Research Report

Risk Evaluation while Driving by Using Hazard Information

Hiroyuki Konishi, Mitsuteru Kokubun, Kazunori Higuchi, Tetsuo Kurahashi, Yoshiyuki Umemura

Abstract

If the number of road traffic accidents is to be reduced, it is essential that drivers be able to accurately assess the risks presented by their surroundings. This research aimed to develop a model that would be capable of estimating the risks presented by a scene depicting an actual driving situation. We manually input hazard information such as the other cars and pedestrians appearing in a scene, and then used the accumulated data to devise a multiple regression formula to estimate the risk. In addition to the hazard information, we devised a multiple regression formula that also considers whether a vehicle is in an intersection, as well as the speed of the vehicle. We asked a team of driving instructors to evaluate the risks, and used their evaluations as standard risk values. Using 96 variables in the multiple regression formula, we obtained a correlation coefficient of 0.973. For the hazard information, we found that the coefficients for other vehicles and elderly pedestrians were given approximately the same weighting, while a parking vehicle was afforded about twice that.

Keywords Road traffic accidents, Risk perception, Hazard, Multiple regression model

1. Introduction

Although devices such as ABS, brake assist, and airbags have been developed to reduce the number of road accidents, or at least reduce their impact, in 2003, the number of people dying in road accidents in Japan had reached 7,702, while the number injured was 1,181,431.

Almost 90% of accidents are said to be caused by "human error". If we break the basic driving skill down into the stages of "recognition", "judgment", and then "operation," we find that most accidents are caused by a mistake at the "judgment" stage.[1)] Putting this another way, the driver's "beliefs" come into play. These beliefs give rise to what we call "assumption driving". More precisely, we define this as "a driver's subjective evaluation of the risks involved in a situation on the road being lower than the objective risks".[2)] To evaluate whether "a given driving situation is dangerous," we need to know what is in front of the vehicle and where it is and, based on the results, decide whether the situation is dangerous. To date, however, no studies have attempted a quantitative study of the situation in front of a vehicle. This is because manual measurement would be too great a task, and automatic recognition has so far been unable to provide sufficiently accurate results.

If driver can get correct risk information or change their biased risk perception, we should control our driving behavior, decreasing speed, showing indicator, and so on. It is difficult to obtain the objective risk value.

In the field of traffic psychology,[3)] "risk" is defined as "the likelihood of an accident occurring or the uncertainty of an accident occurring". A "hazard" is defined as a situation, phenomenon, or factor that a driver must face and which increases the possibility of an accident occurring. More specifically, hazards include intersections and curves, as well as traffic participants such as vehicles and pedestrians. The process by which we recognize such hazards is known as "hazard perception".

As yet, however, clear definitions of the different types of hazard have not been set. Similarly, there are no means of evaluating the degree by which the existence of a hazard, including the type and location of that hazard, increases the risk.

In this study, to collect hazard information, we manually measured the positions of objects in the video image of a scene shot through the windshield of a moving vehicle. Then, using the collected hazard information, we attempted to estimate the different levels of risk encountered while driving. As shown in **Table 1**, we divided the hazard information into two categories (Mobile objects, Signs). In categorizing the hazard information, we drew on the "Hazard Perception Training" used in driver education.[4)]

To estimate the risk, we devised a multiple regression model, illustrated in **Fig. 1**. Because it is not possible to extract the degree of risk from any given scene, we asked several driving school instructors to view the scene and evaluate the degree of danger (the risk) in a scene, and then used that value as a standard for estimating the risk from the hazard information recorded on the video.

2. Outline of experiments

2. 1 Video for experiments

Using a video that had previously been shot through the windshield of a vehicle while traveling through a city (nine scenes, five minutes), we

Table 1 Example of hazard objects measured on images.

Mobile objects		Sign
Vehicle	Pedestrian	
Car	Aged	Traffic sign
Motorcycle	Middle-aged	
Bicycle	Child	

manually assigned and input attributes (parked, moving, etc.) to the vehicles, pedestrians, and bicycles contained within each video frame (3 frames/s, **Fig. 2**(a)). As a result, we were able to perform an analysis using a total of 4574 recorded hazards.

2. 2 Acquisition of risk values

To obtain a numerical value to express to risk presented by the driving environment, we asked two driving school instructors to evaluate the video. The two instructors observed nine recorded driving scenes, shot through the windshield of a vehicle as described above, and then assigned a value of 0 to 10 to each scene to indicate the risk presented by that situation (**Fig. 3**). The correlation coefficient, *r*, between the values provided by the two instructors averaged 0.82. We set the averages of the values provided by the two instructors as the "instructor-estimated degree of danger".

2. 3 Extraction hazard information

For a scene like that shown in the photograph part of Fig. 2(a), we measured the positions and sizes of the hazards and then, for the analysis shown below,

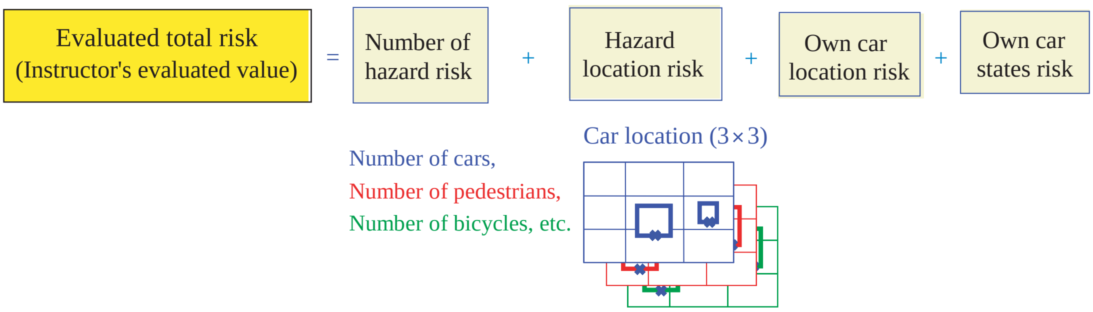

Fig. 1 Schematic view of risk evaluation.

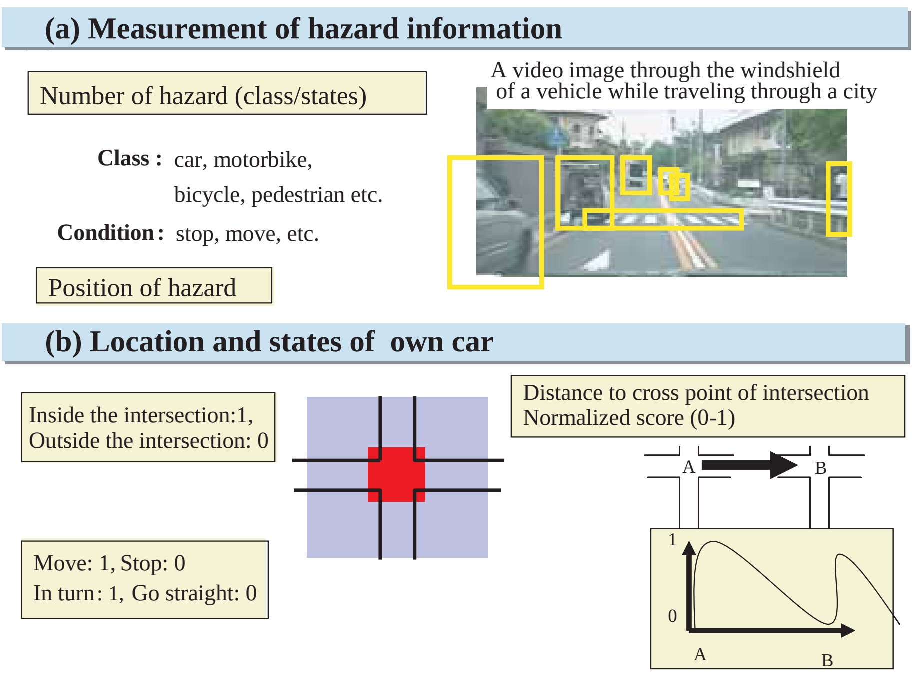

Fig. 2 Valuables for risk evaluation.

applied the variables listed in **Table 2**. As shown in **Fig. 4**, the number of position data item is assigned to nine variables for each object.

3. Result and discussion

3. 1 Evaluation of risk for both operation and state

Figure 5 shows the averages of the risk evaluations, as made by the driving instructors, for four combinations of two orientations of the steering wheel (going straight on and turning) and two movement states (stopping and moving). Two of these states, namely, the angle of the steering wheel and the speed of the vehicle, were measured using onboard sensors. Obviously, the evaluated risk was lower while the vehicle was stopped than when it was in motion, and while going straight ahead than when turning.

Based on this data, we can say that when moving with the vehicle going straight with a large dispersion, states other than the steering wheel states and movement states are significant to estimating the risk.

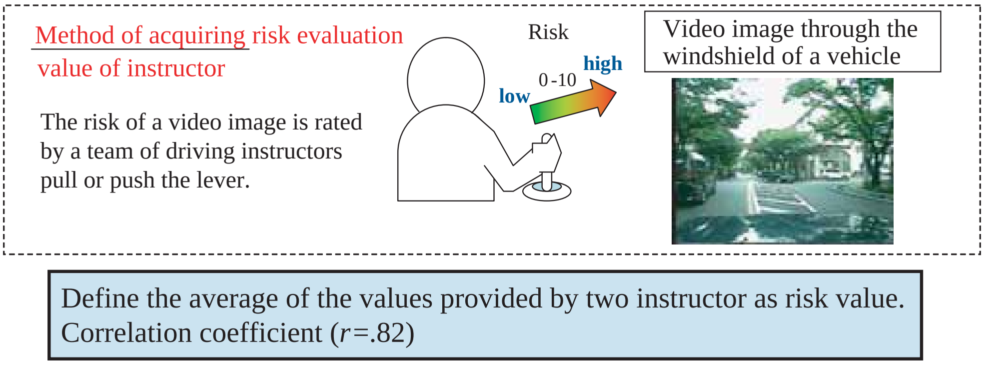

Fig. 3 Method of acquiring risk evaluation value of instructors.

Table 2 Variables for model.

Dependent variable				Instructor's risk value (0-10)			
Independent variable	Own car location			Distance to intersection (m)	Inside intersection (0, 1)		
	Operation and state		Brake and gas pedal	Speed (km/h)	Move (0, 1)	Stop (0, 1)	
			Steer		In turn (0, 1)		
	Hazard	Mobile objects	Automobile	Number of moving cars	Number of moving trucks		
				Number of stopping cars	Number of stopping trucks		
			Two-wheel vehicle	Number of motorbikes	Number of bicycles		
			Pedestrians	Number of pedestrians	Number of aged people	Number of middle aged people	Number of children
		Sign	Traffic sign	Number of pedestrian road	Number of stop sign		

3. 2 Multiple regression analysis

3. 2. 1 Multiple regression analysis using 16 variables

Figure 6 shows the results of multiple regression analysis using the 16 variables listed in **Table 3**. The correlation coefficient, r, varies greatly between scenes, from 0.1 for scene 3 to 0.87 for scene 5. To describe and predict differences in the environment for each scene, we have to eliminate the differences in the predictability between scenes. To achieve this, we must make the states and prediction model more accurate. Table 3 lists the weightings of the variables created using the multiple regression model shown in Fig. 6.

3. 2. 2 Multiple regression analysis using 96 variables

Figure 7 shows the results of performing a multiple regression analysis of the risk using the 96 variables (listed in **Tables 4**(a) and (b)) (r = 0.979). Of these 96 variables, 16 were the same as those in Fig. 6 and Table 3, another 71 (listed in Table 4(b)) were obtained by digitizing the speed of the vehicle while the video was being recorded, and the likes of the positions of other cars in the video, and the remaining 9 were scene type numbers (Table 4(a)). To obtain the data used in the multiple regression analysis, we used all nine of the recorded

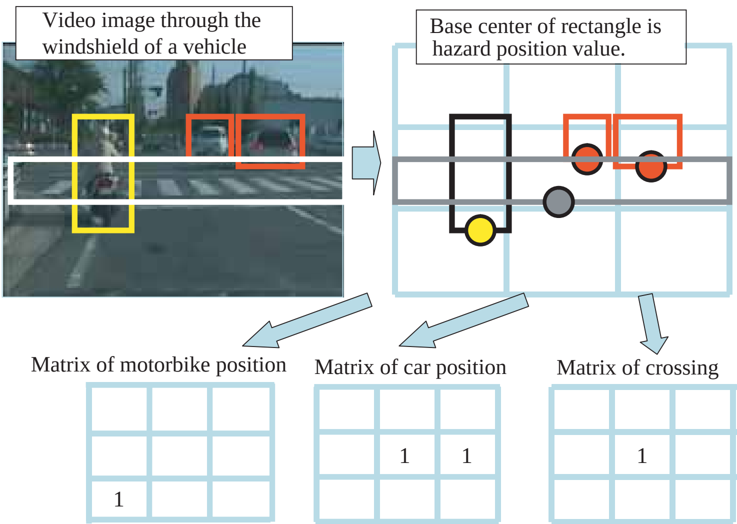

Fig. 4 Example of hazard position data.

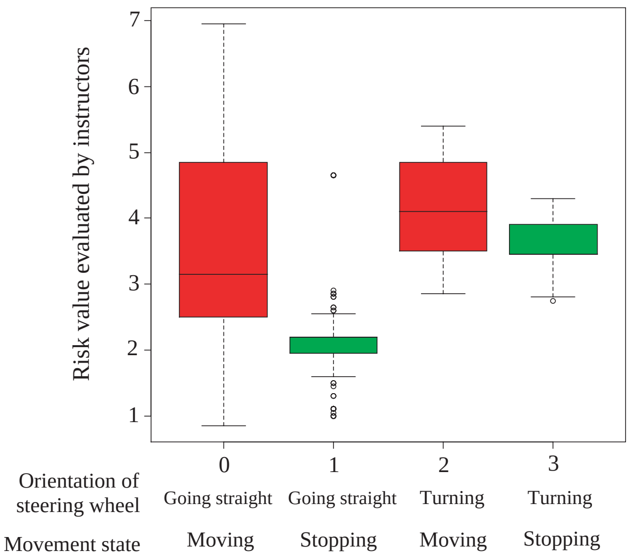

Fig. 5 Mean and standard deviation of risk value.

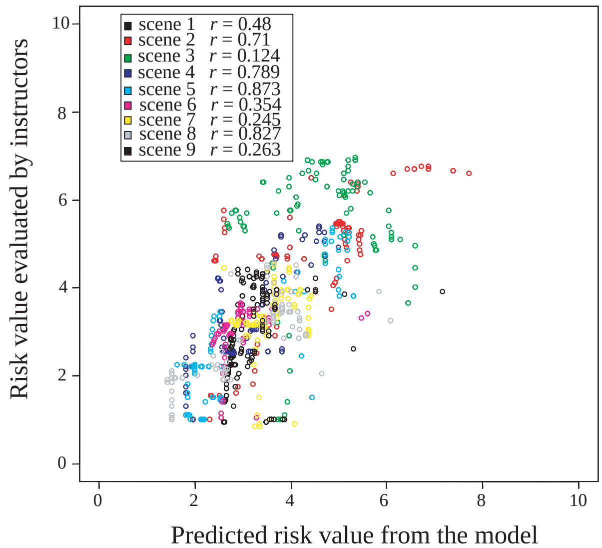

Fig. 6 Multiple regression model using 16 variables (r =.933). The correlation coefficient of nine scenes in the legend.

scenes.

If we look at Table 4(a), we find that the "Number of stopping cars (standardized partial regression coefficient = 0.65)," is given double the risk weighting coefficient of "Number of moving cars (standardized partial coefficient of regression = 0.27)". In addition, if we look at the "Number of aged pedestrians (standardized partial regression coefficient = 0.32)" and the "Number of children (standardized partial coefficient of regression = 0.27)", we find that pedestrians are afforded a similar risk weighting to other vehicles. If, however, we consider the weighting applied to a given location, we find that the risk weighting afforded to "Number of stopping cars: position" is always negative, which goes against what common sense would tell us.

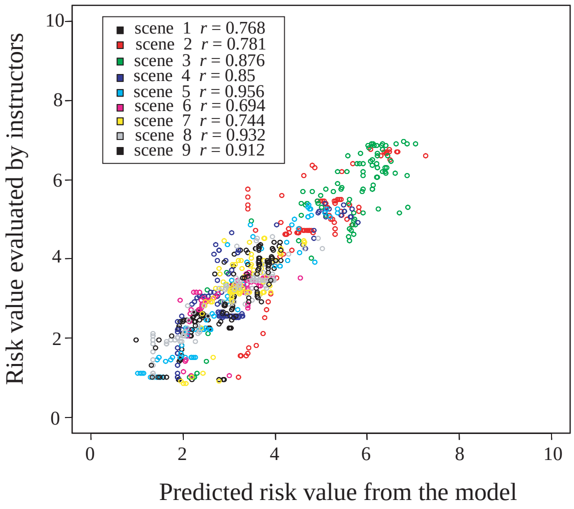

Fig. 7 Multiple regression model using 96 variables (r =.979). The correlation coefficient of nine scenes in the legend.

3. 3 Comparison with multiple regression model using AIC

So far, we have performed multiple regression analysis using either 16 or 96 variables. To investigate the quality of the model, we performed a comparison using the Akaike's Information Criteria (**Table 5**). As a result, we found that the multiple

Table 3 Multiple regression analysis using 16 variables. Gray background shows significant variables.

	Estimate	Std. error	t value	Pr (>|t|)	Standardized partial regression coefficient
Inside the intersection (0,1)	0.174	0.160	1.087	0.277	0.484
Move (0,1)	2.584	0.142	18.221	< 2e-16	0.470
Stop (0,1)	2.044	0.163	12.511	< 2e-16	0.473
In turn (0,1)	0.617	0.152	4.059	0.000	0.451
Standardized distance to the intersection (0~1)	-0.565	0.160	-3.544	0.000	0.358
Number of moving car	0.148	0.030	4.940	0.000	1.320
Number of moving truck	-0.178	0.056	-3.200	0.001	0.638
Number of stopping car	0.295	0.028	10.436	< 2e-16	1.283
Number of stopping truck	-0.095	0.063	-1.516	0.130	0.679
Number of motorbike	0.195	0.104	1.872	0.061	0.334
Number of crossing line	0.142	0.039	3.645	0.000	1.043
Number of pedestrian	-0.058	0.047	-1.234	0.217	0.875
Number of aged pedestrian	2.844	0.171	16.595	< 2e-16	0.205
Number of child	0.380	0.130	2.918	0.004	0.261
Number of bicycle	1.116	0.070	15.953	< 2e-16	0.548
Number of stop line	0.008	0.204	0.040	0.968	0.164

Table 4 Multiple regression analysis using 96 variables.

(a) Gray back-ground shows significant weight coefficient. Coefficients of multiple regression analysis using objects location data.

	Estimate	Std. error	t value	Pr (>|t|)	Standardized partial regression coefficient
Speed	-0.019	0.005	-3.507	0.000	-0.223
Inside intersection (0,1)	-0.229	0.127	-1.805	0.071	-0.079
Move (0,1)	0.145	0.203	0.714	0.476	0.048
Stop (0,1)	-0.251	0.208	-1.210	0.227	-0.084
In tern (0,1)	1.002	0.134	7.494	0.000	0.321
Standardized distance to the intersection (0-1)	-1.062	0.145	-7.346	0.000	-0.270
Scene 1	3.712	0.413	8.995	< 2e -16	0.792
Scene 2	4.319	0.274	15.770	< 2e -16	0.922
Scene 3	6.186	0.304	20.332	< 2e -16	1.320
Scene 4	3.956	0.293	13.505	< 2e -16	0.844
Scene 5	2.831	0.294	9.621	< 2e -16	0.749
Scene 6	3.408	0.288	11.827	< 2e -16	0.702
Scene 7	2.974	0.278	10.714	< 2e -16	0.641
Scene 8	3.008	0.294	10.233	< 2e -16	0.737
Scene 9	2.466	0.291	8.489	< 2e -16	0.524
Number of moving car	0.292	0.116	2.508	0.012	0.273
Number of moving truck	0.261	0.160	1.638	0.102	0.118
Number of stopping car	0.720	0.111	6.492	0.000	0.655
Number of stopping truck	0.691	0.316	2.188	0.029	0.333
Number of motorbike	-0.669	0.411	-1.627	0.104	-0.159
Number of crossing line	-0.204	0.121	-1.677	0.094	-0.151
Number of pedestrian	-0.566	0.380	-1.490	0.137	-0.352
Number of aged pedestrian	2.248	0.248	9.058	< 2e -16	0.328
Number of child	1.456	0.453	3.215	0.001	0.270
Number of bicycle	0.464	0.338	1.373	0.170	0.181
Number of stop line	-0.129	0.227	-0.570	0.569	-0.015

regression analysis using 96 variables was superior.

4. Conclusion

Using hazard information, we examined the description of driving scenes and the quantitative evaluation of risk. Using the data obtained through the multiple regression analysis described in **Chap. 3**, we obtained weighted coefficients for the risk associated with a range of states (including hazard information).

From the results of our analysis, as well as the standard evaluations provided by the driving instructors, we can say that there is

Table 5 Akaike's Information Criterion (AIC) of multiple regression model.

Number of variables in the model	Correlation coefficient r Evaluated value and raw data	AIC
16 variables	0.933	2517.201
96 variables	0.979	1622.994

(b) Coefficients of multiple regression analysis using objects location data. Gray background shows significant weight coefficient.

Number of moving car: position

0.00	-1.19	
-0.31	-0.19	0.08
-0.30	-0.17	0.07

Number of motorbike : position

0.17	0.39	0.67
-0.18	0.69	

Number of all pedestrians : position

0.13	-0.13	
0.59		-0.19

Number of moving truck : position

	0.52	
-0.60	-0.08	0.16
-0.54	0.04	0.09

Number of bicycle : position

	0.13	
-0.29		
-0.24		

Number of pedestrian : position

-0.24	1.38	
0.28	-0.07	0.28
-0.17	0.53	0.38

Number of stopping car : position

	-0.63	-1.09
-0.65	-0.62	-0.51
-0.87	-0.33	-0.12

Number of stop line : position

		0.52

Number of aged pedestrian : position

0.03		

Number of stopping truck : position

	-0.27	-0.14
1.26	-1.59	-0.35

Number of crossing line : position

0.45	0.14	-0.10
0.87	0.26	0.03

Number of children pedestrian : position

-1.16		

a need for further research into the characteristic of risk perception and, particularly, "How great a risk should a driver feel upon observing an elderly person or child?" Future research should carefully examine how risk can be mechanically inferred, as well as the logic needed to issue warnings to a driver using hazard information for the road ahead of that driver's vehicle. While a model capable of quantitatively examining the driving scene ahead of a vehicle and then estimating the related risk has not yet been developed, we believe that research like ours will contribute to the creation of a model capable of quantitatively extracting the risk from a detailed description of the actual scene and the state.

This research actually used an insufficient number of environment states, and incorporated objects that would be difficult for machine recognition to handle. It would seem, however, that it will become possible to statistically classify and analyze driving environments, instead of relying on the researchers' experience and memory as has been the case to date.

References

1) ITARDA : "*Hito wa Don-na Misu wo Shite Koutsuu-Jiko wo Okosunoka?* ; *Kiiwaado wa 'Omoikomi'* ", ITARDA Information (in Japanese), No.33(2001)
2) Kokubun, M., et al. : "Analysis of Drivers' Characteristics in Risk Perception", Trans. of Human Interface Soc. (in Japanese), **5**-1(2003), 27-36
3) Ogawa, K. : "Danger Prediction Training", Road Safety Study (in Japanese), (2000), 242-253, Osaka Traffic Sci. Soc.
4) Chubu Nippon Driver School : "Danger Prediction Master" (in Japanese), (1999)

(Report received on Apr. 23, 2004)

Hiroyuki Konishi
Year of birth : 1967
Division : Human Error Prevention Lab.
Research fields : Psychology (behavior and emotion), Traffic psychology, Human error analysis
Academic society : Jpn. Psychol. Assoc., Inst. Electron. Inf. Commun. Eng.
Awards : IEEE IECON-2000 Conf. Paper Award (2000)

Mitsuteru Kokubun
Year of birth : 1971
Division : Human Error Prevention Lab.
Research fields : Cognitive psychology, Traffic psychology, Psychophysiology
Academic society : Human Interface Soc., Jpn. Ergon. Soc., Jpn. Psychol. Assoc.
Awards : Kamitake Academic Incentive Award (Inst. Psychol., Univ. Tsukuba) (2002), Outstanding Young Scientist Award (IEEE Reliab. Soc. Jpn. Chapter)(2004)

Kazunori Higuchi
Year of birth : 1962
Division : Human Factors Lab.
Research fields : Visual information Processing, Human factors
Academic society : Inst. Electron. Inf. Commun. Eng., Inst. Electr. Eng. Jpn.
Awards : R&D100 Awards (2001)

Tetsuo Kurahashi
Year of birth : 1969
Division : Accident Prevention Lab.
Research fields : Active safety, Human factors
Academic society : Jpn. Soc. Mech. Eng., Inst. Electron. Inf. Commun. Eng.

Yoshiyuki Umemura
Year of birth : 1957
Division : Human Factors Lab.
Research fields : Human factors, Traffic safety
Academic society : Soc. Autom. Eng. Jpn., Inf. Process. Soc. Jpn., Acoust. Soc. Jpn., Assoc. Nat. Lang. Process.