UTDrive: THE SMART VEHICLE PROJECT

Pongtep Angkititrakul and John H.L. Hansen
Center for Robust Speech Systems (CRSS)
Erik Jonsson School of Engineering and Computer Science
The University of Texas at Dallas, Texas, USA
angkitit@utdallas.edu, John.Hansen@utdallas.edu

ABSTRACT

This paper presents research activities of UTDrive: the smart vehicle project. The objectives of the UTDrive project are to collect and research rich multi-modal data recorded in actual car environments for analyzing and modeling driver behavior. The models of driver behavior under normal and distracted driving conditions can be used to create improved human-machine interactive systems and reduce vehicle accidents on the road. The UTDrive corpus consists of audio, video, brake/gas pedal pressure, head distance, GPS information (e.g., position, velocity), and CAN-Bus information (e.g., steering-wheel angle, brake position, throttle position, and vehicle speed). Here, we describe our in-vehicle data collection framework, data collection protocol, dialog and secondary task demands, data analysis, and preliminary experimental results. Finally, we discuss our proposed multi-layer data transcription procedure for in-vehicle data collection and future research directions.

1. INTRODUCTION

There has been significant interest in development of effective human-machine interactive systems in diverse environmental conditions. Voice based route navigation, entertainment control, and information access (voice mail, etc.) represent domains for voice dialog systems in the car environment. The in-vehicle speech-based interactive systems allow the driver to stay focused on the road. Several studies [2, 5] have shown that drivers can achieve better and safer driving performance while using speech interactive systems to operate an in-vehicle system compared to manual interfaces. Although better interfaces can be incorporated, operating a speech interactive system will still divert a driver's attention from the primary driving task with varying degrees of distraction. Ideally, drivers should pay primary attention to driving versus managing other secondary tasks that are not immediately relevant to the primary driving task. With current life styles and advancement for in-vehicle technology, it is inevitable that drivers will perform secondary tasks, or operate driver assistance and entertainment systems while driving. In general, the common tasks such as operating the speech interactive systems in a driving environment include cell-phone dialing, navigation/destination interaction, e-mail processing, music retrieval, and generic command and control for in-vehicle telematics system. If such secondary tasks or distractions fall within the limit of the amount of spare cognitive load for the driver, he or she can still focus on driving. Therefore, the design of safe speech interactive systems for in-vehicle environments should take into account factors from the driver's cognitive capacity, driving skills, and their degree of proficiency for the cognitive application load. With knowledge of the human factors, an effective driver behavior model with real-time driving signals can be integrated into a smart vehicle to support or control driver assistance systems to manage driver distractions (e.g., suspend or adapt applications in a situation of heavy driving workload, provide alert if the distraction level is higher than a safety threshold).

Driving is a multitasking activity that comprises discrete and continuous nature of drivers' control actions to manage various driving and non-driving tasks. Over the past several decades, modeling driver behavior has drawn much research attention. A number of studies have shown that driver behavior can be modeled and anticipated by the patterns of driver's control of steering angle, steering velocity, car velocity, and car acceleration [10], as well as driver identity itself [4, 13]. Miyajima, et al. [9] efficiently employed spectral-based features of the raw pedal pressure signal with Gaussian mixture model (GMM) framework to model driver characteristics. Building effective driver behavior recognition framework requires a thorough understanding of human behavior and the construction of a mathematical model capable of both explaining and predicting the drivers' behavioral characteristics. In recent studies, several researchers have defined different performance measures to understand driving characteristics and to evaluate their studies. Such measures include driving performance, driver behavior, task performance, etc. Driving performance measures consist of driver inputs to the vehicle or measures of how well the vehicle was driven along its intended path [1]. Driving performance measures can be defined by longitudinal velocity and acceleration, standard deviation of steering-wheel angle and its velocity, standard deviation of the vehicle's lateral position (lane keeping), mean following distance (or head distance), response time to brake, etc. Driver behavior measures can be defined by glance time, number of glances, awareness of drivers, etc. Task performance measures can be defined by the time to complete a task and the quality of the completed task (e.g., do drivers acquire information they need from cell-phone calling). Therefore, multi-modal data acquisition is very important to these studies.

UTDrive is part of a NEDO-supported international collaboration between universities in Japan, Italy, Singapore, Turkey, and USA. The UTDrive (USA) project has been designed specifically to:

a) Collect rich multi-modal data recorded in a car environment (i.e., audio, video, gas/brake pedal pressures,

following distance, GPS information, and CAN-Bus information including vehicle speed, steering degree, engine speed, and brake position),

b) Assess the effect of speech interactive system on driver behavior,

c) Formulate better algorithms to increase accuracy for in-vehicle ASR systems,

d) Design dialog management which is capable of adapting itself to support a driver's cognitive capacity, and

e) Develop a framework for smart inter-vehicle communications.

The results of this project will help to develop a framework for building effective models of driver behavior and driver-to-machine interactions for safe driving. This paper is organized as follows. Section 2 discusses the details of the multimodal data acquisition in actual car driving environment. Section 3 describes data collection protocol of the UTDrive project. Section 4 is devoted to the driving signals. Driver distraction is discussed in Section 5. Section 6 concentrates on driver behavior modeling. Section 7 discusses the multilayer transcription for in-vehicle corpus. Finally, Section 8 concludes the paper with future work.

2. MULTIMODAL DATA ACQUISITION

In this section, we describe an overview of the hardware setup for multimodal data acquisition system.

2.1. Audio

A custom designed five microphone array with omni-directional Knowles microphones was constructed on top of the windshield next to the sunlight visors to capture audio signals inside the vehicle. Each microphone was mounted in a small movable box individually attached to an optical rail, as show in Fig. 1. This particular design allows the spacing between each microphone component to be adjustable into various scales (e.g., linear, logarithmic, etc.) across the width of the windshield. In addition, the driver speech signal is also captured by a close-talk microphone and throat microphone. This microphone provides a reference for the speaker (driver), and allows the driver to move their head freely while driving the data-collection vehicle.

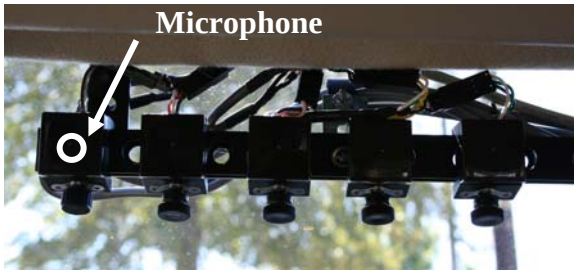


Figure 1: Custom-designed adjustable-spaced microphone array

Since there are different kinds of noise (e.g., A/C, engine, turn signals, passing vehicles) present in the driving environment, the microphone array configuration allows us to apply beam-forming algorithms to enhance the quality of input speech signals [7, 14]. Another aspect present in the car environment is a variety of background noises that effect the quality of the input acoustic signal for the speech interface. More importantly, drivers have to modify their vocal effort to overcome perceived noise levels, namely the Lombard effect [8]. Such effects on speech production (e.g., speech under stress) can degrade the performance of automatic speech recognition (ASR) system more than the ambient noise itself [6]. At a higher level, interacting with an ASR system when focused on driving may result in a speaker missing audio prompts, using incomplete grammar, adding extra pauses or fillers, or extended time delays in a dialog system. Desirable dialog management should be able to employ multi-modal information to handle errors and adapt its context depending on the driving situations.

2.2. Video

Two Firewire cameras are used to capture visual information of driver's face region and front-view of the vehicle, as show in Fig. 2. Visual cues of driver characteristics such as head movement, mouth shape, and eye glance are essential for studying driver behavior. In addition, several studies have shown that combining audio and video information from the driver can improve ASR accuracy for low SNR speech [3, 15]. Integrating both visual and audio content allows us to reject unintended speech prior to speech recognition and significantly improve in-vehicle human-machine dialog system performance [15] (e.g., determining the movement of the driver's mouth, body, and head positions).

2.3. CAN-Bus Information

As automotive electronics advance and government required standards evolve, control devices that meet these requirements have been embracing modern vehicle design, resulting in the deployment of a number of electronic control systems. The Controller Area Network (CAN) is a serial, asynchronous, multi-master communications protocol suited for networking vehicle's electronic control systems, sensors, and actuators. The CAN-Bus signal contains real-time vehicle information in the form of messages integrating many modules, which interact with the environment and process high and low speed information. In the UTDrive project, we obtain the CAN signals from the OBD-2 port through the 16 points J1962. Information captured from CAN while the driver is operating the vehicle (e.g., steering wheel angle, brake position, engine speed, and vehicle) are desirable in studying driver behavior.

2.4. Transducers and Extensive Components

In addition, the following transducers and sensors are included in the UTDrive framework:

- Brake and gas pedal pressure sensors: provide continuous measurement of pressure the driver puts on the pedals.
- Distance sensor: provides the following distance to the next vehicle.
- GPS: provides standard time and position of the vehicle.
- Hands-free car kit: provides safety during data collection and allows audio signals from both channels to be recorded.
- Biometrics: heart-rate and blood pressure measurement.

2.5. Data Acquisition Unit (DAC)

The key component of effective multimodal data collection is synchronization of the data. In our data collection, we use a fully integrated commercial data acquisition unit. With a very high sampling rate of 100 MHz, the DAC is capable of synchronously recording multi-range input data (i.e., 16 analog inputs, 2 CAN-Bus interfaces, 8 digital inputs, 2 encoders, and 2 video cameras), and yet allows acquisition rate for each input channel to be set individually. The DAC can also export all recording data as a video clip in one output screen, or individual data in its proper format with synchronous time stamps. The output video stream can be encoded to reduce its size, and then transcribed and segmented with an annotation tool. Fig. 2 shows a snapshot of a recording video clip with all data displayed on the screen (e.g., audio channels on top, two camera screens in the middle, sensors and CAN-Bus information on the left bottom, and GPS information on the right bottom).

Figure 2: A Snapshot from a recording screen.

In order to avoid signal interference, power cables and signal cables were wired separately on both sides of the car. The data acquisition unit is mounted on a customized platform on the backseat behind the driver. The power inverter and supplier units are designed to be housed in the trunk space. Fig. 3 shows the UTDrive data-collection vehicle and its components.

3. DATA COLLECTION PROTOCOL

For data collection protocol, each participant drives the UTDrive vehicle using two different routes in the neighborhood areas of Richardson-Dallas, TX; the first route represents a residential area environment and the second route represents a business-district environment. Each route takes 10-15 minutes. The participant drives the vehicle following each route twice; with the first being neutral driving and second is driving and performing secondary tasks. Due to safety concerns, the assigned tasks are common tasks with mild to moderate degrees of cognitive load (e.g., interacting with commercial automatic speech recognition (ASR) dialog system, reading signs on the street, tuning radio, having a conversation with the passenger, reporting activities, changing lanes, etc.) The participants are encouraged to drive the vehicle up to three sessions with at least one week separation between sessions, in order to achieve

Figure. 3: UTDrive vehicle and its components.

session-to-session variability. Figure 4 shows a map of the driving route and the assigned tasks. The assigned tasks are performed along each individual street and are alternated for three driving sessions. For example, a driver is requested to interact with a commercial voice portal while driving on one leg of the entire first session route. For the second and the third sessions, the driver is asked to interact with another commercial voice portal and have conversation with passenger while driving along the same leg of the entire route, respectively. This will allow us to compare different distraction levels with constant route driving conditions.

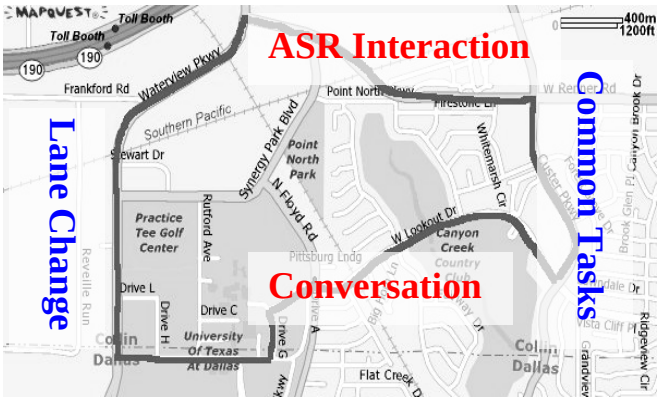


Figure 4: A driving routes and the assigned tasks.

4. DRIVING SIGNALS

A variety of observable driving signals and sensory data have been applied to analyze and characterize driver behavior; for example, brake and gas pedal pressures, steering-wheel degree, velocity of vehicle, velocity of vehicle in front, acceleration, engine speed, lateral position, following distance, yaw angle (the angle between a vehicle's heading and a reference heading) are several under evaluation dimension.

Our preliminary study focuses on four driving signals extracted from the CAN-Bus information: acceleration (RPM), brake position, steering wheel angle, and vehicle speed. Fig. 5 shows the plots, in normalized scales, of these four driving signals for 5 minutes of driving. Positive slope of steering degree plot represents counter-clockwise steering movement, while negative slope represents clockwise steering maneuver.

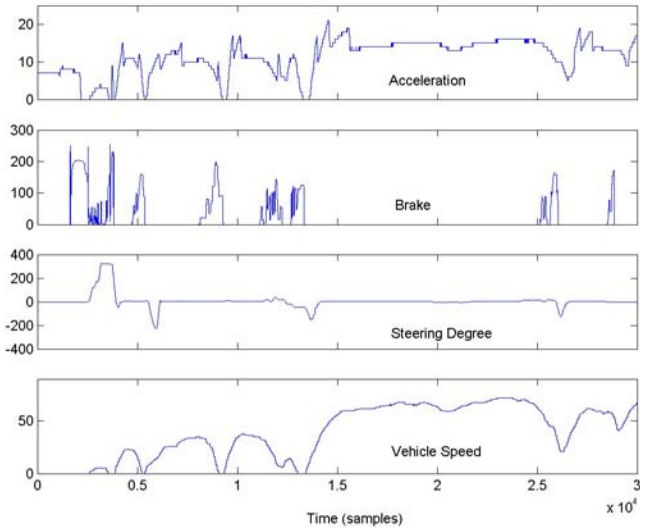
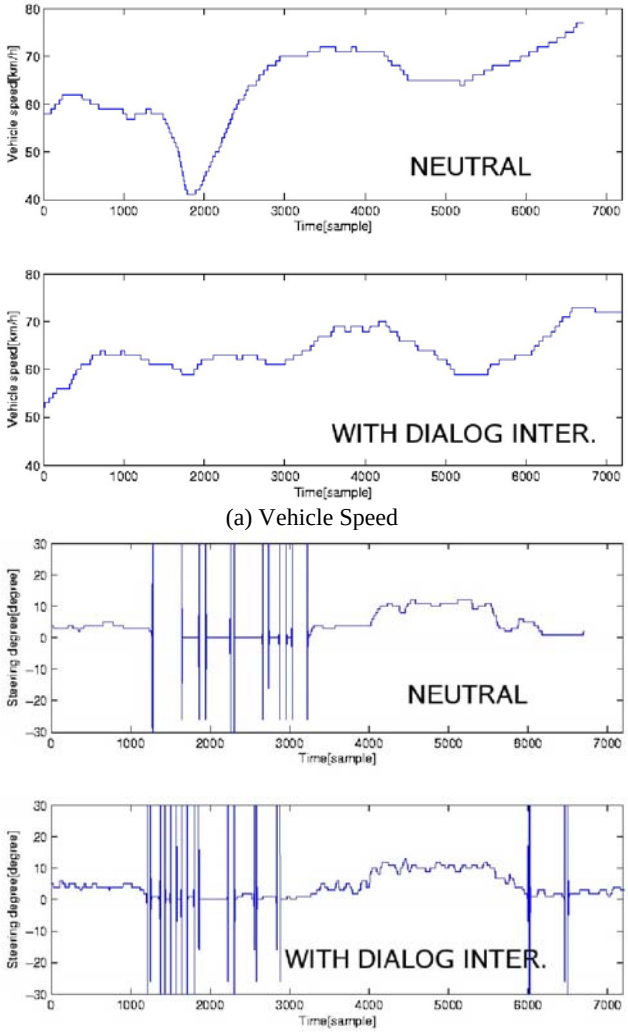


Figure. 5: Vehicle CAN-Bus signals.

5. DRIVER DISTRACTION

Driver awareness has been a major safety concern since the invention of the automobile. According to the National Highway Traffic Safety Administration (NHTSA), there are four distinct types of driver distraction: visual, auditory, bio-mechanical (physical), and cognitive distractions. Although these four modes of distraction are separately classified, they are not mutually exclusive. For example, operating a mobile phone while driving may include all four types of driver distraction: dialing the phone (physical distraction), looking at the phone (visual distraction), holding a conversation (auditory distraction), and focusing on the conversation topic (cognitive distraction) [11]. Common sources of driver distraction are eating or drinking, focusing on the other objects off the road, adjusting radio, talking with passengers, moving the other objects, dialing and talking on a cell-phone, and others.

One approach to distraction detection is based on measurements of driving performance including steering wheel movement, lateral lane position, longitudinal speed, lateral and longitudinal of acceleration and velocity, following distance, vehicle braking, and response time. Under distracted driving, drivers are likely to lose their smooth driving patterns (e.g., slow down or speed up vehicle speed, make excessive steering-wheel maneuver for lane keeping). Figure 6 shows plots of (a) vehicle speed and (b) normalized steering-wheel angle of a driver on the same route twice (under very light traffic). The neutral driving (do nothing) is shown on the top of each plot, and the driving while interacting with a spoken dialog system is shown on the bottom. The vertical line in plot (b) illustrates the sharp maneuver of steering wheel between left and right. As we can see, the driver maintains a smoother driving pattern under the neutral condition with a secondary distraction task..



(a) Vehicle Speed

(b) Steering-wheel Angle

Figure 6: Comparison of neutral and distracted driving of a driver drives on the same road (under very light traffic).

6. DRIVER BEHAVIOR MODELING

Driver behavior consists of lower-level components (e.g., eye movement and steering degree during lane keeping and lane changing) and higher-level cognitive components (e.g., maintaining situation awareness, determining strategies for navigation, managing the other tasks, etc.) [12]. Therefore, effectively modeling driver behavior needs multidisciplinary knowledge of signal processing, control theory, information theory, cognitive psychology, physiology, and machine learning.

A driver behavior model can be developed to characterize different aspects of the driving tasks. For example:

- *Action Classification/Prediction*: Driver behavior model can be used to predict and categorize driver long-term behaviors such as turning, lane changing, stopping, and normal driving.

- *Driver Verification/Identification*: The goal here would be to recognize the driver by their driving behavior characteristics.
- *Distraction Detection*: The objective here is to identify whether the driver is under distraction due to performance of secondary tasks.

7. TRANSCRIPTION CONVENTION

One of the major challenges facing our efforts on utilizing rich multimodal data is a unified transcription protocol. Such protocols do not exist in the community. Multi-layer transcription is necessary for this study. For example:

- *Audio*: different types of background noise inside and outside the vehicle, passengers' speech, radio and music, ring tone, and other audio noise types.
- *Driving Environment*: type of roads (number of lanes, curve or straight, highway or local, speed limit), traffic (traffic situation, traffic light, surrounding vehicles), road condition, etc.
- *Driver Activity*: look away from the road, talk to passengers, dial a phone, talk to a phone, look at rear mirror, look at control panel, sleepy, day-dreaming, etc.
- *Vehicle Mode*: left or right turn, left or right lane change, U-turn, stop and go, stop, etc.

The ability to formulate an effective transcription convention is critical in driving future directions for smart vehicle research. The transcription convention used will lead to better algorithm development which reduces cognitive loads on drivers for smart vehicle systems.

8. CONCLUSIONS AND FUTURE WORK

This paper described research activities of the UTDrive project and vehicle setup for real-time multimodal data acquisition in an actual driving environment. Example profiles using analysis of CAN-Bus information illustrates the range of research possible with the UTDrive corpus. However, robust and reliable driver-behavior modeling systems need to employ the other modalities of data such as video and driver's biometric information to better integrate the driver and system designs of the future.

9. REFERENCES

[1] A. Baron, P. Green, "Safety and Usability of Speech Interfaces for In-Vehicle Tasks while Driving: A Brief Literature Review," *UMTRI-2006-5: The University of Michigan, Transportation Research Institute*, pp.1-8, Ann Arbor, February 2006.

[2] C. Carter and R. Graham, "Experimental Comparison of Manual and Voice Controls for the Operation of In-Vehicle Systems," in *Proceedings of the IEA2000/HFES2000 Congress*, Santa Monica, CA

[3] T. Chen, "Audio-visual speech processing," *IEEE Sig. Proc. Magazine*, vol. 18, no. 1, pp9-21, 2001.

[4] H. Erdogan, A. Ercil, H.K. Ekenel, S.Y. Bilgin, I. Eden, M. Kirisci, H. Abut, "Multimodal person recognition for vehicular applications," N.C. Oza et al. (Edts.): MCS-2005, LNCS-3541, pp. 366-375, Monterey, CA, Jun. 2005.

[5] C. Forlines, B. Schmidt-Nielsen, B. Raj, P. Wittenburg, and P. Wolf, "Comparison between Spoken Queries and Menu-based interfaces for In-Car Digital Music Selection," *TR2005-020*, Cambridge, MA: Mitsubishi Electric Research Laboratories.

[6] J.H.L. Hansen, "Analysis and Compensation of Speech under Stress and Noise for Environmental Robustness in Speech Recognition," *Speech Communications: Special Issue on Speech under Stress*, vol. 20(2), pp. 151-170, Nov. 1996.

[7] T.B. Hughes, H.S. Kim, J.H. DiBiase, and H.F. Silverman, "Performance of an HMM speech recognizer using a real-time tracking microphone array as input," *IEEE Trans Speech and Audio Process*., vol. 7, no. 3, pp. 346-349, 1999.

[8] E. Lombard, "Le signe de l'elevation de la voix," *Ann. Maladies Oreille Larynx*, Nez, Pharynx, vol. 37, pp. 101-119, 1911.

[9] C. Miyajima, Y. Nishiwaki, K. Ozawa, T. Wakita, K. Itou, K. Takeda, and F. Itakura, "Analysis and Modeling of Personality in Driving Behavior and Its Application to Driver Identification," *Proc. of the IEEE*, vol. 95, No. 2, pp. 427-437, Feb 2007.

[10] A. Pentland and A Liu, "Modeling and Prediction of Human Behavior," *Neural Computation*, vol. 11, pp. 229-242, 1999.

[11] M. Pettitt, G. Burnett, A. Stevens, "Defining Driver Distraction," World Congress on Intelligent Transport Systems, San Francisco, Nov. 2005.

[12] D. Salvucci, E.R. Boer, and A. Liu, "Toward an Integrated Model of Driver Behavior in a Cognitive Architecture," *Transportation Research Record*, No. 1779, pp. 9-16, 2001.

[13] A. Wahab, T.C. Keong, H. Abut, and K. Takeda, "Driver Recognition system using FNN and statistical methods," chapter 3 in Advances for in-vehicle and mobile systems, Abut, Hansen, Takeda (Edts.), Springer, New York, 2007.

[14] X.X. Zhang and J.H.L. Hansen, "CSA-BF: A Constrained Switched Adaptive Beamformer for Speech Enhancement and Recognition in Real Car Environment," *IEEE Trans. Speech & Audio Proc*., vol. 11, no. 6, pp. 733-745, Nov. 2004

[15] X.X. Zhang, K. Takeda, J.H.L. Hansen, and T. Maeno, "Audio-Visual Speaker Localization for Car Navigation Systems," in *INTERSPEECH-2004*, Jeju Island, Korea, 2004.