A Preliminary Assessment of Perceived and Objectively Scaled Workload of a Voice-Based Driver Interface

Bryan Reimer, Ph.D.  
MIT AgeLab & New England University Transportation Center

Driver Assessment  
Bolton Landing, NY

June 20, 2013
Acknowledgments

- **Co-authors**
  - Bruce Mehler
  - Hale McAnulty
  - Daniel Munger
  - Alea Mehler
  - Enrique Abdon Garcia Perez
  - Thomas Manhardt
  - Joseph Coughlin

- **Sponsors**
  - Santos Family Foundation
  - Region One New England University Transportation Center
  - Toyota’s Collaborative Safety Research Center
Literature on the Demands of Voice Interfaces

While **A LOT** of research has been conducted on experimental and “wizard of oz” voice interactions, as well as handheld systems, more limited research address the demand of embedded production level vehicle systems:

- Carter and Graham (2000) – Jaguar S-Type (n=32; tracking task)
- Chang et al. (2005) – 2004 Accord & 2005 Acura RL (n=10x2; field)
- Harbluk et al. (2007) – 2005 Acura-TL (n=16 x 2; LCT)
- Shutko et al. (2009) – 2008 model year SYNC (n-25; sim)
- Owens et al. (2010a, b)* - 2010 Mercury Mariner with SYNC (n=21; field / test track)

*Only research with use of physical production integrated system
Research Questions

While voice interfaces offer the promise of reducing the time a driver’s eyes are drawn away from the roadway, a number of questions remain:

• How do we effectively assess the amount of non-visual demand associated with voice interfaces?

• In what conditions does voice control reduce demand over traditional methods of interaction?

• How do different simple vs. more complex voice interactions affect drivers?

• Do age and gender impact drivers’ perceptions and use of voice interfaces?
The MIT n-back

An Emerging International Method for Inducing Graded Cognitive Workload

• Series of 10 single digit numbers (0-9) presented in random order aurally at 2.25 sec intervals

• Subject instructed to respond with nth digit back

• Across levels
  — Auditory demands constant
  — Vocal demands “relatively” constant

• Aims to manipulate secondary cognitive demand

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>6 9 1 7 0 8 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-back Response</td>
<td>6 9 1 7 0 8 4</td>
</tr>
<tr>
<td>1-back Response</td>
<td>- 6 9 1 7 0 8</td>
</tr>
<tr>
<td>2-back Response</td>
<td>- - 6 9 1 7 0</td>
</tr>
</tbody>
</table>

Heart Rate (2 min periods)

(Mehler, Reimer Dusek & Coughlin, 2011)
Instrumented Vehicle

- **Physiology (HR, GSR)**
- **Lane tracking**
- **Audio recording**
- **Eye tracking**
- **Vehicle telemetry via CAN bus link**

- **Cameras for monitoring driver & traffic**
- **GPS**
- **PC based data acquisition system**
- **Rear seat for overseeing experiment**
Interface Tasks

Extensive parking lot training and driving evaluation (x2)

• Visual manual task (radio tuning)
  › Single press preset selection
  › CAMP style manual radio tuning (Angell et al., 2006)

• Voice interface tasks
  › Preset selection (manual preset selection equivalent)
  › Tuning to a station (manual radio tuning equivalent)
  › Full address destination entry
  › Cancel navigation
  › Song selection (success and failure (1 replication))
  › Contact dialing
Experimental Route and Key Protocol Steps

1. Briefing, Consent Questionnaire N-back training
2. Vehicle Setup Safety Briefing Task Training
3. RT 93 N. (20 min) Habituation to Vehicle
4. RT 495 S. Data Collection Period 1
5. Rest Stop Questionnaires Task Training 2
6. RT 495 N. Data Collection Period 2
7. RT 93 S. Data Collection Period 3 (Phone Task)
8. End of Experiment Questionnaires Workload Scale
Subjects

Two age groups: 20-29 and 60-69 years; 102 recruited

<table>
<thead>
<tr>
<th>N</th>
<th>Reason for Exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>Unreliable/unusable ECG (heart rate) recording</td>
</tr>
<tr>
<td>6</td>
<td>Subject did not pass MoCA screening</td>
</tr>
<tr>
<td>6</td>
<td>Equipment failure</td>
</tr>
<tr>
<td>4</td>
<td>Scheduling problems prevented completion of study</td>
</tr>
<tr>
<td>4</td>
<td>Subject demonstrated unsafe driving behavior</td>
</tr>
<tr>
<td>4</td>
<td>Subject was part of the pilot version of the study</td>
</tr>
<tr>
<td>2</td>
<td>System did not recognize subject’s voice (determined in parking lot before drive)</td>
</tr>
<tr>
<td>1</td>
<td>Unsafe conditions on road</td>
</tr>
<tr>
<td>1</td>
<td>Research assistant executed experiment protocol incorrectly</td>
</tr>
</tbody>
</table>

42 Total Excluded

60 case analysis sample equally distributed across age and gender
Self-Reported Workload

<table>
<thead>
<tr>
<th>Task</th>
<th>Workload (out of 10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-Back</td>
<td>1.5 ± 0.5</td>
</tr>
<tr>
<td>Nav Cancel (V)</td>
<td>2.0 ± 0.5</td>
</tr>
<tr>
<td>Radio Easy (M)</td>
<td>2.5 ± 0.5</td>
</tr>
<tr>
<td>Contact Dialing (V)</td>
<td>3.0 ± 0.5</td>
</tr>
<tr>
<td>1-Back</td>
<td>3.5 ± 0.5</td>
</tr>
<tr>
<td>Radio Easy (V)</td>
<td>4.0 ± 0.5</td>
</tr>
<tr>
<td>Radio Hard (V)</td>
<td>4.5 ± 0.5</td>
</tr>
<tr>
<td>Song Select (V)</td>
<td>5.0 ± 0.5</td>
</tr>
<tr>
<td>Nav Entry (V)</td>
<td>5.5 ± 0.5</td>
</tr>
<tr>
<td>Radio Hard (M)</td>
<td>6.0 ± 0.5</td>
</tr>
<tr>
<td>2-Back</td>
<td>6.5 ± 0.5</td>
</tr>
<tr>
<td>Song Fail (V)</td>
<td>7.0 ± 0.5</td>
</tr>
</tbody>
</table>
Heart Rate

Beats per minute

Baseline  | Radio Easy (M)  | Nav Cancel (V)  | 0-Back  | Song Select (V)  | Song Fail (V)  | Radio Easy (V)  | Nav Entry (V)  | Radio Hard (V)  | Contact Dialing (V)  | Radio Hard (M)  | 1-Back  | 2-Back

© 2013
Skin Conductance

Micromhos (microsiemens)

Baseline  Radio Easy (M)  Nav Cancel (V)  0-Back  Radio Easy (V)  Radio Hard (V)  Contact Dialing (V)  Nav Entry (V)  Radio Hard (M)  Song Select (V)  Song Fail (V)  1-Back  2-Back
Mean Velocity

km/hour

Radio Easy (V)  Radio Hard (M)  Radio Hard (V)  Song Select (V)  Radio Easy (M)  Nav Entry (V)  Contact Dialing (V)  Song Fail (V)  Nav Cancel (V)  2-Back  Baseline  0-Back  1-Back
Visual Attention

One needs to question the validity of eye tracking data

• Goal -> faceLab data reduced to categorize eyes on road / eyes to task
• Result -> multiple inconsistencies suggest eye tracking data not reliable
• Outcome -> double coded, mediated data forthcoming
A Link Between Phone Use and Risky Driving

It's clear [from the scientific literature] that cell phones in and of themselves impair the ability to manage the demands of driving, but the fundamental problem may be a broader pattern of behavior of individuals who are willing to pickup the technology.

(Zhao, Mehler, Reimer, D'Ambrosio, Mehler & Coughlin, 2012)
Conclusions on the Demand of Voice Interfaces

Benefits and Cautions

• Cognitive processing demands (as assessed by physiological activation) lower than expectations
  › Largely less than manual radio tuning benchmark
  › Aligned with 1-back cognitive benchmark

• Provide access to higher complexity activity (nav entry) at cognitive demand levels below accepted benchmarks

• Shows high perceived workload and SCL in failure condition

• Clear compensatory slowing to manage workload

• Results encouraging, but visual attention to HMI / eyes-off-road (TBD)
Limitations to Generalizability

- Tasks
- Road environment
- Vehicle selection
- Experimental “pacing” of tasks
- “Demand” -> is not clearly linked to safety
Questions