Sternberg Surrogate Method for Assessing Task Effects on Driving

Dave Benedict
Vehicle Performance Development
Toyota Technical Center, U.S.A

Linda S. Angell
Crash Avoidance System Development
Vehicle Structure & Safety Integration Center
General Motors

Driver Metrics Workshop
San Antonio, Texas

June 2, 2008
Presentation Overview

1. Sternberg Test Methodology
   - Experimental method
   - Task Metrics
2. CAMP Driver Workload Metrics Study
   - Study Results
   - Comparison to On-road results
3. Toyota/ GM Follow-On Study
   - Comparison to DWM Study
   - New Findings
4. Summary/ Advantages of Sternberg Method

Sternberg Surrogate Method for Assessing Task Effects on Driving
1. Sternberg Test Methodology

- Sternberg methodology (Sternberg, 1966) originally developed to study human short-term memory capacity.
- Methodology modified and applied by CAMP DWM project as a laboratory surrogate method.

Like a peripheral detection task
but with a memory loading component

Intended to evaluate effects of visual and/or cognitive loads of in-vehicle tasks
Sample Sternberg Stimuli

Verbal Stimuli
- 38
- 47
- 96

Spatial Stimuli
-  ─
- Y
- +
Sternberg Memory Search Task - Spatial & Verbal (CAMP DWM)

Methodology:

- “Memory set” displayed, memorized, then removed
- Secondary task is initiated (e.g. tune radio)
- During task, Sternberg probes presented every 2 -10 sec. until task is completed
- If probe is from memory set:
  - Press “Yes”;
- If not from memory set:
  - Press “No”
Experimental Setup

Adapted-Sternberg Display

“Yes” and “No” Response Buttons Set on both sides of steering wheel
Example of Positive Trial

Memory Set: ("Remember these as if they are highways on which you need to turn during this trip")

54 47 38

Probe: 47

Response: “Yes, this route marker was in the memory set.”
Another Example

Memory Set: (“Remember these as if they are highways on which you need to turn during this trip”)

54 47 38

Probe: 96

Response: “No, this route marker was not in the memory set.”
Sternberg Memory Task Metrics

% Missed Detections: Participant did not detect, or at least respond, to a probe when presented.

# Probes not detected
# Presented

# of Incorrect Responses, Given a Detection: Participant detected probe, but responded incorrectly.

Overall % Error:
# Probes Not Detected + # Incorrect
# Presented

Combined Dual Task Performance Decrement:
% Missed Detections + % Decrement on In-Vehicle Task Performance
The Combined Decrement Score Was Used Because . . .

While doing the Sternberg test:

- Performance decrements on in-vehicle tasks were observed ALONG WITH

- Performance decrements on the Sternberg test

- The degree to which performance suffered on the in-vehicle task seems to have depended on the task

Combined Decrement = (Proportion Missed Sternberg Detections + Proportion Decrement on Secondary Task)
2. Driver Workload Metrics Project

CAMP Driver Workload Metrics Consortium

Under the IVI Light Vehicle Enabling Research Program

Technical Team
Linda Angell  Louis Tijerina
Steve Kiger    Al Austria
Wayne Biever  Jack Auflick
Tuhin Diptiman Dev Kochhar
Jim Hogsett

And others . . .

CAMP Management Committee
Richard Deering
David Benedict
Michael Shulman
Ron Heft

Available at NHTSA Website: http://www-nrd.nhtsa.dot.gov/departments/nrd-13/newDriverDistraction.html
DWM Experimental Approach

➡️ Phased testing of 234 licensed drivers
  - Males & Females - 21 to 79 yrs old

➡️ Each participant tested in one of three settings:
  - Lab (surrogate method) n = 57
  - On-road (instrumented vehicles) n = 108
  - Test track (instrumented vehicles) n = 69

➡️ In each setting, participants were asked to perform a variety of in-vehicle tasks, following formal training
Variety of Tasks Included:

- Current/conventional tasks
- Artificial tasks
- New and emerging tasks
- Also incl. Just Drive Task

### Task Types

#### Visual-Manual
- HVAC
- Radio (Easy)
- Radio (Hard)
- Cassette
- CD/Track 7
- Coins

#### Auditory-Vocal
- Sports Broadcast
- Biographical Q&A
- BookOnTape Listen
- BookOnTape Summary

- Manual Dial
- Read (Easy)
- Read (Hard)
- Map (Easy)
- Map (Hard)
- Route Tracing
- Destination Entry

- Route Instructions
- Route Orientation
- Travel Computations

### Relative Workload

- Lower
- Higher

### Exploratory (mixed mode tasks):
- Voice Dial (mixed-mode)
- Delta Flightline (mixed-mode)
Driving Performance Metrics

- **Visual Allocation**
  - Glance Duration
  - No. of Glances
  - Glance Location

- **Vehicle Control**
  - Lane Keeping
  - Car Following
  - Speed Control

- **Object & Event Detection**
  - % Missed Events
  - Response Times

Sternberg Surrogate Method for Assessing Task Effects on Driving
Selected Findings Using The Sternberg Surrogate
Sternberg Proportion on Missed Detection

Task Completion Status
- Fully Successful
- Partially Successful

Tasks
- Visual-Manual
- Auditory-Vocal

Sternberg Surrogate Method for Assessing Task Effects on Driving
Sternberg Surrogate Method for Assessing Task Effects on Driving

Sternberg Combined Decrement Score

[“Cost of Concurrence”]

- Combined Decrement Score

Tasks

Visual-Manual

Auditory-Vocal

CD Track 7 Cassette Coins HVAC Manual Dial Radio Easy Radio Hard Read Easy Read Hard Map Easy Map Hard Route Tracking Voice Dial Delta Flightline Route Instructns Route Orientatn Sports Broadcast Travel Computtn Biographical Q&A BOT Listen BOT Summary Just Sternberg
Sternberg Surrogate Method for Assessing Task Effects on Driving

- Sternberg pattern similar to pattern of CHMSL On Road
- 64% of CHMSL variance accounted for by Sternberg

Comparison of On-Road Results & Sternberg for Central Events (CHMSLs)
Sternberg Surrogate Method for Assessing Task Effects on Driving

- Sternberg pattern similar to pattern of On Road though higher level
- 72% of variance accounted for by this Sternberg measure

Comparison of On-Road Results & Sternberg for Peripheral Events (FVTS)

Tasks

- CDTrack
- Cassette
- Coins
- HVAC
- ManualDial
- RadioEasy
- RadioHard
- VoiceDial
- RouteInstruct
- RouteOrient
- Sports
- Travel/Comp
- BiographicQA
- BOTListen
- BOTSsummary
- JustDrive

Comparison of proportion detected:

- FVTS On Road Dectects
- Sternberg Dectects In Lab
**Validity:** Sternberg Measures Were Related To ON ROAD Event Detection Measures

<table>
<thead>
<tr>
<th>Sternberg Metric</th>
<th>Mdn Task Dur</th>
<th>Mdn SDLP</th>
<th>Mdn Speed Diff</th>
<th>Pct Cross Trials</th>
<th>Pct LV Decel Miss Rate</th>
<th>Pct CHMISL Miss Rate</th>
<th>Pct FVTS Miss Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion Missed Detections</td>
<td>-0.731</td>
<td>-0.701</td>
<td>-0.649</td>
<td>-0.385</td>
<td>0.721</td>
<td>0.834</td>
<td>0.889</td>
</tr>
<tr>
<td>Proportion All Errors</td>
<td>-0.009</td>
<td>-0.009</td>
<td>-0.099</td>
<td>-0.349</td>
<td>0.714</td>
<td>0.799</td>
<td>0.868</td>
</tr>
<tr>
<td>RT to Correct Responses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>-0.791</td>
<td>-0.683</td>
<td>-0.903</td>
<td>-0.549</td>
<td>0.751</td>
<td>0.819</td>
<td>0.751</td>
</tr>
<tr>
<td>Median</td>
<td>-0.819</td>
<td>-0.603</td>
<td>-0.999</td>
<td>-0.574</td>
<td>0.717</td>
<td>0.809</td>
<td>0.769</td>
</tr>
<tr>
<td>RT to All Responses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>-0.799</td>
<td>-0.615</td>
<td>-0.999</td>
<td>-0.596</td>
<td>0.748</td>
<td>0.991</td>
<td>0.746</td>
</tr>
<tr>
<td>Median</td>
<td>-0.805</td>
<td>-0.631</td>
<td>-0.999</td>
<td>-0.502</td>
<td>0.754</td>
<td>0.902</td>
<td>0.740</td>
</tr>
<tr>
<td>Combined Decrement Score</td>
<td>-0.244</td>
<td>-0.360</td>
<td>-0.121</td>
<td>-0.005</td>
<td>0.404</td>
<td>0.750</td>
<td></td>
</tr>
</tbody>
</table>

Pearson r correlations $>0.707$
Validity:
Sternberg RT Measures Are Related To Two Main Types Of Eye Glance Metrics

- Glance Rate Metrics (which correlate with event detection)
- Eyes-off-road-and-on-task metrics
Sternberg Measures Were Repeatable
(based on split-half correlations)

<table>
<thead>
<tr>
<th>Sternberg Metric</th>
<th>Split-Half Repeatability, Pearson r</th>
<th>P-Value Significance Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion Missed Detections</td>
<td>0.969</td>
<td>0.000</td>
</tr>
<tr>
<td>Proportion Error, Given Detection</td>
<td>0.170</td>
<td>ns</td>
</tr>
<tr>
<td>Proportion All Errors</td>
<td>0.944</td>
<td>0.000</td>
</tr>
<tr>
<td>RT to Correct Responses</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.791</td>
<td>0.000</td>
</tr>
<tr>
<td>Median</td>
<td>0.822</td>
<td>0.000</td>
</tr>
<tr>
<td>RT to Incorrect Responses</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.028</td>
<td>ns</td>
</tr>
<tr>
<td>Median</td>
<td>0.041</td>
<td>ns</td>
</tr>
<tr>
<td>RT to All Responses</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.767</td>
<td>0.000</td>
</tr>
<tr>
<td>Median</td>
<td>0.798</td>
<td>0.000</td>
</tr>
<tr>
<td>Combined Decrement Score</td>
<td>0.941</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Pearson $r$ correlations ➢ 0.707 (when squared, give $R^2$ of 50% or greater)
3. Toyota-GM Follow-On Study

Sternberg Surrogate Method for Assessing Task Effects on Driving
Purpose of the Follow-On Study

- To replicate prior findings

- To extend previous research, especially as pertains to:
  - Auditory-vocal tasks
  - Modality of Sternberg probes (visual or auditory)
  - Task length issues
  - Integration of Sternberg probes with the driving scene (more naturalistic but still easy surrogate method)
Key New Variable: Visual vs. Auditory Presentation of Sternberg Stimuli

- **VISUAL condition**
  - Memory set was shown on a screen
  - Sternberg probes were displayed visually, superimposed on a driving scene

- **AUDITORY condition**
  - Memory set were presented auditorily
  - Sternberg probes were heard, while looking at the driving scene
Visual Presentation of Memory Set
Visual Presentation of Sternberg Probes
Auditory Presentation of Memory Set

“47, 83, 74)”
Auditory Presentation of Sternberg Probes

"54"
Experimental Method

<table>
<thead>
<tr>
<th>Task Type</th>
<th>Task Num</th>
<th>Task Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>1</td>
<td>Stern Only</td>
</tr>
<tr>
<td>SURT V-M Long</td>
<td>2</td>
<td>SURT Easy 75-5</td>
</tr>
<tr>
<td>V-M Low</td>
<td>3</td>
<td>SURT Hard 115-3</td>
</tr>
<tr>
<td>V-M High</td>
<td>4</td>
<td>Radio Hard</td>
</tr>
<tr>
<td>A-V Short</td>
<td>5</td>
<td>Coins</td>
</tr>
<tr>
<td>A-V Long</td>
<td>6</td>
<td>Manual Dial Phone</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>Map Hard</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>Bio QA *</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>Travel Comp *</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>Sports Broadcast</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>Delta</td>
</tr>
</tbody>
</table>

N = 32 panelists
Balanced for gender & age (<35, >55)
Findings From Statistical Analysis

➤ Significant Effects:

• **TASKS X STIMULUS TYPE:** Differences between tasks BUT the magnitude & nature of the differences depended on which type of Sternberg stimulus was used (auditory or verbal)

• **AGE:** Sternberg task more difficult for older subjects across all tasks, as in prior study
Task Effects Replicated for Visually Presented Sternberg Probes

Sternberg Proportion of Missed Detections

- SuRT 75/5
- SuRT 115/3
- Radio
- Coins
- Cell Phone
- Map
- Sports
- Delta
- Bio QA
- Travel Comp

Sternberg Surrogate Method for Assessing Task Effects on Driving
Cost of Concurrence Effects: Replicated (related to less peripheral event detection on road)
Auditory Modality Disrupted Sternberg Performance

Mean Proportion Missed Sternberg Detections

Visual
Auditory

Sternberg Surrogate Method for Assessing Task Effects on Driving

Sternberg Task
SuRT 75/5
SuRT 115/3
Radio
Coins
Cell Phone
Map
Bio QA
Travel Comp
Sports
Delta

Visual-Manual Tasks
Auditory-Vocal Tasks

GM
TOYOTA TECHNICAL CENTER USA INC
Interpretation of Auditory Effects

- To do the Sternberg task requires rehearsing the highway marker numerals throughout the task (to keep them in memory) – while hearing & making decisions about each probe.
- Rehearsal in working memory is generally thought be done by “repeating to yourself” – using what is called the “articulatory loop” (cf. Baddeley).
- This is believed to use “auditory-vocal” resources.
- Performing this rehearsal process may have used the auditory channel so fully that the Sternberg probes literally did not get “heard” (a ‘listening without hearing’ analogy to ‘looking but not seeing’).
- This resulted in a high missed detection rate (~80%).
- The miss rate was so high that any additional effects of auditory mode on auditory tasks were difficult to see.
• Consistent results for visual-presented probes between two studies.
• Auditory presentation of probes – much higher misdetection & limited differentiation of tasks.
• Consistent results for visual-presented probes between two studies.
• Auditory presentation of probes – much higher misdetection & limited differentiation of tasks.
• Auditory Sternberg has greater effect on tasks with substantial working memory component.
Follow-On Study Summary

- Prior findings were replicated
- Auditory presentation of Sternberg probes appeared to overload the auditory channel.
- Use of simple driving scene did not appear to provide any benefit.
- Benefit of using driving simulator & Sternberg?
4.1 Summary Observations of Sternberg Method

- Has validity, accounting for 64%-72% of the variance in on-road event detection
- Yields multiple metrics that are repeatable (based on split-half evaluations)
- Identified key effects observed in the on-road event response data
4.2 Advantages of Sternberg Method

- Valid for event detection
- Provides a memory loading component
- Can be used for Auditory-Vocal tasks as well as Visual-Manual tasks
- Easy to use, like Occlusion method – Only needs a laptop computer
- Can be administered easily during product development

However, criteria need to be established for identifying tasks which need re-design.
Sternberg Surrogate Method for Assessing Task Effects on Driving
Material for Q & A
Comparison of Sternberg & PDT Methods

Data from CAMP Drive Workload Metrics Final Report

<table>
<thead>
<tr>
<th>Task Type Measure</th>
<th>CHMSL Centrally-Located Event</th>
<th>FVTS Peripherally-Located Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDT Alone</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>PDT w/ STISIM</td>
<td>.806</td>
<td>.810</td>
</tr>
<tr>
<td>Sternberg % Miss</td>
<td>.947</td>
<td>.831</td>
</tr>
<tr>
<td>Sternberg % All Error</td>
<td>.969</td>
<td>.862</td>
</tr>
<tr>
<td>Sternberg Comb. Dec.</td>
<td>.914</td>
<td>.840</td>
</tr>
<tr>
<td>Sternberg Correct RT</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Sternberg All RT</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

Pearson Correlations over $r = .707$ are shown

Hypothesis: Load on Working Memory and Central Attention allowed Sternberg to be effective for a broader set of task types and conditions.
Surrogate Reference Task

Visual easy

Visual medium

Visual difficult

Motoric easy

Motoric difficult
End of Presentation

Note:
Please refer to the CAMP Driver Workload Metrics Final Report, Appendix R, and Chapter 8, for a discussion of the modified Sternberg method, prior findings, and specific caveats and issues that would benefit from future work.

Issue of Task Length – Still Unclear

Results from Visually-Presented Sternberg Probes
Task Effects Replicated for Visually Presented Sternberg Probes (For Same Tasks)

Sternberg Proportion Missed Detections

- Just Stern
- Radio
- Coins
- Manual Dial
- Map
- Bio QA
- Travel Comp
- Sports
- Delta

Sternberg Surrogate Method for Assessing Task Effects on Driving