MITSUBISHI ELECTRIC RESEARCH LABORATORIES
Cambridge, Massachusetts

Voice Search “Shortcuts” for In-Car Infotainment

Bret Harsham
Principal Member of Research Staff
Mitsubishi Electric Research Labs (MERL) Speech Group

Last minute speaker change courtesy of volcanic activity, originally

Garrett Weinberg
Member of Research Staff
Mitsubishi Electric Research Labs (MERL) Speech Group

Context of this talk

- MERL is one of the research labs of Mitsubishi Electric
  - (Not Mitsubishi Motors!)
- MERL speech group focus:
  - Pre/Post processing of speech and audio
  - UI design that can improve the usability of speech interfaces.
- Usability != ASR performance
  - ASR performance affects UI design.
  - UI Design affects ASR performance.

Focus on Push-to-Talk

MITSUBISHI ELECTRIC
Changes for the Better

State of the Art

- Say anything, anytime? Not exactly.
- System must be told when to listen
- Single push-to-talk (PTT) button launches a dialog
- Content “silos” improve accuracy
  - Single-shot approaches are emerging, but thus far are domain-restricted.

Remix of image by Wikimedia contributor “Cerafino”

What Else are Buttons Good For?

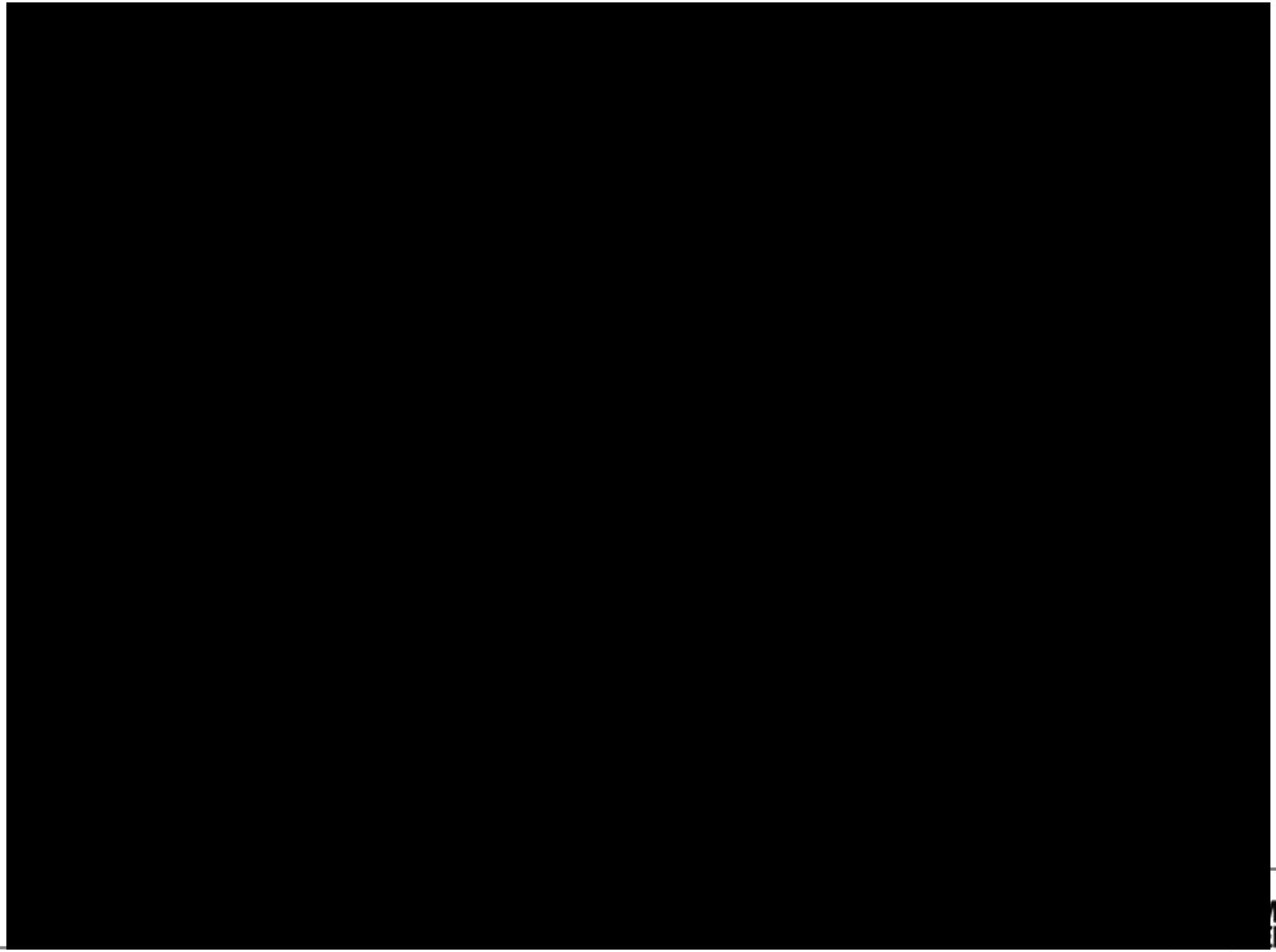
MITSUBISHI ELECTRIC
Changes for the Better

Contextual Push-to-Talk

- *Any* button can be a “listen” button
  - Just tap twice instead of once
- Choice of button provides context to ASR engine
  - Search *what*
  - Shuffle *what*
  - Call *whom*
- Advantages:
  - Skip dialog steps
  - Exploit motor memory
  - Augment rather than replace existing UI

Experimental Validation

- Research Question:
  - Are multiple, contextual push-to-talk buttons better than the conventional, single-PTT approach?
- Hypotheses:
  - Multi-PTT allows for faster task completion.
  - Driving performance and visual attention using Multi-PTT are no worse than Single-PTT.
- Setup:
  - MERL Driving Simulator
  - Eye tracker (Seeing Machines' FaceLAB)
  - 18 subjects (all tried both interfaces and drove a control session)



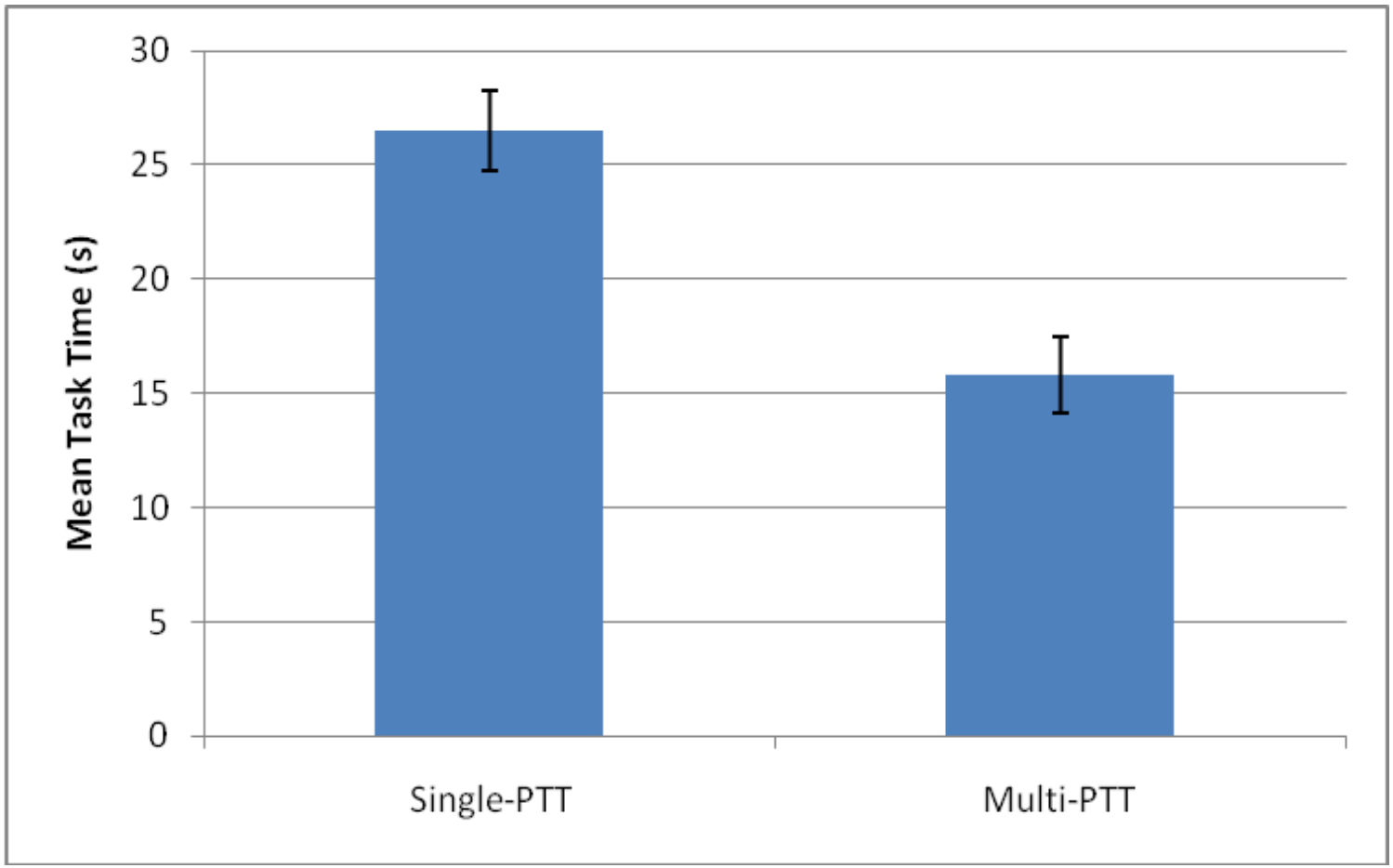
Experimental Results

- No significant difference between Single- and Multi-PTT in steering or lane keeping
- Significant differences found:

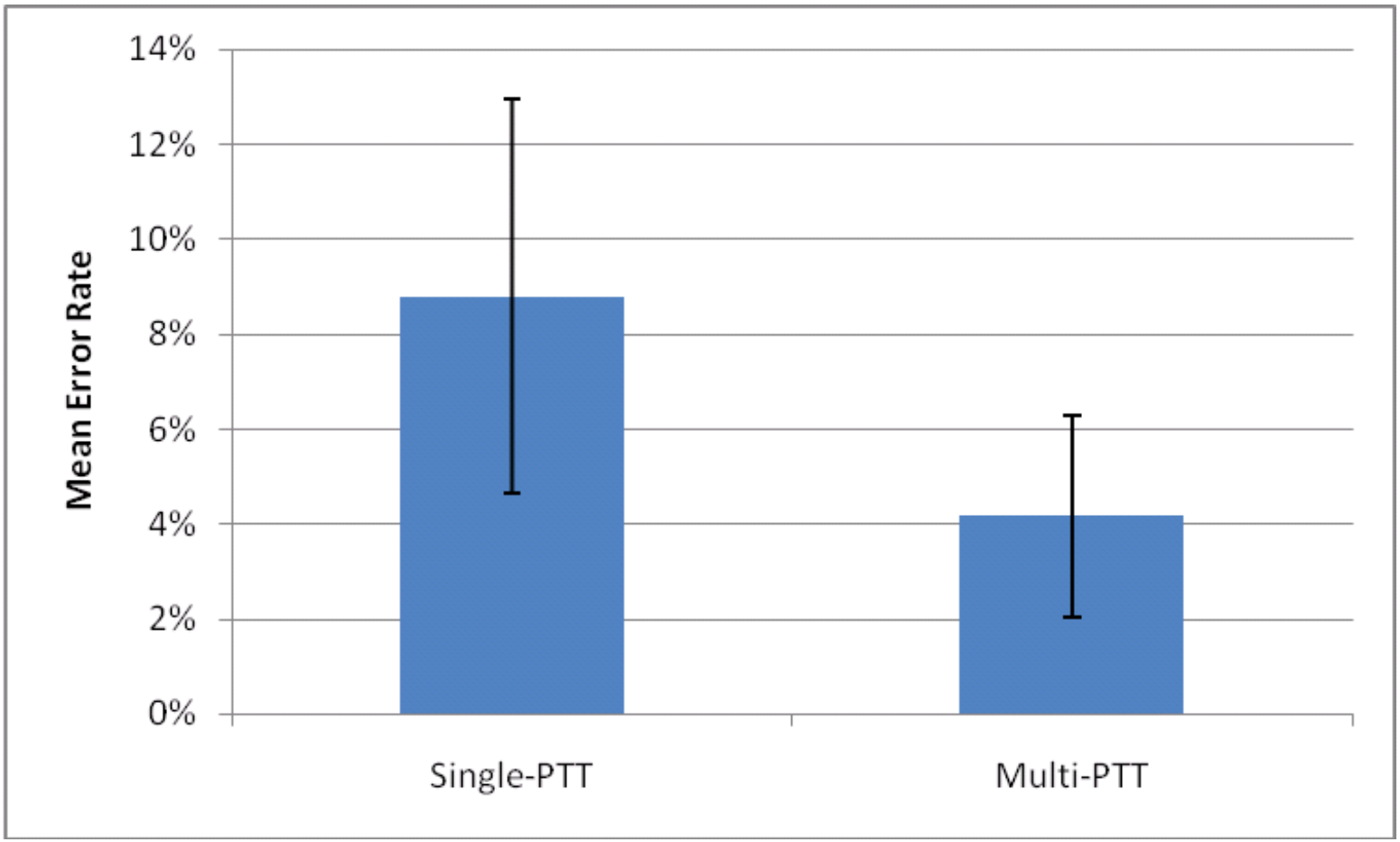
Measure	Single-PTT	Multi-PTT
Variance in following distance (lower is better)	106.7	42.3
Glances away from forward roadway (avg. per task)	11.2	7.9
Workload (NASA Task Load Index) → control session (driving only) rated 29.5	55.2	49.1
Mean task time (seconds)	**26.5**	**15.8**
Mean task error rate	**8.8%**	**4.2%**

- Questionnaire results: 78% of subjects would prefer using the Multi-PTT variant in their cars

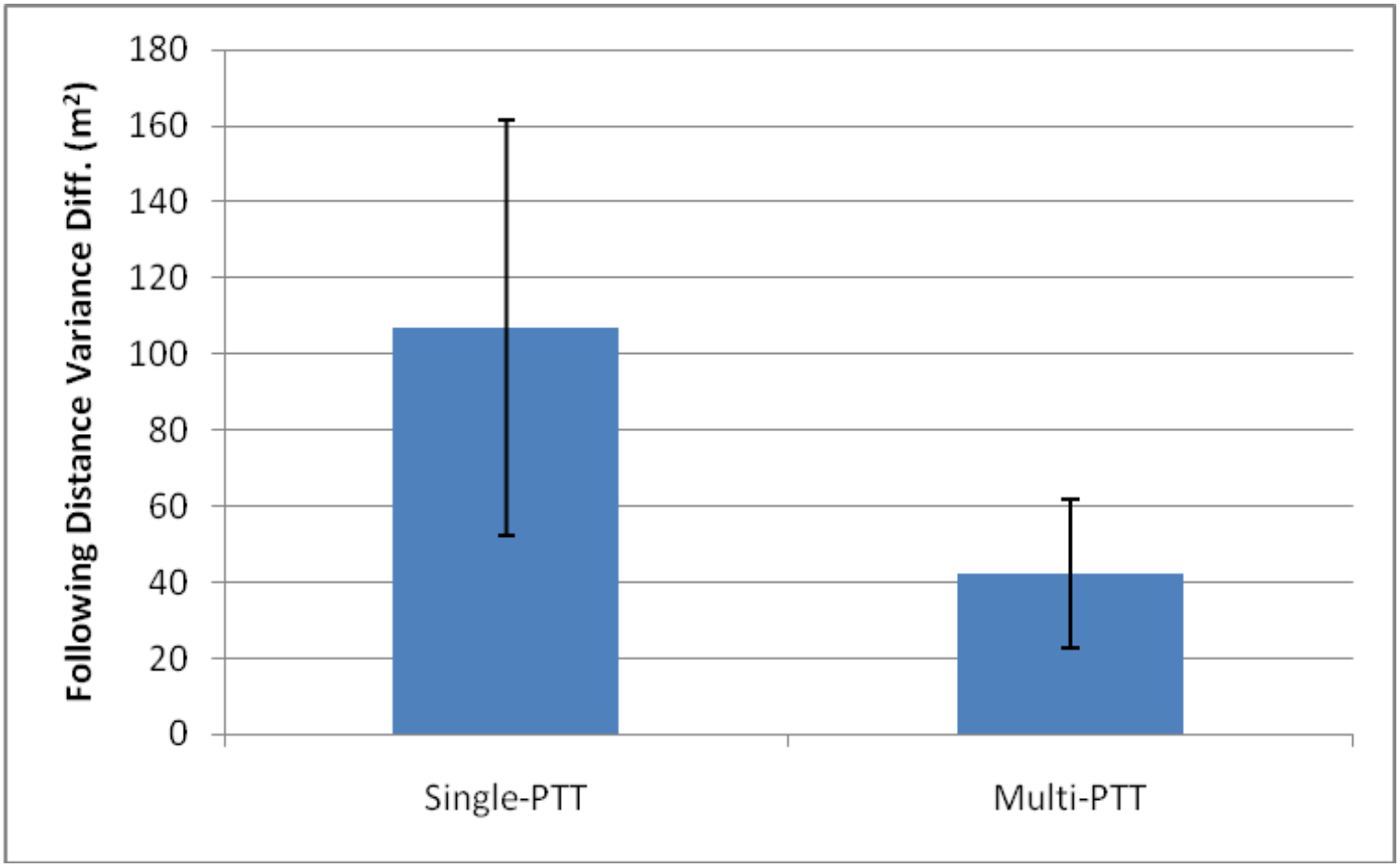
Experimental Results – Mean Task Time



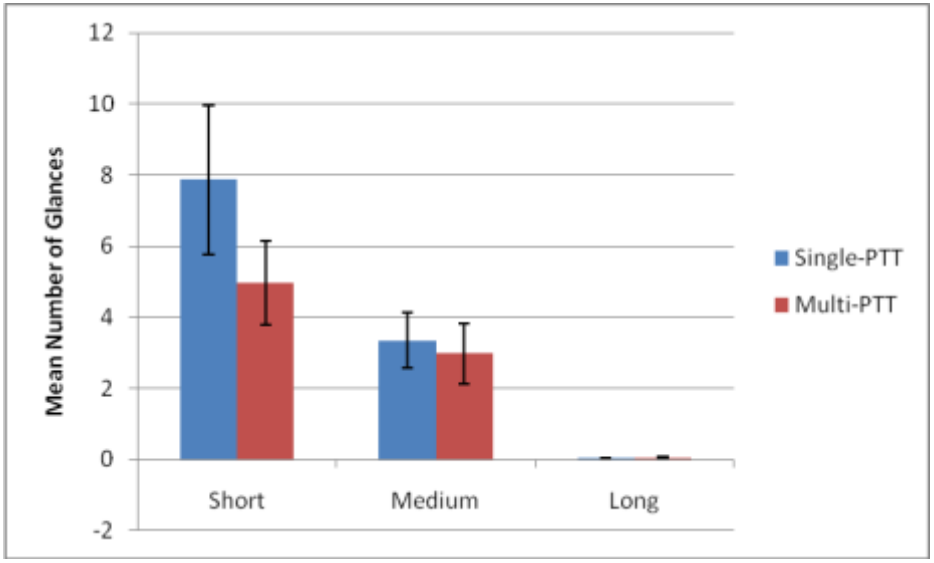
Experimental Results – Mean Task Error Rate

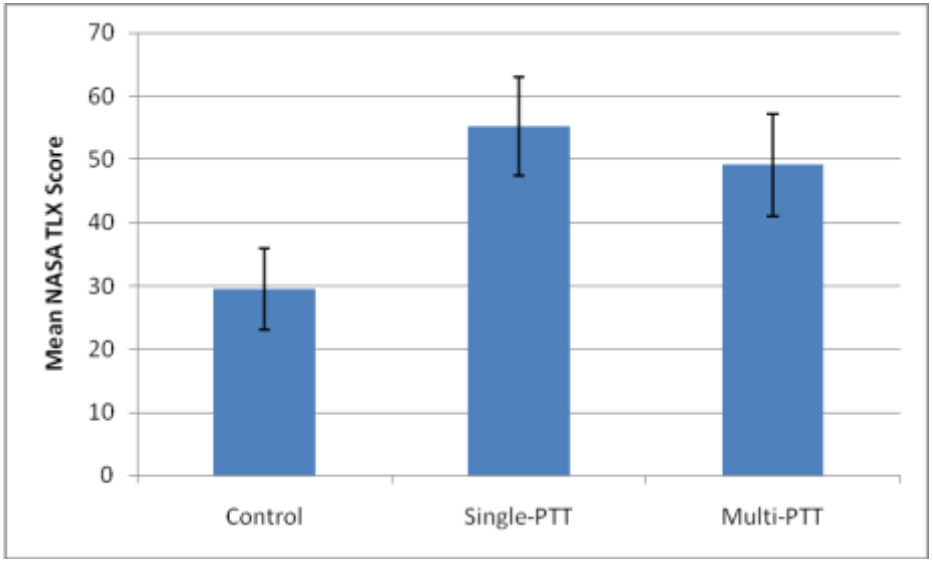


Experimental Results – Variability in Following Distance



Experimental Results – Eye Gaze and Workload





Summary

- We proposed and validated multiple, contextual PTT buttons.
- Interaction durations were reduced, and more attention was paid to the driving task at hand.
- The design is simple and intuitive.

Acknowledgements

- Collaborators Bret Harsham (MERL), Zeljko Medenica (U. New Hampshire), and Cliff Forlines (MERL)

- Albrecht Schmidt & Dagmar Kern (U. Duisburg-Essen) for use of their photos from the 2007 Frankfurt Auto Show

- Kent Wittenburg (MERL) for advice and guidance

For more details, please see our upcoming publication:

Weinberg, G. et al. “Contextual Push-to-Talk: Shortening Voice Dialogs to Improve Driving Performance.” To appear in *Proceedings of the 12th International Conference on Human-Computer Interaction with Mobile Devices and Services* (MobileHCI 2010).