

A Study of Driver Behavior Under Potential Threats in Vehicle Traffic

Lucas Malta, *Student Member, IEEE*, Chiyomi Miyajima, *Member, IEEE*, and Kazuya Takeda, *Member, IEEE*

Abstract—Although, in recent years, significant developments have been made in road safety, traffic statistics indicate that we still need significant improvements in the field. Since traffic accidents usually reflect human factors, in this paper, we focus on clarifying the understanding of driver behaviors under hazardous scenarios. Brake pedal signals or driver speech, or both, are utilized to detect incidents from a real-world driving database of 373 drivers. Results are then analyzed to address the individuality in driver behaviors, the multimodality of driver reactions, and the detection of potentially dangerous locations. All of the existing 25 potentially hazardous scenes in the database are hand labeled and categorized. Based on the joint histograms of behavioral signals and their time derivatives, a detection feature is proposed and satisfactorily applied to the indication of anomalies in driving behavior. Seventeen scenes, where a reaction utilizing the brake pedal was observed, are detected with a true positive (TP) rate of 100% and a false positive (FP) rate of 4.1%. We demonstrate the relevance of considering behavior individuality. During 11 scenes, the drivers verbally reacted. Scenes that included high-energy words are adequately detected by the speech-based method, which achieved a TP rate of 54% for an FP rate of 6.4%. The integration of different behavior modalities satisfactorily boosts the detection of the most subjectively hazardous situations, which suggests the importance of considering multimodal reactions. Finally, a strong relationship is presented between locations where potentially hazardous situations occurred and areas of frequent strong braking.

Index Terms—Driver behavior, human factor, multimedia databases, safety systems.

I. INTRODUCTION

OVER THE last decade, experts from academia and industry have actively been involved in road safety. Efforts that promote safer vehicle traffic have mainly been concentrated in two areas: accident prevention and injury reduction. For example, the preventive approach anticipates future problems concerning infrastructure [1] or the lack of driver awareness [2]. On the other hand, injury-reduction measures have primarily focused on enhancing vehicular safety systems [3], [4]. Although encouraging transportation improvements have been made, the number of road fatalities remains unacceptably high. Since the responsibility for almost three quarters of all traffic

accidents falls on human shoulders [5], better understanding of driver behavior is a decisive step toward safer and more efficient driving.

In this paper, we propose a method for detecting such situations from a large real-world driving database to increase the understanding of driver behaviors during potential threats. Results addressed the individuality in the behavior of drivers, the multimodality of their reactions to potentially hazardous conditions, and the detection of potentially dangerous locations.

Recently, a few vehicular systems have stressed models that consider driver behavior individuality [6], [7]. Nevertheless, each individual driver is likely to perceive traffic conditions differently and behave based on such subjective judgments.

In addition, to accurately interpret different driving contexts, safety systems have to recognize the multimodal nature of driver behaviors. During threatening situations, not only do we need to know how drivers use the pedals, but we also need to know that their speech and gestures may change. Currently, the interpretation of hazardous situations mostly relies on a single reaction modality [8], [9].

In addition, when investigating dangerous situations, interesting correlations between driver behaviors and external factors may surface. In this paper, we focus on detecting potentially hazardous locations and found evidence that such locations can be detected without a large amount of traffic accident statistics. Section VIII presents our approach to the problem.

Results from the proposed detection of potentially hazardous situations from a database can be utilized to improve the understanding of driver behavior as well as for proactively promoting safety. Data recorded from vehicles equipped with such detection capabilities and a Global Positioning System (GPS) could be effective, for example, when frequently updating a map of potentially hazardous sites or a list of reckless drivers. In addition, detection results can also be utilized to increase the hand-labeling speed and to create a database of real-world potentially hazardous situations. Currently, a large number of experiments concerning safety are performed with simulations or with a small amount of real-world data.

Our implemented method utilized multimedia driving behavior signals, namely, force on the brake pedal or speech or both, to perform detection tasks. Hand labeling and categorization of potentially hazardous scenes in the database were also done and presented in Section II. Brake pedal and speech-based detection methods are presented in Sections III and VI, respectively. The integration of these two features is addressed in Section VII. We then offer a discussion in Section IX.

Manuscript received October 16, 2007; revised June 10, 2008 and November 6, 2008. First published April 17, 2009; current version published June 2, 2009. The Associate Editor for this paper was Y. Liu.

The authors are with the Graduate School of Information Science, Nagoya University, Nagoya 464-8603, Japan (e-mail: malta@sp.m.is.nagoya-u.ac.jp; miyajima@is.nagoya-u.ac.jp; kazuya.takeda@nagoya-u.jp).

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.2009.2018321

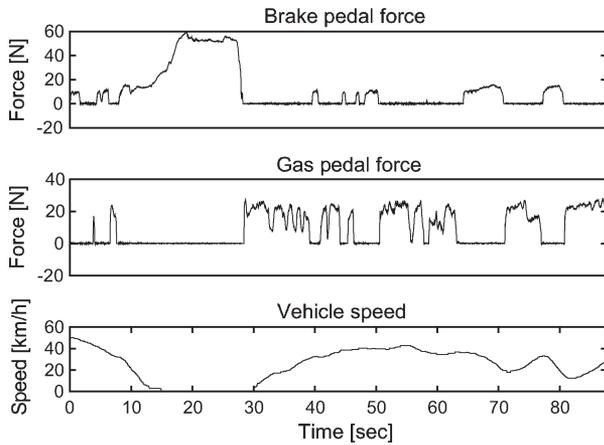


Fig. 1. Examples of driving behavior signals.

II. DATABASE AND PREPARATION

A. CIAIR Driving Corpus

The driving database utilized in this paper was obtained from the Center for Integrated Acoustic Information Research (CIAIR) [10], Nagoya University. This database is unique because a large amount of data were recorded in real-world driving conditions.

Multimodal information was collected in a vehicle under both driving and idling conditions. The database is composed of images, driving behavior, and location signals synchronously recorded with speech. Fig. 1 shows examples of recorded driving behavior signals. Drivers interacted with a human operator and performed such simple speech tasks as asking for weather information or restaurant locations while driving on a public road. For this study, we utilized brake pedal force, speed, and speech signals from 373 drivers (34 h) recorded from November 2000 to March 2002. Of the drivers, 67% were men, and 33% were women. They were, on average, 29 years old (range 20–65 years) and had held a driver's license for a mean period of ten years (range 1–40 years).

Brake pedal force was recorded using the force transducer LPR-B-05KNS1 produced by Kyowa Electronic Instruments Company, Ltd. Speech was recorded using a Sony ECM 77B microphone. Both sensors are commercially available. Only the vehicle velocity sensor was customer made. Recently, a growing number of vehicles have been built with an onboard communication protocol called Controller Area Network (CAN), which makes it much easier to acquire driving signals. Navigation systems with build-in microphones are also becoming increasingly common, facilitating speech recording. Therefore, the basic idea of the proposed method can be applicable to most commercial vehicles.

B. Hand Labeling and Ranking of Potentially Hazardous Scenes

The 25 potentially hazardous scenes that already existed in the database were subjectively selected by taggers from 34 h of video footage and audio with the multimedia data viewer



Fig. 2. Multimedia data browser utilized by taggers for subjective selection and ranking of potentially hazardous scenes from the database. In-car videos are shown at the top, and the speech waveform is shown at the bottom.

shown in Fig. 2. The in-car videos taken from two different viewpoints are shown at the top, whereas the speech waveform is shown at the bottom. Graduate students who had had driver's licenses for more than two years served as volunteer taggers. Since they had no technical skills concerning traffic incidents, the total number of hand-labeled potentially hazardous scenes may have been affected. The data were divided into five groups, and five taggers performed the labeling tasks. Results were then shown to two more taggers to validate the labels.

Of the 24 drivers involved in potentially hazardous scenes, 64% were men and 36% were women. They were, on average, 30 years old (range 21–65 years) and had held a driver's license for a mean period of nine years (range 2–28 years). Only one driver was involved in two different scenes—the other 23 drivers were involved in just one. Forty percent of the potentially hazardous situations occurred at signalized intersections, and 24% occurred at intersections without any traffic control. The weather condition was good during all 25 scenes, varying from cloudy to sunny. The mean duration of hazardous scenes was 8 s.

Although in all the 25 hand-labeled scenes drivers noticed the circumstances that might become hazardous, the risks were considered acceptable in five scenes, and no substantial reactions were observed. Drivers verbally expressed negative feelings in 11 of the selected 25 situations. In 17 of these 25 situations, a reaction utilizing the brake pedal was observed. In eight situations, both reactions were present. In three situations, only a verbal reaction was verified.

In addition, the subjective level of risk was ranked. Six taggers watched all 25 potentially hazardous scenes one time. Then, the scenes were shown again. Taggers were asked to assess risk levels based on video footage and audio and instructed to ignore driver reactions and to only concentrate on the traffic environment. Even if a driver's reaction to a certain hazardous situation was particularly strong, when the scene was subjectively a low hazard, a "low-level risk" label was given to it.

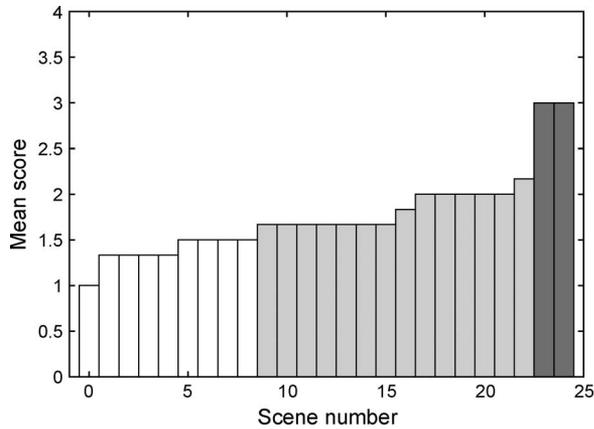


Fig. 3. Mean scores of subjective level of risk for each scene. White, light gray, and dark gray indicate scenes in the low-, medium-, and high-level risk groups, respectively.

Subjective risk was defined as any motion by some other road user, which could possibly develop into a hazard, and for which the driver had to be particularly prepared for taking some evasive action in terms of braking or steering. Therefore, even if the driver avoided a traffic incident by using the brake pedal or the steering wheel, this situation could still be ranked as having a high level of risk, depending on the driving context, for example, presence of pedestrians, speed, and distance among vehicles.

A value was assigned to each level (1 for low, 2 for medium, and 3 for high), and then, the mean score of each scene was calculated. The 25 mean scores were then divided into three groups by utilizing the K-means algorithm [11]. The following were the ranking results: The low-level risk group had nine scenes, the medium-level risk group had 14 scenes, and the high-level risk group had two scenes. Fig. 3 shows the distribution of the mean scores.

III. BRAKE PEDAL FORCE-BASED DETECTION METHOD

This section describes the detection of situations where drivers extraordinarily pushed the brake during a potentially hazardous situation. Since this maneuver is characterized by a strong and sudden use of the pedal, a feature that considered not only compression intensity but also its time derivative was necessary. In this paper, we introduced a statistical representation of the dynamics of brake pedal operation. The detection process utilizing the proposed feature is schematically described in Fig. 4. In the following sections, the signal processing in each block is explained in detail.

A. Stochastic Representation of Brake Pedal Operation

Fig. 5 shows a 6-s interval when the driver strongly applied the brake around 3.2 s. The solid and dotted lines represent the brake pedal force $x(n)$ and its time derivative $\dot{x}(n)$, respectively. Points A and C indicate idling conditions, whereas B indicates the start of moving forward. As shown in the following, for example, the time derivative of a discrete-time signal

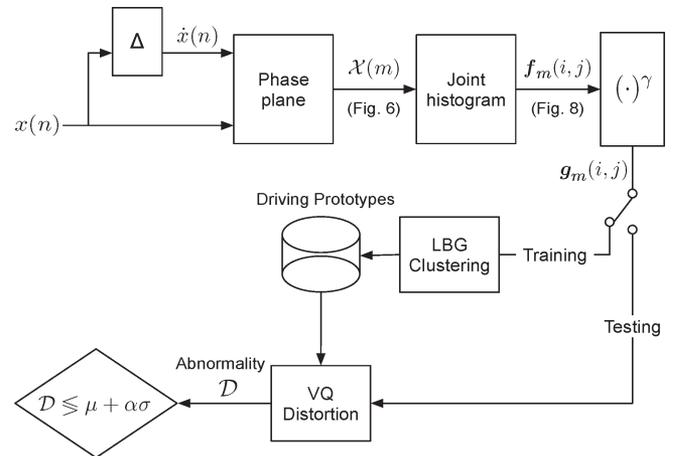


Fig. 4. Block diagram of the detection process.

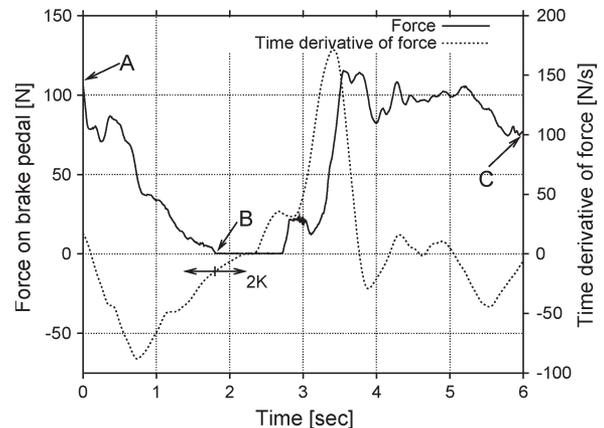


Fig. 5. (Solid line) Six-second interval of brake pedal force signal and (dotted line) its time derivative. The driver extraordinarily compressed the brake pedal around 3.2 s. Points A and C indicate idling conditions, whereas B indicates moving forward. The left right arrow indicates a window $2K$ used to calculate the time derivative.

can be calculated by utilizing linear regression coefficients for signal $x(n)$ with a window of length $2K$:

$$\dot{x}(n) = \frac{\sum_{k=-K}^K k \cdot x(n+k)}{\sum_{k=-K}^K k^2}. \quad (1)$$

Applying (1) causes an essential delay in deciding whether the situation is hazardous or not, since it uses both past and future values to obtain the current derivative; however, this delay is negligible. In our experiments, we used a window of $2K = 800$ ms, as indicated in Fig. 5 by a left right arrow.

The relationship between the two signals illustrated in Fig. 5 can fully be appreciated by plotting them on a single graph, i.e., a phase plane, with the x -axis denoting the force and the y -axis denoting its time derivative. This plot allows us to geometrically interpret the dynamical behavior of brake pedal compression. The dynamics of the m th time period is given by

$$\mathcal{X}(m) = [(x(mT+1), \dot{x}(mT+1)), (x(mT+2), \dot{x}(mT+2)) \dots, (x(mT+T), \dot{x}(mT+T))]. \quad (2)$$

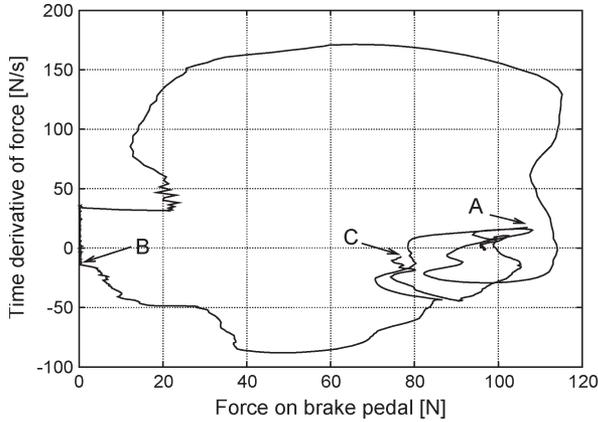


Fig. 6. Joint plot of a 6-s interval of brake pedal force and its time derivative. Each point in this plot represents a different state in time. Points A and C indicate idling conditions, whereas B indicates moving forward.

Fig. 6 shows the dynamics $\mathcal{X}(m)$ on a phase plane. Each point in this graph represents a temporal state of the system in which a temporal development corresponds to a clockwise movement around the curve. Points A, B, and C indicate the same time instants as those shown in Fig. 5. The start of the cyclic process is point A, which is an idling condition. The driver then released the brake and reached the leftmost point B—forward motion had started. Subsequent strong braking again brought the vehicle to an idling condition at point C. The cyclic nature of the process elucidates its dynamical behavior. After scalar quantization of x and \dot{x} , a joint histogram of the brake pedal force and its time derivative $f_m(i, j)$ was calculated for each time period T . A joint histogram is simply a 2-D mapping that counts the number of (x, \dot{x}) observations that fall into disjoint categories, which is known as bins. We define $f_m(i, j)$ as

$$f_m(i, j) = \sum_{k=1}^T \delta_{ij}(x(mT + k), \dot{x}(mT + k))$$

where

$$\delta_{ij}(x, \dot{x}) = \begin{cases} 1, & \theta_i \leq x < \theta_{i+1} \text{ and } \phi_j \leq \dot{x} < \phi_{j+1} \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

θ and ϕ define a 2-D grid of bins represented in Fig. 7. The amplitude of a given bin is the number of (x, \dot{x}) observations that fall into it. The cycle presented in Fig. 6, together with its joint histogram, are shown in Fig. 8. We used this joint histogram as a fundamental stochastic representation of brake-pedal operation dynamics.

B. Histogram Smoothing

The dark areas in Fig. 8, where cycles are concentrated, indicate the values of the brake pedal force and its time derivative that were present most of the time. The light areas indicate the transition from idling to moving and then back to idling again after a strong use of the pedal—the process moves clockwise. Although the relative frequency is low, the transition plays an important role because it explains how changes occurred.

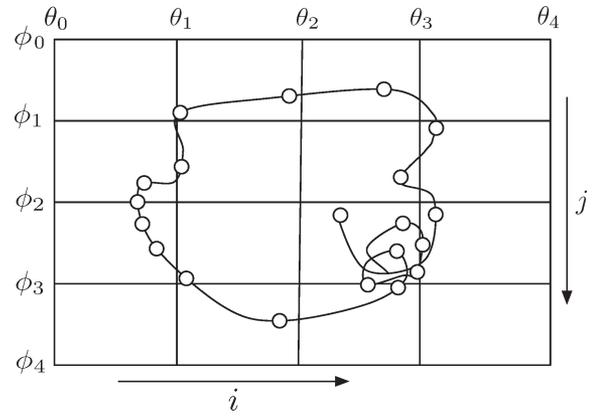


Fig. 7. Two-dimensional grid of bins. The amplitude of a given bin is defined as the number of (x, \dot{x}) observations that fall into it.

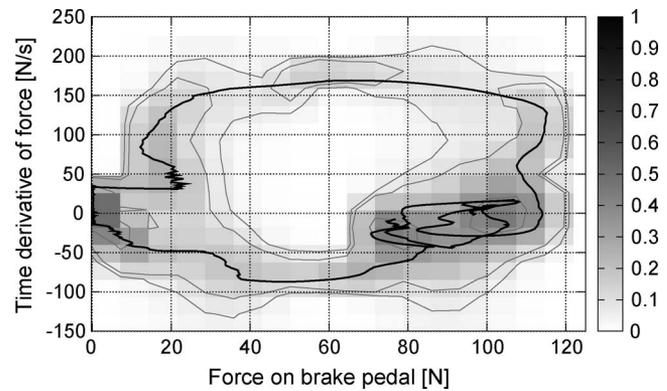


Fig. 8. Joint plot of a 6-s interval of brake pedal force and its time derivative, indicated by a black line, and its joint histogram. Darker areas indicate values presented most of the time.

Therefore, the following enhancement step was proposed to emphasize the difference in low-frequency phenomena. First, the histogram is normalized so that its maximum frequency is 1. Then, the relative frequency is squashed by a nonlinear function

$$g_m(i, j) = f_m(i, j)^\gamma, \quad 0 < \gamma \leq 1. \quad (4)$$

γ is the degree of enhancement. The values in the joint histogram close to 1, which is the maximum, change very little after mapping, whereas the low-amplitude regions are greatly enhanced, depending on γ . The effect of enhancement can be visualized in Fig. 9. This figure shows a line of the joint histogram shown in Fig. 8 for $\gamma = 0.05$ and 0.5.

C. Modeling Normal Driving

An analysis of the phase plane in the long run, as shown in Fig. 10 for a 5-min interval, reveals areas of data concentration. One area with a time derivative of force around 0 N/s and force ranging from approximately 10 to 80 N indicates idling. Another region around (0 N, 0 N/s) represents normal forward moving. These were the normal driving circumstances during most of the 5 min.

In this paper, normal driving was modeled by a set of joint histograms, each of which represents typical driving operation. The Linde–Buzo–Gray (LBG) algorithm [12]

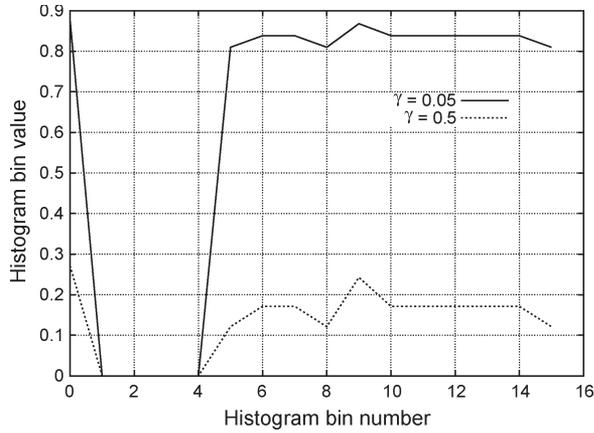


Fig. 9. Line of the joint histogram shown in Fig. 8 for different values of γ .

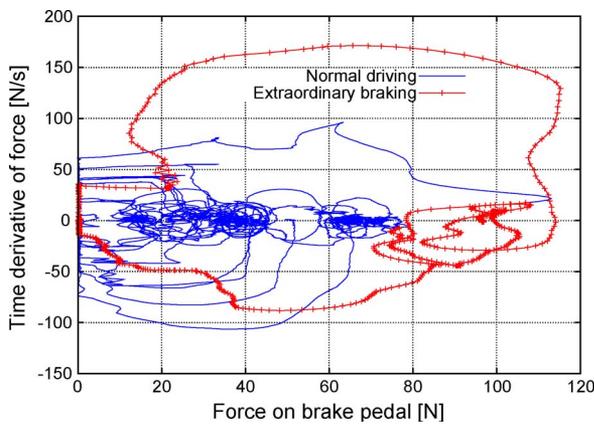


Fig. 10. Joint plot of a 5-min interval of brake pedal force and its time derivative. Line with crosses represents a 6-s interval of extraordinary braking.

was chosen to find an optimal set of prototype histograms through clustering, since it is less sensitive to initial parameters. The LBG algorithm reads a sequence of M vectors v_1, v_2, \dots, v_M and generates a codebook $C_E = \{c_1, c_2, \dots, c_E\}$ containing E codewords. The generation of the codebook can be stated as follows.

Stage 1) When an initial codebook C_1 is not assigned, the initial codeword is obtained from the whole collection of training data as follows:

$$c_1 = \frac{1}{M} \sum_{m=1}^M v_m. \quad (5)$$

Stage 2) After obtaining the first codebook, the algorithm can be described as a finite sequence of steps. At each step, a new codebook, with a total distortion of less or equal to the previous codebook, is created. Each step can be divided into two parts: splitting and optimization.

Stage 3) At a generic step k , the codewords obtained from the previous iteration are split (by adding and subtracting a small random value) to form two new codewords.

Stage 4) During the optimization, the present codebook is used to quantize the training vectors, and the mean

Euclidean distance \mathcal{E} between every training vector and the corresponding codeword is calculated. If the following condition is valid:

$$\left| \frac{\mathcal{E}_{k-1} - \mathcal{E}_k}{\mathcal{E}_k} \right| < \mathcal{E} \quad (6)$$

where \mathcal{E} is a predefined ending condition, then the algorithm goes to Stage 6). If it is not valid, then it goes to Stage 5).

Stage 5) Centroids are recalculated using the results obtained in Stage 4), and the codebook is updated. The algorithm then goes back to Stage 4).

Stage 6) If the desired number of codewords E was achieved, then the algorithm returns. Otherwise, it goes to Stage 3).

Although the number of prototype histograms is important, we still do not have any consistent method for determining this number. The modeling of normal driving can be understood as the training stage of the detection process.

D. VQ Distortion

Scrutinization of Fig. 10 also reveals a signal anomaly. The line with cross symbols over it represents the 6-s interval of Fig. 6, i.e., when the driver slammed on his brakes. To detect such data that deviate from normal driving conditions, VQ distortion \mathcal{D} from the normal driving model—the distance from the nearest histogram prototype—was used as an abnormality, i.e.,

$$\mathcal{D}(m) = \min_p \{d(\mathbf{g}_m, \mathbf{g}_p)\}, \quad 1 \leq p \leq P \quad (7)$$

where P is the number of prototypes. \mathbf{g}_p and \mathbf{g}_m represent one prototype and the current frame histograms, respectively. Frobenius norm $\|\cdot\|_F$ was used as the distance measure between two joint histograms \mathbf{g}_m and \mathbf{g}_p , which are represented as $I \times J$ matrices

$$d(\mathbf{g}_m, \mathbf{g}_p) = \frac{1}{I \times J} \|\mathbf{g}_m - \mathbf{g}_p\|_F^2. \quad (8)$$

The sparsity of potentially hazardous scenes in the database was a significant limitation, which makes it difficult to model them or to apply discriminative analysis such as support vector machines (SVMs) [13]. Accordingly, only normal driving conditions were modeled.

E. Detection

Depending on the threshold selection, driver-dependent and -independent scenarios can be implemented for detection. In the driver-dependent scenario, the mean (μ) and the standard deviation (σ) of abnormalities \mathcal{D} are individually calculated, whereas in the driver-independent scenario, μ and σ are calculated using data from all drivers. The detection is then set to

$$\mathcal{D} \lesssim \mu + \alpha\sigma \quad (9)$$

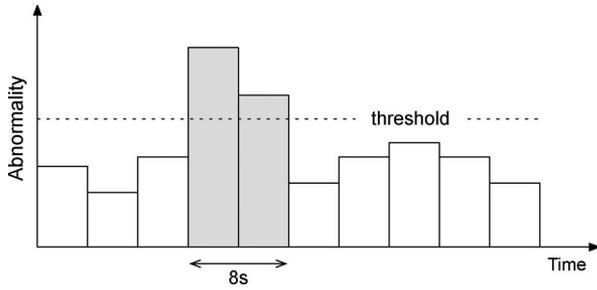


Fig. 11. Evaluation process. The 8-s interval that follows detection is labeled as a single potentially hazardous situation. Eight seconds is the mean duration of hand labels.

where α is identical among all drivers. Values of D must be computed in advance so that μ and σ can be calculated.

In addition, to avoid detecting strong brake pedal compressions while idling, the vehicle velocity was also considered to be an input. Frame intervals whose mean velocity did not overcome a certain threshold were labeled not hazardous (velocity threshold).

IV. EXPERIMENTAL EVALUATION

An experiment evaluated the detection method applied to brake-pedal operation.

A. Evaluation Data

We trained individual normal driving models using all the data of each driver. To increase the amount of training data, we also proposed a model trained with data from all drivers together. This model was identical among all drivers. All of the data of each driver were utilized for testing.

B. Evaluation Framework

When abnormality D overcame the threshold barrier, the next 8 s was considered one potentially hazardous scene; therefore, even if multiple hits inside this interval were observed, only one valid detection of a potentially hazardous scene was counted. An 8-s interval, which is chosen based on the hand-labeled results, was the mean duration of potentially hazardous situations. Fig. 11 illustrates the evaluation process. When detection was observed inside the hand-labeled limits of a potentially hazardous scene, a true positive (TP) detection was counted. Results are presented utilizing receiver operating characteristic (ROC) graphs [14]. The following definitions were also utilized to display ROC graphs.

- 1) Total positives: The number of hand-labeled potentially hazardous scenes in the test data. When evaluating the brake pedal-based detection method, the number of total positives was set to 17, which is the number of scenes where a reaction utilizing the brake pedal was observed.
- 2) Total negatives: The number of 8-s frames inside the test data minus the number of 8-s frames inside the hand-labeled hazardous situations.

Experiments for the brake-pedal force-based method were performed for different values of enhancement γ (0.05, 0.1,

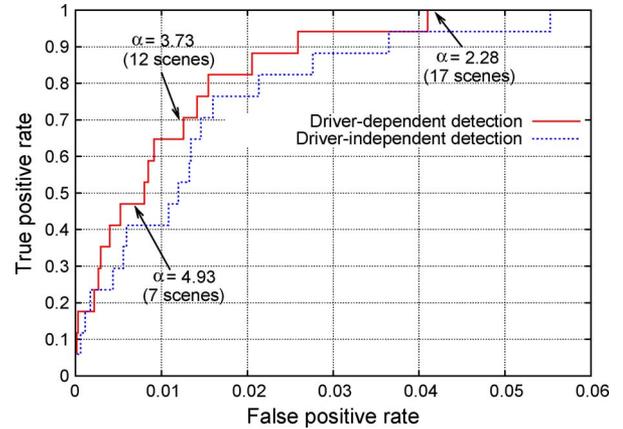


Fig. 12. (Solid line) Best results for brake pedal force-based detection utilizing driver-dependent and (dotted line) driver-independent approaches. Values of α are presented for the driver-dependent detection of seven, 12, and 17 scenes.

and 0.2), histogram bins (12×12 , 16×16 , and 24×24), number of prototype histograms in the normal driving model (2, 4, and 8), frame length (2, 4, and 8 s), frame shift (1, 2, 4, and 8 s), and velocity threshold (0–8 km/h). A delta feature window of 800 ms was set for all the experiments, since this is the most effective value for driver modeling using driving behavior signals [7].

The best parameter configuration was achieved by first selecting initial parameter values and possible ranges based on practice. We then changed one parameter at a time—keeping the other fixed—in the following order: frame length and shift, number of prototype histograms, histogram bins, γ , and velocity threshold. At each step, we selected the optimal value for the parameter being changed. Although we can only guarantee that the achieved configuration is a local maximum, we still do not have a consistent method for efficiently searching throughout the entire parameter space.

V. BRAKE-PEDAL FORCE-BASED DETECTION RESULTS

The best result for the brake-pedal force-based method, which achieved fewer false positive (FP) detections, was obtained with 16×16 bins, two prototypes, frame length and shift of 4.0 s, and enhancement $\gamma = 0.05$. The optimal velocity threshold was 4 km/h.

The model of normal driving that is trained with data from all drivers together and detected with a driver-dependent threshold attained the best performance. The worst performance was verified when both model and threshold were identical among all drivers, which suggests the importance of considering individuality in reactions. Both best and worst results are shown in Fig. 12 as ROC graphs, which were obtained by varying the detection threshold. Values of α that were used in the driver-dependent detection of seven, 12, and 17 scenes were also presented in Fig. 12. Fig. 13 shows the influence of enhancement γ on the best result. Emphasizing low-frequency signals effectively improves the detection performance.

VI. SPEECH-BASED DETECTION METHOD AND RESULTS

The possibilities are varied for the multimodal analysis of driver reactions during hazardous circumstances. For example,

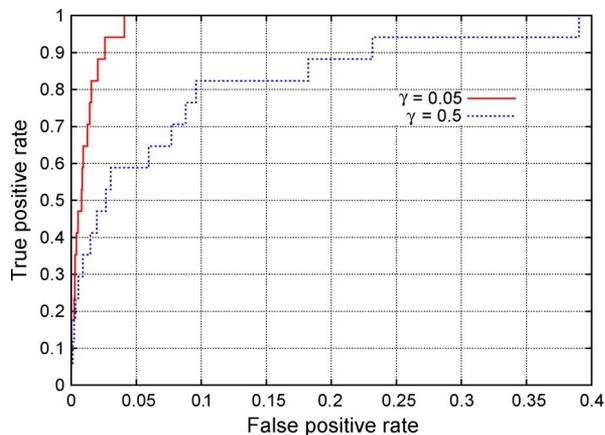


Fig. 13. Effect of γ in best result. $\gamma = 0.05$ is optimal value.

slamming on the brakes and sharply turning the steering wheel are intuitive responses in a dangerous traffic situation. However, under certain conditions, such as driving on the highway, suddenly pressing the brake pedal or rapid steering wheel movement are unsafe practices. Therefore, in this scenario, natural reactions such as uttering words or nonverbal sounds to express negative feelings about an adverse condition are relevant facets for analysis.

The analysis of hand-labeled potentially hazardous situations stressed the advantages of using speech as a feature in this paper. Scenes in which sharp turning of the steering wheel was observed also involved strong brake pedal use. The same was not true for verbal responses. Therefore, since more correlation between braking and steering was observed, we adopted speech as an additional feature.

This method followed analogous detection and the evaluation approaches, as did the brake pedal-based approach. Sudden and high-energy speech utterances presented similar anomalous characteristics as sudden and strong braking; therefore, joint histograms of speech energy and its time derivative were utilized as features, along with the LBG algorithm as a clustering strategy. Moreover, energy was computed as the log of the signal energy, that is, for speech samples $\{s(n), n = 1, \dots, N\}$, we have

$$Energy = \log \sum_{n=1}^N s(n)^2. \quad (10)$$

Experiments for this method were performed by changing one parameter at a time and keeping the others fixed. Different values of enhancement γ (0.1, 0.5, and 1.0), histogram bins (7×7 , 8×8 , and 12×12), number of prototype histograms modeling normal utterances (2, 4, 8, and 16), frame length (1 and 2 s), and frame shift (0.5 and 1.0 s) were utilized. A delta feature window of 960 ms and a driver-dependent threshold were set for all the experiments.

For speech-based detection, the best result, which achieved fewer FPs, was obtained with 8×8 bins, four prototypes, a frame length of 1.0 s, a shift of 0.5 s, and enhancement $\gamma = 0.5$. The model of normal speech trained with individual data attained the best performance. Since this method focused on verifying the detection of the 11 hand-labeled scenes in

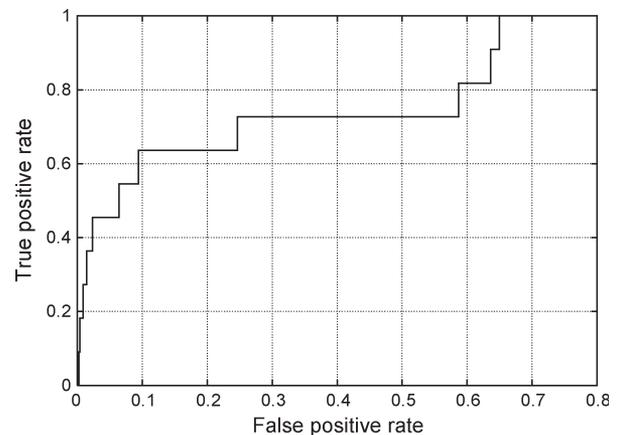


Fig. 14. Best speech-based detection results.

which verbal reactions were observed, the number of *total positives* was set to 11. Results are shown in Fig. 14 as ROC graphs, which were obtained by varying the detection threshold.

VII. INTEGRATION-BASED DETECTION METHOD AND RESULTS

This section detects potentially hazardous situations by integrating two different sources of information: brake-pedal force and speech. One possible strategy is to combine these two pieces of information at the feature level by constructing a large feature vector. The problem with feature-level fusion is that, during a hazardous situation, different behavioral changes do not necessarily occur concurrently. Analysis of hand-labeled potentially hazardous situations indicated that they are more likely to occur at different timings. Accordingly, the two modalities were integrated at the abnormality level, as described in the following.

Inside an 8-s window, the abnormalities of brake and speech signals \hat{D}_B and \hat{D}_S , respectively, were integrated by utilizing

$$\beta \hat{D}_B + (1 - \beta) \hat{D}_S. \quad (11)$$

\hat{D}_B and \hat{D}_S represent the local maxima of abnormality inside the 8-s window, which was calculated after zero-mean normalization. The parameter β was called the fusion factor and ranged from $0 \leq \beta \leq 1$. A zero-mean normalization of distances was required to have a parameter β that equally favors both decision methods when set to 0.5. Fig. 15 illustrates the integration process.

Experiments were performed by utilizing data that were divided into three groups based on subjective levels of risk. Consequently, the relationship between driver reactions and subjective levels of risk could be verified. In addition, different values of β were utilized: 0.00, 0.25, 0.50, 0.75, and 1.00. The following parameters were set to optimal values for each single modal detection: γ , number of histogram bins, number of prototype histograms modeling normality, and frame length and shift. The 8-s window was concurrently shifted.

The best results were achieved for the high-level risk group with $\beta = 0.25$, for the medium-level risk group with $\beta = 0.75$, and for the low-level risk group with $\beta = 1.00$, that is, only the brake pedal was used. The second best detection method for

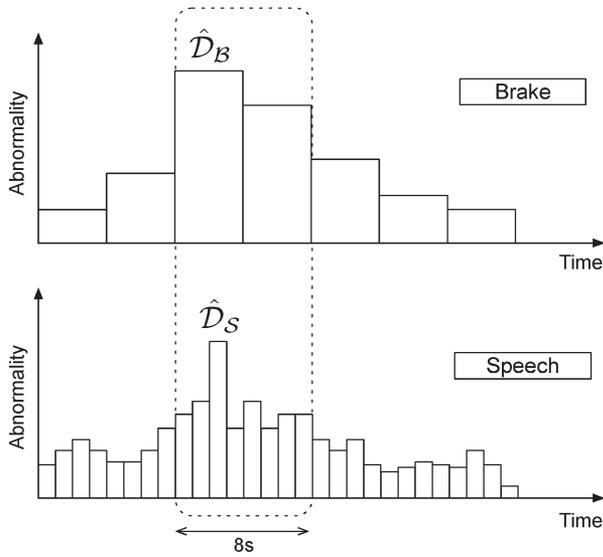


Fig. 15. Integration-based detection method. Highest abnormalities inside an 8-s window are added using a weighted sum.

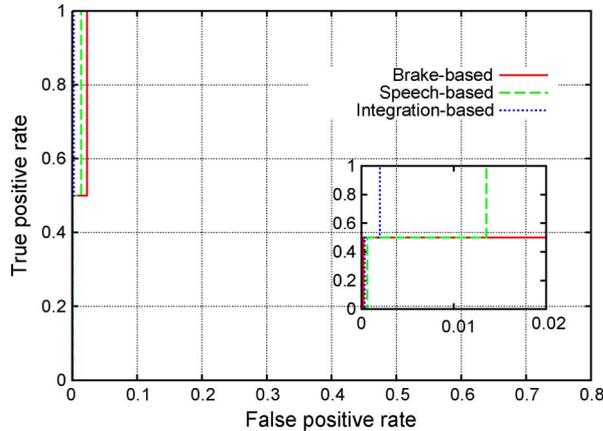


Fig. 16. Brake-pedal force, speech, and best integration-based ($\beta = 0.25$) results for the high-level risk group.

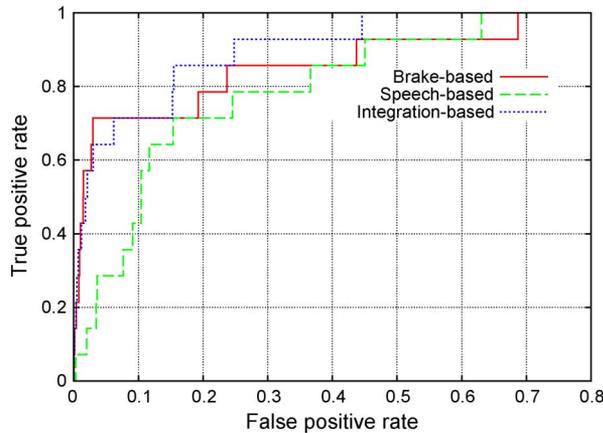


Fig. 17. Brake-pedal force, speech, and best integration-based ($\beta = 0.75$) results for the medium-level risk group.

the low-level risk group was the integration based on $\beta = 0.75$. The corresponding ROC graphs, which were obtained by varying the detection threshold, are shown in Figs. 16–18 for the high-, medium-, and low-level risk groups, respectively. The

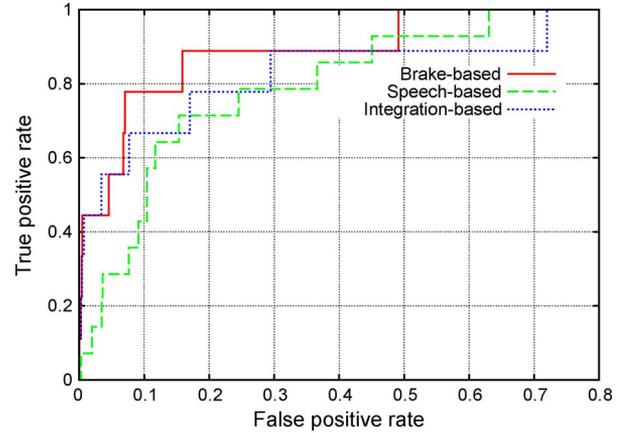


Fig. 18. Brake-pedal force, speech, and best integration-based ($\beta = 0.75$) results for the low-level risk group.

results for $\beta = 1.00$ (brake), $\beta = 0.00$ (speech), and best $0.00 < \beta < 1.00$ (integration) are displayed. When scenes in the medium-level risk group were detected using the integration-based method, the FP rate for 100% detection decreased to 33% when compared with the brake-based approach and was 27% when compared with the speech-based approach. These decreases in FP rate for scenes in the high-level risk group were 95% and 92%, respectively. The combination of different modalities boosted the detection of the most subjectively hazardous scenes.

Only 25 potentially hazardous situations were available, which makes it difficult to quantitatively discuss about generality and accuracy of results. Nevertheless, since a large database of 373 drivers was used in the experiments, we believe that the proposed method would also be effective in a real scenario.

VIII. AUTOMATIC DETECTION OF POTENTIALLY DANGEROUS LOCATIONS

Using new technologies alone does not achieve much safer urban traffic. Other measures, including infrastructure improvements and enforcing safety measures, are also extremely important to overall road safety and must be implemented without delay. In this context, the automatic detection of potentially dangerous locations plays a fundamental role in accelerating the process of improvement, saving time, and reducing costs. The detection and characterization of dangerous locations have mainly been done with traffic accident statistics [1], [15]. An inevitable drawback is that accidents must occur before safety measures can be considered.

Different traffic environments have different effects on drivers, which suggests that locations where potentially hazardous situations frequently occur may also have a distinguishing impact on driver behavior. Fig. 19 shows hand-labeled potentially hazardous situations plotted as small squares over routes followed by subjects. A concentration of hazardous events at certain sites, which were marked as A, B, C, and D, was observed. Particularly, intersections at sites B and D received two labels each. After adjusting the brake pedal force-based detection threshold, the first 100 FP detections were

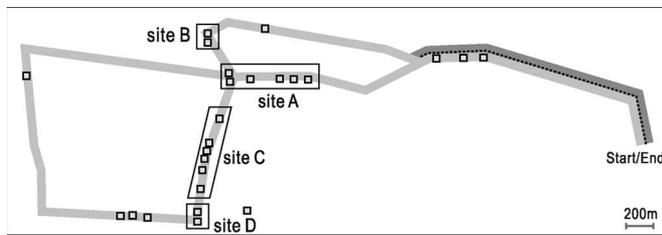


Fig. 19. Location of hand-labeled hazardous situations marked as squares over routes, indicated by a thick line, followed by subjects. Labels A–D show areas where these situations are concentrated. Dashed line indicates the only road where subjects drove twice (in different directions).

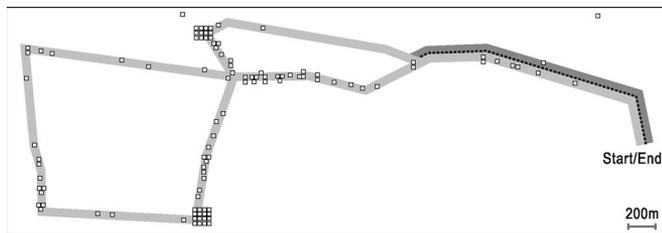


Fig. 20. Location of first 100 FPs marked as squares over the route, indicated by a thick line, followed by subjects. Dashed line indicates the only road where subjects drove twice (in different directions).

selected, and their locations were plotted over the route map, as shown in Fig. 20.

The comparison between these two figures provides an intuitive yet revealing relationship between locations where sudden and strong braking was frequently observed and areas where potentially hazardous situations occurred. The following is a preliminary characterization of sites A–D:

- 1) Area A: a high-speed dense-traffic environment;
- 2) Area B: a dense-traffic intersection where drivers had to perform a right turn;
- 3) Area C: a narrow two-way street with intense pedestrian and commercial vehicle traffic on which parking is also allowed, which frequently disturbs vehicle trajectories;
- 4) Area D: an intersection where drivers are frequently forced to drastically slow down to turn left onto a much narrower street (site C) or stop at a red light before turning.

All four selected areas have characteristics that may occasionally lead to higher risks. This topic deserves more extensive field observation and further in-depth investigation, because the results are promising and give valuable insight into the influence of external factors on driver reactions.

IX. DISCUSSION

In this paper, we proposed a method to detect potentially hazardous situations from a large real-world database and utilized the results to provide a clearer understanding of driver behavior during potential threats. Multimodal driving signals were obtained from 373 drivers.

The results demonstrated that the brake pedal force-based method attained satisfactory results: a TP rate of 100% for an FP rate of 4.1%. This method focused on verifying the detection of the 17 hand-labeled scenes in which reactions utilizing the

brake pedal were observed. Results from this paper also indicated that future advancements in this area should consider the uniqueness of driving behavior signals for improved retrieval of hazardous situations and to develop other safety systems. An assistant system, which correctly interprets driver responses, would be far more efficient and interactive than the current systems.

In 11 of the 25 hand-labeled scenes, the drivers verbally reacted, which can broadly be divided into two groups: high-energy words and whispered speech. This division was clear in the results. To detect these 11 situations, the speech-based method obtained a TP rate of 54% (six scenes) for an FP rate of 6.4%. However, to detect the remaining five scenes characterized by a low-energy verbal reaction and to achieve a TP rate of 100%, an FP rate of 65% was observed. Whispered speech could not adequately be detected, because it was, most of the time, mistaken for silence, which is an ordinary condition. The detection of whispered responses caused a high increase in the FP rate. Pitch, formant, and timing-related features, which effectively detect emotion from speech and whispers, must be considered for more efficient detection in future research. Moreover, expletives uttered during hazardous situations are often difficult to recognize, which suggests that a low-speech recognition rate is also a promising feature.

Our integration-based detection method improved the retrieval results for the high- and medium-level risk groups, suggesting the advantages of a multimodal interpretation of driver reactions. The low-level risk group was detected better with brake pedal force-based detection. Scenes in the high-level risk group were far more easily detected than the others. During these scenes, drivers reacted with particularly high-energy speech, which accounted for a comparatively low optimal $\beta = 0.25$. Vigorous verbal reactions were also present in scenes ranked as having a medium-level risk; however, since, in most of these situations, drivers only reacted by utilizing the brake, a $\beta = 0.75$ was found to be optimal. The responses in the low-level risk group were mainly characterized by only the brake pedal or low-energy speech.

In five of the 25 hand-labeled potentially hazardous scenes, no substantial reactions from drivers were verified, and in three scenes, the drivers only reacted verbally. These situations were divided between low- and medium-level risk groups, which increased the number of FPs for a TP rate of 100%. Using new features concerning vehicle surroundings, physiological information, and more efficient speech-based detection is a necessary facet to provide both more accurate hand labeling and retrieval of all situations. Other reaction modalities not addressed in this paper, such as driver gestures, facial expressions, and steering angle, also need to be considered in future work. The proposed detection method is very scalable; therefore, it can be applicable to other modalities as well.

In addition, in this paper, a strong relationship was observed between areas where potentially hazardous situations occurred and locations where brakes were frequently strongly compressed. Since this finding provides genuine insight into the topic, it deserves further in-depth investigation.

This paper focused on a better understanding of driver reactions during potentially hazardous situations. We found

evidence indicating that further analysis of driving behavior signal processing must consider the individuality of driver reactions and the integration of multimodal responses to hazards. Nevertheless, additional methodological improvements are required, along with a larger number of potentially hazardous scenes. Hence, in this context, future research will be able to better characterize dangerous situations in vehicle urban traffic. This paper's findings provide a realistic understanding of driver responses to potentially hazardous conditions and can mainly be utilized to improve systems that proactively promote safety.

REFERENCES

- [1] Road Bureau, Ministry of Land, Infrastructure and Transport, *The Next Stage for Roads in Japan*, 2007. [Online]. Available: mlit.go.jp/road/road_e/index_e.html
- [2] P. I. J. Wouters and J. M. J. Bos, "Traffic accident reduction by monitoring driver behaviour with in-car data recorders," *Accident Anal. Prev.*, vol. 32, no. 5, pp. 643–650, Sep. 2000.
- [3] Y. Sugimoto and C. Sauer, "Effectiveness estimation method for advanced driver assistance system and its application to collision mitigation brake system," in *Proc. 19th Int. Tech. Conf. Enhanced Safety Vehicles*, 2005, pp. 1–8. no. 05-0148-O.
- [4] R. Labayrade, C. Royere, and D. Aubert, "A collision mitigation system using laser scanner and stereovision fusion and its assessment," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2005, pp. 441–446.
- [5] A. Vits and I. Hodac, *Final Report of the eSafety Working Group on Road Safety*, Nov. 2002, Inform. Soc. Technol. Eur. Comm.
- [6] S. Tezuka, H. Soma, and K. Tanifuji, "A study of driver behavior inference model at time of lane change using Bayesian networks," in *Proc. IEEE ICIT*, Dec. 2006, pp. 2308–2313.
- [7] 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.
- [8] F. Sagberg and T. Bjornskau, "Hazard perception and driving experience among novice drivers," *Accident Anal. Prev.*, vol. 38, no. 2, pp. 407–414, Mar. 2006.
- [9] M. Munenori, U. Tetsuya, and T. Masaki, "Development of drive recorder (OBVIOUS recorder)," *Fujitsu Ten Tech. J.*, vol. 27, pp. 3–9, Jul. 2006.
- [10] N. Kawaguchi, K. Takeda, and F. Itakura, "Multimedia corpus of in-car speech communication," *J. VLSI Signal Process.*, vol. 36, no. 2/3, pp. 153–159, Feb./Mar. 2004. [Online]. Available: <http://dx.doi.org/10.1023/B:VLSI.0000015094.60008.dc>
- [11] J. B. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. Math. Stat. Probability*, 1967, pp. 281–297.
- [12] A. Gersho and R. Gray, *Vector Quantization and Signal Compression*. New York: Springer-Verlag, Nov. 1991.
- [13] V. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1996.
- [14] T. Fawcett, *ROC Graphs: Notes and Practical Considerations for Data Mining Researchers*. Palo Alto, CA: Intell. Enterprise Technol. Lab., HP Lab., Jan. 2003.
- [15] P. Greibe, "Accident prediction models for urban roads," *Accident Anal. Prev.*, vol. 35, no. 2, pp. 273–285, Mar. 2003.



Lucas Malta (S'08) received the B.E. degree in electrical engineering in 2004 from the Federal University of Minas Gerais, Belo Horizonte, Brazil, and the M.E. degree in information science in 2007 from Nagoya University, Nagoya, Japan, where he is currently working toward the Ph.D. degree.

His research focuses on driver behavior, human-centered driver assistance, and speech processing.



Chiyomi Miyajima (M'06) received the B.E. degree in computer science and the M.E. and Doctorate 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. She is currently 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.



Kazuya Takeda (M'97) received the B.E., M.E., and Doctor of engineering degrees from Nagoya University, Nagoya, Japan, in 1983, 1985, and 1994, respectively.

In 1986, he was with the Advanced Telecommunication Research Laboratories (ATR), Kyoto, Japan, where he was involved in two major projects: speech database construction and speech synthesis system development. In 1989, he was with KDD R&D Laboratories, Japan, where he participated in the construction of voice-activated telephone extension systems. Since 1995, he has been with Nagoya University. His research interest covers a wide range of digital signal processing applications such as acoustic, speech, music, and human behavior signals.

Dr. Takeda is the Editor-in-Chief of the *IEICE Transactions on Information and Systems*. He is a Board Member of the Acoustical Society of Japan.