International Large-Scale Vehicle Corpora for Research on Driver Behavior on the Road

Kazuya Takeda, *Senior Member, IEEE*, John H. L. Hansen, *Fellow, IEEE*, Pınar Boyraz, *Member, IEEE*, Lucas Malta, Chiyomi Miyajima, *Member, IEEE*, and Hüseyin Abut, *Life Senior Member, IEEE*

***Abstract*—This paper considers a comprehensive and collaborative project to collect large amounts of driving data on the road for use in a wide range of areas of vehicle-related research centered on driving behavior. Unlike previous data collection efforts, the corpora collected here contain both human and vehicle sensor data, together with rich and continuous transcriptions. While most efforts on in-vehicle research are generally focused within individual countries, this effort links a collaborative team from three diverse regions (i.e., Asia, American, and Europe). Details relating to the data collection paradigm, such as sensors, driver information, routes, and transcription protocols, are discussed, and a preliminary analysis of the data across the three data collection sites from the U.S. (Dallas), Japan (Nagoya), and Turkey (Istanbul) is provided. The usability of the corpora has been experimentally verified with a Cohen's kappa coefficient of 0.74 for transcription reliability, as well as being successfully exploited for several in-vehicle applications. Most importantly, the corpora are publicly available for research use and represent one of the first multination efforts to share resources and understand driver characteristics. Future work on distributing the corpora to the wider research community is also discussed.**

***Index Terms*—Driving behavior signal, on-road data collection, road vehicles, tagging.**

I. Introduction and Background

SIGNIFICANT progress has been made over the past two decades and continues at an even faster pace today in all aspects of transportation systems. In particular, vehicles equipped with improved safety capabilities and infotainments have been deployed as a result of extensive scientific investigation and subsequent technological development in Asia, America, and Europe. However, there is a considerable disconnect between countries, automobile manufacturers, third-party after-market technology add-ons, and transportation infrastructure engineers. Because of the competitive nature of the automotive industry, research advances made within automobile companies may not be made available to the general research community. In addition, research advancements in one country may not benefit researchers in other countries as much if data and results are not shared.

In particular, the information processing aspects of in-vehicle technologies are becoming increasingly important as the number of sensors and Engine Control Unit (ECUs) within vehicles increases. How to integrate the data streams associated with driving is one of the central issues in advanced vehicle technologies. Therefore, the lack of common research resources including in-vehicle data is a crucial issue in the research community. This paper presents a comprehensive and collaborative study on large-scale on-road multimedia data collection within vehicles in an attempt to begin addressing this problem.

As will be discussed in Section II, prior research efforts to collect in-vehicle data can be roughly divided into three groups in terms of their aims: 1) speech corpora (which can include speech, video, etc.); 2) accident analysis; and 3) field operational tests (FOTs).

In the first domain, speech corpora are collected for developing automatic speech recognition (ASR) systems. Since the stochastic pattern recognition approach has proven its potential in speech recognition research, various speech corpora have been built for developing and testing ASR systems. Most of them consist of high-quality recordings of controlled utterances and their phonetic transcriptions. Speech data in cars are specifically needed to capture the various noise conditions, driver task load/stress, and vocabulary appropriate for in-vehicle applications. In general, *human sensor* data and its *transcription* are the main components of such corpora.

The second type of data collected is comprised of accident analysis corpora, which aims primarily at recording traffic accidents and studying accident causation mechanisms. Due to their nature, data collection in such studies must be performed using naturalistic and low intervention methods involving large numbers of drivers. In general, these corpora contain *video* and *vehicle sensor* data. Transcriptions are carried out solely on discrete events (e.g., near crashes), rather than being continuously provided throughout the data. Hence, most of the collected data has no tags.

The third group of corpora is FOT data. FOT is an experimental validation of advanced driving assistance systems under real traffic conditions. Predetermined hypotheses, such as "Using Forward Collision Warning systems will decrease the number of front-end crashes, near crashes, and incidents,"

Manuscript received February 25, 2011; revised July 18, 2011; accepted August 15, 2011. Date of publication October 14, 2011; date of current version December 5, 2011. The Associate Editor for this paper was Y. Liu.

K. Takeda is with the Department of Media Science, Graduate School of Information Science, Nagoya University, Nagoya 464-8603, Japan.

J. H. L. Hansen is with the Center for Robust Speech Systems (CRSS), Department of Electrical Engineering, The University of Texas at Dallas, Richardson, TX 75080-3021 USA.

P. Boyraz is with the Mechanical Engineering Department, Istanbul Technical University, 34437 Istanbul, Turkey.

L. Malta is with the Volvo Technology Corporation, 40508 Göteborg, Sweden.

C. Miyajima is with the Graduate School of Information Science, Nagoya University, Nagoya 464-8603, Japan.

H. Abut is with Bogaziçi University, 34342 Istanbul, Turkey.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2011.2167680

1524-9050/$26.00 © 2011 IEEE

TABLE I
COMPREHENSIVE COMPARISON TABLE FOR MULTISENSOR IN-VEHICLE CORPORA

Corpus	Year	Location	Signal Content	Accessibility
SPEECHDAT-CAR	1999	Europe	Speech	Open
CU-Move	1999	USA	Speech	Partial
CIAIR	1999	Japan	Speech, Video, Brake/Gas Pedal, SWA, GPS	Partial
Stern et al.	1999	USA	EOG, SWA, Gas/Clutch	Open
ACAS	1999	USA	Multi-sensor	Open
DARPA Communicator	2000	USA	Speech	Restricted
RWCW	2001	USA	Multi-sensor	Open
AVICAR	2004	USA	Speech, Video	Open
IVBSS	2005	USA	Multi-sensor	Open
100-Car Study	2006	USA	Video, CAN-Bus, Head distance, GPS, Partial Speech	Partial
euroFOT	2008	Europe	Multi-sensor	Partial
SHRP-2	2010	USA	Multi-sensor, enhanced	Partial

are tested based on the measured variables to evaluate the effectiveness of the deployed technologies (FCW in the case above) in a comparative manner. In general, FOT corpora not only consist of *video* and *vehicle sensor* data but include *human sensor* data such as gaze directions as well. However, data collected in FOTs are often limited because it is collected to achieve specifically predefined goals. Furthermore, the data are not fully open to the public. Therefore, FOT corpora are not always suitable for general research by the scientific community.

In summary, a relatively small part of the existing in-vehicle data corpora can be used in studies focusing on interactions between humans, vehicles, and the environment. The goal of this research is to build such real-world corpora. The corpora should contain video and driver/vehicle sensor signals as well as information on live road traffic. The corpora should also be large and consist of multicountry data. More importantly, they should be accessible throughout the research community.

The aim of this paper is to report on the development of such corpora as follows. In Section II, a detailed survey of the existing in-vehicle signal corpora is given. In Section III, data collection efforts toward corpus development are discussed. Next, in Section IV, a unique transcription protocol for data mining is proposed. The usefulness of the corpora will be evaluated by showing the findings from a few applications in Section V. Unresolved issues for distributing the corpus to a wider population are discussed in Section VI. Finally, a brief summary and conclusions are presented in Section VII.

II. RELATED WORKS

There have been a number of in-vehicle data collection and analysis efforts prior to this research. Those efforts can be roughly divided into three groups as mentioned earlier: 1) in-vehicle speech; 2) accident analysis; and 3) FOT corpora. Examples of these efforts are summarized in a chronological order in Table I.

A. In-Vehicle Speech Corpora

Most in-vehicle speech corpora consist of high-quality recordings of driver's utterances and their phonetic transcriptions for developing in-vehicle ASR systems. In the late 1990s, three major activities have been embarked upon in three different continents, namely 1) SPEECHDAT-CAR [1] in Europe, 2) CU-Move [2]–[4] in the U.S., and 3) Center for Integrated Acoustic Information Research (CIAIR) [5] in Nagoya, Japan. These investigations focused mostly on studying naturalistic conversational speech during driving, except for CIAIR, which included video and other signals in addition to the multichannel speech data.

The SPEECHDAT-CAR effort is the first international program to form a multilanguage speech corpus for car applications. It began in 1998 with the participation of 23 research groups from nine European countries, which later also included a single U.S. site. The purpose was to develop a set of speech databases to support training and testing of multilingual speech recognition for hands-free applications in automobiles.

The CU-Move research team at the University of Colorado[1] [2]–[4] has collected natural conversational speech from 500 speakers within six cities across the U.S. using an instrumented vehicle. The CU-Move corpus consists of digit strings, route navigation expressions, street and location information, phonetically balanced sentences, and route navigation in a human-to-Wizard-of-Oz-like scenario.

The CIAIR [5] corpus is also designed primarily for speech research, but it was one of the first corpora that also included synchronously recorded multichannel audio/video, brake and gas pedal pressure readings, steering wheel angle (SWA), and location signals from over 800 Japanese drivers.

Subsequently, AT&T under the Defense Advanced Research Projects Agency Communicator program collected acoustic data from 27 microphones distributed throughout a car starting

[1]The research team overseeing CU-Move Corpus (2002–2005) development moved to the University of Texas at Dallas in 2005 and later continues with the UTDrive Corpus.

in the year 2000 [6]. Later, in 2004, AVICAR was built for audio-visual speech recognition research. Audio and video signals from 100 drivers were collected and used for developing bimodal speech recognition technologies [7], [8].

These in-vehicle speech corpora are primarily intended to be used for developing and testing basic speech algorithms. Therefore, they are recorded under controlled conditions (e.g., the route and secondary tasks are fixed). The recorded signals are carefully transcribed so that the corpora can be shared for various research purposes. Other types of human sensor data, such as eye-blinking frequency and eye-related measurements performed by Duke *et al.* [9], have also been collected. However, the size and type of signal modalities of corpora such as that of Duke *et al.* are quite limited.

B. Accident Analysis Corpora

Different types of vehicle signal corpora have been collected in connection with traffic incidents and accidents so that researchers can analyze causation mechanisms, such as distraction. This type of data collection must be conducted in a naturalistic and nonintrusive manner. In general, these corpora consist of *video* and *vehicle sensor* data.

The "100-Car Study" [10], [11], undertaken by the Virginia Tech Transportation Institute and funded by the AAA Foundation for Traffic Safety, was one of the most important of these studies. It includes naturalistic, continuous multisensor data [five channels of video, front and rear radar sensors, accelerometers, lane tracker, and Global Positioning System (GPS)] collected from 350 (109 primary and 241 secondary) drivers. The study aimed at 1) assessing the frequency of crashes, near crashes, and incidents where drivers were engaging in potentially risky driving behaviors and 2) establishing models for risky driving behaviors by computing odds ratios to estimate the relative risk of involvement in a crash or near crash. The experiments lasted over a period of 12–13 months and were conducted in the Washington, D.C., area [12].

In the U.S., as part of the SHRP2 project of the Transportation Research Board, Project S07 has begun collecting driving data at multiple sites and the data will be made available to multiple researchers [13].

C. FOT Corpora

Since the mid 2000s, FOT studies have been one of the primary subjects of European Union and U.S. traffic safety research. The automotive collision avoidance system FOT (ACAS/FOT) of General Motors and the U.S. Department of Transportation National Highway Transportation Administration is known as one of the earliest FOT studies. This FOT involved exposing a fleet of 11 ACAS-equipped passenger cars to 12 mo of naturalistic driving by ordinary drivers on public roads in the U.S. The goal here was to examine the functionality of safety technologies under development and their suitability for widespread deployment from the perspectives of both driving safety and driver acceptance [14].

Starting in 2000, the number of studies and their extent has increased. For example, the road departure crash warning system FOT (RDCW FOT) gained insight into the suitability of RDCW systems for widespread deployment within the U.S. passenger vehicle fleet. Eleven passenger cars equipped with crash warning functionalities—the RDCW system—were involved in this FOT [15].

More recent studies include the FOT of the integrated vehicle-based safety systems (IVBSSs), which was launched in the U.S. in 2005. IVBSS was a two-phase multiyear cooperative research effort involving a fleet of 16 passenger cars and 10 heavy trucks with the goal of evaluating the safety benefits and driver acceptance of prototype integrated crash warning systems [16].

In addition to the FOTs in the U.S., Europe launched its first FOT on active safety systems (euroFOT) in 2005, involving 28 organizations. The field tests have focused in particular on eight distinct functions that assist the driver in detecting hazards, preventing accidents, and make driving more efficient. Over the course of one year, hundreds of cars and trucks equipped with a range of different intelligent technologies are being tested on European roads across France, Germany, Italy, and Sweden [17].

III. Corpus Development: An International Framework

To build a firm research foundation on which technologies for enhanced vehicle and driver safety can be built, there are three crucial components: 1) on-road data; 2) international perspective; and 3) careful experimental design. Many challenges are inevitable in both implementing the measurement systems operating in the real world on human subjects and carrying out internationally collaborative data collection experiments with multiple similarly instrumented vehicles driven on city streets as well as highways. Faced with these challenges, data collection has been carried out in three large metropolitan areas, involving different road and traffic conditions, climates, traffic rules, law enforcement, and driving habits. This section is divided into three subsections: 1) vehicles and sensors; 2) routes; and 3) secondary tasks. In addition to these three main parts, statistics on the driver population are also presented.

A. Data Collection Setup: Vehicles and Sensors

Three vehicles, one at each site, have been instrumented to collect driver, driving, vehicular, road, and traffic data under real-world conditions. The three sites have agreed upon a framework for a desirable data set, which can be broken down into common/core signal sets, and secondary optional data to be collected at each site. All collected signals are listed in Table II together with their corresponding A/D sampling rates. Multimedia data has been acquired and logged either with a commercially available data logger, enabling the system to record the data synchronously (Japan and the U.S.), or semi-custom-designed data acquisition systems (Japan and Turkey).

The "NUDrive Vehicle" is used for data collection in Nagoya, Japan. It is an automatic van with a 2.36-cc hybrid engine built for countries with the driver's seat on the right (where driving is on the left).

TABLE II
LIST OF SENSORS AND DATA SPECIFICATIONS (J: NAGOYA; US: DALLAS; TR: ISTANBUL)

Sensor	Common	Optional	Sampling Rate
Microphones	close talking, distant	array (J/US)	16kHz / 16bit
Video	frontal view, driver face	omni-view (J);stereo driver face (TR)	30Hz
Brake Pedal	analog (0-500N)	CAN-Bus (US/TR)	16kHz / 16bit; CAN-Bus: 10 or 32 Hz.
Gas Pedal	analog (0-500N)	CAN-Bus (US/TR)	16kHz / 16bit; CAN-Bus: 10 or 32 Hz.
Steering Wheel Angle	analog (±720deg)	CAN-Bus (US/TR)	16kHz / 16bit; CAN-Bus: 10 or 32 Hz.
Vehicle Speed	analog (J)	CAN-Bus (US/TR)	CAN-Bus: 10 or 32 Hz.
GPS	GPS (J/US)	DGPS (TR)	Device-dependent
Range Distance	head distance	Ultrasonic near range; Laser scanner (TR)	CAN-Bus: 10 or 32 Hz; TR; 181 X-Y readings/sec.
Perspiration	-	Palm, Feet (J)	16kHz / 16bit
Heart Rate	-	Chest (J)	average interval in ms

A multichannel synchronous recording system configured with CORINS, MVR-303, and System design (DASBOX) have been used for synchronous recording of the signals listed in Table II. MVR-303 has a synchronous control unit and a system control PC, which can record multichannel synchronous videos and analog signals. Each PC node can store 240 GB of video data (1.4 million pixels at 29.4118 fps), corresponding to 90 min of video. Operations of the gas and brake pedals are captured using custom-designed pressure sensors. The amount of perspiration and skin potential are measured with a system from SKINOS. The perspiration sensors are attached on the left palm and left foot of the driver. GPS and 3-D accelerometer signals are also recorded (see [18] and [19] for more details). The vehicle and a subset of the sensors are shown in Fig. 1 (top).

The "UTDrive Vehicle" is used for data collection in Dallas, TX. The vehicle is an automatic sports utility vehicle (SUV) with a 2.4-cc engine, with the driver's seat on the left (driving is on the right).

A fully integrated data acquisition system, namely, Dewetron DA-121, is used in the UTDrive SUV. This is a portable system, which has eliminated the need for additional car batteries to meet the increased power demand. With a very high sampling rate of 100 MHz, DA-121 is capable of synchronously recording multirange input data (i.e., 16 analog inputs, 2 CAN-Bus interfaces, eight digital inputs, two encoders, and two video cameras) and yet allows the sampling rate for each data channel to be set individually. DA-121 can export all recorded data as a video clip in one output screen, or individual data channels can be output in their respective proper data formats (e.g., .wav, .avi, .txt, .mat, etc.) with synchronous time stamps (see [20] for more details). The vehicle and some of the sensors are shown in Fig. 1 (center).

"UYANIK (AWAKE)" is used for data collection in Istanbul, Turkey. It is a stick-shift sedan passenger car with a 2.5-cc diesel engine built for European roads with a driver's seat on the left (again, driving is on the right).

Data are collected by an ensemble of three acquisition systems synchronized using the clock of one of the networked computers. Video from three cameras is captured with a semi-custom StreamPix digital video recorder from NORPIX. For audio recordings, an ADAT HD24 data acquisition system from Alesis is used. Four microphone channels and a sync signal between the two acquisition systems were sampled at 48 kHz and 24 bits per sample. Later, these are converted to 16 kHz and 16 bits off-line in ".wav" format. Finally, the acquisition of CAN-Bus signals, laser scanner, GPS receiver, gas and brake pedal sensors, and IMU 3-D accelerometer are realized over the universal serial bus and RS232 PCMCIA ports of a notebook computer using custom software developed at the Mekar Laboratories of the Istanbul Technical University (ITU) (see [21] for more details). The vehicle and some of the sensors are shown in Fig. 1 (bottom).

B. Routes

The Nagoya route is 27.7 km long and starts and ends at Nagoya University, Nagoya, Japan. It includes both city roads and an expressway segment [18]. The Dallas route consists of two different shorter routes ($\sim$15 km each, closed loops) in the neighborhood of the University of Texas at Dallas. The first route is in a residential area, and the second is in a business district. Drivers performed multiple rounds of driving and repeated the process on up to three separate days. The route begins and ends at the same location [20]. The Istanbul route is 25.6 km long and starts and ends in front of the Automotive Research Center (OTAM) on the campus of ITU. It includes both city roads and expressway segments [21]. Route and data collection equipment were the same for all drivers at each site.

C. Secondary Tasks

Drivers along these three routes perform up to four common (similar) tasks, and there are also segments where no tasks are performed (reference driving). These driving conditions can be generally grouped as follows:

1) no task (reference driving): no secondary task performed;
2) signboard reading: reading signboards seen by the driver;

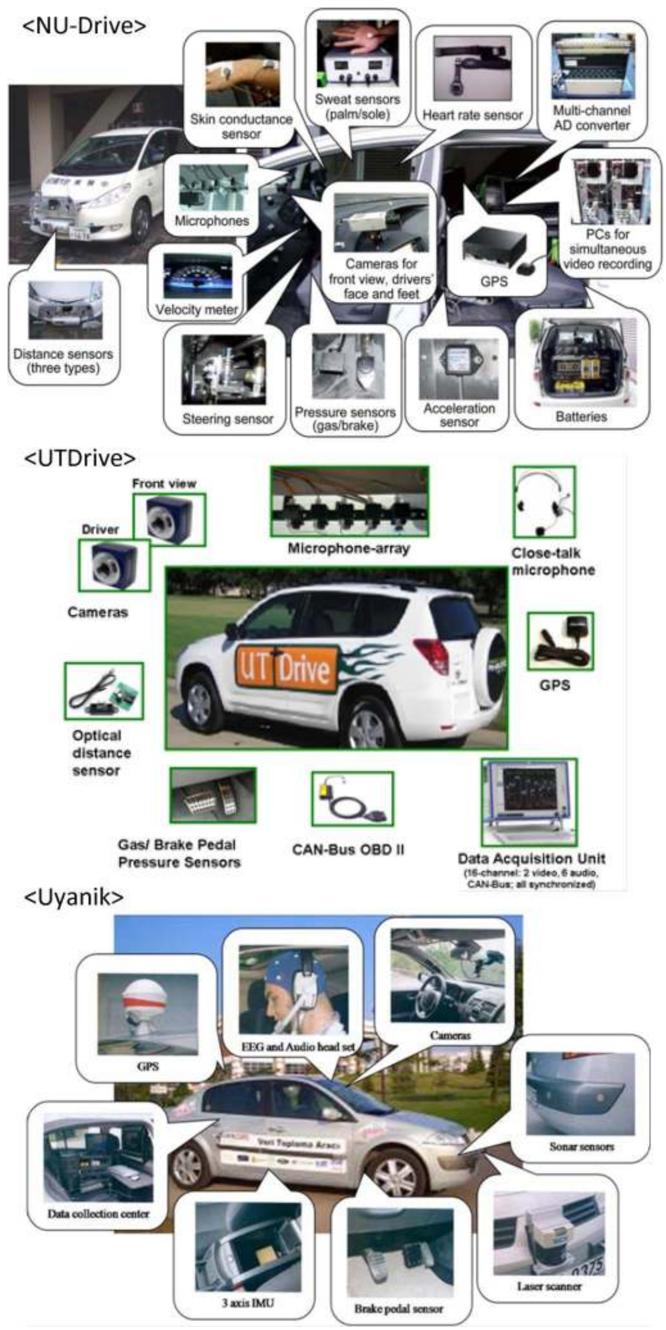

Fig. 1. Data collection vehicles, sensors, and instruments. (Top) NUDrive. (middle) UTDrive. (Bottom) Uyanik.

3) navigation dialog: exchanging navigation information over a mobile phone.

In addition, there are site-specific secondary tasks such as music retrieval in Nagoya, airline arrival/departure information and city weather information retrieval in Dallas, and on-line banking in Istanbul. Several other cognitive load tasks are also performed while driving. As an example, we present the detailed secondary task scheme for Dallas in Table III.

D. Driver Profiles

The content of the corpora in terms of driver profiles is presented here to demonstrate that the corpora cover a very wide range of demographic categories with statistically significant sample sizes in each category. The categories are gender, age, driving experience, and driving frequency.

TABLE III
SESSIONS AND DRIVERS TASKS USED IN UTDRIVE DATA COLLECTION

	Part	Secondary Tasks		
		A	B	C
Route 1	1	Lane Changing	Common tasks (radio, AC etc.)	Sign Reading
	2	Cell phone dialogue	Cell phone dialogue	Conversation
	3	Common Tasks	Sign Reading	Spontaneous
	4	Conversation	Spontaneous	Cell phone dialogue
Route 2	1	Sign Reading	Lane Changing	Common tasks (radio, AC etc.)
	2	Cell phone dialogue	Cell phone dialogue	Conversation
	3	Common tasks (radio, AC etc.)	Sign Reading	Lane changing
	4	Spontaneous	Conversation	Sign Reading

Nagoya statistics: Over a period of one year, more than 500 drivers have participated in data collection at the three sites. In Nagoya, 312 drivers have participated in the driving experiments, with each driver required to drive about 1 h to complete the tasks, which resulted in a total of approximately 300 h of recorded data. The size of the Nagoya database is 4.2 TB. Distributions of age, gender, and years of driving experience are shown in Fig. 2.

Dallas statistics: A total of 77 drivers (40 male and 37 female) participated in the Dallas experiments, which included multiple recording sessions across several days (i.e., over 150 data collection runs). To give the reader an idea of the richness of the metadata (apart from the collected multisensor continuous synchronous data, which occupy 2 TB in compressed form) the mobile phone usage frequency is presented in Fig. 2. The UTDrive site also included an incentive for the drivers to repeat the data collection procedure up to three times, with typically 1 week between each session. This allows the Dallas data to be used for intersession variability assessment for these drivers. Each session lasted approximately 45–60 min.

Istanbul statistics: In this data collection effort, 108 drivers, mostly from the academic community, drove the vehicle along a 25.6-km route. There were 19 female and 89 male participants. The time required to complete the route varied from 40 min to 1 h, depending on traffic conditions. The age range for female drivers was 21–48, and the corresponding male range was 22–61. Since uncompressed stereo video was needed in the vision-based studies undertaken by some of the participants, the overall database was more than 24 TB. However, after compressing and downsampling, the data storage volume decreased to approximately 4.1 TB.

Gender

	Male	Female
US	40	37
JPN	184	128
Turkey	89	19

Age

	-20	-25	-30	-35	-40	-45	-50	50+	n/a
US	5	16	8	9	6	4	2	10	17
JPN	4	73	44	57	42	45	28	19	-

Driving Experience

	-3	-5	-10	-15	-20	20+	n/a
US	6	9	12	8	6	18	18
JPN	22	46	53	57	53	81	-

Cell phone usage

	<	Seldom		Very often	>	n/a
US	18	12	6	6	16	19

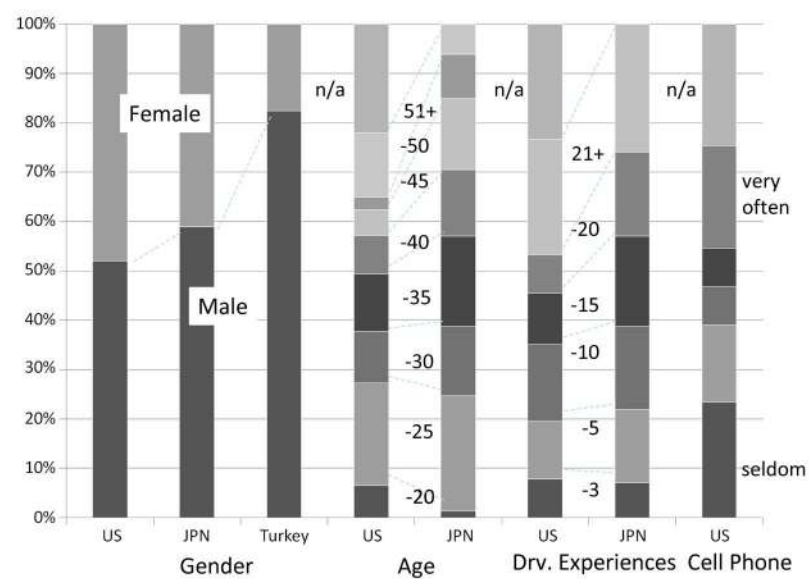

Fig. 2. Driver population at three sites (Note: age and driving experience in years; n/a implies data was not available).

E. Preliminary Data Comparison Among Three Sites

Comparative analysis on the driving behavior signals collected from the three sites is one of the most important aspects of our data collection. However, in this paper, we do not focus on the different driving behaviors observed at the three sites in detail since the main objective here is to confirm the usability of the corpora. In this section, we compare the statistical distributions of the brake/gas pedal signals as an example of site-dependent characteristics.

In Fig. 3, it can be observed that both histograms of gas and brake pedal signals show a bimodal distribution for all three countries, despite the significant variety of vehicles, sensors, drivers, road, and environmental conditions. (It is important to note again that the brake/gas pedal signals were measured using pressure sensors in Nagoya and Istanbul while they were captured from CAN-Bus in Dallas.) In contrast, the data distributions of vehicle velocity differed between the Dallas site and the other two sites possibly because velocity is highly dependent on the driving route and traffic conditions, and the driving environment in Dallas may be unique in some respects.

Our observation that the data collected from each site showed a similar distribution leads us to believe that the amount of data collected is large enough to conduct reliable statistical analyses of driving behavior in different countries.

IV. Novel Transcription Protocol: Opportunities for Data Mining

A. Transcription Protocol

To properly associate driving situations with recorded data, contextual labeling of multimodal information is critical. We have proposed a novel data transcription protocol that takes into account a comprehensive cross section of factors that could affect drivers and their responses. It is intended that this unique protocol would be an irreplaceable tool for both signal processing scholars and the vehicular technology community, and it could easily benefit other fields beyond the field of driver

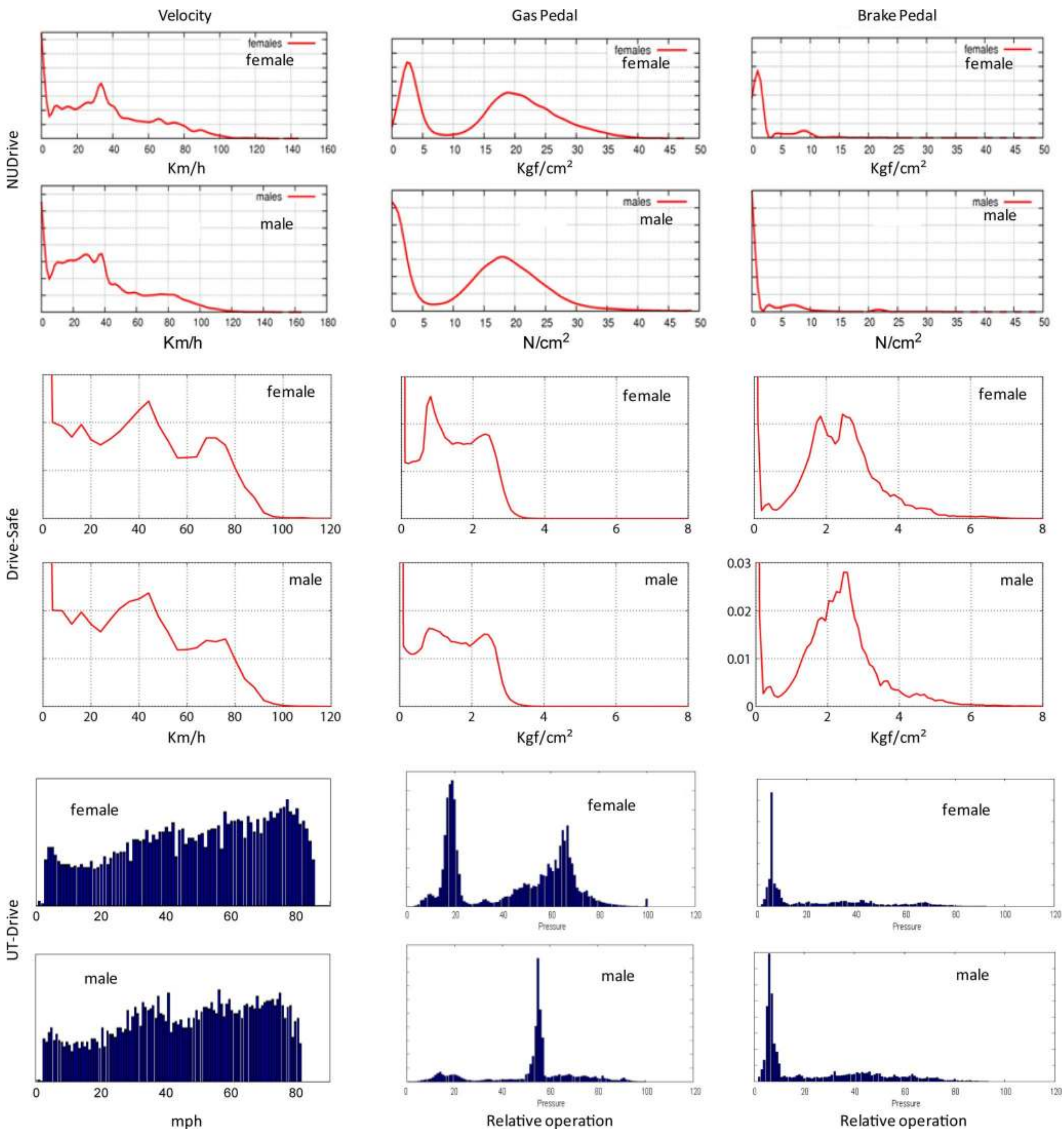

Fig. 3. Comparative driving analysis for male and female drivers. Velocity distributions are shown in Km/hour for NUDrive (Japan) and Drive-Safe (Turkey), and miles/hour for UT-Drive (USA). Gas Pedal and Brake Pedal distribution information is shown in terms of kilograms force/cm2 for NUDrive (Japan) and Drive-Safe (Turkey), and relative pressure for UT-Drive (USA).

behavior recognition. The proposed protocol is comprised of the following six major categories:

1) driver affective/mental state;
2) driver actions;
3) driver's secondary and tertiary tasks;
4) driving environment;
5) vehicle status;
6) speech/background noise.

The designed transcription protocol is comprehensive, providing researchers with segmented/labeled data available for analysis on many research topics. For example, a researcher can see the effect of an in-vehicle navigation system on the driver, as well as examining the change in driving performance with regard to environmental changes/road conditions.

Compared with the annotations made in the 100-Car Study, which are widely used as a basis in many other projects, the present protocol does not focus on explaining the causes of driver distractions or characterizing driver behavior during events such as crashes and near crashes. Rather, we adopted a more objective and systematic approach that provides a rich

Fig. 4. Transcription markings as multiple streams from the Istanbul transcription tool.

TABLE IV
SAMPLE OUTPUT FROM TRANSCRIPTION PROCEDURE

Time Line (seconds)		Variable			Description
from	to	id	name	val.	
2.040	3.037	8	GESTURE HEAD	3	TURN RIGHT
3.038	3.803	8	GESTURE HEAD	1	STRAIGHT
7.200	20.200	6	EYES	1	ON THE ROAD
15.709	26.175	9	GESTURE HANDS	2	OFF THE WHEEL
17.156	27.622	6	EYES	4	OFF ROAD (OTHER)
20.200	31.200	6	EYES	4	OFF ROAD (OTHER)
29.956	38.722	8	GESTURE HEAD	1	STRAIGHT
31.200	38.447	6	EYES	1	ON THE ROAD
34.505	44.259	34	SUDD. & STRG ACC.		
35.060	38.680	3	PRECAUTION		
38.033	42.021	8	GESTURE HEAD	2	TURN LEFT
38.962	47.700	27	TURN	1	LEFT
39.205	44.372	6	EYES	3	SIDE VIEW MIRROR

and continuous description of the current driving situation. This description enriches the raw sensor data and allows new avenues of research, including analysis and model development possibilities.

Transcription markings can be seen as multiple streams of information (the transcription process requires repeatedly viewing signal information). To effectively observe and perform multilayered tagging of audio, video, traffic, and environment transcription tasks, data browsers were developed at each site. We did not use a common browser because each site has to deal with the local language, differences in the work force employed in tagging, and variations in data formats. As an example, we include a screen shot from the Istanbul browser in Fig. 4, and the transcription results are shown in Table IV. We can see three video windows (two focused on the driver and one on the road ahead) and listen to three audio microphones signals (close talking, rearview mirror mounted, and mobile phone). The microphones and video cameras are synchronized, and by using the slide bar in the bottom window, the viewer can go back and forth in time. The time stamp also appears on the top bar of the window. A dynamic bar-type graph of several sensor readings coming from both the CAN-Bus of the vehicle, as well as gas and brake pedal pressures, appears on the middle right of the screen. (It is noted that since the CAN-Bus information is proprietary, the Istanbul and Dallas sites use customized signal decoders.)

Examples of the Nagoya and Dallas browsers can be found in [18] and [20], respectively. The transcription process used in Nagoya, and an evaluation of its consistency and reliability, will be discussed in Section IV-B since it has been studied longer and the analysis of the data for most of the drivers has been completed.

B. Evaluation of Transcription Reliability

Because transcription was performed by several different transcribers, the consistency and reliability of the Nagoya transcriptions was initially tested manually for a subset corpus of 30 drivers using only the front-view video. No audio was provided to avoid bias when labeling the facial data from the video. The annotators were graduate students who had volunteered for the transcription task. For this task, not all of the labels described in the Appendix were needed, and therefore, only a subset was used. This subset is composed of the following: 1) overall facial appearance: positive and negative facial expressions were grouped together, and therefore, this binary feature indicates deviations from a neutral face; 2) turns: left and right turns and turning with interruptions; 3) curves: left and right curves; 4) the presence of obstacles such as pedestrians, bicycles, parked vehicles, and other objects; 5) traffic density (volume); and 6) stops at traffic lights.

The reliability of the transcription information was validated in the following manner. In the first step, each annotator transcribed the same driving data segment by marking start and stop times for all observed events. Next, a ground truth for the transcription was prepared to benchmark the resultant markings of different annotators. Next, the tagging results of each annotator were compared with the benchmark transcription by using a statistical measure of interrater reliability called Cohen's kappa coefficient. When the kappa coefficient is from 0.61 to 0.80, there is substantial agreement, and from 0.81 to 1.0, there is an almost perfect agreement. Mean and standard deviation (STD) of the results from all seven annotators are given in Table V under the "First comparison" column.

Next, necessary adjustments and further transcriber training were performed after transcription of the complete corpus was finished. Data different from those utilized in the first step were transcribed by the annotators, and the new transcriptions were then compared. Values of the mean and STD of the kappa coefficients are shown in Table V under the "Second

TABLE V
TRANSCRIPTION RELIABILITY RESULTS

Variable	First comparison		Second comparison	
	mean	STD	mean	STD
Turn	0.84	0.03	0.94	0.03
Curve	0.72	0.03	0.89	0.07
Traffic density Light	0.35	0.32	0.57	0.24
Traffic density Medium	0.31	0.34	0.52	0.22
Obstructions	0.64	0.23	0.69	0.24
Overall face	0.54	0.23	0.60	0.19
Stop at red-light signal	NA	NA	0.98	0.01

comparison" column. This column is used to determine if the reliability of the transcribers improved.

In the last step, all of the transcription results were checked individually by the authors to ensure a high level of reliability. The results presented in Table V indicate that we can increase transcription reliability by properly training the annotators. Under the "Second comparison" column, all variables had at least moderate agreement with the actual transcription. As a quality control method, the final check may not be practical for a large amount of data, but it could be used as an effective method when training annotators.

It was noted that the transcription process is tedious and time consuming. On the average, 10 min of data takes over 50 min to transcribe. To speed up the process, efforts have been made to improve our corresponding transcription interfaces (i.e. UTDAT, the Nagoya browser, and IVMCTool). Improved interface design is expected to decrease the number of errors. For example, inappropriate overlapping of labels will be prohibited.

Future work would include an automated data transcription method that can facilitate the labeling process. In particular, automated transcription can help filter data to focus on a specific region of interest (i.e. potentially hazardous situations, accidents, or certain driving events could be detected) [13].

V. Sample Applications

Global analysis using this collaborative large-scale on-road driving data indicates trends into driver behavior characteristics associated with several parameters. There are a vast number of potential applications, including driver identification [23], driver distraction detection [24]–[26], emotion and stress detection/estimation [27], [28], mining potentially hazardous situations [29], and route recognition [27]. In this section, we present three sample applications in the areas of driver distraction/cognitive load, frustration estimation, and lane tracking.

The objective of this section is to show the scientific value of our corpora, which results from their unique features. These features include the following: 1) data collected in multiple countries; 2) synchronous data from both human and vehicle sensors; and 3) detailed continuous transcription of the data. As a result, these corpora can be exploited for development of algorithms with a variety of vehicle applications. Note that the focus here is not to develop new algorithms, but to illustrate the unique application opportunities this corpus offers for researchers in the field.

TABLE VI
PROPORTION OF DRIVERS WHOSE MEASUREMENT WHEN PERFORMING THE SECONDARY TASK EXCEEDS BASELINE

Task	z for *MR*		z for *ALS*	
Signal	Force	ΔForce	Force	ΔForce
MAX	0.80	0.82	0.46	0.56
STD	0.72	0.54	0.56	0.56

A. Driver Distraction and Behavior

Dallas experiments on driver distraction: Steering entropy is widely used in driving behavior research [24]. In addition to steering wheel entropy and other SWA-based metrics, it is suggested that additional signal data, such as that contained in our corpora, will be useful for further investigation of this problem. In our previous work, we have proposed assessing distraction and audiocognitive load during lane keeping and curve negotiation maneuvers using several types of data from vehicle and human sensors [25]. In this section, we extend the previous work using hidden Markov models (HMMs) and Gaussian mixture models to perform distraction detection as a binary decision problem [26].

In this paper, the variance in the SWA obtained from CAN-bus signal analysis showed an increase of 40% when the driver operated the radio controls versus nondistracted driving. The SWA variance showed increases of as much as 203% when the driver engaged in a conversation (while it is much higher for mobile phone, the value was also large when conversation was with a passenger), indicating the significant change in vehicle control when the driver attempts to perform multiple tasks. It was observed that when using generic approaches and HMM, the accuracy of distraction detection would range between 80% and 90%; however, using specific driving performance metrics and baseline driving for comparison, it is possible to reach to a much higher accuracy rate.

Nagoya experiments on cognitive load: In this set of experiments, the Nagoya team explored the effect of various speech-related secondary tasks on driver workload. Driving stability, inferred from gas pedal operation, was used to discriminate high from low-workload conditions [29].

The STD and maximum (MAX) of gas pedal force, and their time derivative (rate of change Δ), are used for identifying disrupted driving. More unstable driving is expected to result in higher values of both STD and MAX. Specifically, comparisons were made between the "just driving" baseline condition and conditions where the driver was interacting with an ASR system [Music Retrieval (MR)] or was repeating alphanumeric strings (ALSs). Experiments were carried out with data from 24 female and 26 male drivers. Results for all measures are summarized in Table VI. Numbers in this table refer to the percentage (z) of drivers who drove more unstably while using the MR, or during ALS, when compared with the baseline. For example, 72% of the drivers had a larger STD of normalized gas pedal force while using the MR than while

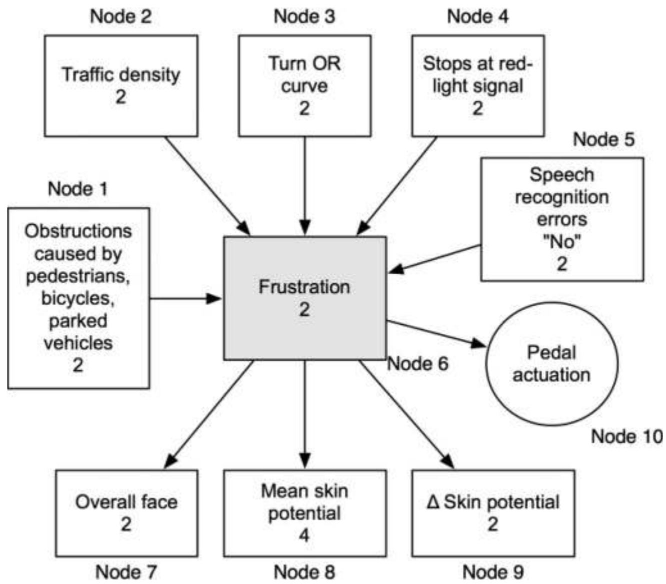

Fig. 5. Probabilistic dependencies between multimedia signals represented by a Bayesian network.

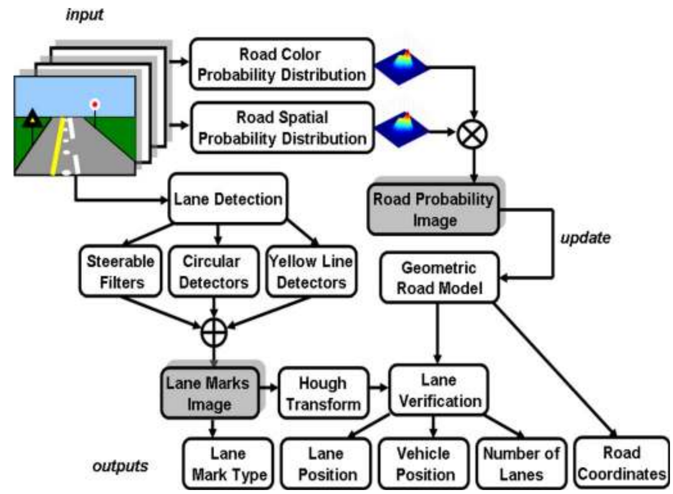

Fig. 6. Robust lane-tracking algorithm and its outputs.

just driving.MR, MAX, and STD of gas pedal force as well as MAX of ΔForce were significantly different from the baseline, while for ALS, no significant difference from the baseline was observed. Overall, compared with the "just driving" baseline, more disrupted behavior was observed when drivers interacted with the speech recognition system (MR). Furthermore, MAX was a more discriminating measure of the difference in pedal behavior than was STD.

Note that the analysis results only confirmed that the driving behavior signals of the corpora can be used for evaluating the work load differences in the predefined speech tasks. It is still not clear that the same measures can be applied to other general secondary tasks. The detailed discussions on this issue are found in [29].

B. Estimation of Driver Frustration

The basic idea here was to associate driver frustration level not to a single signal but to response characteristics, i.e., the relationship between the environmental information (inputs) and driver behavior (outputs) [30]. This novel approach can only be relevant and meaningful with multimedia signals related to driving behavior collected under real-world on-road conditions from real drivers driving real vehicles as we have done.

As shown in Fig. 5, a Bayesian network was employed for representing probabilistic dependency of environmental information and driver frustration level as well as driver frustration level and driver actions [31]. Values for nodes 1–4 and 7 are assigned from the data transcription. Node 5 represents whether the driver has uttered "no" to the ASR system (i.e., error in communication). The mean skin potential signal (sweat measurement) and its time derivative are encoded in nodes 8 and 9, respectively. Node 10 represents a continuous variable obtained through the cepstrum analysis of pedal operation. Frustration is encoded into two binary states by thresholding the objective scores that each driver has assessed while viewing video footage of his/her driving record. Therefore, the ground truth frustration level was measured through self-reporting. After the experiment, which lasted for about 15 min, participants were asked to assess their subjective level of frustration by referring to the front-view and facial videos, as well as the corresponding audio. The detailed procedures of the experiments can be found in [30].

For example, it has been observed from experiments that 80% of the driving data labeled as "frustration" can be accurately detected with a 9% false alarm rate, (i.e., the system has correctly estimated 80% of driver frustration). When drivers were not frustrated, the system made erroneous false positive decision 9% of the time.

C. Lane Tracking

There has been extensive research in the intelligent transportation community to develop robust lane-tracking systems which could be used to assist drivers. These approaches typically include a road model, feature extraction, postprocessing (verification), and tracking modules. However, a great majority of the proposed algorithms do not perform adequately under varying weather, road, visibility, traffic, and driver-related conditions.

Fig. 6 shows a basic diagram of our lane-tracking algorithm, which, like other lane-tracking algorithms, uses a geometric road model consisting of two lines to represent the road edges. The road edge lines are updated using a combination of road color probability distributions and road spatial coordinate probability distributions. These features are generated using road image pixels from nine videos segments, each containing at least 200 frames from the UTDrive corpus [32], [33]. As the training corpus contains different types of roads, such as interstate, urban, and highway, the algorithm developed here is expected to be robust at representing a variety of road profiles.

The detection module of the algorithm employs the following three different operators: 1) steerable filters to detect line-type lane markers [34]; 2) a yellow color filter; and 3) circle detectors. These three operators were chosen to capture the three different types of lane markers observed in the UTDrive

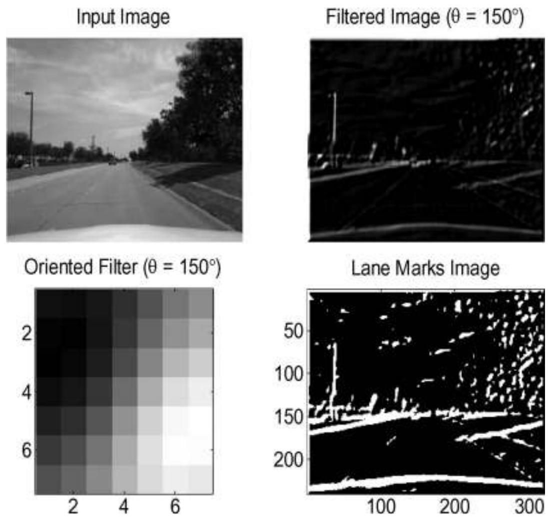

Fig. 7. Lane marker image formation, input, filters, and filtered image.

TABLE VII
MSE AND STD OF THE ERROR IN LANE TRACKING

	MSE	STD
Left Lanes	15.24	3.86
Right Lanes	28.02	4.67

data. However, the algorithm could be extended to include other types of lane markings by adding extra filters. After applying these operators, the resultant images are combined since it is now believed to have all the lane-mark features that could be extracted. The input image, oriented filter, filtered image, and the combined image (lane marker image) are depicted in Fig. 7.

Next, a Hough transform is applied to extract the lines together with their angles and sizes in the image. The extracted lines then go through a postprocessing step (the "lane verification" stage shown in Fig. 6), which involves the elimination of lane candidates not complying with the angle restrictions imposed by the geometric road model. From the lane-tracking algorithm, the following five different outputs are obtained: 1) lane mark type; 2) lane positions; 3) number of lanes; 4) relative vehicle position within its lane; and 5) road coordinates.

To measure the performance of the lane-tracking algorithm, the same corpus, but a different portion of the data, is also used as ground truth. The difference between the lane position coming from the lane tracker and the hand-marked ground truth value is used as the measurement of performance. Table VII shows the mean square error and the STD of the error in lane position. The algorithm, running at close to real time, is able to locate the lanes within an acceptable error range (the STD was 4.67 pixels at worst). Therefore, it is confirmed that collected image data can be used for training feasible image processing algorithms.

VI. Future Work

There are a few technologies that need to be developed to distribute our corpora to the wider research community. First, standard signal processing tools for the corpora should be developed for users. Second, the development of comparative evaluation platforms (which function in a manner similar to how U.S. National Institute of Standards does for the speech community for speaker recognition [38] and language recognition [39]) would be valuable. In addition, building a platform through which researchers can upload as well as download the data they collect would facilitate data sharing, as well as sharing best practices to address logistics concerning human subject IRB guidelines.

In addition, the collected data would be useful for the comparative study of driving behavior in different countries [35]. To achieve this objective, there is a need to perform future studies utilizing the combined data from these three sites and other corpora as they become available. Even though the authors are currently studying many common problems themselves, a larger cross section of the signal processing community and vehicular systems engineers are needed to achieve widespread success [36].

VII. Summary and Conclusion

This paper has provided an overview of efforts to initiate an international collaboration to build a large-scale on-the-road driving data corpora that can be used for a wide range of vehicle-related research areas on driving behavior. Unlike its predecessors, the current corpora contains both human and vehicle sensor data in addition to video recordings. In total, data from more than 500 drivers has been collected in three different countries, using very similar data collection vehicles, sensors, routes, and secondary tasks, which for the first time would allow cross-continent-based comparisons of driver distraction/behavior research advancements.

A novel transcription protocol was proposed to give rich and continuous annotation of the data, something which has been lacking in earlier corpora. The reliability of the transcription has also been verified.

The usefulness of the corpora was clearly demonstrated through three sample applications dealing with the following: 1) analyzing the effect of distractions/work load on driver behavior; 2) detecting frustration; and 3) training a lane-tracking system.

Potential applications of the corpora are numerous [36] in both the scientific arena and for technical development by scientists and engineers in signal processing and active safety fields. Some examples of potential applications include driver metric studies, development of mathematical models of driver dynamics, and improvement of sensor fusion technologies for robust sensing. However, the most unique feature of the corpora is the fact that it can be a common platform for researchers in diverse fields working in vehicle technologies including ergonomics, control, signal processing, electronics, and sensor devices. Needless to say, integrating all those technologies is very critical for developing safe and efficient vehicles.

Finally, it is worth noting that these corpora are publicly available (primarily for the research community) and a small sample data set can be downloaded through our website. Readers who are interested in the corpora are welcome to contact the authors.

TABLE VIII
LIST OF TRANSCRIPTION TAGS

	Possible data used in decision-making	Definition of possible parameters
DRIVER ACTION VARIABLES		
Overall face	Face Video	1. Negative 2. Positive
Head position	Face Video	1. Other than straight
SECONDARY AND TERTIARY TASKS VARIABLES		
Task types	Data log, front view video, speech	1. Music search 2. Dialog on cell phone (w/ human) 3. Dialog on cell phone (w/ machine) 4. Sign reading 5. Instrument Control 6. Conversation with passenger 7. Spontaneous speech 8. Reaching to perform something 9. Receiving route guidance using cell phone 10. Repeating alphanumeric digits 11. Driving without any secondary task
DRIVING ENVIRONMENT VARIABLES		
Road type	Front view video	1. Highway 2. Driveway 3. Two-way 4. Tunnel
Number of lanes	Front view video	
Road design	Front view video	1. Left curve 2. Right curve 3. U-turn 4. Signalized intersection 5. Lane splitting (own lane) 6. Lane splitting (another lane) 7. Lane ends (own lane) 8. Lane ends (another lane)
Traffic volume	Front view video	1. Level of service A or B (light) 2. Level of service C or D (medium) 3. Level of service E or F (heavy)

TABLE VIII
(Continued.) LIST OF TRANSCRIPTION TAGS

Traffic light color change	Front view video (labels are not used during idling, pinpoint label)	1. Green to yellow 2. Yellow to red 3. Red to green
Obstructions due to pedestrians, bicycles, and parked vehicles	Front view video	
Weather condition	Front view video	1. Clear 2. Cloudy 3. Drizzle 4. Heavy rain 5. Fog 6. Snow
VEHICLE STATUS VARIABLES		
Lane change	Front view video, steering wheel angle	1. Left 2. Right
Stop	Front view video, brake, vehicle speed	
Stop at red light	Front view video, brake, vehicle speed	
Turn	Front view video, brake, vehicle speed	1.Turn left 2. Turn right
Turn with interruption due to bicycles, pedestrians (foot traffic)	Front view video, steering-wheel angle, brake, gas	1. Left: Left turn is started but breaks continuity before reaching end once or more than once 2. Right: same strategy for right
Turn with Interruption due to red light	Front view video, steering-wheel angle, brake, gas	1. Left: same as above but the type of interruption is classified using front view video 2. Right: same strategy for right
Turn with interruption due to vehicles	Front view video, steering-wheel angle, brake, gas	1. Left: same as above but the type of interruption is classified using front view video 2. Right: same strategy for right
Vehicle following	Front view video	
Current lane position	Front view video	
Overtaking	Front view video, brake, vehicle speed	
SPEECH/BACKGROUND NOISE VARIABLES		
Speech	Speech	1. Driver 2. Operator 3. ASR System

TABLE VIII
(Continued.) LIST OF TRANSCRIPTION TAGS

Background Noise	Background Noise	1. Stationary (engine, A/C,...) 2. Intermittent (music, ring tone,...) 3. Emergency vehicles
DRIVER AFFECTIVE/MENTAL STATE VARIABLES		
Frustration	Face video, driver speech	Min: 0; Max: 30

APPENDIX A

A list of transcription tags is shown in Table VIII.

ACKNOWLEDGMENT

The authors would like to thank the substantial contributions of many collaborators who have participated in this 3-year New Energy and Industrial Technology Development Organization (NEDO), Japan-supported research on driver behavior modeling, particularly P. Angkititrakul (formerly of Center for Robust Speech Systems at the University of Texas (CRSS-UT) Dallas, now at Nagoya University), A. Sathyanarayana and X. Yang (formerly of CRSS-UT Dallas), H. Erdogan, R. Ficici, and E. Cokelek (Sabanci University, Turkey), E. Erzin (Koc University, Turkey), OYAK-Renault, Turkcell, OPET, SATKO, OTAM, and MEKAR Laboratories (all in Istanbul, Turkey), A. Wahab (formerly at NTU Singapore, now at IIUM in Malaysia), and J.C. De Martin (Torino, Italy). Without their valuable contributions, the advancements stemming from this international collaboration would not have been possible. Finally, the authors would also like to thank and recognize all those who participated in the 2003, 2005, 2007, and 2009 Biennial Workshops on DSP for In-Vehicle Systems and those involved in the three subsequently edited textbooks. Their involvement, discussions, and research contributions have helped shape the direction of the next generation of intelligent in-vehicle systems.

REFERENCES

[1] H. van den Heuvel, J. Boudy, R. Comeyne, S. Euler, A. Moreno, and G. Richard, "The SPEECHDAT-CAR multilingual speech databases for in-car applications: Some first validation results," in *Proc. EUROSPEECH*, Budapest, Hungary, 1999, pp. 2279–2282.

[2] J. H. L. Hansen, P. Angkititrakul, J. Plucienkowski, S. Gallant, U. Yapanel, B. Pellom, W. Ward, and R. Cole, "CU-move: Analysis and corpus development for interactive in-vehicle speech systems," in *Proc. ISCA Interspeech*, Aalborg, Denmark, Sep. 2001, pp. 2023–2026.

[3] J. H. L. Hansen, X. Zhang, M. Akbacak, U. H. Yapanel, B. Pellom, W. Ward, and P. Angkititrakul, "CU-MOVE: Advanced in-vehicle speech systems for route navigation," in *DSP for In-Vehicle and Mobile Systems*, H. Abut, J. H. L. Hansen, and K. Takeda, Eds. New York: Springer, 2005, ch. 2.

[4] H. Abut, J. H. L. Hansen, and K. Takeda, Eds., *DSP for In-Vehicle and Mobile Systems*. New York: Springer, 2005.

[5] N. Kawaguchi, S. Matsubara, I. Kishida, Y. Irie, H. Murao, Y. Yamaguchi, K. Takeda, and F. Itakura, "Construction and analysis of a multi-layered in-car spoken dialogue corpus," in *Proc. EUROSPEECH*, Aalborg, Denmark, Sep. 2001, pp. 2027–2030.

[6] E. Levin, S. Narayanan, R. Pieraccini, K. Biatov, E. Bocchieri, G. Di Fabbrizio, W. Eckert, S. Lee, A. Pokrovsky, M. Rahim, P. Ruscitti, and M. Walker, "The AT&T–DARPA communicator mixed-initiative spoken dialogue system," in *Proc. ICSLP*, 2000, pp. 122–125.

[7] B. Lee, M. Hasegawa-Johnson, C. Goudeseune, S. Kamdar, J. Boley, S. Borys, M. Liu, and T. Huang, "AVICAR: Audio-visual speech corpus in a car environment," in *Proc. INTERSPEECH*, Jeju Island, Korea, Oct. 4–8, 2004.

[8] M. Hasegawa-Johnson *et al.*, AVICAR: Audio-visual speech corpus in a car environment. [Online]. Available: http://www.isle.illinois.edu/sst/AVICAR/

[9] T. Duke, Z. Boulos, J. A. Stern, and T. A. Ranney, "Eye-activity measures of fatigue and napping as a fatigue countermeasure," U.S. Dept. Transp., Fed. Hwy. Admin., Washington, DC, Final Rep. FHWA-MC-99-028, Jan. 1999.

[10] T. A. Dingus, S. G. Klauer, V. L. Neal, A. Petersen, S. E. Lee, J. Sudweeks, M. A. Perez, J. Hankey, D. Ramsey, S. Gupta, C. Bucher, Z. R. Doerzaph, J. Jermeland, and R. R. Knipling, "The 100-car naturalistic driving study—Phase II: Results of the 100-car field experiment," Nat. Highway Traffic Safety Admin., Washington, DC, Contract No. DTNH22-00-C-07007, 2006.

[11] S. G. Klauer, T. A. Dingus, V. L. Neale, J. D. Sudweeks, and D. J. Ramsey, "The impact of driver inattention on near crash/crash risk: An analysis using the 100 car naturalistic driving study data," Nat. Highway Traffic Safety Admin., Washington, DC, Contract No. DTNH22-00-C-07007, Task order 23, 2006.

[12] B. Magladry and D. Bruce, "Improved vehicle safety and how technology will get us there, hopefully," in *In-Vehicle Corpus and Signal Processing for Driver Behavior*, K. Takeda, J. H. L. Hansen, H. Erdogan, and H. Abut, Eds. New York: Springer Science+Business and Media, 2008, ch. 1.

[13] [Online]. Available: http://www.trb.org/StrategicHighwayResearchProgram2SHRP2/Public/Blank2.aspx

[14] "Automotive collision avoidance system field operational test report: Methodology and results," Nat. Highway Traffic Safety Admin., Washington, DC, Dep. Transp., HS 809 900, 2005.

[15] D. LeBlanc, J. Sayer, C. Winkler, R. Ervin, S. Bogard, J. Devonshire, M. Mefford, M. Hagan, Z. Bareket, R. Goodsell, and T. Gordon, "Road departure crash warning system field operational test: Methodology and results," Univ. Michigan Transp. Res. Inst., Ann Arbor, MI, 2006.

[16] J. Sayer, D. LeBlanc, S. Bogard, D. Funkhouser, S. Bao, M. L. Buonarosa, and A. Blankespoor, "Integrated Vehicle-Based Safety Systems (IVBSS) field operational test final program report," Univ. Michigan Transp. Res. Inst. (UMTRI), Ann Arbor, MI, Dep. Transp., HS 811 482, UMTRI-2010-36, 2010.

[17] euroFOT: The First European Field Operational Test on Active Safety System. [Online]. Available: http://www.eurofot-ip.eu/

[18] C. Miyajima, T. Kusakawa, T. Nishino, K. Itou, and K. Takeda, "On-going data collection for driving behavior signals," in *In-Vehicle Corpus and Signal Processing for Driver Behavior*, K. Takeda, J. H. L. Hansen, H. Erdogan, and H. Abut, Eds. New York: Springer, 2008, ch. 4.

[19] A. Ozaki, S. Hara, T. Kusakawa, C. Miyajima, T. Nishino, N. Kitaoka, K. Itou, and K. Takeda, "In-car speech data collection along with various multimodal signals," in *Proc. 6th Int. LREC*, Marrakech, Morocco, May 28–30, 2008.

[20] P. Angkititrakul, J. H. L. Hansen, S. Choi, T. Creek, J. Hayes, J. Kim, D. Kwak, L. T. Noecker, and A. Phan, "UTDrive: The smart vehicle project," in *In-Vehicle Corpus and Signal Processing for Driver Behavior*, K. Takeda, J. H. L. Hansen, H. Erdogan, and H. Abut, Eds. New York: Springer, 2008, ch. 5.

[21] H. Abut, H. Erdogan, A. Ercil, B. Cürüklü, H. C. Koman, F. Tas, A. O. Argunsah, S. Cosar, B. Akan, H. Karabalkan, E. Cokelek, R. Ficici, V. Sezer, S. Danis, M. Karaca, M. Abbak, M. G. Uzunbas, K. Ertimen, M. Imamoglu, and C. Karabat, "Data collection with "UYANIK"," in *In-Vehicle Corpus and Signal Processing for Driver Behavior*, K. Takeda, J. H. L. Hansen, H. Erdogan, and H. Abut, Eds. New York: Springer, 2008, ch. 3.

[22] J. Cohen, "A coefficient of agreement for nominal scales," *Educ. Psychol. Meas.*, vol. 20, no. 1, pp. 37–46, Apr. 1960.

[23] C. Miyajima, Y. Nishiwaki, K. Ozawa, T. Wakita, K. Itou, K. Takeda, and F. Itakura, "Driver modeling based on driving behavior and its evaluation in driver identification," *Proc. IEEE*, vol. 95, no. 2, pp. 427–437, Feb. 2007.

[24] E. R. Boer, M. E. Rakauskas, M. J. Ward, and M. A. Goodrich, "Steering entropy revisited," in *Proc. 3rd Int. Driving Symp. Human Factors Driver Assessment, Training Vehicle Des.*, 2005, pp. 25–32.

[25] P. Boyraz, A. Sathayanarayana, and J. H. L Hansen, "Lane keeping performance metrics for assessment of audio-cognitive distractions," Soc. Automotive Eng., Driver Performance Metrics, 2009.

[26] A. Sathyanarayana, P. Boyraz, and J. H. L Hansen, "Driver behaviour analysis and route recognition by hidden Markov models," in *Proc. IEEE Int. Conf. Veh. Electron. Safety*, Columbus, OH, Sep. 22–24, 2008, pp. 276–281.

[27] H. Boril, P. Boyraz, and J. H. L. Hansen, "Towards multi-modal driver's stress detection," in *Proc. 4th Biennial Workshop DSP In-Vehicle Syst. Safety*, Dallas, TX, Jun. 25–27, 2009.

[28] N. Fujiwara, T. Itoh, and K. Araki, "Analysis of changes in dialogue rhythm due to dialogue acts in task-oriented dialogues," in *Proc. TSD*, vol. 4629, *LNAI*, V. Matousek and P. Mautner, Eds., 2007, pp. 564–573.

[29] L. Malta, A. Ozaki, C. Miyajima, and K. Takeda, "Use of on-road data in evaluating driver performance metrics," in *Performance Metrics for Assessing Driver Distraction: The Quest for Improved Road Safety*. Warrendale, PA: Soc. Automotive Eng., 2010, ch. 9.

[30] L. Malta, C. Miyajima, N. Kitaoka, and K. Takeda, "Analysis of real-world driver's frustration," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 1, pp. 109–118, Mar. 2011.

[31] L. Malta, C. Miyajima, and K. Takeda, "Multi-modal real-world driving data collection, transcription, and integration using Bayesian network," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Eindhoven, The Netherlands, Jun. 2008, pp. 150–155.

[32] P. Boyraz, X. Yang, and J. H. L. Hansen, "Computer vision applications for context-aware intelligent vehicles," in *Proc. 4th Biennial Workshop DSP In-Vehicle Syst. Safety*, Dallas, TX, Jun. 25–27, 2009.

[33] P. Boyraz, X. Yang, A. Sathyanarayana, and J. H. L. Hansen, "Context-aware active vehicle safety and driver assistance," in *Proc. 21st Int. Tech. Conf. ESV*, Stuttgart, Germany, Jun. 2009.

[34] M. Jacob and M. Unser, "Design of steerable filters for feature detection using canny-like edge criteria," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 8, pp. 1007–1019, Aug. 2004.

[35] H. Abut, J. H. L. Hansen, and K. Takeda, Eds., *Advances for In-Vehicle and Mobile Systems: Challenges for International Standards*. New York: Springer, 2007.

[36] K. Takeda, J. H. L. Hansen, H. Erdogan, and H. Abut, Eds., *In-Vehicle Corpus and Signal Processing for Driver Behavior*. New York: Springer, 2008.

[37] [Online]. Available: http://www.sp.m.is.nagoya-u.ac.jp/NEDO/Drive-Best/

[38] Nat. Inst. Sci., Technol., SRE-2010: Speaker Recognition Evaluation. [Online]. Available: http://www.itl.nist.gov/iad/mig/tests/sre/2010/index.html

[39] Nat. Inst. Sci., Technol., LRE-2011: Language Recognition Evaluation. [Online]. Available: http://www.nist.gov/itl/iad/mig/lre11.cfm

John H. L. Hansen (S'81–M'82–SM'93–F'07) received the B.S. degree in electrical engineering from Rutgers University, New Brunswick, NJ, and the M.S. and Ph.D. degrees in electrical engineering from Georgia Institute of Technology, Atlanta.

He joined the Erik Jonsson School of Engineering and Computer Science, University of Texas at Dallas (UTD), Richardson, in 2005, where he is a Professor and the Department Head of the Department of Electrical Engineering and holds the Distinguished University Chair in Telecommunications Engineering. He also holds a joint appointment as Professor with the School of Behavioral and Brain Sciences (Speech and Hearing). At UTD, he established the Center for Robust Speech Systems (CRSS), which is part of the Human Language Technology Research Institute. Previously, he served as the Department Chair and a Professor with the Department of Speech, Language, and Hearing Sciences and a Professor with the Department Electrical and Computer Engineering, University of Colorado, Boulder, where he cofounded the Center for Spoken Language Research. In 1988, he established the Robust Speech Processing Laboratory and continues to direct research activities with CRSS at UTD. He has supervised 55 (24 Ph.D., 31 M.S./M.A.) thesis candidates. He is the author or a coauthor of 402 journal and conference papers and eight textbooks in the field of speech processing and language technology, as well as subject assessment/monitoring for in-vehicle applications. His research interests include digital speech processing, analysis and modeling of speech and speaker traits, speech enhancement, feature estimation in noise, robust speech recognition with emphasis on spoken document retrieval, and in-vehicle interactive systems for hands-free human–computer interaction.

Dr. Hansen was named IEEE Fellow in 2007 for his contributions on "robust speech recognition in stress and noise" and International Speech Communication Association Fellow in 2010 for his contributions on "research for speech processing of signals under adverse conditions." He received the Acoustical Society of America's 25 Year Award in 2010. He was also the recipient of the 2005 University of Colorado Teacher Recognition Award, as voted on by the student body. He is currently serving as the Technical Committee Chair and Member of the IEEE Signal Processing Society Speech and Language Processing Technical Committee (2005–2008; 2010–2013; elected IEEE Speech and Language Technical Committee Chair for 2011–2013) and the Educational Technical Committee (2005–2008; 2008–2010). Previously, he has served as the Technical Advisor to the U.S. Delegate for NATO (IST/TG-01), an IEEE Signal Processing Society Distinguished Lecturer (2005/2006), an Associate Editor for the IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING (1992–1999), an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS (1998–2000), and an Editorial Board Member of the *IEEE Signal Processing Magazine* (2001–2003). He also organized and served as the General Chair for ICSLP/Interspeech-2002: the International Conference on Spoken Language Processing, September 16–20, 2002, and served as a Coorganizer and the Technical Program Chair for the 2010 International Conference on Acoustics, Speech, and Signal Processing, Dallas, TX.

Kazuya Takeda (SM'09) received the B.E. and M.E. degrees in electrical engineering and the Dr. Eng. degree from Nagoya University, Nagoya, Japan, in 1983, 1985, and 1994, respectively.

From 1986 to 1989, he was with the Advanced Telecommunication Research (ATR) Laboratories, Osaka, Japan. His main research interest at ATR was corpus-based speech synthesis. He was a Visiting Scientist with the Massachusetts Institute of Technology, Cambridge, from November 1987 to April 1988. From 1989 to 1995, he was a Researcher and Research Supervisor with the KDD Research and Development Laboratories, Kamifukuoka, Japan. From 1995 to 2003, he was an Associate Professor with the Faculty of Engineering, Nagoya University. Since 2003, he has been a Professor with the Department of Media Science, Graduate School of Information Science, Nagoya University. His current research interests are media signal processing and its applications, including spatial audio, robust speech recognition, and driving behavior modeling. He is the author or a coauthor of more than 100 journal papers, six books, and more than 100 conference papers.

Dr. Takeda was a Technical Cochair of the 2007 International Conference on Multimodal Interfaces and the 2009 International Conference on Vehicular Safety and Electronics. He was a cofounder of the Biennial Workshop on digital signal processing for In-Vehicle Systems and Safety in 2003.

Pinar Boyraz (M'08) received the B.S. degrees in textile engineering and mechanical engineering from Istanbul Technical University, Istanbul, Turkey, in 2003 and 2004 and the Ph.D. degree in mechatronics from Loughborough University, Leicestershire, U.K., in July 2008.

She was a Research Associate with the Erik Jonsson School of Engineering and Computer Science, University of Texas at Dallas, Richardson, from February 2008 to January 2010. She is currently an Assistant Professor with the Mechanical Engineering Department, Istanbul Technical University. Her research interests include applications of control theory, robust and optimal control, signal processing, vehicle/system dynamics, image/video processing and analysis, and applications of artificial intelligence.

Lucas Malta received the B.S. degree in electrical engineering from the Federal University of Minas Gerais, Belo Horizonte, Brazil, in 2004 and the M.S. and Ph.D. degrees in engineering from Nagoya University, Nagoya, Japan, in 2007 and 2010, respectively.

He is currently a Systems Engineer with Volvo Technology Corporation, Göteborg, Sweden. His research interest includes driver behavior, driver modeling, and human-centered driver assistance.

Chiyomi Miyajima (M'09) received the B.E. degree in computer science and the M.E. and Ph.D. degrees in electrical and computer engineering from the Nagoya Institute of Technology, Nagoya, Japan, in 1996, 1998, and 2001, respectively.

From 2001 to 2003, she was a Research Associate with the Department of Computer Science, Nagoya Institute of Technology. Currently, she is an Assistant Professor with the Graduate School of Information Science, Nagoya University. Her research interests include speaker recognition, multimodal speech recognition, sign language recognition, and modeling of human behavior.

Hüseyin Abut (LSM'11) received the B.S. degree in electrical engineering from Robert College, Istanbul, Turkey, in 1968 and the M.S. and the Ph.D. degrees from the North Carolina State University, Raleigh, in 1970 and 1972, respectively.

He is a Professor Emeritus with the Department of Electrical and Computer Engineering, San Diego State University, San Diego, CA, where he has served as a Full Professor and Director between 1981 and 2006. At present, he is an Adjunct Professor with Bogaziçi University, Istanbul. He has also held visiting faculty appointments in Turkey, Japan, Singapore, Germany, and Stanford University, Stanford, CA. He served as the President and Vice President for R&D startup companies both in the U.S. and abroad. He was instrumental during the establishment of a foundation university in Turkey in the capacity of Trustee. He has served twice as Senior Advisor for the U.N. Program Development. He is the editor or a coeditor of four books, namely *Vector Quantization* (IEEE Press, 1990); *DSP in Mobile and Vehicular Systems* (Springer, Science-Business, 2005); *Advances for In-Vehicle and Mobile Systems: Challenges for International Standards* (Springer, Science-Business, 2007); and *In-Vehicle Corpus and Signal Processing for Driver Behavior* (Springer, Science-Business, 2009). He is the author or a coauthor of more than 100 publications in journals, books, and book chapters in the fields of speech processing, speaker verification, image processing, image understanding, biometrics, and signal processing for vehicular applications and driver behavior. He is the holder of two patents on speaker verification systems for consumer applications.

Dr. Abut was a Distinguished Lecturer (Class of 2002) for the IEEE Signal Processing Society. He has been associated with numerous international workshops, biennials, and conferences, including nine International Conference on Acoustics, Speech, and Signal Processing Meetings—General Chair 2000. He has served at various capacities as Member, Secretary, Board Member, and Senator with several academic and professional organizations, including the IEEE Signal Processing Society. He has served as a founding Associate Editor for the IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING.