In-vehicle Multi-channel Signal Processing and Analysis in UTDrive Project: Driver Behavior Modeling and Active Safety Systems Development

Pinar Boyraz, Amardeep Sathyanarayana, John H. L. Hansen

Erik Jonsson School of Engineering and Computer Science
University of Texas at Dallas, 75080, Richardson, TX, USA
{pinar.boyraz; axs06300; john.hansen}@utdallas.edu

Abstract

It has been pointed that most of the accidents on the roads are caused by driver faults, inattention and low performance. Therefore, future active safety systems are required to be aware of the driver status to be able to have preventative features. This probe study gives a system structure depending on multi-channel signal processing for three modules: Driver Identification, Route Recognition and Distraction Detection. The novelty lies in personalizing the route recognition and distraction detection systems according to particular driver with the help of driver identification system. The driver ID system also uses multiple modalities to verify the identity of the driver; therefore it can be applied to future smart cars working as car-keys. All the modules are tested using a separate data batch from the training sets using eight drivers' multi-channel driving signals, video and audio. The system was able to identify the driver with 100% accuracy using speech signals of length 30 sec or more and a frontal face image. After identifying the driver, the maneuver/ route recognition was achieved with 100% accuracy and the distraction detection had 72% accuracy in worst case. In overall, system is able to identify the driver, recognize the maneuver being performed at a particular time and able to detect driver distraction with reasonable accuracy.

INTRODUCTION

In order to increase the safety on the roads, current research efforts in in-vehicle systems have three main focus areas: in-vehicle controllers, driver assistance/monitoring systems and environmental risk assessment systems. Ideally, in the near future, these seemingly separate efforts are expected to come together under a decision making hierarchical system structure to reduce accidents caused by dynamical factors related to: vehicle, driver and traffic/infrastructure. The configuration of the integration of these sub-systems may vary from a fully automatic smart car to a semi-autonomous, driver-centered approach. To find the optimal solution for this problem with the least intrusion to the driver, driver behavior models will have crucial importance in developing driver-adaptive, context-aware active safety systems.

Under support from an international NEDO funded consortium, UTDrive project began two years ago, with the formulation of a data collection vehicle, Toyota RAV4, customized with a variety of sensors and transducers for multi-modal data acquisition. The data include audio, video, gas/brake pedal pressures, following distance, CAN-Bus information and GPS information. Signals are synchronously recorded with the help of a commercial data acquisition unit. In the first phase of the project (P1), 100 sessions of multi-channel driving data has been collected from a demographically wide range across 53 participants. Two driving routes in the neighborhood areas of Richardson, TX are chosen; the first route represents a residential scenario and the second represents a business-district scenario. Fundamentally, these two scenarios are quite different in terms of traffic density, infrastructure and attention sources required from the driver. Data collection from both routes includes neutral driving and driving under task distraction. For driving sessions with distraction, manual secondary tasks (adjusting radio, AC/heater, etc.), cognitive tasks (reading road signs, cell-phone dialing Airline flight speech prompted system and discussing with the research team member) and driving maneuvers (lane change, left/right hand turn) were requested from the driver. This extensive database is carefully transcribed to distinguish the time windows of interest (i.e. each particular maneuver, the section including the speech with Airline dialog, etc.) and log this data under a developed protocol. The transcribed multi-sensor data are then analyzed using different state-of-the-art techniques in speech signal processing, such as Hidden Markov Models (HMM) and Gaussian

Mixture Models (GMM) for the purpose of distraction detection. The results obtained so far have led contributions in three book publications [1, 2, 3] compiling the papers in international workshops under the name of DSP for In-Vehicle Systems.

In this paper, the second phase (P2) of the research will be detailed. In P2, three main areas related to driver behavior signal processing and analysis is explored in further depth: multi-sensor driver identification, route recognition and reliable driver distraction detection. First, the formulated driver identification system is explained in detail. It utilizes video (facial features), audio (speaker-dependent features) and CAN-Bus cues (driving performance metrics) of the individual drivers. This system can be classified as a multi-modal biometric identification system aimed at recognizing the driver with the ultimate goal of adapting the car set points and future controllers to the characteristics of the driver for safe operation of the vehicle. The second system is based on a novel idea of building a route model formed by maneuvers and sub-maneuvers in the analogy to speech recognizers working on phonemes (sub-maneuvers), words (maneuvers) and sentence (route) models having a semantic/syntactic language model (context of driving and sequence of driving) . The third system attempts to detect the distraction of the drivers from the multi-sensor data stream using HMMs.

This paper is organized in the following way: First the background on face recognition, speaker identification and CAN-Bus signal processing are mentioned with an emphasis on need for multi-modal systems for in-vehicle driver identification. Second, data collection vehicle, experimental procedure and corpus are mentioned. Next, integration of these three systems is explained in section 'System Integration and Overview' and then three modules are explained in greater detail in 'Driver Identification', 'Route Recognition' and 'Distraction Detection' sections. Finally, further work is recommended for this very promising in-vehicle safety system to be improved. The contribution of the study lies in combining the existing ideas on improving the safety using in-vehicle electronic devices in a system integration and mechatronics approach.

BACKGROUND

The research area this paper addresses is interdisciplinary and builds on multi-modal biometric identification systems employing mainly face and speaker recognition and driver characteristics from CAN-Bus. Recognizing the driver robustly despite of the adverse conditions of in-vehicle environment such as changing illumination and engine noise is very important in adapting the driver assistance and monitoring modules to driver characteristics. Here, brief background is given on face and speaker recognition, multi-modal bio-metric systems, route recognition and distraction detection to understand how these systems can be combined to increase the safety of vehicles.

Face recognition is a mature technology in itself and has been used in commercial systems in authenticity and security applications. A comprehensive literature survey on face recognition algorithms can be found in (4). The in-vehicle application poses extra challenges for face recognition as follows:

- The illumination changes are dramatic and at significant levels
- Drivers cannot be expected to stand still for image acquisition therefore system should use video sequences for recognition
- Video sequences contain face images with varying scale, orientation and non-rigid motion
- Driver appearance may change over time

Most of these issues are addressed in a probabilistic scheme in (5). They applied still-to-video and video-to-video recognition algorithms incorporating the temporal information from the videos in a probabilistic framework. In this paper, our focus is not developing the most capable face recognition system for in-vehicle application; rather we try to include face recognition cues in a multi-modal driver recognition system. In fact, we will be only using principal component analysis (PCA) method for now as it was applied in (6), since our main focus is to develop a multi-modal system for recognition with simplistic modules. Incorporating more robust 3-D, temporal and probabilistic approaches for in-vehicle use deserves a separate investigation in its own right.

The second modality of our system is based on speaker identification cues. For a comprehensive overview on speaker identification (7) is recommended. Here, most widely used MFCC will be employed for feature extraction and GMM will be used to assess the performance of this simplistic speaker identification system.

In our system, the third modality comprises several metrics derived from CAN-Bus signals comprising mainly vehicle speed, steering wheel angle and brake/pedal signals. Use of multi-modal systems for person identification is not a complete novel idea and kinematics of gait; key stroke in typing and several other dynamics of motion have been used for recognition. Although CAN-Bus signals can be used to derive more detailed models of driving models employing control theory, here they will be taken as time series representing a particular motion sequence (i.e. right turn, left turn and lane change). Using Can-BUS information and fusion with two previously mentioned modules is an in-vehicle focused and novel approach to multi-modal person recognition in car driving context. There is very little study on CAN-Bus signal modeling, however, some promising results can be found in (8).

CAN-Bus signals are not forming only the dynamic modality of our recognition system, but they are also the information source for diagnosis system comprising route recognition and distraction detection. We will be employing Hidden Markov Models (HMM) for modeling the maneuvers and detecting the distraction. There is substantial successful work on application of HMM in driver modeling (9, 10). Although these previous studies unleashed the potential of HMM in driver behavior modeling there is still need for extensive studies including larger databases and more real-world driving situations in models in a hierarchical approach.

It should be also noted that multi-modal person recognition with an in-vehicle application has been studied before (11), however, the recognition system has not been connected to maneuver recognition and distraction detection modules to improve their performances. Therefore, in this paper, we are offering an improvement in the performance of maneuver recognition and distraction detection algorithms by recognizing the driver in the beginning of the driving session as well as suggesting an authorization system as the other researchers suggested.

Data collection vehicle, experimental procedure and corpus

The vehicle (Figure 1) is equipped to perform multi-modal data collection with signal channels including:

- Videos: driver cabin and the road scene
- Microphone array and close microphone to record driver's speech
- Distance sensor using laser to measure the distance between ego vehicle and leading vehicle
- GPS for position measurement
- CAN-Bus: vehicle speed, steering wheel angle, brake/gas
- Gas/Brake pedal pressure sensors

These sensors allow collecting dynamic driving data and some physiological cues on driver status in a non-intrusive manner. Since the equipment is visible to the participant and there is an experimenter in the car, the collected data cannot be classified as pure naturalistic driving data; however, the routes, secondary sub-tasks and the scenarios are in good agreement with real driving experience.

Figure 1. Data collection vehicle and incorporated sensors

The driving scenarios include two different routes: residential and commercial areas including right turn, left turn, lane change, cruise and car following segments. Each route is driven by each driver twice: neutral and distracted. These routes can be seen in Figure 2.

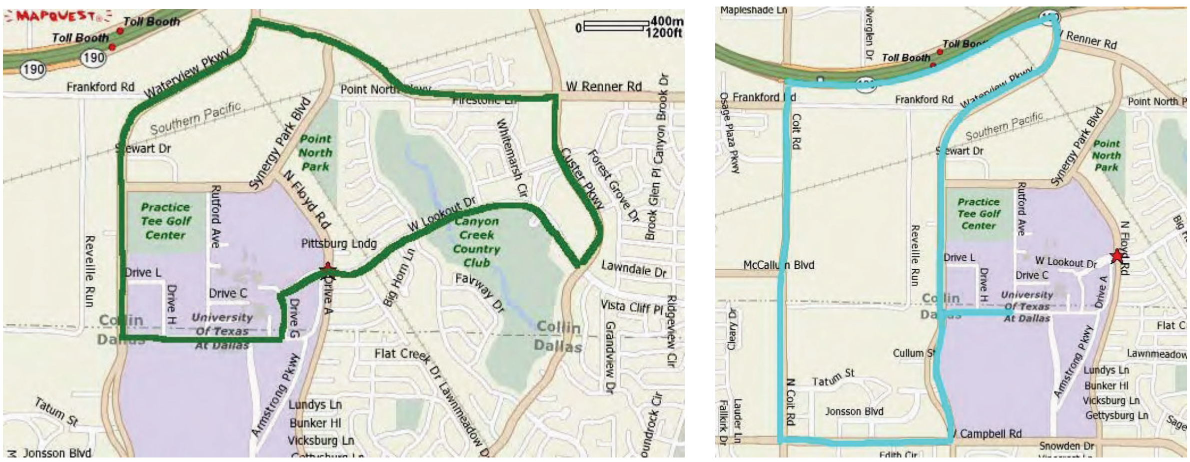

Figure 2. Route 1 (Left) and Route 2 (Right)

UTDrive Corpus includes 40 male 37 female drivers' multi-sensor driving data (each person has three sessions repeated twice giving six sessions in total) and the experiments are in continuation to extend the database. It is close to naturalistic driving data since the routes and the scenarios are from real roads. However, it should be noted as well that it is not completely naturalistic since the driver is aware that he/she is being recorded and there is often nervousness due to using the data collection vehicle which is completely new to participants. In this investigation a narrow data base containing only three drivers will be examined since it reflects the real situation that a vehicle may be used by 3-4 drivers but not more. While this restriction makes it easy for recognition, it comes with a drawback as well: there is limited data or limited number of observations of a maneuver from the same person in our database. Nevertheless, despite this limitation with very limited data we will demonstrate that the recognition system can help other two diagnosis modules increase the overall performance of the safety system. Next session gives the overview of the system integration between multi-modal biometric driver identification, route recognition and distraction detection modules.

System integration and overview

One important concept in mechatronics approach in active safety system design is to have the system integration for boosting the over-all system performance simplifying the structures. Applying this principle we combine the multi-modal biometric driver identification system with route/maneuver recognition and distraction detection systems. Individual systems combined here can work; however, the performances of the systems decrease due to dynamics of driving and personal differences among the drivers. Although systems are trained on a larger database including several drivers, the user might have different driving characteristics which would directly affect the performances of maneuver recognition and distraction modules. These problems can be alleviated by employing a driver identification system and personalization of the system, multi-modal driver identification system authorizes the driver as well as loading driver-characteristic properties. The flow-diagram of the system is shown in Figure 3.

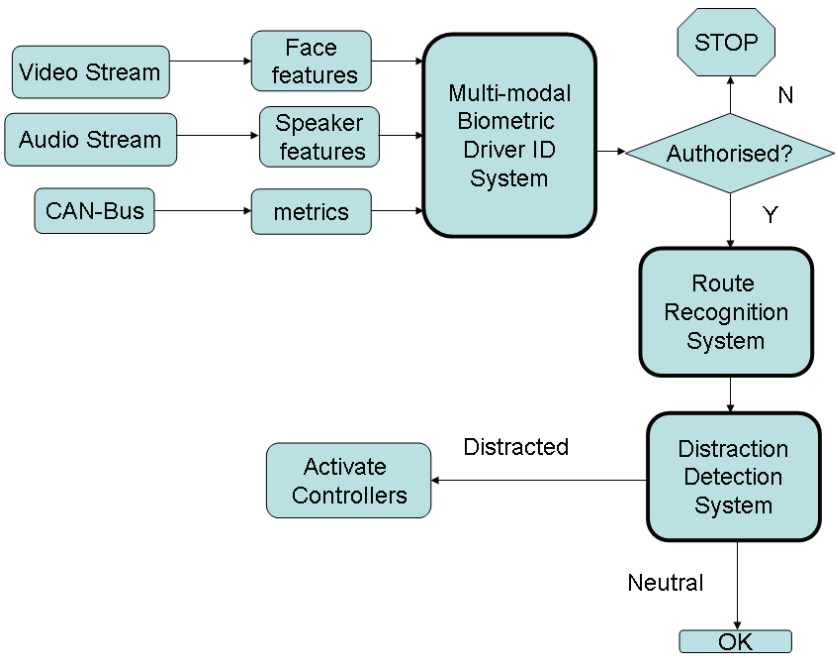

Figure 3. System Integration and flow diagram

In the following sub-sections the individual module development and performances are mentioned.

Driver identification

Face Recognition Modality

Driver identification module uses multi-modal information from the driver: face-recognition and speaker identification cues are used as primary modality while they are connected with and backed up by driving characteristics derived from CAN-Bus. The final identification result is a fusion of decision from these three modalities, however; first the identification results from individual modalities are given here.
First modality uses *eigen-faces* approach employing PCA. Ten images from each of three drivers (total 30) are included for training and 5 images are used for testing. In the resulting PCA analysis first 19 eigen-values and associated eigen-vectors are selected. Results are given for Driver 1 in Figure 4, indicating the reliable weights which give the shortest Euclidean distance between the weights obtained from the test and those obtained from test signals.

Figure 4. Test images weights with 19 eigen-vector subspace, reliable weights for driver I: 2,4,5,6,7,8,11,12,16,17,18,19

Cumulative PCA results can be seen in Table I, there are two failed test images from driver which are the cases when driver had a slight tilt or rotation. These failures can be easily fixed with a more advanced face feature extraction and classification scheme. However, in this application 13 cases of 15 test images were correctly classified, which is satisfactory performance for only one modality. The failures can be corrected by other modalities easily without applying a more advanced method on this modality.

Table I. Cumulative PCA results for face recognition module using 3 driver-database

Driver 1 Test image 1 (ground truth: 1)			Driver 2 Test image 1 (ground truth: 2)			Driver 3 Test image 1 (ground truth: 3)		
Class 1	1.308		Class 1	2.8714		Class 1	3.9308	
Class 2	4.0576		Class 2	0.8057		Class 2	4.5672	
Class3	4.2706		Class3	4.7348		Class3	0.524	
Class4	5.969	OK	Class4	4.4371	OK	Class4	4.9626	OK
Driver 1 Test image 2 (ground truth: 1)			Driver 2 Test image 2 (ground truth: 2)			Driver 3 Test image 2 (ground truth: 3)		
Class 1	1.309		Class 1	3.1135		Class 1	2.3383	
Class 2	3.9824		Class 2	0.778		Class 2	3.8475	
Class3	4.2566		Class3	4.4914		Class3	2.1232	
Class4	5.9447	OK	Class4	4.3241	OK	Class4	5.2668	OK
Driver 1 Test image 3 (ground truth: 1)			Driver 2 Test image 3 (ground truth: 2)			Driver 3 Test image 3 (ground truth: 3)		
Class 1	1.2337		Class 1	3.0956		Class 1	2.2045	
Class 2	4.0102		Class 2	1.2475		Class 2	3.7372	
Class3	4.3019		Class3	4.4251		Class3	2.3235	
Class4	5.9428	OK	Class4	4.0836	OK	Class4	5.1811	fails
Driver 1 Test image 4 (ground truth: 1)			Driver 2 Test image 4 (ground truth: 2)			Driver 3 Test image 4 (ground truth: 3)		
Class 1	1.1886		Class 1	3.2933		Class 1	2.2606	
Class 2	3.9427		Class 2	1.4564		Class 2	3.7417	
Class3	4.2865		Class3	4.9023		Class3	2.2739	
Class4	5.9099	OK	Class4	4.2657	OK	Class4	5.0772	fails
Driver 1 Test image 5 (ground truth: 1)			Driver 2 Test image 5 (ground truth: 2)			Driver 3 Test image 5 (ground truth: 3)		
Class 1	1.2915		Class 1	3.4055		Class 1	2.3884	
Class 2	3.8257		Class 2	1.5501		Class 2	3.8041	
Class3	4.2282		Class3	4.8726		Class3	2.1363	
Class4	5.8824	OK	Class4	3.8343	OK	Class4	4.9167	OK

Speaker Recognition Modality

For developing the speaker recognition module, 8 drivers' speech signals are included in training and testing. The Speaker/driver recognition system consists of three main blocks namely feature extraction, universal background model generation and the speaker/driver dependent model adaptation apart from testing. Feature extraction is front-end processing were distinguishable features of the speech signal are extracted and stored in a feature vector. Mel-frequency cepstral coefficients are very widely used features in speaker recognition domain. We used 19 dimension MFCC feature vectors. The universal background model (UBM) is trained using a large number of drivers' speech data (over 20 hrs of speech data) preferably other than the train and test set of drivers. The driver dependent Gaussian mixture (GMM) model is obtained by MAP adapting the UBM using driver specific feature vector files. An average of around 8 mins worth of speech data is used per driver to MAP adapt the UBM to train the driver dependent GMM. The driver dependent model will then contain only the distribution of a particular driver's speech. 3-6 mins of every driver's speech data (feature vector files) is used for testing. The data is windowed into various lengths for testing to know the best performance of the system with minimal data. Using the log-likelihood scoring these speech signals are scored against all GMM models and UBM. The

highest scores in each row in Table II give the classification result. As can be seen from Table II for full length of test data, the highlighted scores represent the highest scores for the drivers giving a correct classification rate of 100%.
The experiments were repeated for variable length of test data to obtain the minimum length of test utterance required to recognize the driver. Models were scored with 2 min, 1 min, 30sec, 10 sec, 5sec and 2 sec data. The drivers could be recognized using the speech signal with 100% accuracy for 30 sec or longer data lengths. Reducing the test data further to 10 sec, 5 sec and 2 sec length information leads the worst case accuracy dropping to 91%, 86% and 68% respectively. From these results we can draw the conclusion that 30secs of speech data is enough to recognize the driver with very good accuracy.

Table II. Speaker ID recognition test scores using full-length signals (3-6 mins)

FULL	MODELS									
LLR score for raw spkr/driver files		M1	M2	M6	M8	M10	M11	M17	M18	UBM
	1	-110.646	-128.013	-151.448	-131.678	-134.252	-138.781	-123.783	-128.72	-133.702
	2	-109.907	-96.3208	-131.896	-114.587	-114.537	-104.037	-110.249	-112.455	-108.181
	6	-148.673	-151.674	-109.55	-130.734	-146.519	-151.21	-145.097	-143.381	-140.103
	8	-121.752	-119.508	-115.728	-103.815	-112.888	-118.85	-113.64	-112.24	-115.213
	10	-161.914	-156.105	-166.587	-154.393	-124.645	-151.392	-155.965	-143.089	-149.036
	11	-174.27	-154.657	-196.47	-180.338	-167.637	-128.519	-173.834	-167.165	-161.067
	17	-136.965	-139.704	-158.256	-144.038	-140.699	-141.398	-120.451	-134.858	-139.078
	18	-194.236	-197.659	-196.553	-181.062	-182.105	-192.124	-192.782	-155.483	-186.165

CAN-Bus Based Driver Identification

Different from face recognition module, CAN-Bus includes time-varying characteristics of the driver therefore can be considered as less reliable. However, this modality is crucial for finding the nominal behavior of the particular driver and using this baseline to detect the distractions. Here, HMMs are used to model drivers right turn maneuvers. For each driver, a separate HMM is trained using only RT signals collected from that driver, however, the resultant HMMs are tested with RT maneuvers from all the drivers. The maximum log-likelihood of the results are found and correspondent HMM is tracked back to find out the identity of the driver. The cumulative results of this procedure are given in Table III.

Table III. Driver Identification Correct Classification Rates using HMMs trained by only CAN-Bus signals

	Driver 1 HMMs	Driver 2 HMMs	Driver 3 HMMs
Driver 1 RT test signals	--	83 %	69 %
Driver 2 RT test signals	30 %	--	22.2 %
Driver 3 RT test signals	89 %	100 %	--

The results from Table III should be interpreted carefully. For example when HMMs for Driver 1 are tested using Driver 2's signals only 30% of the cases were correctly identified as 'different from Driver 1', so the rejection rate was very low. On the other hand, when the same models are tested with Driver 3's signals 89% of them were correctly rejected. From this table we can see the best performance is observed when Driver 2 HMMs are tested with Driver 3 signals; 100% of them were rejected. This result is showing that drivers might have different characteristics and this can be modeled stochastically, however, they are not necessarily distinguishable in all cases. This makes CAN-Bus based module weaker than vision and audio biometrics. However, as can be seen in route maneuver recognition and distraction sections, the stochastic driver models can be used in those areas with better performance.

Fusion of Audio-Visual-CAN Bus Modalities

The fusion of the modalities can be achieved at different stages. One option is to include the feature vectors from all modalities as a single combined feature vector for that driver and then apply a classification algorithm for identification. The other more common way is to have the modalities completely separate and combine the classification results by using weight factors and belief networks. This process requires careful selection of the weights to have the leverage in overall performance of the identification system. From the individual performances of the modalities, we can say that face recognition and speaker ID systems are the best ones. Since we were not able to have satisfactory classification results from CAN-Bus modality, it is not included in the identification part.

Route/ Maneuver recognition

In order to develop the maneuver recognition system we use the same HMMs trained for each driver individually and test them with different type of maneuvers (lane change (LC) in this investigation). We observed that for Driver 1 and 3 a 100% correct classification was possible whereas for Driver 2 the HMM was not able to distinguish between the maneuvers. The results can be seen in Table IV; when the lane change maneuvers are used to test right turn HMMs, the likelihoods decreased which means system was able to reject lane changes to be classified as right turns. We demonstrate only this example between two maneuvers; however, a more extensive analysis is necessary to include more maneuvers here.

Table IV. Maneuver Recognition Sample Results for Driver 1 and 3

Driver 1 (LC maneuvers used to test RT HMMs)

-34665.8673	-35331.38023
-34834.2951	-35662.83673
-34831.945	-35524.56934
-35032.99553	-35753.50861
-34693.09733	-35433.78337
-34987.87301	-35589.16114

RT ground truth: -33396.7252, 100% recognition,

Driver 3 (LC maneuvers used to test RT HMMs)

-22846.8283	-22633.07504	-22591.839
-22888.5122	-22603.62654	-22603.627
-22884.9639	-22601.89573	-22601.896
-22953.3498	-22646.22121	-22646.221
-22919.6953	-22570.17769	-22570.178
-23033.2274	-22633.07504	-22633.075

RT ground truth: -21513.2232, 100% recognition

Distraction detection

As the maneuver recognition system, distraction detection uses the HMMs trained by neutral RT signals. Distracted RT maneuver signals (21 of them) are used to test these HMMs to see if they are able to distinguish between the neutral and distracted signals. The cumulative results are 72%, 100% and 83% correct classification of distracted signals for three drivers.

CONCLUSIONS and DISCUSSION

This probe study uses a database of eight drivers' audio, video and CAN-Bus signals to develop a preliminary driver identification and monitoring system emphasizing the need to make any driver assistance/ monitoring system driver-adaptive. Video and audio modalities are used to identify the drivers and the individual-specific HMMs are used to recognize the maneuver and detect the distraction of the driver. It is strongly believed that by using individual-based HMMs, the models of the driing behaviour can be more reliable and accurate.

Driver identification part can be used as verification if the smart keys are deployed for security purposes. Identification module is highly static in this sense, however, route recognition and distraction detection monitors the driver dynamically during the driving session and can help to reduce the accidents if it can be connected to preventive active safety systems or warning systems.

Acknowledgement

Authors would like to thank Sanjay Patil for his technical help in speaker ID part, Aravind Chandrasekaran for his help in debugging the HTK code used in speaker ID tests, and Zelam Purohit for her diligent work in transcribing the audio and CAN-Bus signals.

References:

[1] Abut, H., Hansen, J.H.L., Takeda, K. (Edts.), DSP for In-Vehicle and Mobile Systems, Springer, 2005, ISBN: 0387229787

[2] Hansen, J.H.L., Takeda, K., Abut, H. (Edts.), Advances for In-Vehicle and Mobile Systems: Challenges for International Standards, Springer, April 2007, ISBN-10: 038733503X

[3] Takeda, K., Hansen, J.H.L., Erdogan, H., Abut, H., In-Vehicle Corpus and Signal Processing for Driver Behavior: An International Partnership, Springer, in press

[4] Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A., 'Face recognition: A literature survey', ACM Computing Surveys (CSUR), Vol. 35, Issue 4, pp. 399-458, 2003.

[5] Zhou, S., Krueger, V., and Chellappa, R., 'Probabilistic recognition of human faces from video', Computer Vision and Image Understanding, Vol. 91, pp. 214-245, July-August 2003.

[6] M. Turk, A. Pentland, Eigenfaces for Recognition, Journal of Cognitive Neurosicence, Vol. 3, No. 1, 1991, pp. 71-86.

[7] Campbell, J.P, 'Speaker Recognition: A Tutorial', Proceedings of the IEEE, vol. 85, no. 9, pp. 1437-1462, September 1997.

[8] Boyraz, P., Acar, M., Kerr, D., 'Signal Modelling and Hidden Markov Models for Driving Manoeuvre Recognition and Driver Fault Diagnosis in an urban road scenario', IEEE Intelligent Vehicle Symposium, pp. 987-992, 13-15 June 2007, Istanbul, Turkey.

[9] D. Mitrovic, "Reliable Method for Events Recognition", *IEEE Trans.on Intelligent Transp. Syst.,* vol. 6, no. 2, pp. 198-205, June, 2005.

[10] K. Torkkola, S. Venkatesan, H. Liu, "Sensor Sequence Modeling for Driving", FLAIRS Conference, pp. 721-727, Clearwater Beach, Florida, USA, , 2005. [online source: DBLP, http://dblp.uni-trier.de]

[11] Erzin, E., Yemez, Y., Tekalp, M., Ercil, A., Erdogan, H., Abut, H., 'Multimodal person recognition for human-vehicle interaction', IEEE Multimedia, vol. 13, issue 2, pp. 18-31. April-June 2006 .