelt use, sweeping reform will not likely take place until it is mandated by law [19] but the first hurdle we face on the road to improving contemporary

driving safety (as well as regaining a human social presence in the physical world of everyday life) is to determine the contemporary equivalent of “crash testing” human-computer interfaces to better understand how people are likely and able to interleave computer use with other primary life-critical tasks.

Since the dawn of HCI, cognitive modeling has been making progress in providing theory and methods needed to predict the usability of routine behavior on standalone devices for well-defined user tasks with clear goal and task structures [3]. Paper-based cognitive modeling [3] has evolved into computational cognitive modeling [11] which permits increased complexity and predictive power in the models. Eye tracking has emerged as a powerful technique for revealing in greater detail how people are truly interacting with devices [6] and can also reveal nuanced interactions between parallelized task hierarchies [8].

Recent theoretical advances, especially the work of Howes, Lewis, and Vera [9] on cognitively-bounded rational analysis, have opened the door for even more rigorous theoretical exploration and determination of the cognitive strategies that people use when multitasking. In addition to exploring strategies, there is also the challenge of appropriately exploring and adjusting a model’s perceptual, motor, and memory parameters to capture an individual’s “personal equation” [17], though much contemporary cognitive modeling and predictive engineering still presumes a single hypothesized average user.

This paper presents recent advances in all of these areas of cognitive modeling by presenting a rigorous exploration of cognitive strategies and individualized parameters for a highly-interactive multimodal dual task. This project pushes the boundaries of computational cognitive modeling by employing a high-performance computing cluster to explore a large strategy and parameter space. All told, this paper reveals the nuanced strategies with which humans accomplish complex optimized multitasking performance.

THE MULTIMODAL DUAL TASK

The multimodal dual task experiment was originally designed by Ballas, Heitmeyer, and Perez at the Naval Research Laboratory [1], and has already proven useful for exploring multitasking performance through cognitive modeling [7, 12]. The work presented here models the data collected from our replication of the experiment [8], in which eye movement data were collected to inform detailed analysis and modeling.

© 2014 Association for Computing Machinery. ACM acknowledges that this contribution was authored or co-authored by an employee, contractor or affiliate of the United States government. As such, the United States Government retains a nonexclusive, royalty-free right to publish or reproduce this article, or to allow others to do so, for Government purposes only.

CHI 2014, April 26 - May 01 2014, Toronto, ON, Canada
Copyright 2014 ACM 978-1-4503-2473-1/14/04...\$15.00.
<http://dx.doi.org/10.1145/2556288.2557351>

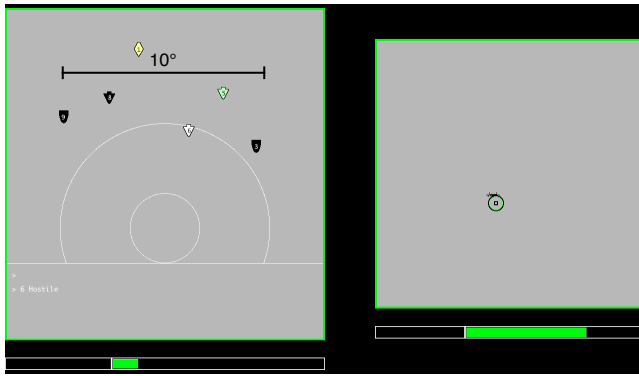


Figure 1. A screenshot of the multimodal dual-task display.

Figure 1 shows a screenshot of the display used in the dual task experiment. On the left is the classification task, and on the right is the tracking task. The classification task requires the participant to monitor icons or “blips” that move down the classification window and to key-in, with the left hand, the blip number and its “hostility” classification as soon as the blip changes color from black to red, green, or yellow. Red indicates hostile, green indicates neutral, and yellow indicates unknown such that the participant needs to further study the blip’s shape, speed, and direction to determine its hostility based on a set of memorized rules. The tracking task requires the participant to keep the cursor (the circle) on a randomly moving target using the joystick with the right hand. When the cursor is within 20 pixels of the target, the cursor is green and the participant earns money. When the cursor is more than 40 pixels away from the target, the cursor is red and the participant loses money. Participants earned financial bonuses in both tasks based on carefully designed payoff schemes that reward fast and accurate performance.

To study the effects of perceptual information on multitasking performance, two factors were manipulated, each with two conditions: (a) *peripheral visible* or *peripheral not visible* and (b) *sound on* or *sound off*. In the peripheral-not-visible conditions, the contents of the window that the participant was not currently looking at were hidden using a gaze-contingent display. This simulates a task environment in which two windows are separated by enough distance such that the contents of the peripheral window cannot be monitored with peripheral vision. In the sound-on conditions, spatialized auditory cues were played to indicate a blip’s appearance and color change in the classification window.

Twelve participants completed the experiment on three consecutive days. Performance improved slightly from day to day, but with no clear indication of shifts in strategy. The data presented here are from the third day, and from the ten participants who achieved good overall performance. More details of this experiment are discussed in [8].

Figure 2 shows the overall dual task performance by plotting the average classification time against the root-mean-square (RMS) tracking error, for each of the four conditions, with a different plot symbol for each of the two

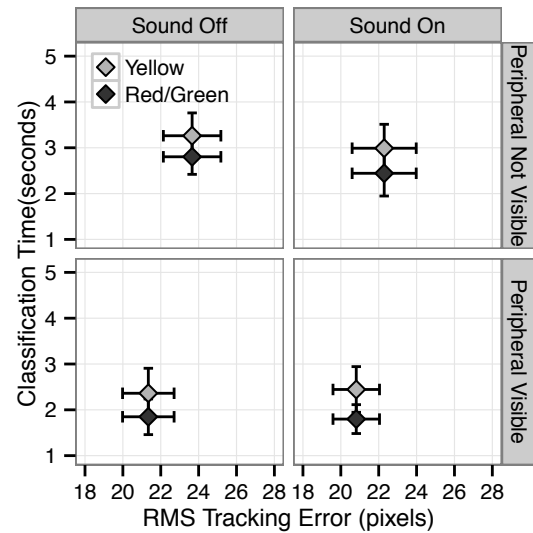


Figure 2. The classification time and tracking error across the four conditions, for yellow and red-and-green blips, and the 95% confidence intervals (CI) of the participant means.

kinds of blips. The general trend is that more perceptual information resulted in better performance in both the classification and the tracking task. For example, note in Figure 2 that the performance shown in the bottom right panel, in which both auditory and visual peripheral cues were present, is better than in the top-left panel, in which there were no peripheral cues. However, the effect of sound is only significant when the peripheral window is not visible, as is suggested in that the performance shown in the top right panel is better than in the top left panel, but the performance shown across the two bottom panels is roughly the same.

Figure 3 summarizes the most important eye movement data from the experiment. These data measure the time spent to complete each of the four steps involved in classifying a blip: (a) Initiate the eye movement from the tracking window to the classification window (the radar); (b) once on the classification window, find the target and move the eyes to it; (c) keep the eyes on the blip until its visual features are identified and then move the eyes back to tracking; and (d) after the eyes are back on tracking, key-in the classification. The main differences in the eye movement data across the four conditions are: (a) The eyes-to-radar time and the eyes-to-target time are shorter in the peripheral visible conditions; and (b) the eyes-to-radar time is shorter when sound is on in the peripheral-not-visible conditions. These differences could no doubt contribute to the effects observed in Figure 2, which we set out to investigate through cognitive modeling.

THE COMPUTATIONAL COGNITIVE MODEL

The models presented in this paper are built using the EPIC (Executive Process-Interactive Control) cognitive architecture [11]. EPIC provides a general software framework for simulating humans interacting with a task environment. It models a human as an information processing system with a cognitive processor and various perceptual and motor processors. The cognitive processor

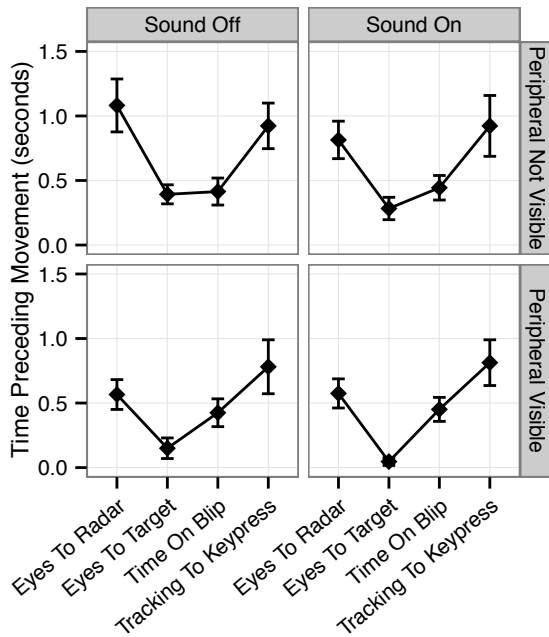


Figure 3. The time preceding the eye and hand movements needed to classify a blip. (The radar is the classification window.)

decides what actions to take based on perceptual information and preprogrammed task strategies which are written by an analyst as a set of IF-THEN statements called production rules. EPIC is particularly well-suited for exploring multitasking strategies; it only enforces sequential processing for motor activities, such as to constrain the hands to execute only one movement at a time. Perceptual information flows into the visual and auditory processors in parallel, though visual acuity degrades as objects move farther away from the center of the gaze. The cognitive processor can execute multiple production rules in a single 50 ms cycle. This view on multitasking—that serial processing bottlenecks only exist in the motor processing stage—is supported by previous psychological studies (e.g., [18]).

Figure 4 illustrates the basic structure of the multitasking strategy that we identified for the multimodal dual task (in [7]). The strategy consists mainly of two independent sets of production rules, represented abstractly by the top and the bottom state transition diagrams in the figure. One set of production rules summarized in the figure passes control of the ocular motor (eye movement) processor back and forth between the two tasks, while another set of rules passes control of the manual motor (hand movement) processor back and forth. This independent interleaving of ocular and manual motor processing was found (in [7]) to explain performance better than a hierarchical task decomposition, and is thus used across all strategies and all conditions.

The transition of the ocular motor processing from tracking to classification does not mean that the eyes are instructed at that point in time to move to the classification window. The decision of when to actually move the eyes is handled

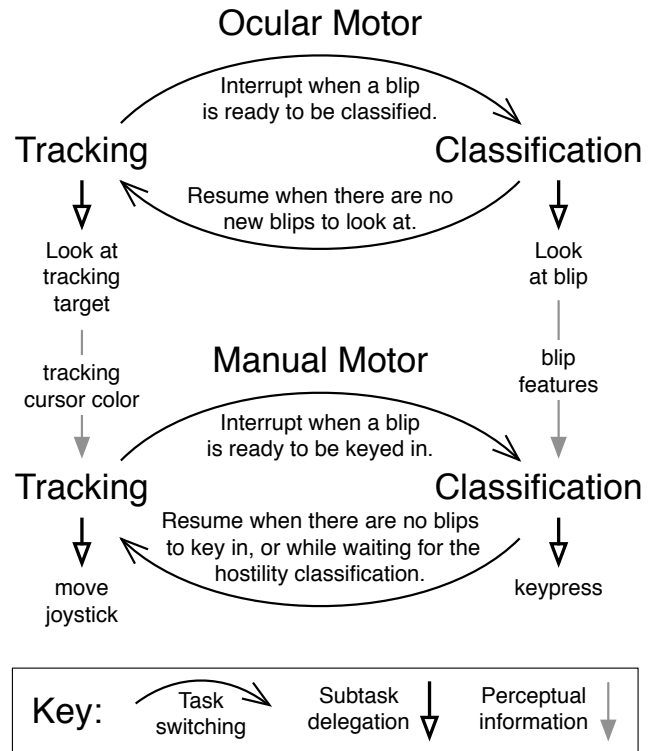


Figure 4. State transition diagrams showing the independent interleaving of the ocular motor (eye movement) and manual motor (hand movement) processing in the dual task model.

by decisions within the strategies which will be discussed later. Across all strategies, however, the decision of when to transition the ocular motor processing to classification is handled differently in different experimental conditions. The transition occurs in the peripheral-visible conditions when a blip color-change is detected, in the sound-on conditions when an auditory cue is detected, and in the sound-off and peripheral-not-visible condition when a self-paced periodic glance is initiated.

Some aspects of this basic strategy can be modified to represent variations in participant behaviors, such as to prioritize one task over the other, and this paper explores these variant strategies. In a previous study [7], we demonstrated that one variant strategy could fit the average participant performance. The present study goes beyond the previous study and introduces a modeling technique which provides a new principled approach to addressing some general theoretical modeling issues such as how to (a) systematically find strategies that explain the data, (b) minimize the model's degrees of freedom, and (c) explore a large strategy-and-parameter space in a reasonable amount of time.

STRATEGY AND PARAMETER SPACE EXPLORATION

A Systematic Approach for Exploring Strategies

For many tasks, a reasonable high-level structure for the task, such as that shown in Figure 4, can be identified through straightforward task analysis. But for many such tasks, the lower-level subtasks might be accomplished with numerous different substrategies, or *microstrategies* [5].

This view, that people exert cognitive control to select alternative task strategies, stands in contrast with modeling work (e.g., [16]) that assumes fixed strategies. As demonstrated by Gray and Boehm-Davis [5], people use different microstrategies even for tasks as simple as mouse-clicking under different task settings. Thus, it is necessary to explore many microstrategies to determine which strategies best explain human multitasking performance.

Conventional cognitive modeling studies (such as [16]) often present only one model that fits the human data. This is problematic because to show that a certain task strategy is necessary to explain the human data, one also needs to show that other strategies *cannot* fit the data [14]. With little or no information given about what other strategies were explored and why some strategies were ruled out, the modeling results and the conclusions based on them cannot be properly evaluated.

In the last decade, cognitive modeling researchers have started to explore strategies more systematically. Some studies (e.g., [2, 9, 10]) have explored strategies that can be numerically parameterized, such as in deciding how many keys to press before switching tasks, by sampling these parameters at regular intervals and executing all of the resulting strategies. Though this method is effective, many strategies that people actually employ include alternatives that cannot be parameterized numerically, such as in deciding whether to move the mouse before or after finding a target. To explore the effects of these sorts of alternative task strategies, Kieras and Meyer [12] proposed a bracketing heuristic in which a fastest-possible model and a slowest-reasonable model are used to “bracket” a reasonable range of task performance. However, the fastest and the slowest models may be hard to determine because the effects of all possible substrategies can be hard to predict without actually trying them all.

This paper proposes a systematic approach for exploring strategies, and applies it to the multimodal dual task model. The main idea of this approach is to organize plausible strategies into a well-defined multi-dimensional space with each dimension representing one aspect of the task that can be accomplished with alternative microstrategies. These dimensions can include alternatives that cannot be parameterized numerically. The strategy space can be thought of as a multidimensional grid, with each point in the grid representing a specific combination of microstrategies, with each microstrategy from a different dimension. By exploring all the strategies defined by the strategy space, we can see the complete range of the model’s predicted task performance.

Table 1 shows the strategy space defined for the multimodal dual task. The strategy space consists of five dimensions, each of which contains two to three alternative microstrategies, and each of which relates to the task switching and subtask delegation shown in Figure 4. Dimension 1 controls how the model delegates the tracking subtask. Dimensions 2, 3, and 4 control when the model moves the eyes from the tracking window to the

Dimension 1. How judiciously to track the target.

Joystick-Nonstop (JNS). Track whenever possible.

Joystick-To-Green (JTG). Track whenever necessary to keep the tracking cursor green.

Dimension 2. When to move the eyes to the classification window after seeing a blip change color.

Immediate-Eyes-To-Blip (IEB). Move immediately.

Track-Then-Eyes-To-Blip (TEB). Continue tracking until the tracking cursor is green and then move.

Dimension 3. When to move the eyes to the classification window after hearing an auditory cue.

Immediate-Eyes-To-Sound (IES). Move immediately.

Track-Then-Eyes-To-Sound (TES). Track until the auditory cue is classified and then move the eyes.

Dimension 4. In the peripheral-not-visible conditions, where exactly to put the eyes in the classification window when switching to the classification task.

Look-Window-Center (LWC). Go to the center.

Look-Sound-Location (LSL). Go to the exact location of the spatialized sound.

Look-Prior-Blip-Location (LBL). Go to a blip location recalled from a previous visit.

Dimension 5. What to do with the hands after acquiring the visual features of a yellow blip but while waiting for the hostility classification to be retrieved from memory.

Keypad-Then-Joystick (KTJ). Wait for the retrieval, key in the response, and then resume tracking.

Joystick-Until-Classification (JUC). Track while waiting.

Interruptible-Tracking-Preparation (ITP). Track while waiting but permit an immediate interruption upon retrieval.

Table 1. The strategy space explored in the dual task model. The space includes five strategic dimensions with each dimension having two or three alternative microstrategies.

classification window, and where exactly it puts the eyes in the classification window. Dimension 5 controls how the model will sometimes switch manual motor from classification to tracking in the middle of a classification.

These dimensions relate to some important general issues in multitasking. For example, Dimension 2 explores the effect of prioritizing one task over another. Dimensions 3 and 4 explore to what degree can people utilize information in auditory cues to assist with a secondary task. Dimension 5 examines what extent of overlapping is possible between the cognitive processes of one task (classifying a blip) and the manual motor processes of another (manual tracking). All of these issues are critical for understanding human performance in a wide range of multitasking contexts such as coordinating with specialists while piloting a drone, operating a navigation system while driving a car, or using a smartphone while walking down the street. By exploring this strategy space, we may learn something about the strategies that people use for general classes of time-critical, multimodal multitasking.

A Principled Approach for Setting Model Parameters

Aside from strategy exploration, another hurdle in conducting principled cognitive modeling is finding parameter settings that accurately characterize task-specific cognitive, perceptual, and motor capabilities. Default settings have been established for many parameters based on past psychological research, but for some tasks, parameters have not been established. To set these unknown parameters, a common practice is to sample many different settings until the model fits the data, but this approach increases the model’s degrees of freedom and weakens its predictive power.

A more principled way to set the model parameters, advocated by Howes et al. [9], is to calibrate them using data sets that are not used to evaluate the model’s goodness of fit. This will reduce the variability of the model when fitting the evaluation data, and hence increase the model’s predictive power. We adopted this approach and calibrated the following parameters for our dual task model:

- *Text recoding time.* This parameter determines how long it takes to recognize the blip ID. It was calibrated using the fixation time on red and green blips.
- *Speed and direction recoding time.* This parameter determines how long it takes to acquire the yellow blips’ visual features for hostility classification. It was calibrated using the fixation time on yellow blips.
- *Tracking coefficients.* These parameters determine the tracking movement time for any tracking distance. They were calibrated by fitting a Fitts’ law equation to the long tracking movements extracted from the tracking error data (see [20] for a full discussion of this).
- *Glance interval.* This parameter determines how often the model should check for new blips in the sound-off and peripheral-not-visible condition. It was calibrated using the eye movement data from that condition.

None of these parameters were fitted directly to the evaluation data, which are the classification time and the RMS tracking error. This greatly reduces the model’s degrees of freedom. However, not all parameters could be calibrated. Our model still has one free parameter—the hostility recoding time for yellow blips. Based on some preliminary exploration, we determined that the hostility recoding time could range from roughly 100 ms to 1200 ms, and so we sampled the parameter at 100 ms intervals to find the best-fitting setting for each participant. The parameter settings were explored in conjunction with strategy exploration using our parallelized cognitive modeling system, which we introduce next.

A Parallelized System for Cognitive Modeling

Fully exploring strategic dimensions and parameter settings requires a large number of models. In our case, there are 72 different strategies (the combinatorial product of the five strategy dimensions) and 12 settings for the hostility recoding time parameter, resulting in 864 models. Running one dual task model takes about 1 minute on a single core 1.8 GHz machine, and so running 864 models would

require about 14 hours. If we were to model individual participant performance, this number would be multiplied by the number of participants. This long execution time makes it impractical to explore a large strategy and parameter space using a traditional modeling system on a contemporary desktop or laptop computer.

To solve this computational problem, we built a parallelized cognitive modeling (PCM) system that utilizes a computer cluster to speed up this large scale exploration. The specific cluster used here is part of the University of Oregon’s Applied Computational Instrument for Scientific Synthesis. Though previous research [4] used a computer cluster to explore model parameters, our system pushes the boundary further by using a cluster to also explore task strategies.

Figure 5 shows the components of the PCM system and illustrates how it generates and parallelizes multiple models at once. The system takes three files as inputs: the basic model, the parameter space, and the strategy space. The basic model defines the basic structure of the strategy, such as the one shown in Figure 4, using production rules; it also includes slots that will be filled in later with specific parameter and strategy settings to generate complete models. The parameter space defines the ranges and sampling intervals for the model’s free parameters. The strategy space includes instructions about how to modify

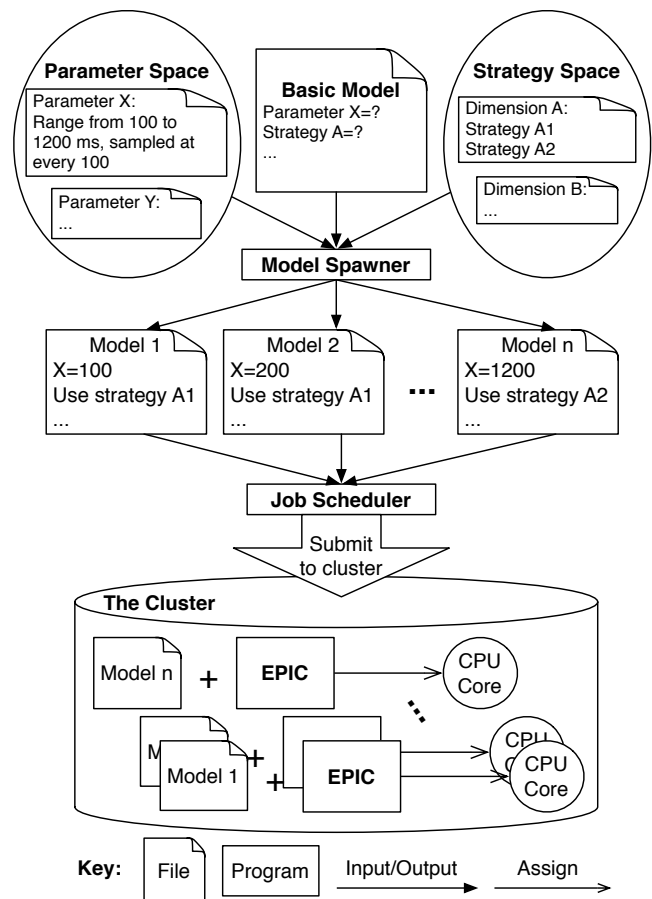


Figure 5. The components of the Parallelized Cognitive Modeling (PCM) system.

the basic model to implement each microstrategy. The instructions include which production rules to modify, and what conditions and actions to add and delete. The three files are fed into the *model spawner*, which parses the parameter definitions and the strategy modification instructions. The spawner then modifies the basic model to generate versions of the model that together span across all possible strategy and parameter settings. The generated models are then arranged by the *job scheduler* which assigns each model to run on a CPU core in the cluster. By using the PCM system, the original 14-hour running time on a desktop machine is reduced to less than 20 minutes. This enables us to conduct truly large scale modeling and to collect much more comprehensive results than has been feasible in previous cognitive modeling studies.

RESULTS

Modeling The Average Human Performance

This section presents the results of using the parallelized cognitive modeling (PCM) system to explore the strategy and parameter space defined earlier for the multimodal dual task. Figure 6 compares the predictions of all 864 models with the aggregated human data. As in Figure 2, each panel plots the average classification time against the RMS tracking error for one of the four conditions. The diamonds show the human data. The data clouds show the model predictions, with one data point per model. The results for yellow blips and for red-and-green blips are shown separately. The more the clouds fall within the human data brackets, the better the models explain the observed data.

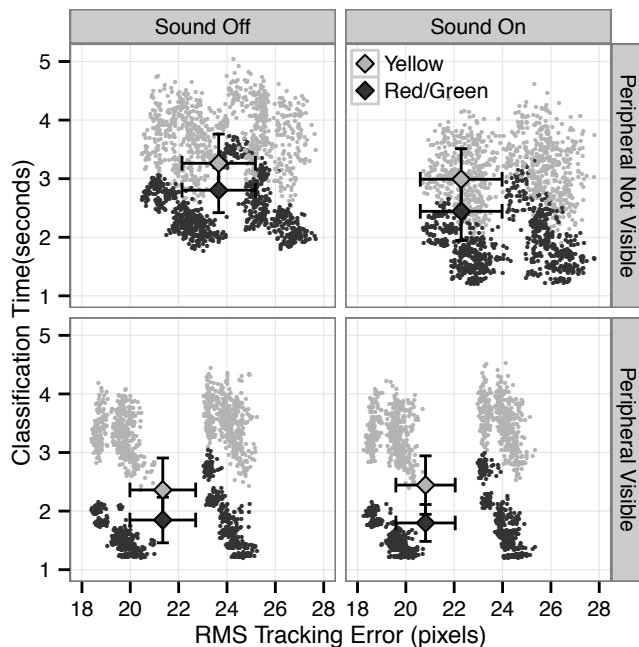


Figure 6. The observed and predicted classification time and RMS tracking error for each of the four experimental conditions. The diamonds show the participant means and the 95% CIs. The clouds show the predictions from the 864 different models, with the light gray plot symbols for yellow blips and the dark gray symbols for red and green blips.

Though the clouds in Figure 6 span large areas, the clouds still show trends that are similar to those observed in the human data. These trends include: (a) The clouds in the bottom panels are generally lower and more to the left than those in the top panels indicating that, in the peripheral-visible condition, the models predict better performance for both classification and tracking; (b) the clouds in the top-left panel are higher and more to the right than those in the top-right panel indicating that, in the peripheral-not-visible conditions, the model predicts better performance for both tasks when auditory cues are present; and (c) the clouds in the bottom two panels look virtually the same suggesting that, when the peripheral window is visible, the model predicts that auditory cues have no effect. All of these effects are consistent with the human data, suggesting that the model’s basic strategy captures the main effects of the experimental conditions.

Though the models predict the effects of auditory and visual cues, they do not fit the human data in all conditions. Particularly in the peripheral-visible conditions, the human data fall between the two parts of the clouds as opposed to on the clouds. Figure 7 shows that the left and right parts of the clouds shown in each frame in Figure 6 result from the two Dimension 1 microstrategies. The left part of the cloud is produced by the Joystick-Nonstop (JNS) microstrategy, and the right part is produced by the Joystick-To-Green (JTG) microstrategy. The JNS microstrategy clearly produces a smaller tracking error than the JTG microstrategy. In all conditions, the human data are closer to the predictions of JNS than to those of JTG, which suggests that the participants tried hard to optimize tracking

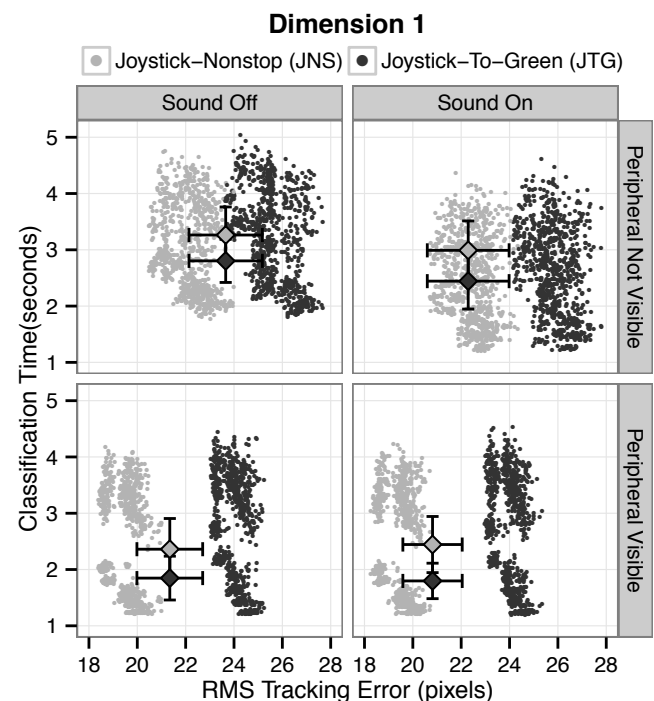


Figure 7. The effect of the microstrategies of Dimension 1. The data are the same as in Figure 6, but the clouds are shaded differently to distinguish the two microstrategies.

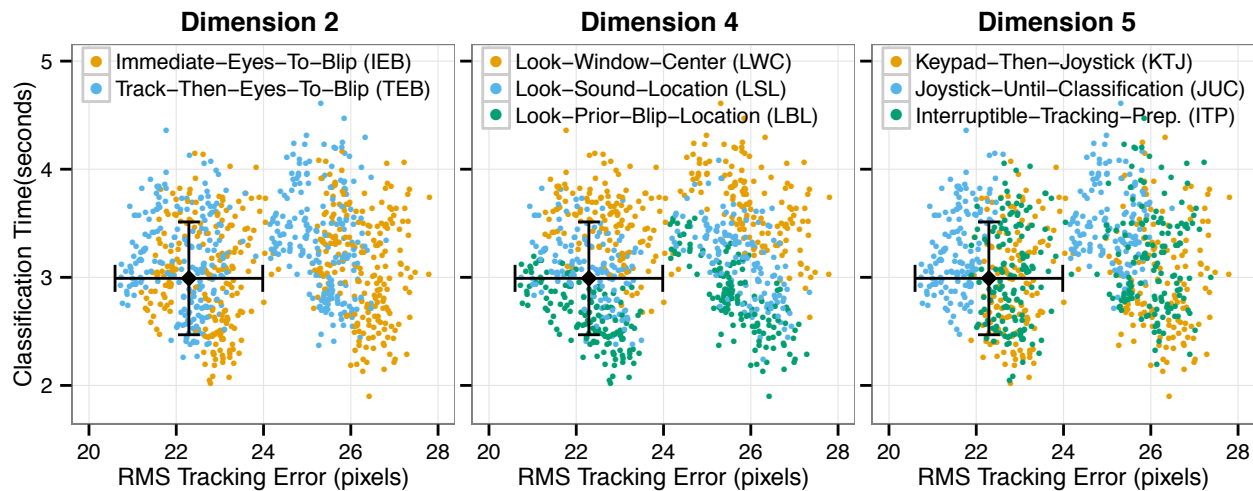


Figure 8. The effects of different microstrategies from Dimensions 2, 4, and 5 shown in different colors. Each panel shows the human data and model predictions for the sound-on and peripheral-not-visible condition.

performance. The difference in the cloud densities between the top two panels and the bottom two panels suggests an interaction between peripheral visibility and some other aspect of the model.

Figure 8 continues the examination of the effects of microstrategies for Dimensions 2, 4, and 5 to find out which strategies explain the data better. The results for Dimension 3 are not shown because the microstrategies in that dimension did not generate a range of different predictions. Each panel in Figure 8 shows only the results from the sound-on and peripheral-not-visible condition, and for yellow blips. The effects of microstrategies in other experimental conditions are similar.

The first panel in Figure 8 shows the effect of Dimension 2. There is a small difference between the Immediate-Eyes-To-Blip (IEB) microstrategy, which immediately moves the eyes to the classification window, and the Track-Then-Eyes-To-Blip (TEB) microstrategy. The prediction cloud of TEB is generally higher and more to the left than that of IEB, which illustrates that, between the two microstrategies, TEB prioritizes tracking whereas IEB prioritizes classification. However, judging from this graph alone, it is unclear which strategy leads to a better overall payoff. Answering this question requires a more comprehensive analysis, presented later.

The second panel in Figure 8 shows the effect of the Dimension 4 microstrategies, which control where to look in the classification window. Because the Look-Window-Center microstrategy usually requires additional visual search to find the blip, it is not surprising that it predicts the longest classification time (the yellow cloud). It is, however, surprising that the Look-Sound-Location (LSL) microstrategy (the blue cloud) takes longer to classify the blip than the Look-Prior-Blip-Location (LBL) microstrategy (the green cloud), which performs the best in both tasks. Further examination of the model shows that LSL is slower because extracting the location information from sound requires roughly 200 ms under the EPIC

cognitive architecture, and this delay exceeds the average time to put the eyes on the target by selecting a prior blip location from memory.

The third panel in Figure 8 shows the effect of the Dimension 5 microstrategies, which decide what to do with the hands after a blip has been fixated but before the hostility classification has been retrieved such that the blip's classification could be keyed-in. The Joystick-Until-Classification (JUC) and Interruptible-Tracking-Preparation (ITP) microstrategies use this time to perform the tracking task, whereas the Keypad-Then-Joystick microstrategy stays idle until the classification is retrieved. Intuitively, Strategies JUC and ITP should outperform ITP in both the classification and tracking tasks because they perform the two tasks simultaneously. However, the graph shows that although Strategies JUC and ITP predict better tracking performance, they also predict slightly longer classification time. Further examination of the model shows that prioritizing manual tracking sometimes postpones the classification response. This low-level interaction between the two tasks may be overlooked by theory-based qualitative reasoning but is easily captured by quantitative computational modeling.

In all three panels of Figure 8, the human data (the brackets) have a large variance across the predictions of all of the microstrategies. This again suggests that all participants did not use the same microstrategies. Individualized models are needed to explore how different participants likely selected different microstrategies and how this likely led to individual performance variations.

Modeling The Individual Performance

Figure 9 illustrates how the average-performance model fails to capture individual participant performance variations. The graph superimposes the observed data from each of the ten participants on the model predictions. To simplify the presentation, only the results for yellow blips are shown. As can be seen in the top panels, the observations for individual participants fall mostly on the

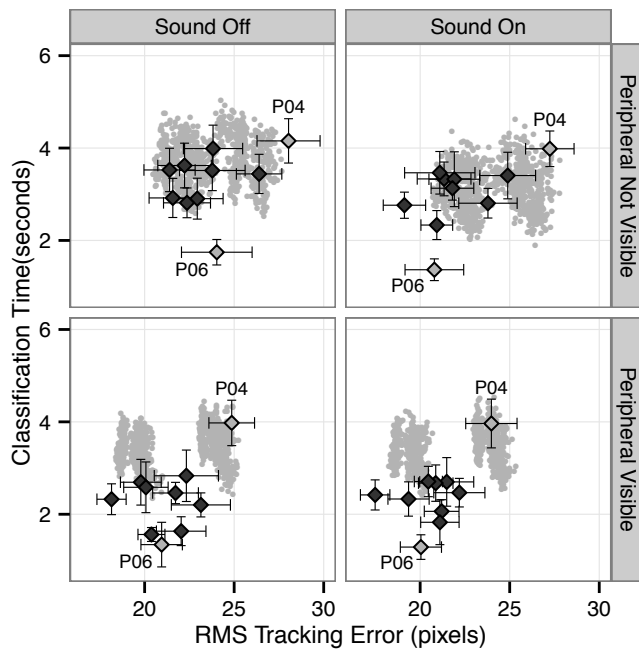


Figure 9. The individual data for each of the ten participants (the diamond-shaped plot symbols) compared to the predictions of the model built for the aggregated human data (the gray clouds). P04 and P06's data are indicated.

clouds, and in the bottom panels mostly between and below the clouds. That the clouds do not envelop the human data across all conditions shows that the model as it stands cannot fully account for individual participant performance.

Individualized models are needed to explain individual performance. Participants have different cognitive, perceptual, and motor capabilities and to generate accurate predictions for each participant these differences need to be reflected in the model's parameter settings for each participant. Such individualized parameter settings can be obtained by calibrating the parameters using individual data rather than aggregated data. Then, the strategy and parameter space can be explored in the same manner as was done for the average-performance model, but this time to find the best-fitting model for each individual participant.

Figure 10 shows the results of this individualized modeling for the top performer P06 and the bottom performer P04. (Again, only the results for yellow blips are shown.) Note how the prediction clouds for each condition shift across the two participants. When compared to Figure 9, it can be seen that the individualized-parameter models explain the individual data better than the model that was parameterized based on the aggregated human data. The plots for P06 in Figure 10 show that the clouds now approach P06's data more closely than the clouds in Figure 9. The top two panels for P04 in Figure 10 show the prediction clouds readily encompassing P04's data whereas, in the top two panels in Figure 9, P04 fell at the edge of the clouds. It is only through individualized modeling that we are able to account for the divergent behaviors in the top and bottom performers.

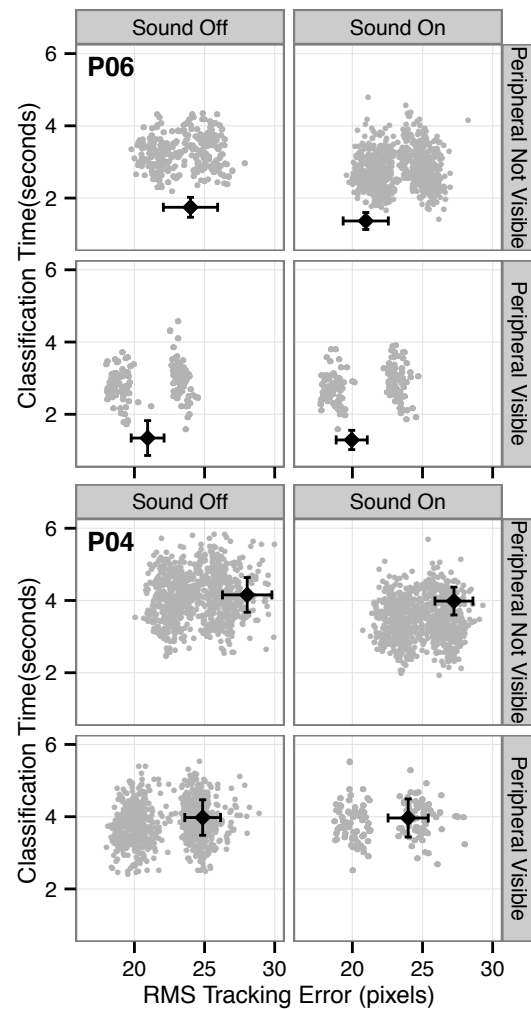


Figure 10. The observed data and predictions of the individualized models for P06 and P04.

The Best-Fitting Individualized Models

One benefit to building individualized models is that it helps to reveal the strategies that lead to good task performance, which has implications for training and personnel selection. This section thus compares the strategies of the top and bottom performers to identify the optimal microstrategies for this task.

Table 2 shows the best strategy and parameter settings produced by the PCM system to account for the performance of each of the ten participants, and to account for the average performance. The individualized models are sorted from best to worst participant performance as determined by the bonus that each participant earned. The best fitting strategy and parameter settings were determined by finding the model with the lowest average absolute error (AAE) in its prediction of each participant's classification response time (RT) and tracking error (TE). Table 2 illustrates which microstrategy selections appear to have contributed to good overall performance, and to some participants performing better than others.

	Participant	Strategy Dimension					Hostility Recoding Time (ms)	RT and TE (AAE)	Participant Bonus
		Dim1	Dim2	Dim3	Dim4	Dim5			
Top Performers	P06	JNS	IEB	IES	LBL	ITP	200	14%	\$17.86
	P20	JNS	IEB	IES	LSL	JUC	300	6%	\$17.85
	P07	JNS	IEB	IES	LBL	ITP	200	13%	\$17.63
	P17	JNS	TEB	TES	LBL	KTJ	100	7%	\$17.16
	P10	JNS	IEB	IES	LSL	JUC	500	8%	\$16.61
Bottom Performers	P11	JNS	TEB	TES	LWC	ITP	100	15%	\$16.45
	P09	JNS	TEB	TES	LSL	JUC	200	9%	\$14.75
	P12	JNS	TEB	TES	LWC	ITP	200	7%	\$14.75
	P16	JTG	IEB	IES	LWC	JUC	500	8%	\$14.13
	P04	JTG	TEB	TES	LWC	JUC	900	5%	\$11.94
AP Model	JNS	TEB	TES	LWC	ITP	100	8%		

Table 2. The best fitting strategies and hostility recoding times for the individualized participant models and for the average-performance (AP) model.

For Dimensions 1 and 5, nearly all participants selected the same high-payoff microstrategies. For Dimension 1, nearly all used the Joystick-Nonstop (JNS) microstrategy and, for Dimension 5, the Joystick-Until-Classification (JUC) and Interruptible-Tracking-Preparation (ITP) microstrategies. As discussed earlier and as illustrated in Figures 7 and 8, these microstrategies lead to better tracking performance and roughly the same level of classification performance compared to the other microstrategies in Dimensions 1 and 5. That nearly all participants appeared to employ these high-payoff microstrategy selections suggests that, for some strategic dimensions, given appropriate motivation and feedback, all participants are able to optimize rather than just satisfice in their performance, as in [9].

However, for Dimensions 2, 3, and 4, it seems that some participants were more successful than others at finding high-payoff microstrategies. As can be seen in Table 2, the top five performers tended to use, for Dimension 2, the Immediate-Eyes-To-Blip (IEB) microstrategy; for Dimension 3, Immediate-Eyes-To-Sound (IES); and for Dimension 4, Look-Prior-Blip-Location (LBL) and Look-Sound-Location (LSL). Whereas the bottom five performers tended to use, for Dimension 2, Track-Then-Eyes-To-Blip (TEB); for Dimension 3, Track-Then-Eyes-To-Sound (TES); and for Dimension 4, Look-Window-Center (LWC). It appears as if at least some of the observed differences in multitasking performance can be explained in terms of microstrategy selection.

The top performers somehow found the high-payoff microstrategies. For Dimensions 2 and 3, IEB and IES are better than TEB and TES because the first two prioritize classification over tracking whenever there is a blip to classify, which is a good decision given how little the tracking target typically drifts in the time that is required to key-in a classification, and how quickly classification-delay penalties accrue; that the IES microstrategy appears to have been used by top performers suggests that it is not worth spending the time required to decode the auditory cue to

assist with classification—it is better to just look at the blip as soon as possible. For Dimension 4, LSL and LBL perform better than LWC because LSL uses sound and LBL uses memory to find the active blip, both evidently more quickly than the LWC approach of essentially saying “I’ll figure it out when I get there.” It appears as if the bottom performers failed to adopt these more efficient microstrategies, perhaps because they took more work, or perhaps because the benefit of adopting them was not clear.

Comparing the strategies of the average-performance model (the bottom row in Table 2) with those of the individualized models shows that the conventional approach of modeling the average human performance may not lead to the most accurate model. As can be seen in Table 2, the microstrategies of the average-performance model are more similar to those of the bottom performers than those of the top performers, and the average-performance model’s estimated hostility recoding time (100 ms) is shorter than most of the participants’ estimated times. If we were to base our conclusions on the results of the average-performance model, we might conclude that the participants did not find the optimal microstrategies (such as LBL and LSL in Dimension 4) and that participants seemed to be very fast at classifying hostility. The individualized models, however, suggest that the participants were not very fast at hostility classification (the average encoding time is 320 ms) but were able to compensate for it with faster strategies. Fitting a model to average human data will sometimes lead to incorrect conclusions. Individualized modeling may be necessary to understand human multitasking performance.

DISCUSSION

This paper demonstrates that multitasking performance might be best understood with highly parallelized microstrategy exploration in the context of individualized computational cognitive modeling. By running and analyzing 8,640 models on a cluster-based parallelized cognitive modeling system, we showed that the difference

between the top and bottom performers did not appear to result from their intrinsic perceptual, cognitive, and motor capabilities but rather from their skill at orchestrating a complex symphony of microstrategy selection, coordination, and execution. The effect of microstrategy selection on overall task performance suggests that multitasking is not accomplished by simply interleaving subtasks but rather with a controlled process that involves many micro-decisions which together determine overall task performance.

The cognitive modeling work presented here has powerful implications for the design of user interfaces intended for use in multitasking scenarios which, in the age of mobile computing, includes pretty much all computer interfaces. First, our modeling results suggest that effective multitasking might only occur when manual motor processing can be managed independently of ocular motor processing across two tasks or devices. Given the high demand of both ocular and manual motor processing when using touchscreen interfaces, it is questionable whether contemporary smartphones lend themselves to effective multitasking. Second, that microstrategies can so dramatically impact multitasking performance suggests that interface designers, if they wish to design a highly effective user interface, should thoughtfully consider how their designs will support a range of efficient microstrategies. Parallelized strategy exploration and individualized cognitive modeling may be key to understanding and ultimately accommodating human performance in time-pressured multitasking settings.

ACKNOWLEDGMENTS

This project was funded in part by NSF grant IIS-1017593 and, for the computer cluster, NSF grant OCI-0960354. The opinions expressed are strictly those of the authors.

REFERENCES

1. Ballas, J., Heitmeyer, C., and Perez-Quinones, M. (1992). Evaluating two aspects of direct manipulation in advanced cockpits. In *Proceedings of CHI*, 127–134.
2. Brumby, D. P., Salvucci, D. D., and Howes, A. (2009). Focus on driving: How cognitive constraints shape the adaptation of strategy when dialing while driving. In *Proceedings of CHI*, 1629–1638.
3. Card, S. K., Moran, T. P., and Newell, A. (1983). *The psychology of human-computer interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates.
4. Gluck, K. A., Stanley, C. T., Moore, L. R., Reitter, D., and Halbrügge, M. (2010). Exploration for understanding in cognitive modeling. *Journal of Artificial General Intelligence*, 2(2), 88–107.
5. Gray, W. D. and Boehm-Davis, D. A. (2000). Milliseconds matter: An introduction to microstrategies and to their use in describing and predicting interactive behavior. *J. Exp. Psychol.: Applied*, 6(4), 322–335.
6. Halverson, T. and Hornof, A. J. (2011). A computational model of "active vision" for visual search in human-computer interaction. *Human-Computer Interaction*, 26, 285–314.
7. Hornof, A. J. and Zhang, Y. (2010). Task-constrained interleaving of perceptual and motor processes in a time-critical dual task as revealed through eye tracking. In *Proceedings of the 10th International Conference on Cognitive Modeling*, 97–102.
8. Hornof, Zhang, and Halverson (2010). Knowing where and when to look in a time-critical multimodal dual task. In *Proceedings of CHI*, 2103–2112.
9. Howes, A., Lewis, R. L., and Vera, A. (2009). Rational adaptation under task and processing constraints: Implications for testing theories of cognition and action. *Psychological Review*, 116(4), 717–751.
10. Janssen, C., Brumby, D., Dowell, J., Chater, N., and Howes, A. (2011). Identifying optimum performance trade-offs using a cognitively bounded rational analysis model of discretionary task interleaving. *Topics in Cognitive Science*, 3(1), 123–139.
11. Kieras, D. E. and Meyer, D. E. (1997). An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Human-Computer Interaction*, 12(4), 391–438.
12. Kieras, D. E. and Meyer, D. E. (2000). The role of cognitive task analysis in the application of predictive models of human performance. In J. M. Schraagen, S. F. Chipman, V. L. Shalin, & V. Shalin (Eds.), *Cognitive task analysis* (pp. 237–260). Mahwah, NJ: Erlbaum.
13. Olson, R. L., Hanowski, R. J., Hickman, J. S., and Bocanegra, J. (2009). *Driver distraction in commercial vehicle operations*. Technical report #FMCSA-RRR-09-045. U.S. Department of Transportation.
14. Roberts, S. and Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, 107(2), 358–367.
15. Richtel, M. (2009, December 6). Promoting the car phone, despite risks. *New York Times*. Retrieved from <http://www.nytimes.com>.
16. Salvucci, D. and Taatgen, N. (2008). Threaded cognition: An integrated theory of concurrent multitasking. *Psychological Review*, 115(1), 101–130.
17. Sanford, E. C. (1888). Personal equation. *The American Journal of Psychology*, 2(1), pp. 3–38.
18. Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, 3(2), 159–177.
19. Williams, A. F. and Lund A. K. (1986). Seat belt use laws and occupant crash protection in the United States. *Am. J. Public Health*, 76(12), 1438–1442.
20. Zhang, Y. and Hornof, A. (2012). A discrete movement model for cursor tracking derived from moment-to-moment tracking data and the modeling of a dual-task experiment. In *Proceedings of the 56th Human Factors and Ergonomics Society Annual Meeting*, 1000–1024.