運転者の発話と運転行動を用いた危険な状況の検出

マルタルーカス[†]　宮島千代美[†]　伊藤　克亘[††]　武田　一哉[†]

† 名古屋大学大学院情報科学研究科
†† 法政大学
E-mail: † malta@sp.m.is.nagoya-u.ac.jp, {miyajima, takeda}@is.nagoya-u.ac.jp,
†† itou@k.hosei.ac.jp

あらまし　運転者のブレーキ操作や発話内容から，運転中の危険な状況を検出する手法について検討した．ブレーキに基づく検出では，ペダル踏力とその時間変化の 2 次元ヒストグラムを用いて，通常と分布が異なる箇所を検出した．発話に基づく検出では，危険な状況で発すると考えられる単語を音声の書き起こしテキストから検出した．CIAIR 対話音声・運転行動信号コーパスのうち，人間と対話中のテータ 438 名分に対して，人手でビデオ映像と運転行動信号を確認しながら危険なシーンのラベル付けを行った結果，計 25 箇所の危険なシーンが見つかった．これらのうち，ブレーキ信号，あるいは発話内容に何らかの異常を伴うシーンがそれぞれ 17 箇所，11 箇所存在した．ブレーキ，発話それぞれに基づいて検出を行った結果，80%の正解シーンを検出するために必要な誤検出数はブレーキで 473 シーン，発話で 33 シーンであった．また，Woz システム，音声対話システムの対話中のデータについても同様の実験を行った．

Analysis of changes in driving behavior signals for the detection of potentially hazardous situations in vehicle traffic

Lucas MALTA[†], Chiyomi MIYAJIMA[†], Katunobu ITOU[††], and Kazuya TAKEDA[†]

† Graduate School of Information Science, Nagoya University
†† Graduate School of Information Science, Hosei University
E-mail: † malta@sp.m.is.nagoya-u.ac.jp, {miyajima, takeda}@is.nagoya-u.ac.jp,
†† itou@k.hosei.ac.jp

Abstract　We introduce a method for automatically detecting potentially dangerous situations in motor vehicle traffic using driving behavior signals. Our proposed approach focuses on changes in a driver's behavior, which we detect through brake pedal operation as well as driver speech. Experiments were performed using a large multimedia driving database obtained from the CIAIR project at Nagoya University. We analyzed data from 438 drivers who interacted verbally with a human operator. In eleven of the 25 situations we hand labeled as potentially hazardous, drivers uttered expletive words to express negative feelings. In 17, sudden and intense compression of the brake pedal was observed. For the detection of 80% of these 17 scenes, the proposed method based on 2D-histograms of brake pressure and its dynamics also detected 473 false positives. As for the other eleven scenes, using our lexicographical speech feature-based method, a detection rate of 80% was achieved for 33 false alarms. We also present an analysis of data recorded while drivers interacted with a machine and a Wizard of Oz system.

1. Introduction

In recent years the concept of Intelligent Transportation Systems (ITS) has been a growing research area within the automotive industry. Various types of active safety systems, which aim at safer and more efficient transport, have been developed and evaluated mainly with computer-aided car crash simulations [1] [2]. ITS research has shown that most accidents are partly due to human factors [3]. Apparently, there is often a mismatch between the driver's skills and the complexity of potentially dangerous situations in traffic. This illustrates that there is a clear need to better understand patterns of information concerning driver reactions in hazardous circumstances. Such patterns have yet to be fully identified and exploited in order to develop more effective safety systems and intelligent interfaces. To gain insight into driver behaviors, we investigated two possible reactions during a hazardous situation, namely the sudden and intense compression of the brake pedal and use of specific words to express feelings about traffic situations.

Automatic incident detection is one of the major challenges in urban free-way operations. Previous published works have focused mainly on car crash detection systems, which do not take into account driving behavior signals [4]. To boost performance and reliability, such signals could be added to current safety systems, which currently rely only on a vehicle's status or position.

Studies into and the modeling of driving signals such as gas and brake pedal operation, velocity, and following distance play an increasingly important role in developing intelligent interactive vehicles. Driving signals have been applied to a wide range of fields. Driver individuality modeling and driver identification with Gaussian Mixture Models (GMM) was proposed in [5] and with Neural Networks in [6]. [7] presents results on a drowsy and drunken driving condition detector based on eye movements. Predictions on a vehicle's future state [8] and the cognitive modeling of drivers [9] have also been studied. These studies discovered new knowledge about driving signals and successfully proposed new paradigms and applications.

Many brake assistant systems [10] have been proposed to provide security in hazardous situations in which drivers fail to press the pedal [3]. However, slamming on the brakes and sharply turning the steering wheel are two very intuitive reactions in a dangerous traffic situation. It is natural then to focus on the sudden and intense compression of the brake pedal or the sharp turning of the steering wheel when trying to detect potentially dangerous situations in motor vehicle traffic. In this study, we use pressure on the brake pedal and its dynamics as detector inputs. When driving on a highway, suddenly pressing the brake pedal is an unsafe practice. Uttering certain words and nonverbal sounds to express negative feelings is a common reaction in dangerous traffic situations. In our method, a lexical search of driver speech transcriptions was designed to detect these types of verbal reactions.

We performed the experiments using the prerecorded multimodal driving data of more than 430 drivers taken from the Centre for Integrated Acoustic Information Research (CIAIR) [11] project. For this database, drivers were asked to perform speech tasks while driving, and the control signals (driving), video footage, the vehicle location and speech were recorded synchronously.

Table 1 Specifications of the database

Partner	Human	WOz	Machine
Recorded speech [h]	37.29	37.92	32.23
Driver/Partner breakdown [%]	40/60	39/61	21/79
Vocabulary size [words]	5,001	3,216	1,839
Mean pressure on the brake pedal [N]	19.60	19.14	19.34
Mean velocity [km/h]	22.97	23.39	22.83

In the following sections, a brief introduction of the database and driving signals is given, followed by a definition of typically dangerous situations and their labeling. We then offer descriptions of the driving signals-based and speech-based detection methods. Finally, we present the experimental results and discussion.

2. Database and Preparation

2.1 CIAIR Database

The driving data used in this work was obtained from the In-car Signal Corpus hosted by the CIAIR [11]. Multimodal information was collected in a vehicle under both driving and idling conditions. The database is composed of images and control (driving) and location signals that were recorded synchronously with speech. Drivers were asked to interact with three different dialogue systems while driving (a human operator, a machine, and a Wizard of Oz dialogue system) and perform simple speech tasks such as asking information about weather or restaurant locations. Currently, 800 subjects have been involved in the data collection, with a total recording time of over 600 hours.

In this research, only brake pedal pressure and recorded speech utterances were used. The control signals and velocity were both recorded at 1 kHz (16 bits), and further low-pass filtered and down-sampled to 100 Hz. A brief description of the parts of the CIAIR database we used is shown in Table 1.

2.2 Labeling and Data Preparation

Judging if a given traffic situation is dangerous is quite subjective. In many cases, if no collision occurs, this task becomes particularly tough. Nevertheless, when the following behaviors are observed while driving, it is more likely that something hazardous has occurred:

- ”Sudden and strong use” of the brake pedal
- ”Sharply turning” of the steering wheel
- Expletive words
- Repetition of expletive words
- Anxious facial expressions

Taking the above items as key elements, all potentially dangerous situations in the database were hand-labeled and categorized. In this research, we focused on the sudden and strong braking as well as expletive words and their repetition.

The 45 dangerous scenes that already existed in the database were labeled in the following way: the start

Table 2 Number of hand-labeled potentially dangerous scenes.

Partner	Human	WOz	Machine
Hand-labeled potentially dangerous situations	25	12	8
Scenes where sudden and strong use of the brake pedal was observed	17	12	8
Scenes where the use of expletive words was observed	11	7	2
None of the above behaviors was observed	5	0	0

point was the initial change in driver behavior, detected subjectively by watching the video or analyzing the pedal and steering signals. The end point was set when the "normal" condition, observed before the start point, returned. We included a margin of 1s before the start point and after the end point of each hand-labeled potentially hazardous situation.

Labeling a scene as potentially hazardous does not necessarily guarantee that at that moment the driver was feeling dangerous. To avoid trying to detect hazardous scenes which we could not make sure the driver was aware of, we searched within the labeling results for scenes where neither the use of brake pedal nor a verbal reaction from the driver was observed.

In some situations drivers only reacted verbally. For example, in one of our hand-labeled scenes the driver was stopped at a red light and in-car equipment distracted him from realizing both the traffic light change and other vehicles coming up fast. His only reaction was verbal. Scenes where changes in driving behavior were only detected through brake pedal operation were also identified among the hand-labeled potentially dangerous situations. Table 2 shows the breakdown of the labeling results. In five scenes, no reaction from drivers was observed. In all of them, drivers seemed not to be aware or caring about the potentially hazardous condition.

Only scenes where the use of brake pedal was observed were taken into account for the driving signals-based method. On the other hand, only scenes where a verbal reaction was observed were taken into account for the speech-based detection method.

3. Proposed Method

3.1 Driving Signal-Based Detection

A key to detecting potentially dangerous maneuvers is to evaluate the dynamic behavior of driving signals. One of the most common forms of this kind of measurement is the estimation of linear regression coefficients, which are calculated in the following way for a signal $x(n)$ with a window of length $2K$:

$$\Delta x(n) = \frac{\sum_{k=-K}^{K} kx(n+k)}{\sum_{k=-K}^{K} k^2}. \qquad (1)$$

In our calculations, we used a window shift of 10ms and length $2K$ of 800ms. In addition, we performed frame analysis of a brake pedal pressure signal and its dynamics. The same interval of time (frame) was then analyzed for these two different signals.

Two detection approaches were proposed. In the first one, frames of brake dynamics where the summation of the signal's interval overcame a threshold barrier, defined differently for each driver, received a label. A "scene" was defined as a fixed number of consecutive frames. If a label was attached to all frames in a scene, it was considered potentially dangerous. If one or more detected scenes lay completely inside the hand-labeled limits of a dangerous situation then the detection was considered valid. The frame analysis for this approach was performed with a window shift of 50ms and length of 100ms.

In the second approach, the pressure signal was used together with its dynamics. Figure 1 shows a six-second interval when the driver strongly used the brake pedal. The solid and dashed lines indicate brake pedal pressure and its dynamics, respectively. The relationship between these two signals can be fully appreciated by plotting them on a single graph, with the x-axis denoting the pressure and the y-axis denoting the dynamics. Figure 2 shows the joint plot. Each point in this graph represents a state of the system in time, which changes if we travel clockwise around the curve. The cyclic nature of the process elucidates the dynamical behavior of these two signals.

To automatically extract features of interest, a joint histogram of brake pedal pressure and its dynamics was calculated for each frame, followed by clustering performed with the LBG algorithm [12]. Two clusters for each driver were generated in order to represent the most common situations while driving, namely idling and moving without using the brake pedal. We then measured the distortion from clusters to each frame. The role of this measurement is very important, since frames with high distortion tend to represent uncommon driving conditions. Two different distortion measures were used. For the first one, we calculated the Euclidean distance from each frame to clusters ($d1$ and $d2$) and adopted as a feature the smallest of them $\min(d1, d2)$. For the second, the same two distances were calculated but we used instead, their multiplication ($d1*d2$) as feature, since it tends to be bigger when the frame is uncorrelated with both clusters. Frames whose distortion overcame a threshold barrier, defined differently for each driver, were labeled as potentially dangerous scene. Also in this approach, if one or more detected scenes lay completely inside the hand-labeled limits of a dangerous situation then the detection was considered valid.

Figure 3 shows an example of joint histogram calculated for 256 bins, correspondent to the data present in Fig. 2. Dark areas, where cycles concentrate, indicate values of brake and its dynamics that were present in most of time during the six seconds interval. The light areas indicate a movement from idling to moving con-

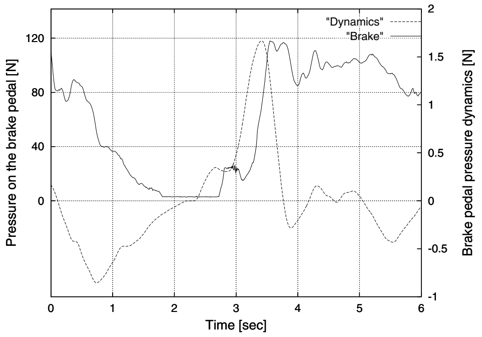

Fig. 1 Six-second interval of brake pedal pressure signal (solid line) and its dynamics (dashed line).

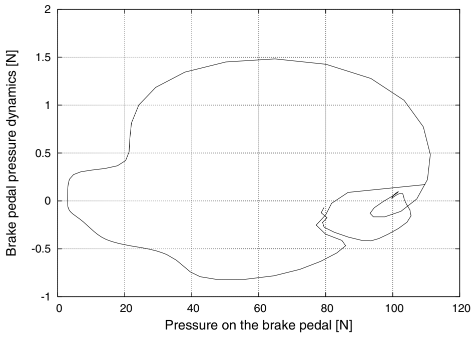

Fig. 2 Joint plot of brake pedal pressure and its dynamics (smoothed with Bezier smoothing). The state of this system is a point which travels clockwise around the curve.

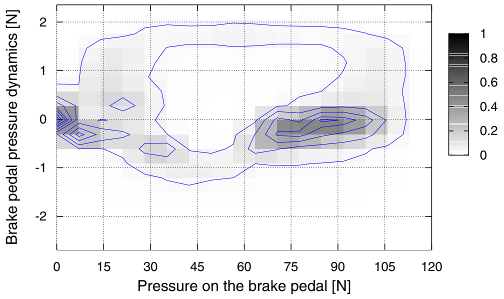

Fig. 3 Joint histogram of brake pedal pressure and its dynamics.

dition (brake and its dynamics equal zero) and then back again to the idling condition after a strong use of the pedal (see Fig. 1). Remember that the process moves clockwise.

These light areas play a fundamental role, since they tell us how the change between conditions occurred. To better represent the light areas as a feature, we applied an enhancement step before the LBG clustering stage. A normalization process, which makes the maximum value in the histogram equals one, and the following mapping comprise this step:

$$y = x^{\alpha}, \qquad (2)$$

where x is the original histogram value and y is the mapped one, used as feature for the detection. α is the degree of enhancement. Values close to one (maximum) do not considerably change after the mapping, while low amplitude regions can be greatly enhanced, depending on α. Experiments using the second approach were performed for different values of α (0.05, 0.1 and 0.2), histogram bins (256 and 1024) and non-overlapping frame (length of 2s, 4s, 6s and 8s).

3. 2 Speech-Based Detection

A transcription of driver and human operator utterances was manually annotated and labeled. Thirty keywords that might be spoken in dangerous situations were selected in advance with the help of car industry experts and a student survey. A lexical analyzer then labeled, as indicating a potentially dangerous scene, the transcription files in all places where more than two keywords were encountered within two lines, a number we defined experimentally. In 20 of the 45 situations we hand-labeled as being potentially hazardous, a verbal reaction from the driver was observed. In 11 of them, the driver was interacting with a human operator; in 7 with a Wizard of Oz system and in 2 with a machine. The effect of automatic speech recognition errors was ignored.

4. Experimental Results and Discussion

4. 1 Driving Dignal-Based Detection Results

The best result was achieved for joint histogram-based approach with 256 bins, 4s frames, $\alpha = 0.05$ and min($d1$,$d2$) as a distortion measure, where d1 and d1 are the distances from the current frame to clusters 1 and 2 respectively.

Comparatively, the second approach which utilizes only brake pressure dynamics presented a coarse result. Using data recorded while drivers interact with a human operator, a reduction from 23,423 to 4,843 in the total number of false positive scenes was observed. Figure 4 shows this result as a ROC curve, obtained by varying the threshold relative to the minimum distortion that a frame must have to be considered a dangerous scene. This threshold was adjusted individually to eliminate differences in the driving style of different drivers. Besides, using the same best parameter configuration, histogram-based detection was performed for the Wizard of Oz (527 false positives for 80% of detection and 1,577 for 100%) and machine (471 false positives for 80% of detection and 951 for 100%) data.

The driving signals-based detection relied only on the brake pedal pressure and its dynamics. Potentially hazardous situations can, however, be strongly related to vehicle speed and steering angle operation as well. For example sharp turn of the steering wheel and strong use of the brake pedal at high speed are intuitively linked with dangerousness. Such extension will be explored in future work.

4. 2 Speech-Based Detection Results

A total of 15 dangerous situations in which a verbal reaction from the driver was observed could be detected

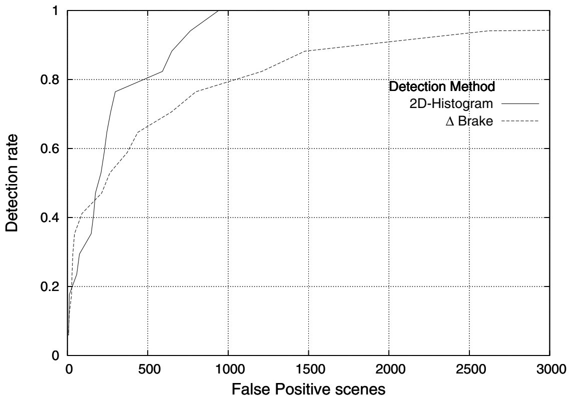

Fig. 4 ROC curve for the brake pedal dynamics-based detection (dashed line) and 2D-histogram-based detection (solid line) using data recorded while drivers interacted with a human partner.

(11 while drivers interacted with a human operator and 4 with a Wizard of Oz system). In all the other five scenes that were not detected, drivers expressed their feelings with just one keyword. By decreasing the number of keywords from 30 to 0, the ROC curve shown in Fig. 5 was obtained. The solid line indicates the detection using data recorded while drivers interacted with a human operator, and the dashed while interacting with a Wizard of Woz system. The number of false positives is lower compared to Fig. 4. If we analyze the data recorded while drivers interact with a human operator, in the detection of nine scenes (about 80%), only 33 false positives were present.

The curve in Fig. 5 shows the clear tendency of drivers to utter a group of specific words while in dangerous situations. Figure 6 shows the most common keywords divided into five groups. In Table 3 we present examples from each of them.

The use of data fusion methods to perform a combination of speech and driving signals-based detections is promising and might provide an effective reduction in the number of false positives. There are many ways of integrating these two sources of information. They will be carefully studied in future work.

4.3 Causes of Dangerous Situations

The causes of dangerous traffic situations in the 45 hand-labeled scenes are listed below:

- Driver negligence (sixteen situations)
- Unexpected behavior from other vehicles (sixteen situations)
- Errors caused due to distraction (e.g. operation of in-car equipment) (seven situations)
- Unfamiliarity with location (three situations)
- External conditions (e.g., limited vision due obstruction or sun-blindness) (three situations)

35% of the scenes labeled as dangerous, drivers failed to exercise the necessary care, such as failing to make a visual check before changing lanes. No statistical differences in the causes of dangerous situations while driver interacted with different partners were found.

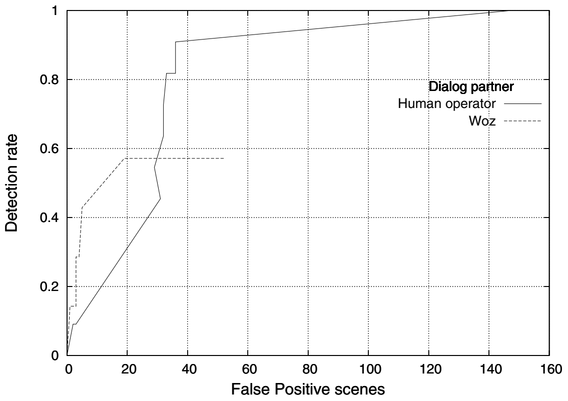

Fig. 5 ROC curve for the speech-based detection using data recorded while drivers interacted with different partners, a human operator and a Wizard of Woz system.

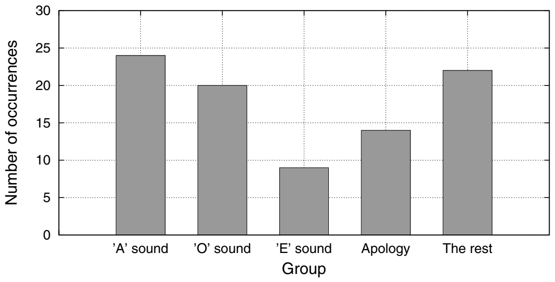

Fig. 6 Most common keywords divided into five groups.

Table 3 Examples of the most common keywords.

’A’ sound	’O’ sound	’E’ sound	Apology	The rest
あっ, あー, あらー	おい, おっ, おっと	えー, ええ, えっ	ごめん	危ない 怖い
(ahh)	(’o’ in old)	(’e’ in get)	(sorry...)	(damn it)

5. Discussion

In this work, we presented and discussed a new approach for detecting potentially hazardous situations in vehicle traffic. In our approach, driving behavior signals, namely pressure on the brake pedal and speech utterances, were used to detect a chain of changes in driver status and to retrieve incidents from a large real-world driving database. Although many accident databases relate collision incidents to a small set of maneuvers, a hazardous situation is often due multiple factors that we have yet to properly identify and model. When we can complete such modeling, it will be possible to evaluate the existing safety systems and devise more intelligent ones.

In this work, two types of detection methods were proposed. The first one was based on automatic detection of the sudden and strong use of the brake pedal. We have shown how brake pedal pressure and its dynamics can be used together to reduce the number of false alarms. In order to perform a more efficient detection, we still need to discover and extract a better feature to represent driving behavior. Vehicle velocity and steering angle operation are promising features and will be evaluated. Studies on driver recognition have shown that spectral analysis using "cepstrum" is very useful for representing driving behaviors. We will also consider this analysis in a future work.

Speech-based detection showed a better performance. However, 100% of detection rate was not achieved, since scenes where drivers uttered less than 2 keywords were not considered valid. Besides, in the Japanese language, there is a strong tendency to use nonverbal sounds from the "A sound group", which makes the number of false positives increase drastically. These drawbacks suggest that an analysis based not only on lexicographical speech features, but also on acoustic correlates of expletive words might help decreasing false positives and increasing generalization and performance of our method. The combination of driving signals and speech for classification will also be explored. The fusion of different sources of information is still an open question, but satisfactory results are often obtained.

It also is tempting, but difficult to compare the detection results for different speech task data. A careful analysis of the trajectory where drivers traveled while performing such tasks has to made before making any comparisons between them. We observed, however, the presence of specific course locations where dangerous situations happened more frequently. Searching for these locations is a promising application of our research. These trajectory-related questions will be analyzed in future work. We also observed a tendency that intuitively more dangerous situations need less false positives in order to be detected when compared to less dangerous ones, which suggests that a rank of the hand-labeled dangerous scenes based on dangerousness would help analyzing trends in our results.

Since dangerous scenes do not have clear boundaries (even the taggers who hand-labeled the database were often confused at figuring out if a scene was dangerous or not) we also need to devise and explore a method which deals with vague boundaries. Data sparsity is also a significant problem in dangerous scene detection. Generating hazardous situation in practice is not a simple task, so any detection would suffer from lack of necessary patterns in the learning stage. However, new data is being collected and other types of driver information, such as heart beat and eye gaze information will soon be available.

In conclusion, there is still a lot of information to be discovered and analyzed. A final detector would be multimodal, taking into account all associations, anomalies, and statistically significant structures in driving behavior data. Knowledge from different areas such as pattern recognition, signal processing, image understanding, and computer vision would be gathered to perform the nontrivial extraction of potentially useful implicit information.

References

[1] R. Labayrade, C. Royere, and D. Aubert, "A colision mitigation system using laser scanner and stereovision fusion and its assessment," *Proceedings of the IEEE Intelligent Vehicles Symposium*, pp. 441–446, 2005.

[2] Y. Sugimoto and C. Sauer, "Effectiveness estimation method for advanced driver assistance system and its application to collision mitigation brake system," *Proceedings of the 19th International Technical Conference on the Enhanced Safety of Vehicles, Washington DC, United States*, 2005.

[3] *Final Report of the eSafety Working Group on Road Safety.* Information Society Technologies and European Commission, November, 2002.

[4] A. Hoess, *Multifunctional automotive radar network.* European Conference on Intelligent Road Vehicles, Clermont-Ferrand, France, June, 2001.

[5] T. Wakita, K. Ozawa, *et al.*, "Driver identification using driving behavior signals," *IEICE Trans. Information and Systems*, vol. E88-D, Number 3, pp. 1188–1194, 2006.

[6] H. Erdogan, A. Ozyagci, T. Eskil, M. Rodoper, A. Ercil, and H. Abut, "Experiments on decision fusion for driver recognition," *Biennial on DSP for in-vehicle and mobile systems, Sesimbra Portugal*, September, 2005.

[7] P. Smith, M. Shah, and N. da V. Lobo, "Monitoring head/eye motion for driver alertness with one camera," *Proceedings of the ICPR*, vol. 4, pp. 636–642, Sept. 2000.

[8] N. Oliver and A. Pentland, "Driver behavior recognition and prediction in a smartcar," *Proc. SPIE Aerosense 200, Enhanced and Synthetic Vision*, April, 2000.

[9] D. Salvucci, E. Boer, and A. Liu, "Towards an integrated model of driver behavior in a cognitive architecture," *Transportation Research Record, 1779*, pp. 9–16, 2001.

[10] S. Tokoro, K. Kuroda, A. Kawakubo, K. Fujita, and H. Fujinami, "Electronically scanned millimeter-wave radar for pre-crash safety and adaptive cruise control system," *Proceedings of the IEEE Intelligent Vehicles Symposium*, pp. 9–11, 2003.

[11] N. Kawaguchi, K. Takeda, F. Itakura, *et al.*, "Construction and analysis of the multi-layered in-car spoken dialogue corpus," *DSP in Vehicular and Mobile Systems*, 2003.

[12] A. Gersho and R. Gray, "Quantization and signal compression," *Norwell, Massachusetts: Kluwer Academic Publishers*, 1992.