Assessing Attentional Focus with Modern Multi-Modal Driver Vehicle Interfaces: Selected Results from Field and Simulation Studies

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

Lifesavers, New Findings from Distracted Driving Research
Nashville Tennessee

April 28, 2014

Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu

Acknowledgments

- Multiple contributors
- Sponsors
  - Toyota Collaborative Safety Research Center
  - Insurance Institute of Highway Safety (IIHS)
  - Santos Family Foundation
  - Region One New England University Transportation Center

Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu

The Future is Larger Screens

Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu

The Future is More Devices

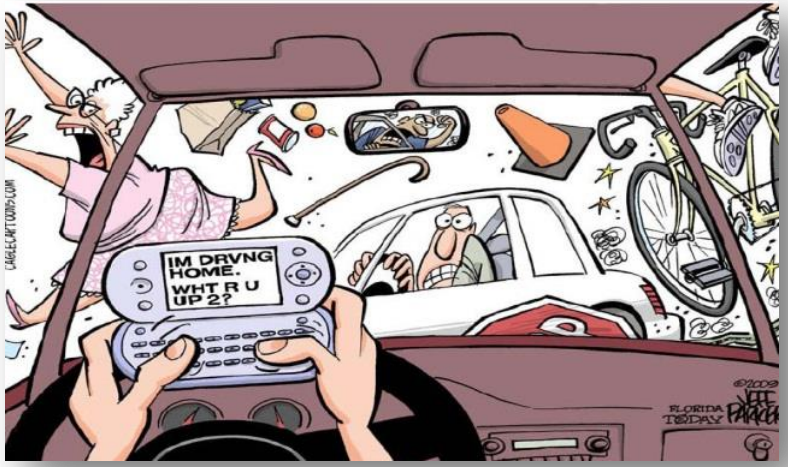
Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu

The Future is More Information

Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu

The Challenge

For Safety Professionals, Regulators and Manufacturers



How to develop a safe vehicle interface that provides drivers with enjoyable and easy access to vehicle systems and applications that they are coming to expect, while maximizing driver focus on the road?

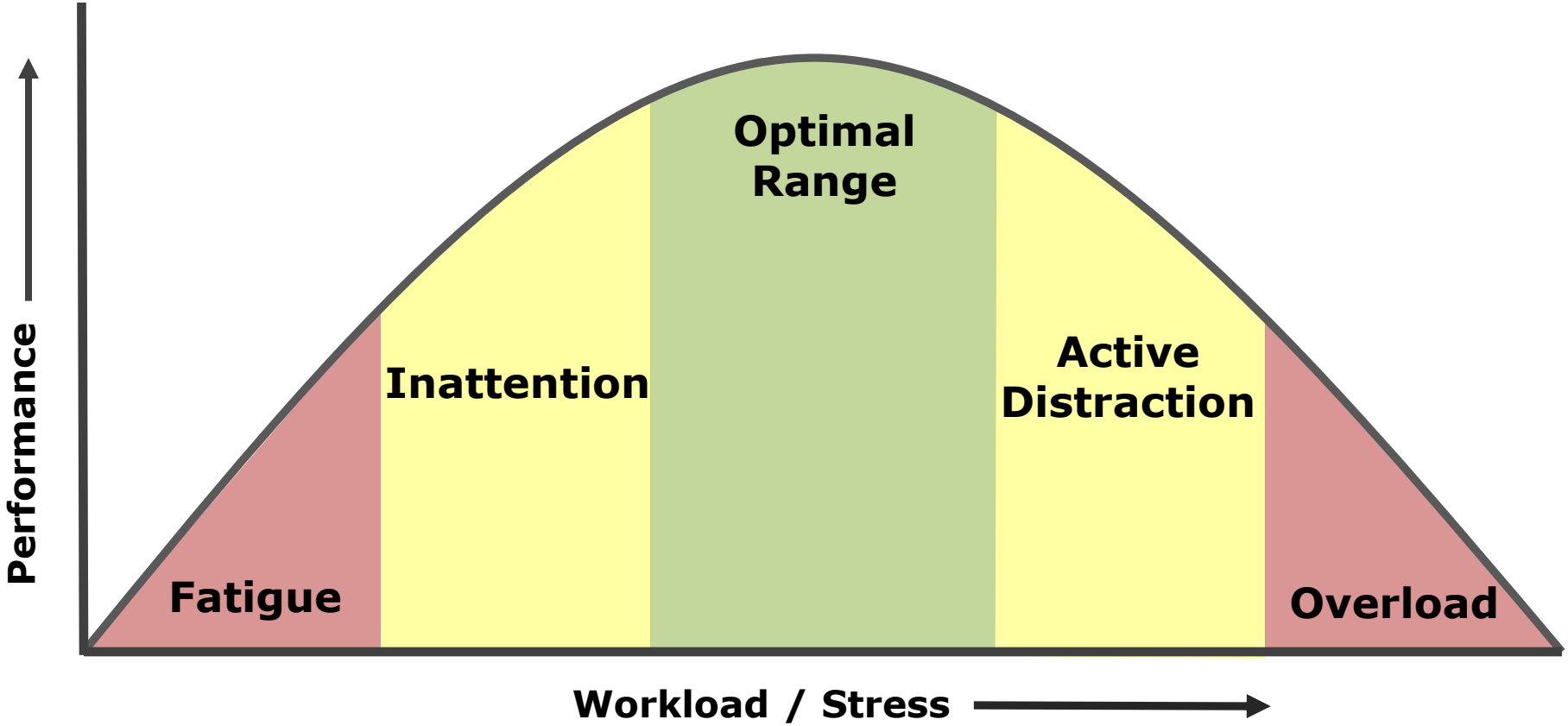
Minimizing total demand on the driver will require a better understanding of how loads interact across multiple modalities and interfaces.

Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu

Workload & Performance

Yerkes-Dodson Law

The relationship between performance and physiological or mental arousal

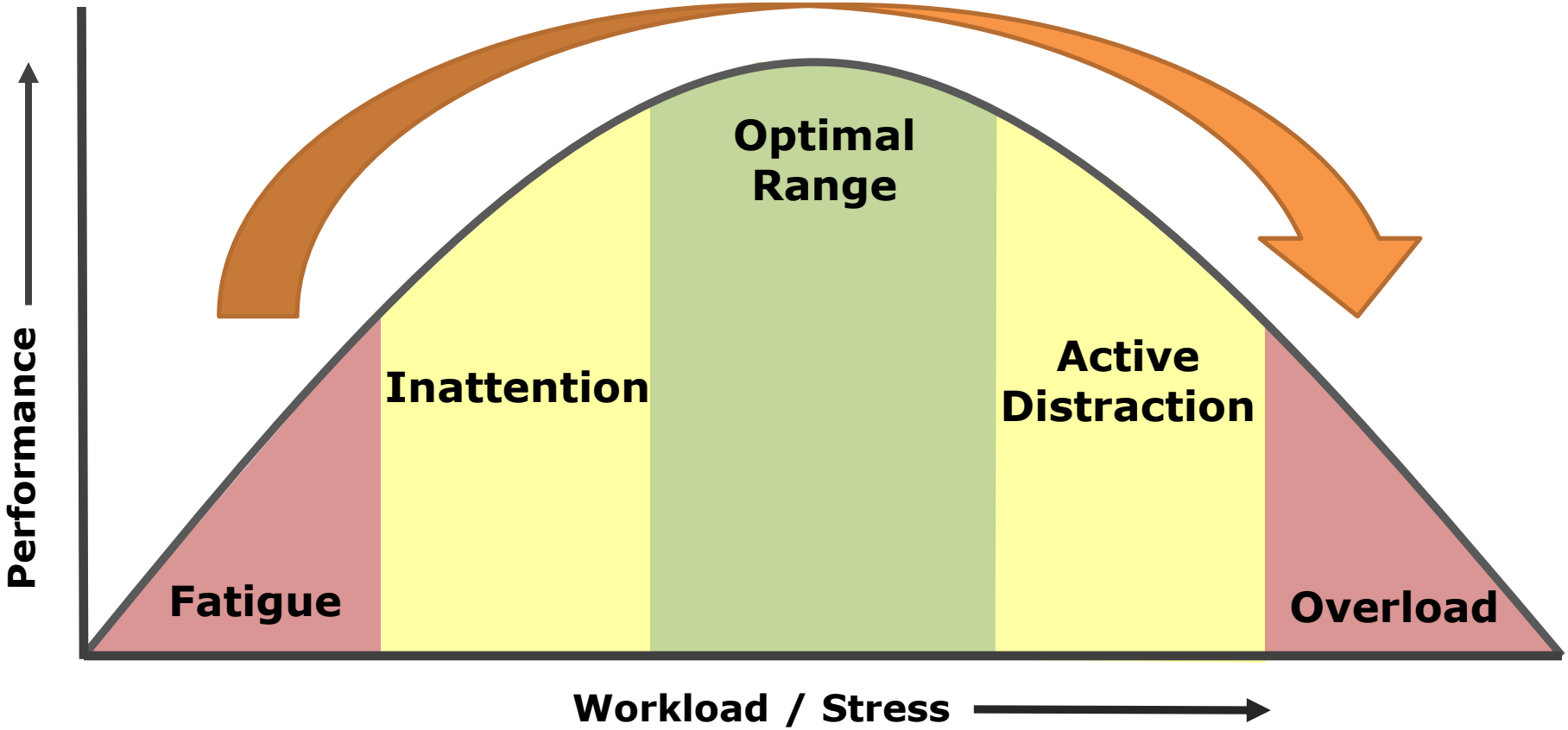


(Source: Coughlin, Reimer & Mehler, 2011)

Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu

Workload & Performance

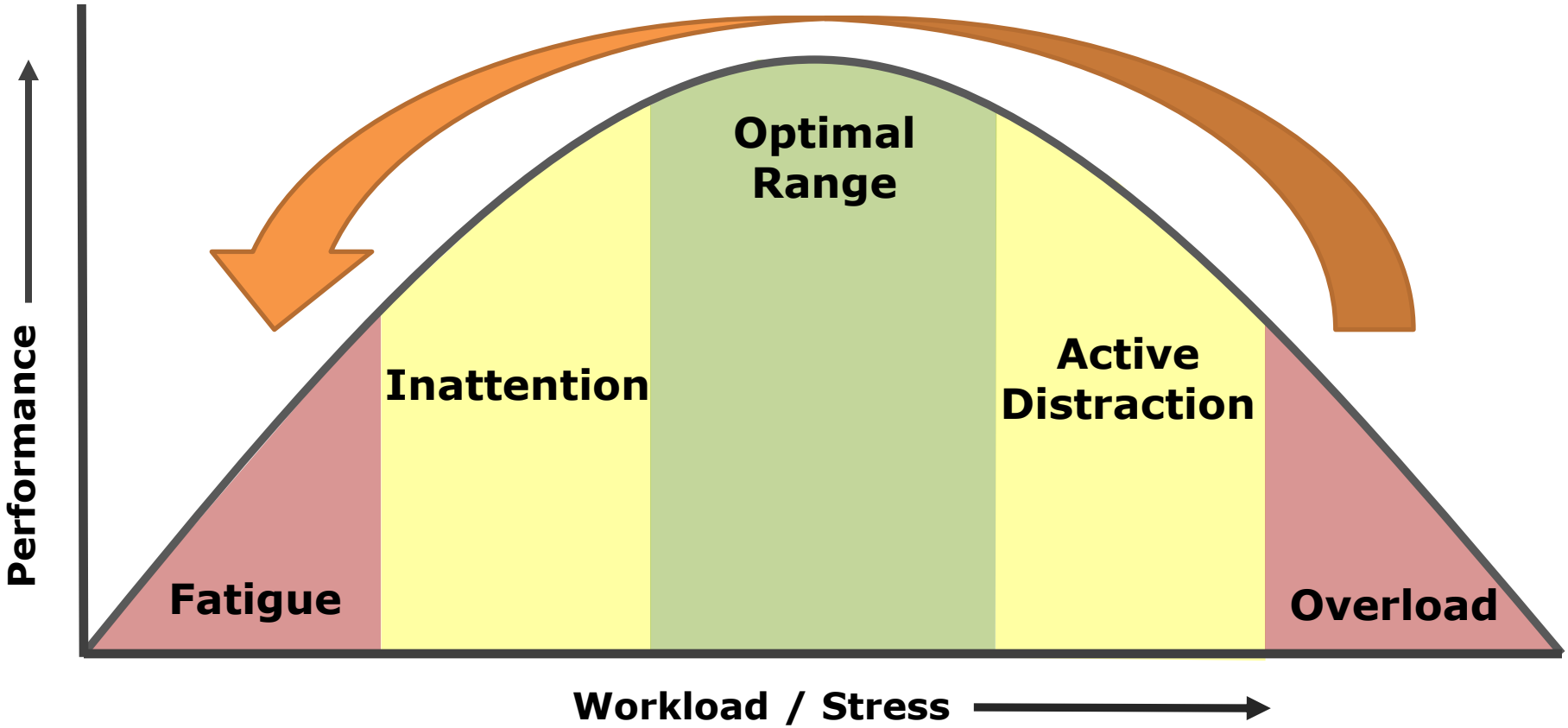
More Information in the Vehicle Tends to Increase Workload



Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu

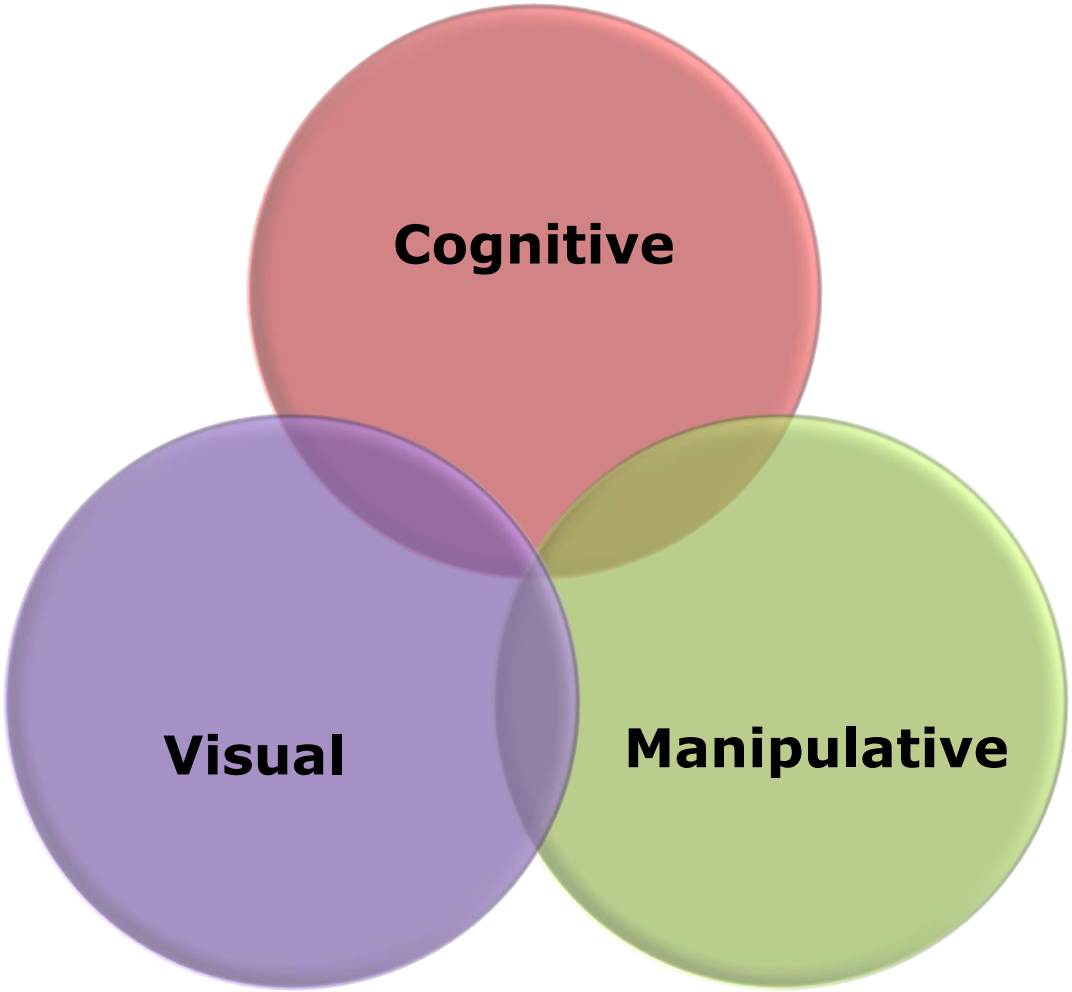
Workload & Performance

Automation Tends to Lower Workload



Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu

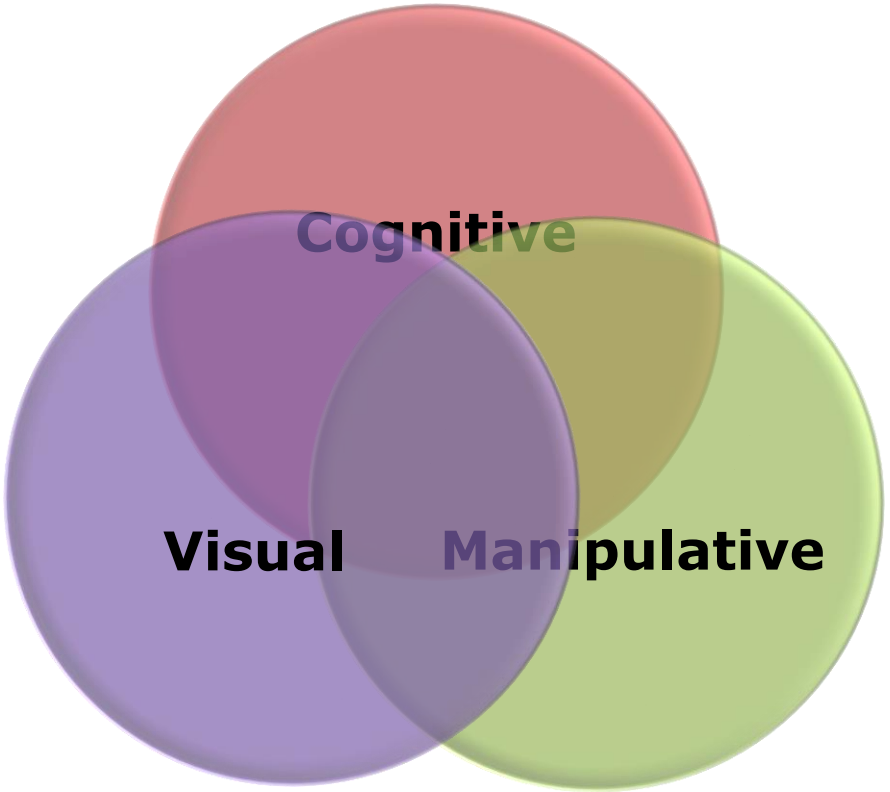
A Common Perception of the Three Major Pillars of Distraction



Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu

In reality

..... the pillars are highly overlapping



Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu

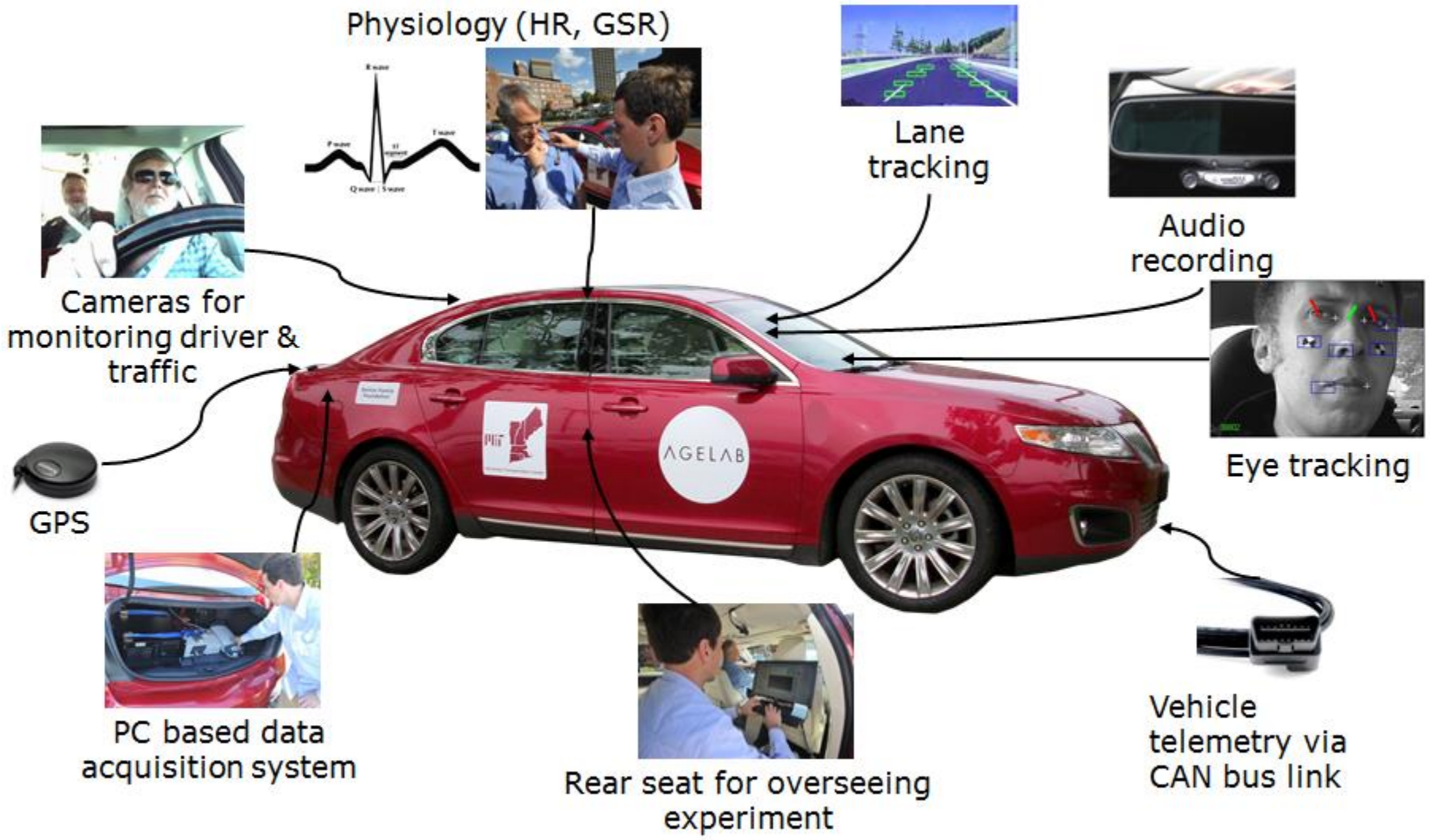
Selected Results

Insights from around 300 participants

- Three field studies of a production level voice interface
  - Assessed the demands associated with a voice interface
  - Considered the impact of structured vs. self-guided training
  - Evaluated an "experienced" user mode vs. the "default" mode
- A series of simulation studies looking at full address entry
  - Visual-manual entry using an iPod with iOS 5 "Google Maps" vs. Garmin GPS
  - Visual-manual entry using a Samsung Galaxy 4 vs. two voice-modes
  - Voice-command entry using a Samsung Galaxy 4 vs. Google Glass

Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu

Instrumented Vehicle



Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu

Interface Tasks

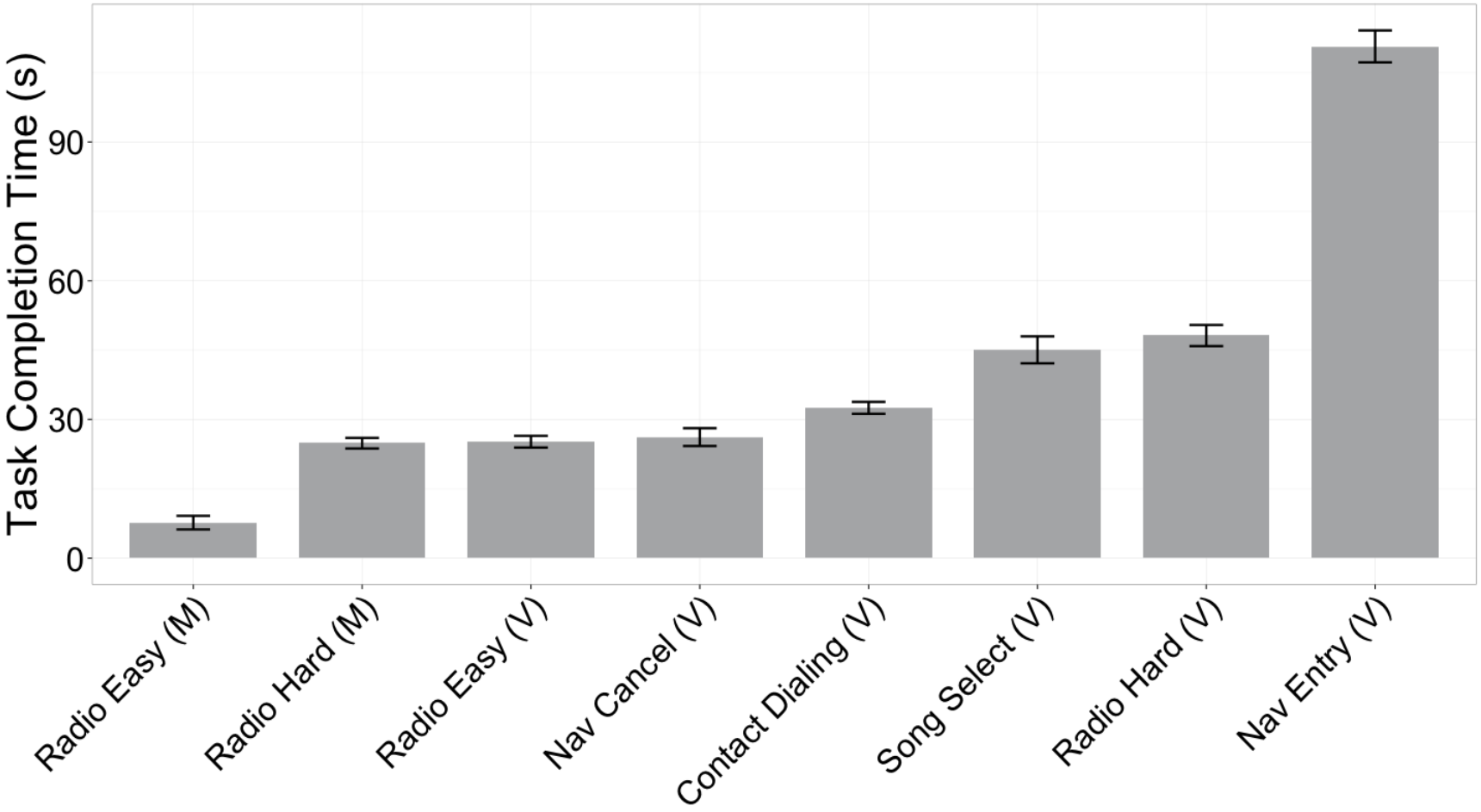
Extensive parking lot training and driving evaluation (x2)

- Visual-manual task (radio tuning)
  - Single press preset selection – Radio Easy (M)
  - Manual radio tuning to a specified station (i.e. FM 98.5) – Radio Hard (M)
- Voice interface tasks
  - Preset selection (manual preset selection equivalent) – Radio Easy (V)
  - Tuning to a station (manual radio tuning equivalent) – Radio Hard (V)
  - Full address destination entry – Nav Entry (V)
  - Cancel navigation – Nav Cancel (V)
  - Simple Pre-set phone contact dialing – Contact Dialing (V)
  - Song selection – Song Select (V)
  - Song selection failure (1 experience) – Song Fail (V)

Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu

Study 1: Time on Task

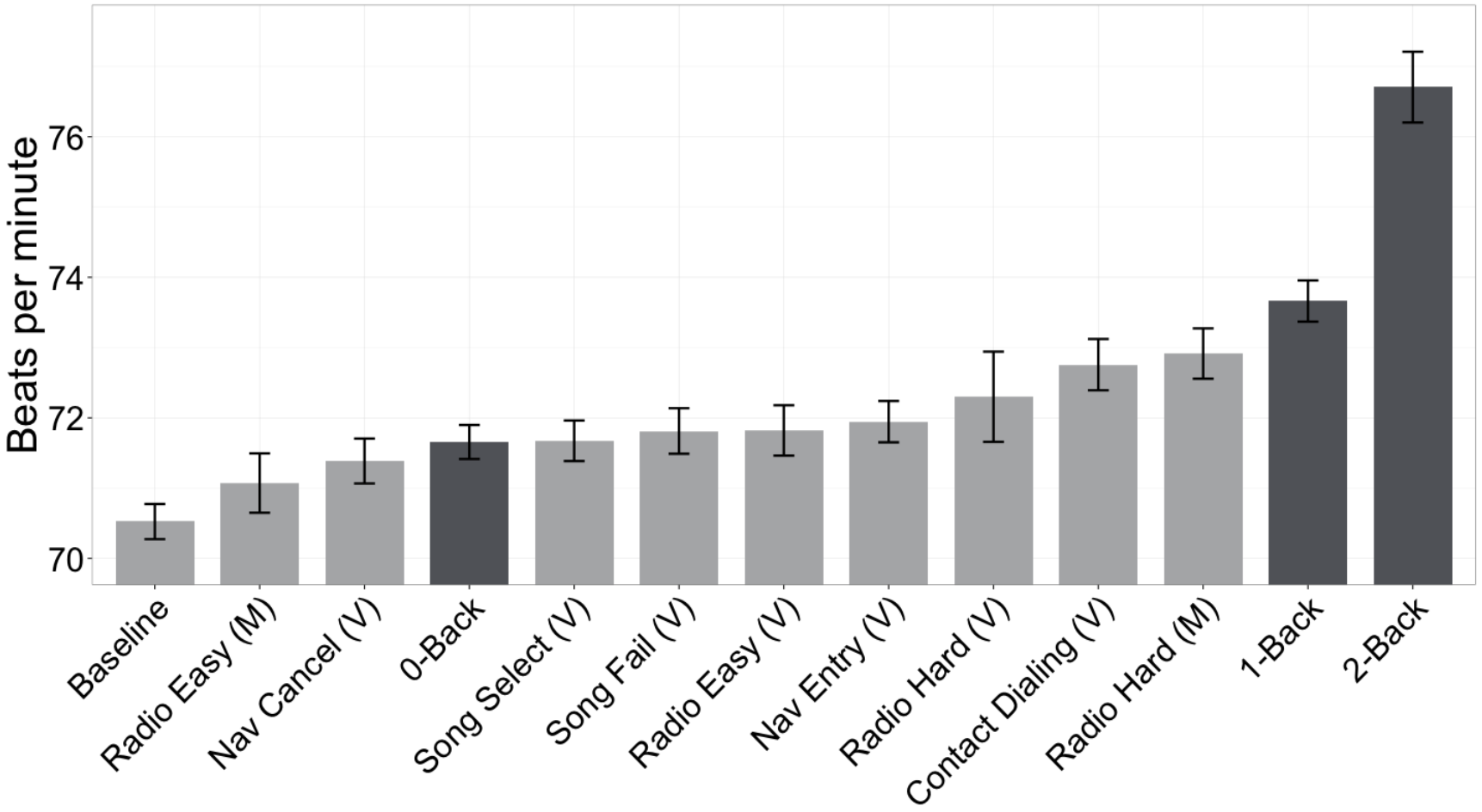
The Voice-based Navigation Entry task took much longer to complete than any other task ($p < .001$)



Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu

Study 1: Heart Rate

There is an overall main effect of task ($p < .001$)

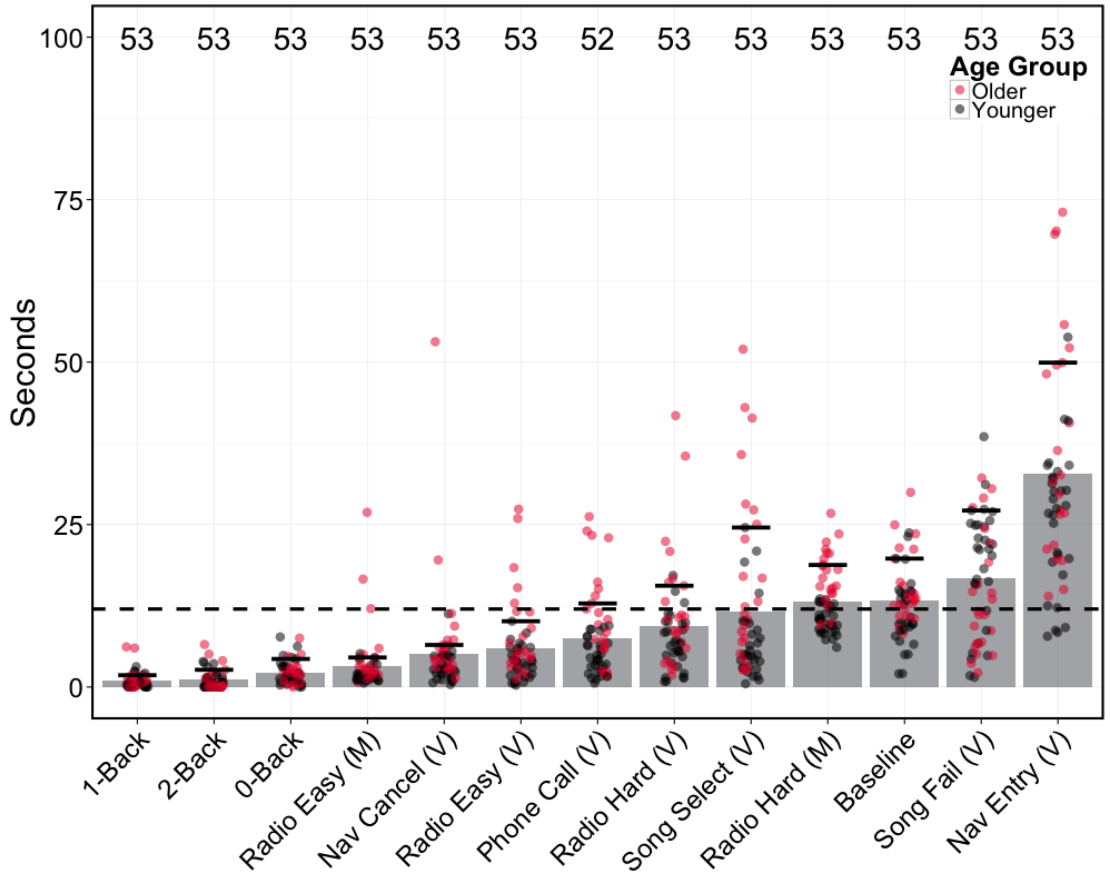


Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu

Study 1: Total Off-Road Glance Time

Longest for Voice Navigation Entry

Voice Radio Hard was lower than Manual Radio Hard



12 second threshold shown as a dashed line. The longer individual line above each bar represents the 85% point in the sample distribution for each task.

Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu

Study 2: Key Methods

Were consistent with Study 1 except that:

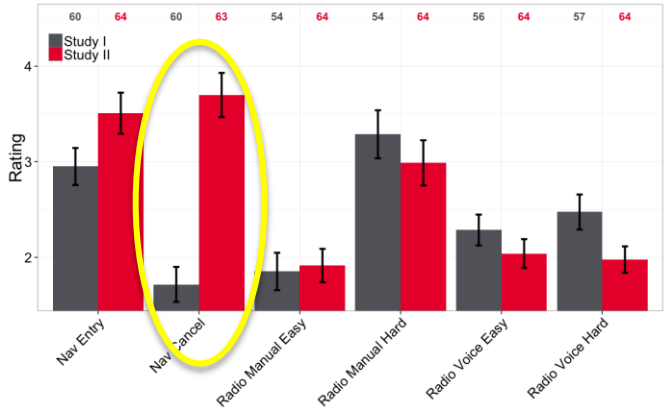
- **Half of the drivers** were **formally "trained"** as in Study 1, while **the rest** were exposed to a **"self-guided" learning** experience
- Drivers were equally distributed across **four age groups**: 18-24, 25-39, 40-54, and 55+ **following new NHTSA guidelines**
- The **radio manual tuning task** was adjusted to better conform to the reference task recommended by The Alliance and NHTSA
- **The prompt for the route cancel task during the assessment period was altered to remove an explicit reminder of how to execute the command**

Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu

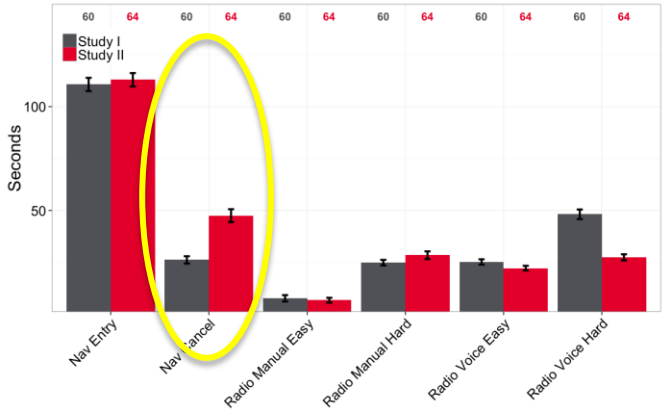
Relationships Between Studies 1 & 2

Notable Differences associated with Navigation Cancel

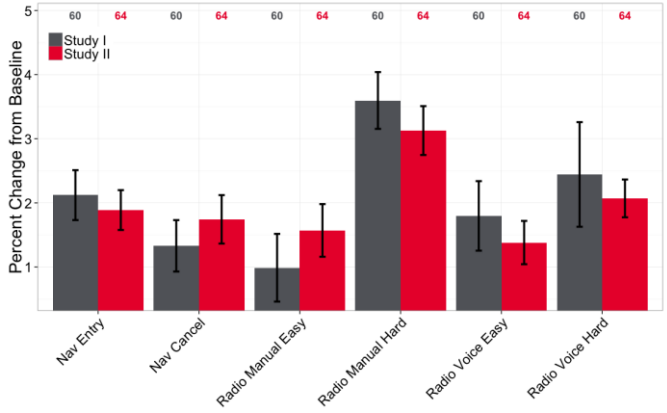
Reported Workload



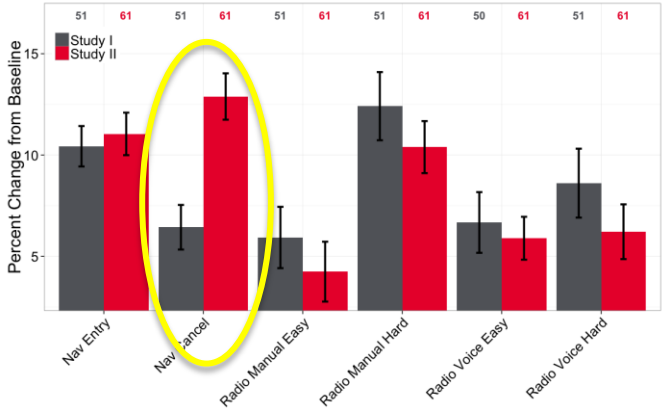
Task Time



Heart Rate



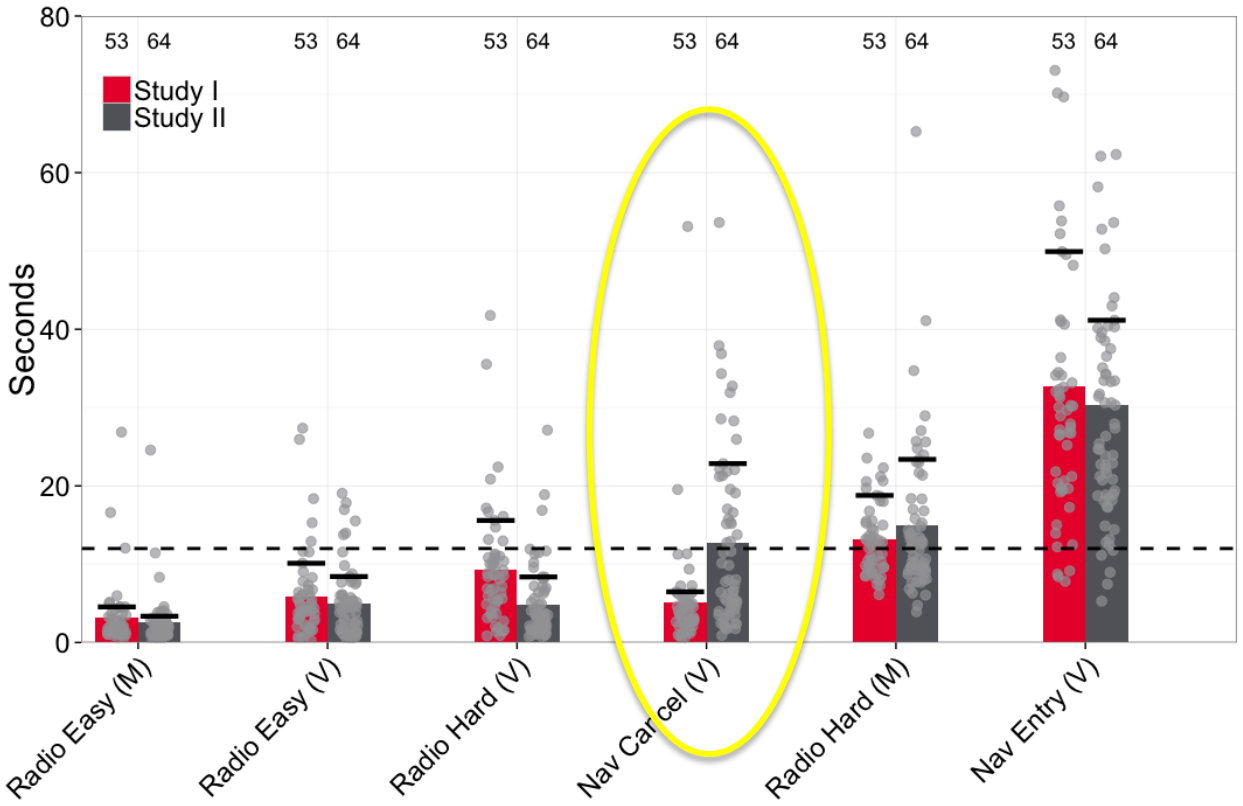
Skin Conductance



Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu

Total Off Road Glance Time - Studies 1 & 2

Clear Difference for Navigation Cancel Task (as expected). Nominally lower values in Study 2 for other Activities Except Radio Manual Hard. Sample Age Effect?



12 second threshold shown as a dashed line. The longer individual line above each bar represents the 85% point in the sample distribution for each task.

Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu

Study 3: Key Methods

Identical to the formally training group in Study 2, except that the voice system's settings were adjusted from the "default" mode to an "expert" mode.

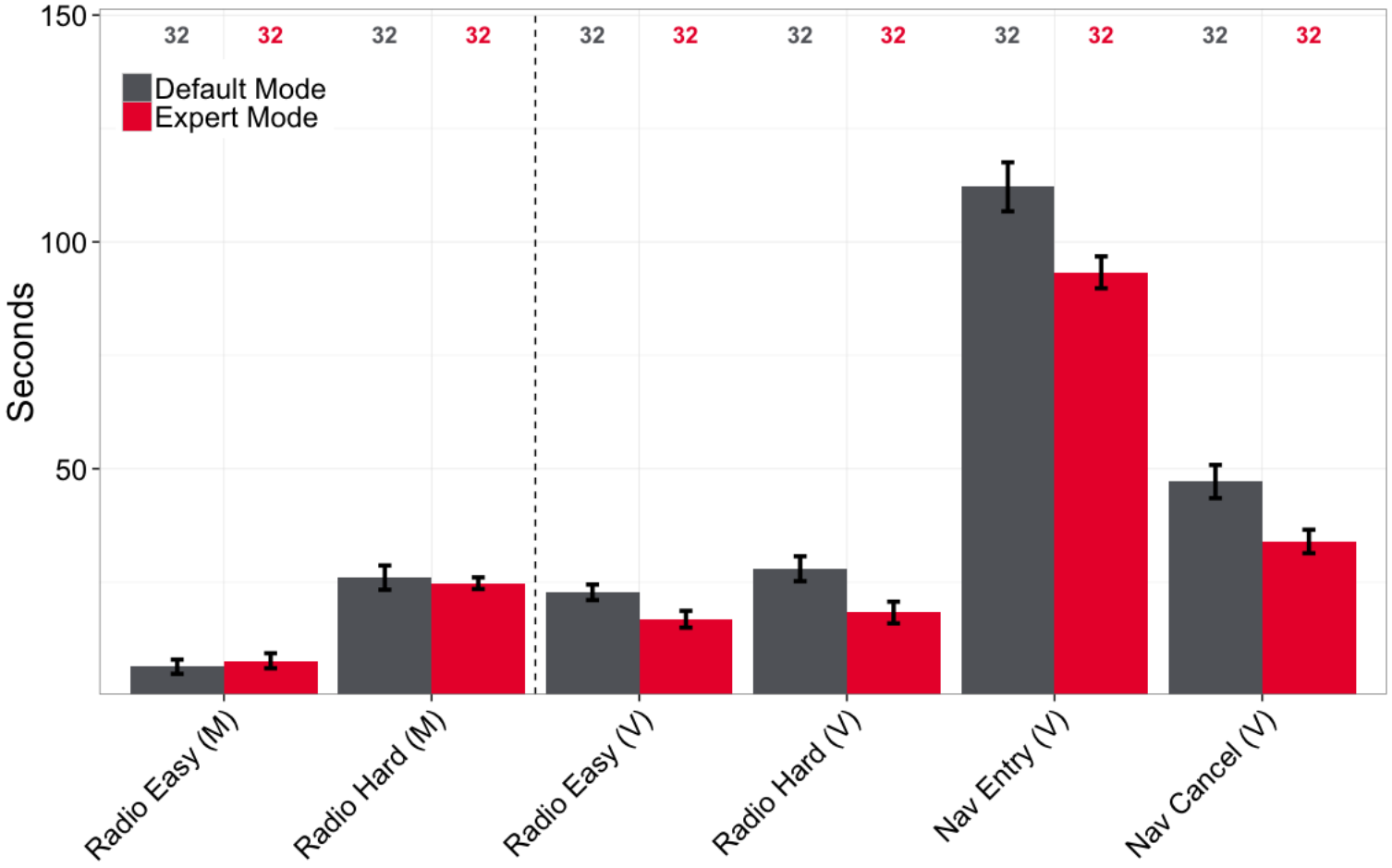
- When the "push-to-talk" button is pressed, expert mode only sounds a tone instead of saying, "Please say a command."
- When a command is recognized by the system, it is normally accepted by the system without a voice asking the driver to confirm that this is what they want to do by saying "yes".

The **net result** in the expert mode is **a reduction in the amount of audio content listening time** required of the driver and a **reduction in the number of confirmatory responses** required.

Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu

Study 3: Task Time

Mode Affected "Voice" Task Completion Time ($p < .001$)
with Expert Mode Taking Less Time

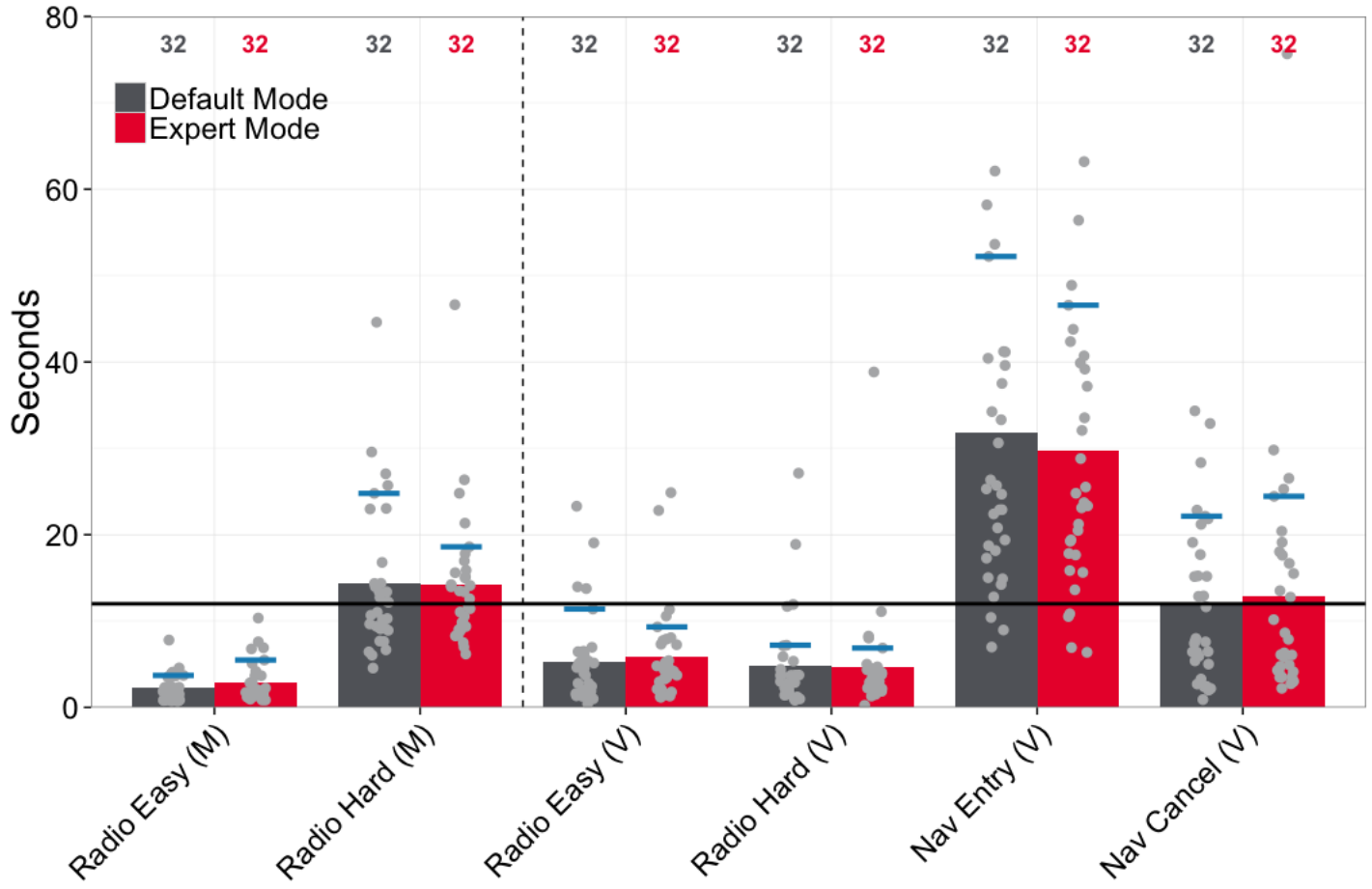


Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu

Study 3: Total Off-Road Glance Time (TEORT)

Did Not Differ Between Modes for "Voice" Tasks (p = 0.99).

Suggesting that drivers were not unnecessarily looking at the confirmation message screen in the default mode.



12 second threshold shown as a solid line. The longer individual line above each bar represents the 85% point in the sample distribution for each task.

Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu

Some General Conclusions

- The voice-command interface showed advantages in lower workload and visual engagement in some activates (e.g. radio tuning)
- **Cognitive load** for the voice-command tasks studied was **generally lower than expected** (based on self-report, physiology, driving performance)
- **Visual demand** for some voice-command tasks was **higher than might be expected**
- Voice recognition was higher than expected with only 6 of 193 subjects being "dropped" for issues
- Reducing the amount of audio content listening time required and confirmatory responses (expert mode) shortened task time but did not appreciably reduce visual demand

Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu

Limitations & Caveats Studies 1-3

- Generalizability
  - Tasks
  - Road environment
  - Vehicle selection
  - Experimental "pacing" of tasks
- The "demands" observed were not clearly linked to overt safety issues in driving performance metrics.
  - While conceptual concerns can be raised and are worthy of study, they should be interpreted cautiously
- Voice interfaces are rapidly evolving and the system tested here may not be representative of current technologies.

Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu

iPod iOS 5 “Google Maps” vs. Garmin GPS

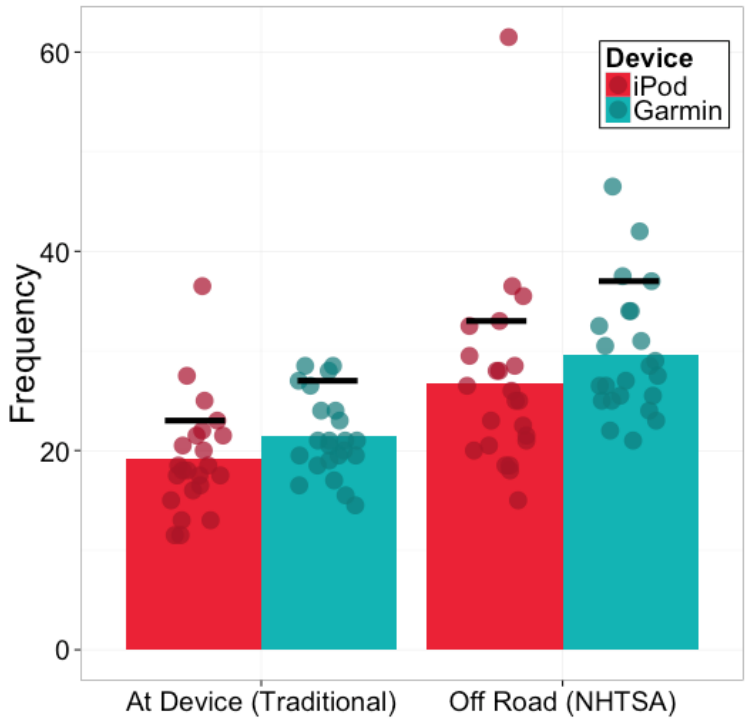
Brief description of a simulation study

- Compared performance on a visual-manual full address destination entry task using:
  - iPod iOS 5 “Google Maps”
  - Garmin portable GPS
- A total of 23 participants between 20 and 34 years (M=26) drawn from the greater Boston area were included in the analysis sample

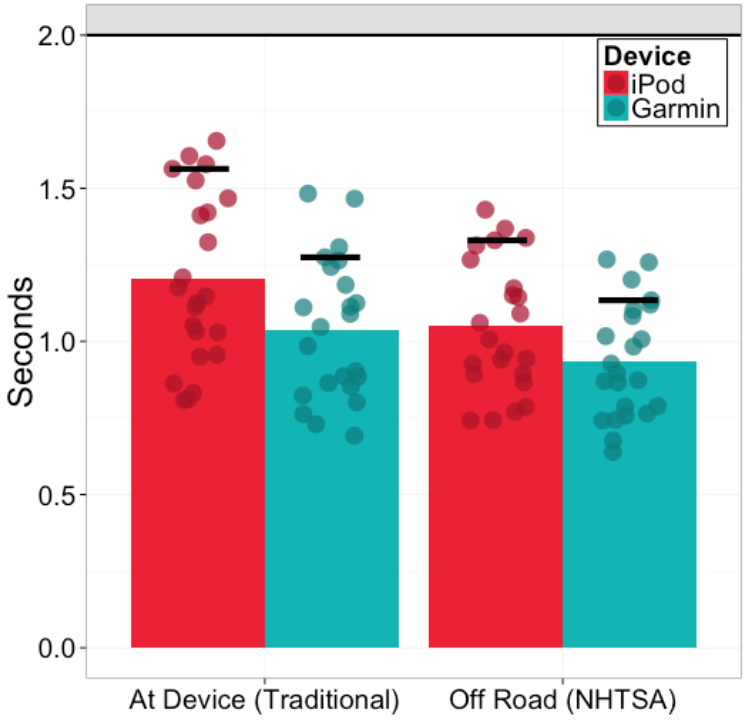
Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu

Glance Behaviors (I)

Glance Frequency



Mean Single Glance Time

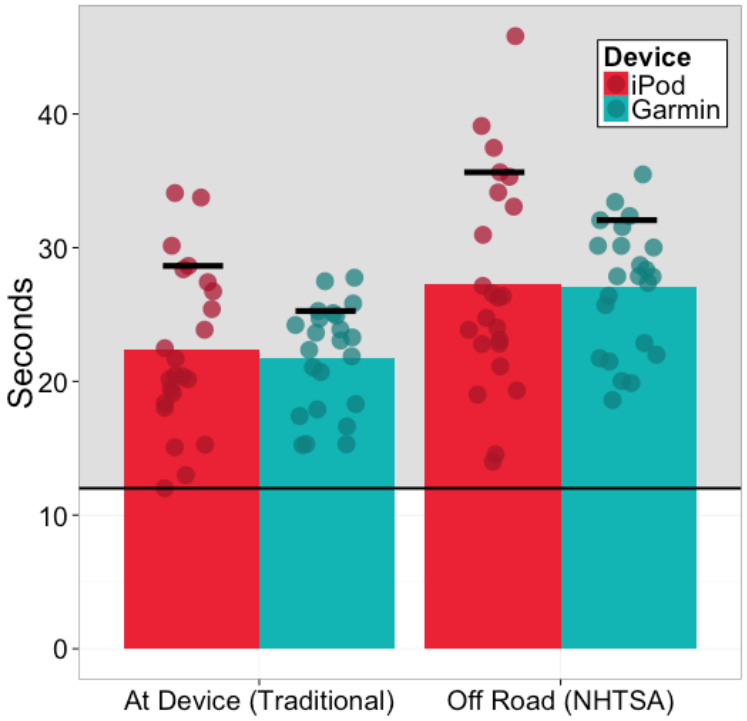


Entry with the Garmin required significantly **more frequent but shorter glances** than the iPod

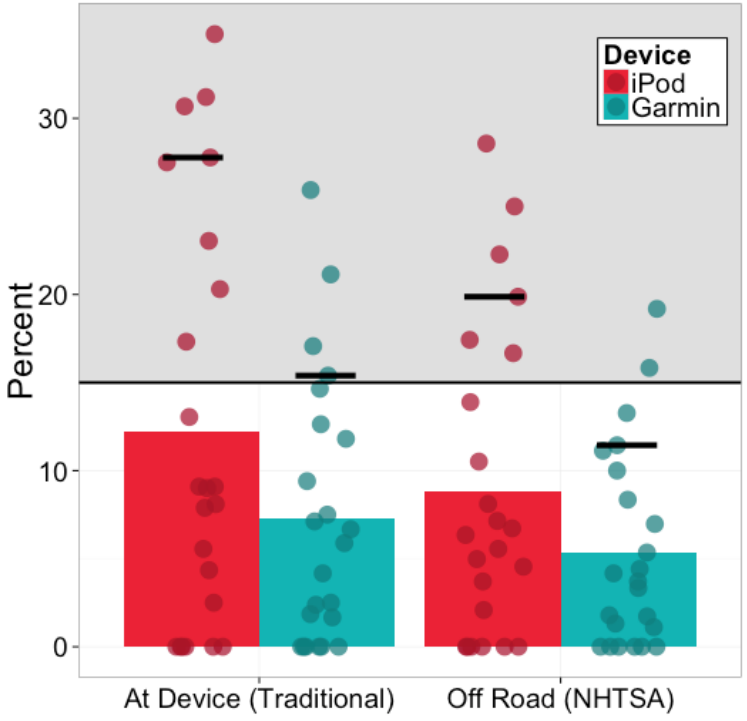
Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu

Glance Behaviors (II)

Total Glance Time



Long Duration Glances



No statistical **difference** in **total glance time by device** but clearly more subjects exhibited **longer duration glances with the iPod**

Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu

Samsung Galaxy Mode Comparisons

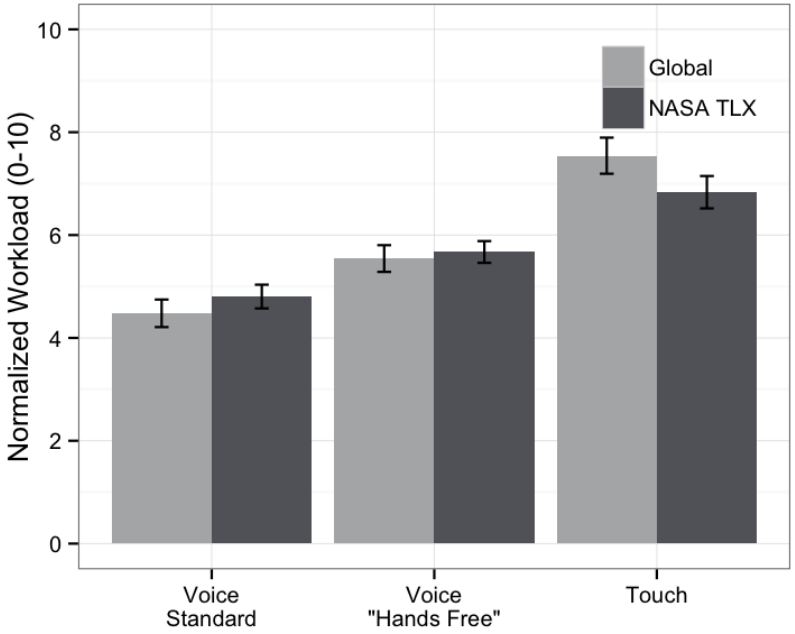
Brief description of a simulation study

- Compared performance of a full address destination entry task using several interaction modes available on the Samsung Galaxy smartphone:
  - Touch
  - Voice
  - "Hands Free" Voice
- A total of 22 participants have been drawn to date from two age groups, equally balanced between genders

Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu

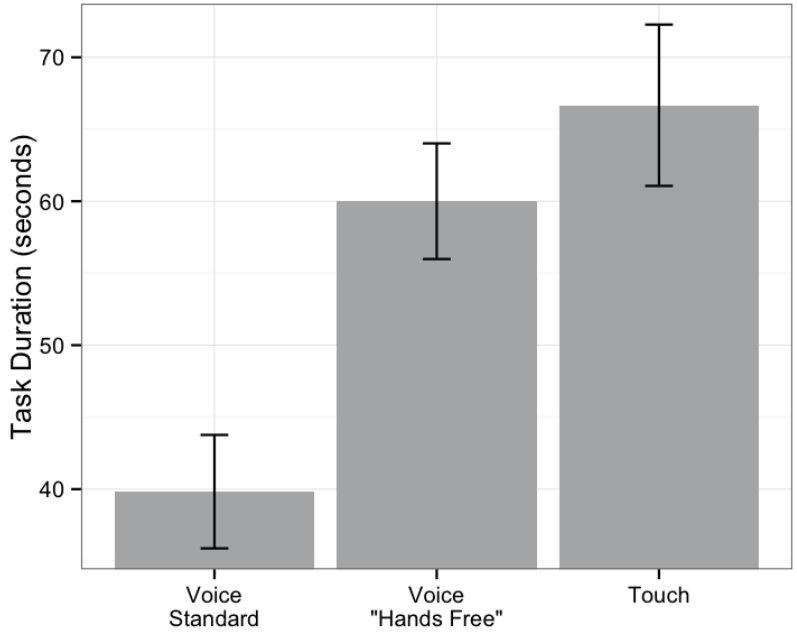
Results (I)

Workload



$p < .001$ for both scales

Task Time



$p < .001$

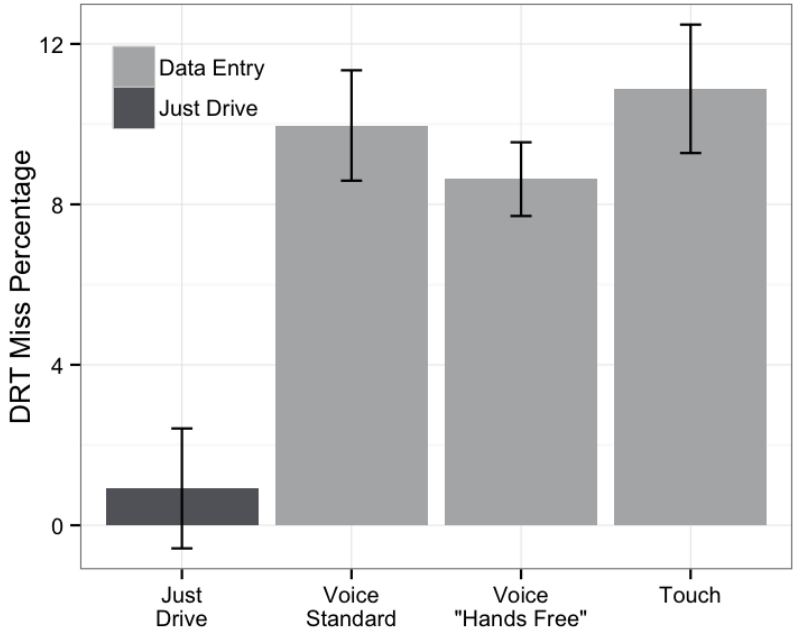
The **standard non-driving voice mode shows clear statistical advantages** over the "hands free" mode

(all voice pairwise p-values<.05)

Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu

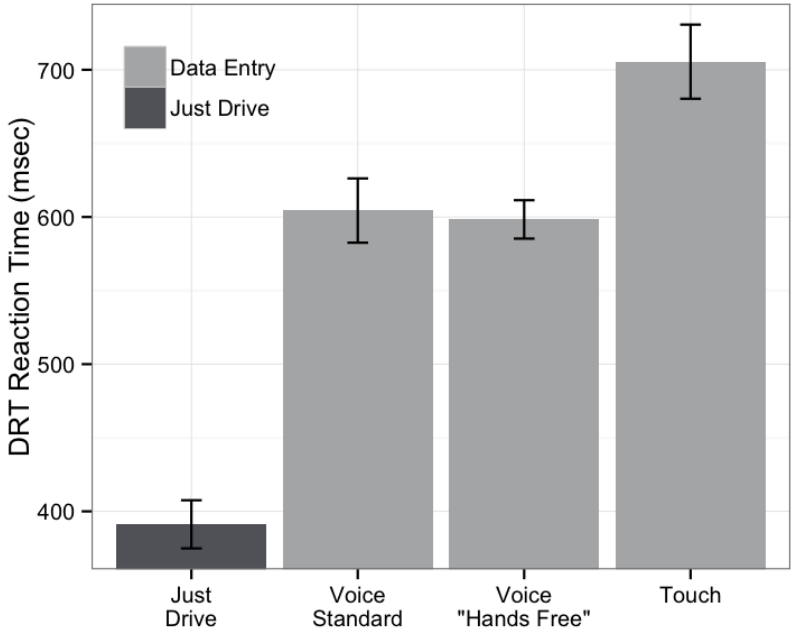
Results (II)

DRT Misses



$p < .01$

DRT Reaction Time



$p < .001$

The **DRT** shows **no statistical discrimination** between the two voice interface modes.

Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu

Google Glass vs. Samsung Galaxy

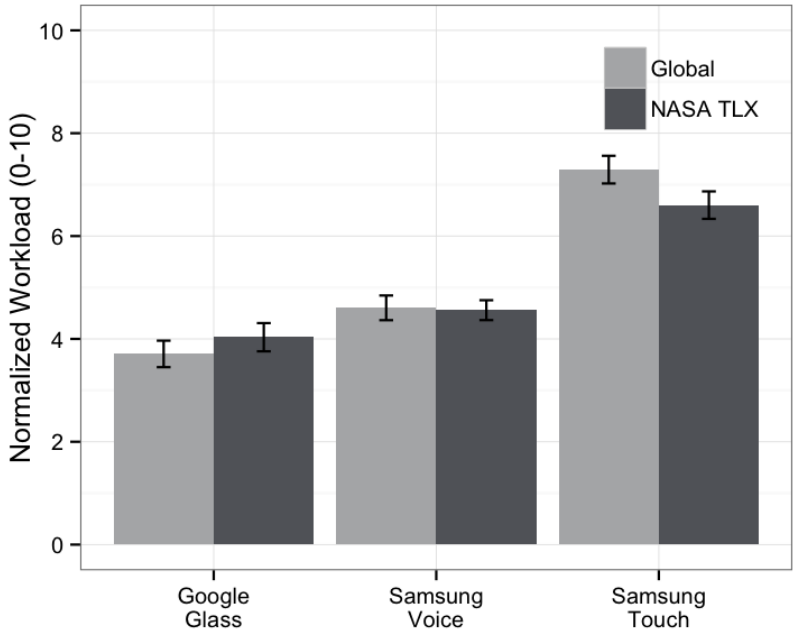
Brief description of a simulation study

- Compared performance of a full alphanumeric destination entry task using:
  - Google Glass
  - "Driver mode" voice interface of a Samsung Galaxy smartphone
  - Touch interface of a Samsung Galaxy smartphone
- A total of 24 participants were drawn from a college-age sample (mean age 25.0 years)
  - Native English speakers
  - Technologically experienced (considered as a best case example of technology early adopters likely to use the Glass system)

Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu

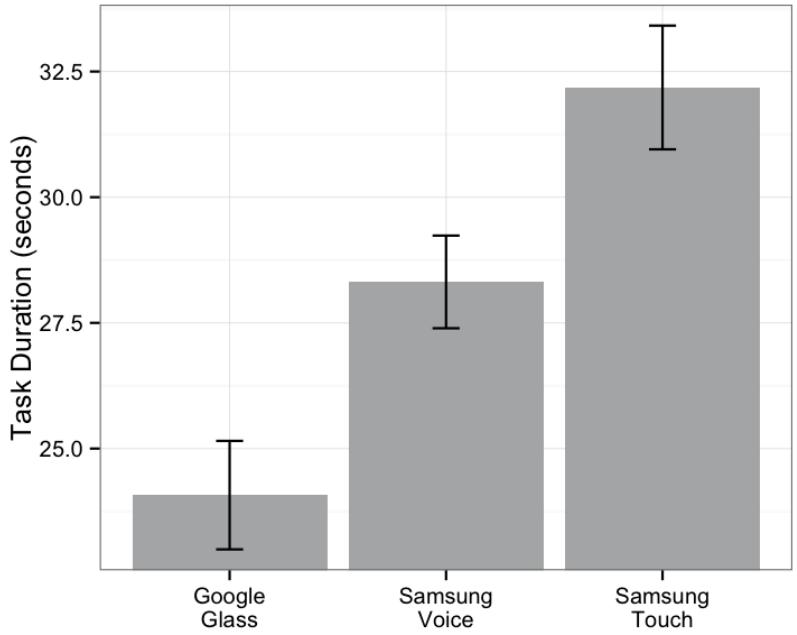
Results (I)

Workload



$p < .001$ for both scales

Task Time



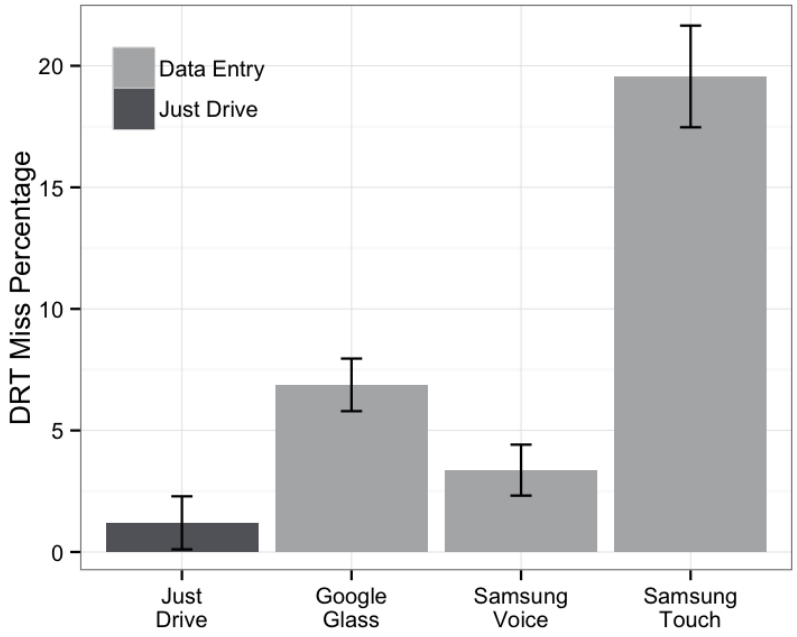
$p < .001$

Both **voice-interfaces** were given **lower workload** ratings than the Samsung touch interface. The **Google Glass dialog structure** resulted in a **shorter interaction**.

Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu

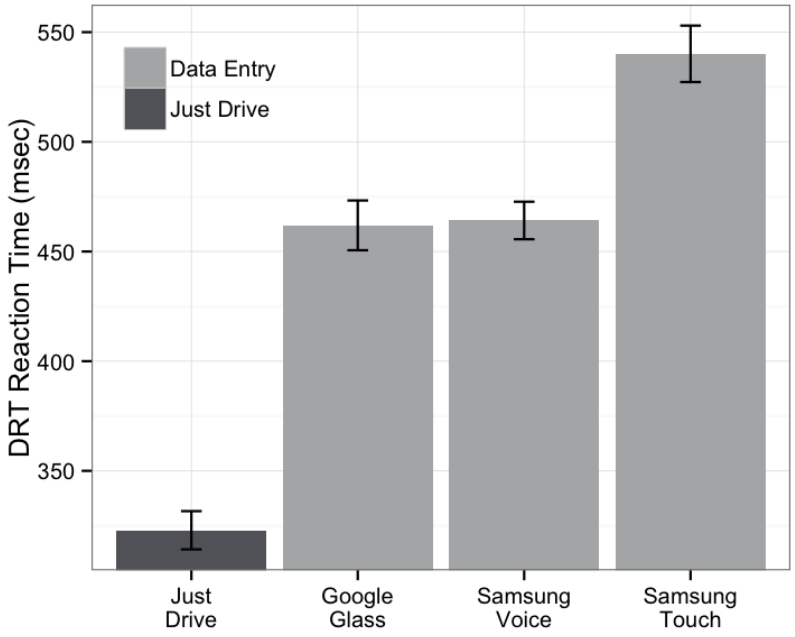
Results (II)

DRT Misses



$p < .001$

DRT Reaction Time



$p < .001$

The **DRT** reaction time shows **no statistical discrimination** between the two voice interface modes **but the miss percentages are different,** clear **advantages to voice over touch**.

Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu

Key Observations

- Results with a number of different interfaces illustrate that with **modern DVI's** the **attentional draws can be highly multimodal** (combinations of visual, manual, auditory, vocal, haptic, etc.)
- Thus, future **designs need to be** developed and **evaluated considering all resource demands** that may arise from interactions with these systems
- While **demand is not easily linked to safety**, current research indicates that the strongest connection is between increased off-road visual attention (especially long duration glances) and adverse road events - suggesting that the assessment of visual demand is key when assessing "voice" interfaces
- **Safety efforts may need to focus more on defining what is an acceptable activity to complete underway vs. the cost benefits of different implementations that remain difficult to holistically compare using the best available scientific methods**

Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu

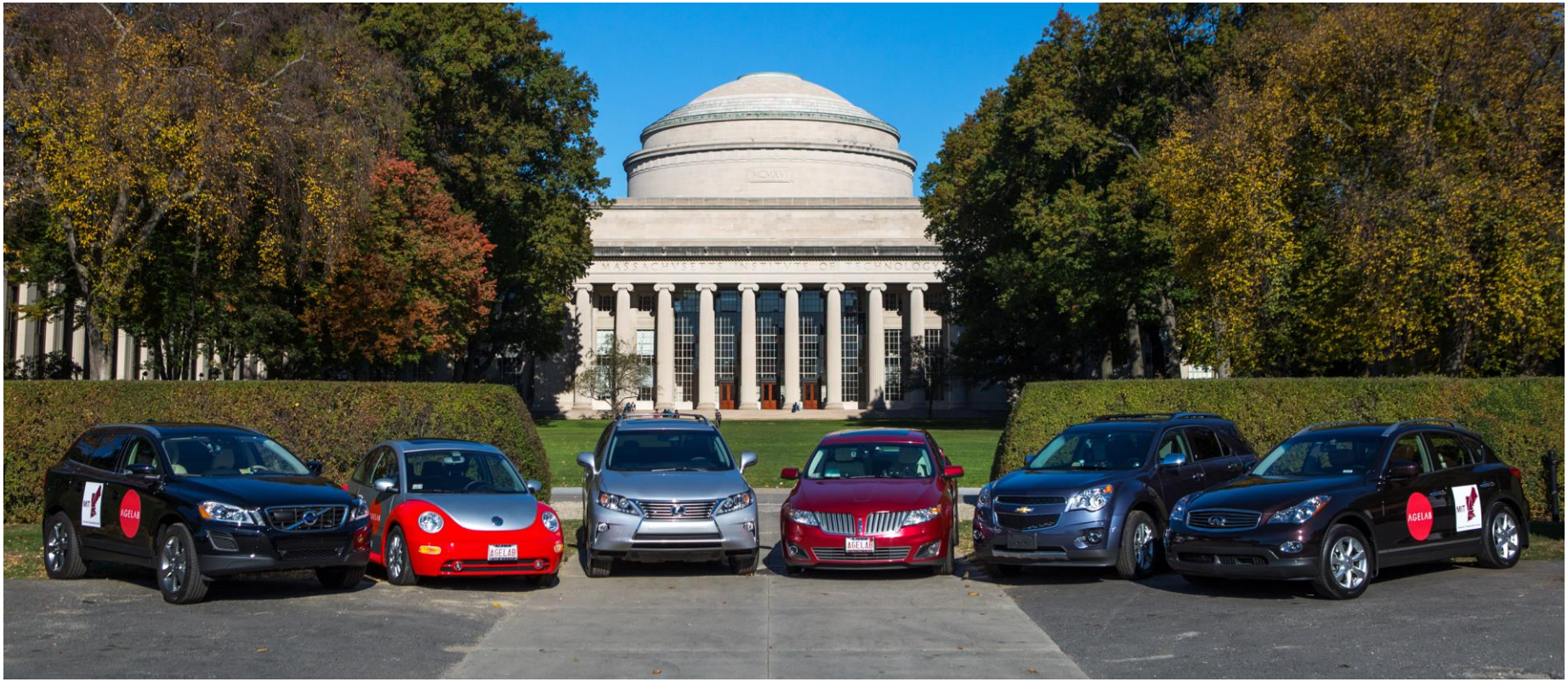
Next Steps

Looking at other interfaces

- A field study assessing differences in multi-modal demand between two new vehicle systems and a smartphone (data collection complete)
- A set of field studies will assess different operational characteristics of production level automotive voice systems in at least three additional vehicles
- Multiple studies exploring modality differences among hand-held technologies (experiments ongoing)
- An broad industry effort to develop comprehensive measures that can be used to better assess and contribute to understanding of methods that can optimize multi-modal demands

Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu

Questions



Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu

For Additional Details See

Dopart, C., Häggman, A., Thornberry, C., Mehler, B., Dobres, J. & Reimer, B. (2013). A Driving Simulation Study Examining Destination Entry with iPhone iOS 5 Google Maps and a Garmin Portable GPS System. Proceedings of *the 57th Annual Meeting of the Human Factors and Ergonomics Society.* San Diego, CA. pp. 1889-1893.

Reimer, B. & Mehler, B. (2013). The Effects of a Production Level "Voice-Command" Interface on Driver Behavior: Summary Findings on Reported Workload, Physiology, Visual Attention, and Driving Performance. MIT AgeLab White Paper No. 2013-18A. Massachusetts Institute of Technology, Cambridge, MA.

Reimer, B., Mehler, B., Dobres, J. & Coughlin, J.F. (2013). The Effects of a Production Level "Voice-Command" Interface on Driver Behavior: Reported Workload, Physiology, Visual Attention, and Driving Performance. MIT AgeLab Technical Report No. 2013-17A. Massachusetts Institute of Technology, Cambridge, MA.

Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu