New **E**ngland
University
Transportation
Center

NE University Transportation Center
77 Massachusetts Avenue, E40-279
Cambridge, MA 02139
Phone: 617-253-0753
Fax: 617-258-7570
web.mit.edu/utc

Principal Investigator:	Jim Glass	**Co-Principal Investigator:**	
Title:	Senior Research Scientist (CSAIL)	**Title:**	
University:	MIT	**University:**	
Email:	glass@mit.edu	**Email:**	
Phone:	(617) 253-1640	**Phone:**	

Final Report

Project Title:

Evaluation of a Natural Speech Based Informational Inquiry System as a Potential Means to Increase Transit Utilization

Project Number:	Project End Date:	Submission Date:
MITR22-6	3/31/2013	5/10/2013

This project proposed to explore the potential of a user friendly, natural speech based information inquiry application as one means of increasing public transit utilization. It suggested that a key challenge to expanding transit ridership is to encourage people who have not used a system to develop familiarity with it. The initial phase of the project focused on gathering strategic information from potential users on implementations of a speech interface they would be most likely to use and the types of information they perceive as being most useful in supporting and/or increasing their use of the transit system. Attention was focused on how demographic factors such as age and gender, as well as technology experience, impact on preferences and likely willingness to consider using various implementations of such systems. Parallel to the start of this project, the Massachusetts Bay Transportation Authority (MBTA) and Massachusetts Department of Transportation developed broad accessibility to real-time data for the commercial development of applications for traffic and public transit availability. Google and other third party developers quickly developed robust applications utilizing this information that have undergone rapid optimization making the development of a research based evaluation difficult. In essence, the state of the technology was moving too fast to take a snapshot of an application for an extended assessment of its impact on mobility. The project was therefore extended somewhat to include a broader examination of the impact of speech-based interfaces in a wider variety of transportation-related scenarios.

In the first phase of this effort, research was directed towards expanding the transit information capability of the CityBrowser spoken dialogue prototype, and incorporating crowdsourcing methods for query collection to enhance the robustness of the system. The CityBrowser content database was expanded to include information about subways, and initial investigation was also undertaken to incorporate transit information that had recently been made available for the Boston area, but these data ultimately proved too unreliable to incorporate into the system. Subsequent versions of CityBrowser incorporated directions provided by a Google API. Crowdsourcing-based methods were investigated to expand the amount of queries available to train the semantic understanding component of CityBrowser. Amazon Mechanical Turk was used to collect queries, and have users interact with the system (Liu, Glass, Seneff, 2010).

In a second component of this effort (Reimer, Mehler, McAnulty, Mehler, Garcia Perez, Manhardt, & Coughlin, 2013), a 2010 Lincoln MKS with a SYNC™ voice interface was assessed. The vehicle was instrumented for time synchronized recording of vehicle information from the controller area network (CAN) bus, a MEDAC System/3 physiology monitoring unit, FaceLAB® 5.0 eye tracking, cameras for capturing driver behavior and vehicle surroundings, and GPS tracking. Subjects were asked to complete six in-vehicle task areas: manual control of the radio, voice command control of the radio, navigation system destination entry, song selection (from an MP3 storage device), stored phone number dialing, and an auditory presentation / verbal response calibration task (n-back). Each task was presented twice. Data from 60 subjects distributed evenly across two age groups (20-29 and 60-69) and equally across gender was assessed.

An initial consideration of the data on perceived workload (subjective, self-report), compensatory behavior (relative speed), and objective estimates of cognitive workload (heart rate and SCL) suggests that there may be both potential benefits and cautions in the implementation the interface. Various task demands were considered in reference to the Radio-

Manual Hard tuning task recommended in the Alliance guidelines (2006) as an upper bound for an acceptable level of secondary demand on a driver. The subjective workload rating of all other HMI interactions (except for the Song Fail task, a deliberate test of a failure condition) was lower. The Radio-Manual Hard tuning task resulted in the largest compensatory speed reduction and was associated with the highest levels of heart rate and a high level of skin conductance among the vehicle interface functions assessed. When the voice interface was used for the same task, it showed a lower subjective workload rating and comparable or nominally less speed compensation and physiological arousal measures.

While manual entry of destination information is considered by many to exceed the acceptable level of visual demand while driving and locked-out, voice interfaces are viewed largely as an acceptable method of entering a destination. Relative to the Radio-Manual Hard task, the voice entry of a destination using the Sync system produced lower subjective workload, less reduction in speed (compensatory behavior) and a lower increase in arousal (heart rate / skin conductance). Interestingly, selecting a pre-set radio station was given a higher subjective workload rating when operated using the voice-command interface and resulted in nominally higher increase in physiological arousal. The location of the 3-levels of the n-back task across the various measures aligns well for its proposed use as a cognitive workload calibration metric (Mehler, Reimer & Coughlin, 2012). While Tom Ranney (NHTSA, 2011) suggested in his report on driver distraction that the most difficult demand level (2-back) should be considered as a first stage threshold for an acceptable dose of cognitive demand produced by an in-vehicle system, this line of research suggests that the 1-back level may be a more suitable threshold.

Reimer, B. Mehler, B., McAnulty, H., Munger, D., Mehler, A., Garcia Perez, E.A., Manhardt, T. & Coughlin, J.F. (2013). A Preliminary Assessment of Perceived and Objective Scaled Workload of a Voice-Based Driver Vehicle Interface. Paper to appear in the proceedings of the 7th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design, Bolton Landing, New York, June 18-19, 2013.

Abstract: Interaction with a voice-command interface for radio control, destination entry, MP3 song selection, and phone dialing was assessed along with traditional manual radio control and a multi-level audio–verbal calibration task (n-back) on-road in 60 drivers. Subjective workload, compensatory behavior, and physiological indices of cognitive workload suggest that there may be both potential benefits and cautions in the implementation of a representative production level interface.

Liu, S., Glass, J. and Seneff, S. (2010). A Collective Data Generation Method for Speech Language Models. Proc. IEEE Spoken Language Technologies Workshop., Berkeley, CA, 2010.