

# Evaluation of Six Night Vision Enhancement Systems: Qualitative and Quantitative Support for Intelligent Image Processing

Sascha Mahlke, Berlin University of Technology, Berlin, Germany, Diana Rösler, Chemnitz University of Technology, Chemnitz, Germany, Katharina Seifert, Volkswagen AG Group Research, Wolfsburg, Germany, Josef F. Krems, Chemnitz University of Technology, Chemnitz, Germany, and Manfred Thüring, Berlin University of Technology, Berlin, Germany

**Objective:** An evaluation study was conducted to answer the question of which system properties of night vision enhancement systems (NVESs) provide a benefit for drivers without increasing their workload. **Background:** Different infrared sensor, image processing, and display technologies can be integrated into an NVES to support nighttime driving. Because each of these components has its specific strengths and weaknesses, careful testing is required to determine their best combination. **Method:** Six prototypical systems were assessed in two steps. First, a heuristic evaluation with experts from ergonomics, perception, and traffic psychology was conducted. It produced a broad overview of possible effects of system properties on driving. Based on these results, an experimental field study with 15 experienced drivers was performed. Criteria used to evaluate the development potential of the six prototypes were the usability dimensions of effectiveness, efficiency, and user satisfaction (International Organization for Standardization, 1998). **Results:** Results showed that the intelligibility of information, the easiness with which obstacles could be located in the environment, and the position of the display presenting the output of the system were of crucial importance for the usability of the NVES and its acceptance. **Conclusion:** All relevant requirements are met best by NVESs that are positioned at an unobtrusive location and are equipped with functions for the automatic identification of objects and for event-based warnings. **Application:** These design recommendations and the presented approach to evaluate the systems can be directly incorporated into the development process of future NVESs.

## INTRODUCTION

In the dark, human visual information processing is severely limited (e.g., Levine & Shefner, 2000), but unfortunately, most drivers do not realize that their visual ability is restricted in low light conditions (Tyrrell, Patton, & Brooks, 2004). Leibowitz and Owens (1977, 1986) discussed this impairment as a consequence of visual degradation. They divided the visual processes during driving into two rather separate subcomponents: (a) peripheral visual processes, relevant for lane keeping and lateral control (Mourant & Rock-

well, 1972; Summala, Nieminen, & Punto, 1996); and (b) processes of central vision, indispensable for target detection as well as recognition. In particular, some main functions for target detection, such as perception of contrast, depth, and distance, are severely affected by nighttime driving, whereas peripheral vision and thus lateral control are less impaired. Therefore, it comes as no surprise that many severe and fatal crashes happen at night (e.g., Mariani et al., 2002; Owens & Sivak, 1996; Sullivan & Flannagan, 2001). An analysis of such crashes showed that most of them did not involve any actions to avoid an

impending collision (Tijerina, Browning, Mangold, Madigan, & Pierowicz, 1995). Under conditions of reduced visibility, drivers seem to have almost no opportunity to take preventive measures when an accident is about to happen.

### Need for Night Vision Enhancement Systems

Options for improving drivers' vision at night are rather confined. The usage of high-beam headlights is restricted with oncoming traffic, and technical and legal regulations limit the light output of low-beam headlights to reduce glare for other road users. Moreover, reasonable road illumination is not always realizable or may be too expensive, as on long stretches of country roads. Therefore, car manufacturers are considering alternative solutions for night vision support. One of the most promising applications is using infrared-based night vision enhancement systems (NVESs). Such systems can extend the visibility of objects to levels comparable with high-beam conditions.

NVESs have a long tradition. Since the 1940s they have been primarily used in military contexts (Johnson, 2004; Tsimhoni & Green, 2002), but for about 10 years, there has been a broad discussion about their automotive application. Some studies already have examined detection performance of critical targets with NVESs under realistic driving conditions (e.g., Barham, Oxley, & Ayala, 1998; Barham, Oxley, Thompson, Fish, & Rio, 1999). These studies showed contradictory results. Gish, Staplin, and Perel (1999) concluded from their data that there is a high probability of overlooking critical targets in a visually def-

manding display because both tasks, driving and target detection, depend upon the same visual resources and thus interfere with each other. Gish et al. (1999) found no improvement of performance with the aids they studied. Older drivers made significantly less usage of the system and felt less confident in handling the system than did younger drivers. Kiefer (1995) reviewed the literature about target detection with vision enhancement systems and concluded that drivers obtained better results and reported lower subjective workload without system support. Conflicting evidence comes from a study that showed positive effects in the detection of pedestrians and road elements (Stahl, Oxley, Berntman, & Lind, 1994). Younger drivers benefited more from system usage than older ones, but older drivers found the NVES easy to use and helpful in detecting hazards. General judgments about NVESs, however, are hard to make because of the use of diverse infrared sensor, display, and image-processing technologies in different systems and the still ongoing evolution of these technologies.

*Far versus near infrared sensor technology.* Currently, developmental efforts are focusing mainly on two different types of sensor systems: either the near infrared sensor technology or the far infrared sensor technology (see Figure 1 for a comparison and Schlessinger, 1995, for an overview on infrared sensor technologies). The use of near infrared systems requires active infrared headlights with a radiation range of about 750 to 3,000 nm. A near infrared camera captures the reflections from infrared-reflecting surfaces. The camera output is a monochromatic image, comparable to the real-world scene produced by



Figure 1. Sensor images: near infrared (left) and far infrared (right).

high-beam headlights, with clearly recognizable edges of the infrared-reflecting objects. These characteristics make the image appear familiar and, therefore, facilitate drivers' image processing (Rumar, 2002). Far infrared systems, on the other hand, act as passive sensors of thermal radiation in the range of about 6,000 to 30,000 nm, which is the radiation range emitted by objects. The image is determined by the relative differences in thermal radiation caused by different temperature and/or thermal conductivity of objects. The far infrared sensor has a larger spatial coverage than the near infrared sensor, but its image looks rather unfamiliar.

Some advantages and drawbacks have been mentioned for both far and near infrared sensor systems (e.g., Tsimhoni, Bärgrman, Minoda, & Flannagan, 2004). Tsimhoni et al. compared pedestrian detection performance with near and far infrared videos. Results supported a greater effectiveness in pedestrian detection when far infrared videos were assessed. Because of pedestrians' heat radiation they are more obvious in far infrared sources. Parkes, Ward, and Bossi (1995) indicated that the far infrared sensor technology depends heavily on temperature, and thus system pictures can change in unpredictable ways. In spite of training, military drivers and pilots have shown problems in interpreting far infrared sensor displays (Brickner, 1989). Furthermore, it seems to be difficult for drivers to match the display image with the outside world because important cues for orientation, such as road marking and signs, are missing on the display (Rumar, 2002). With the realistic image of near infrared sensor systems, on the other hand, there is the risk that drivers would rely exclusively on the visual display and use it for driving. In military contexts, the occurrence of this phenomenon has been confirmed (Padmos & van Erp, 1996).

*Display technology.* No industrial standard for night vision displays has been developed yet, and for most car manufacturers, these displays are still in the preproduction stage. Exceptions are five companies that have already started offering NVESs: Cadillac since 2000, Lexus since 2003, Honda since 2004, and Mercedes-Benz and BMW since the end of 2005. NVESs present their information on an additional display: a head-down display taking the place of the conventional instrument panel, a head-up display integrated into the dashboard in front of the driver, or a head-up

display using the windshield for projection (see Gish & Staplin, 1995, for a detailed literature review). The information is usually presented as an analogue video-based image. Such analogue presentations may lead to higher demands on visual and mental resources because drivers have to search for relevant information on the display and compare it with the outside. Therefore, it entails potential impairments of traffic safety. These problems have been discussed at length (e.g., Rumar, 2002; Tijerina et al., 1995; Tsimhoni & Green, 2002), but to date no research has studied them in an experimental design under real traffic conditions.

*Image-processing technology.* Different approaches to automatic image processing have been used to improve camera output. For near infrared sensors, it is necessary to reduce glare effects or to sharpen the contrast of the image (Rumar, 2002). Intelligent automatic image processing, a more advanced approach, tries to compensate for the lower contrast features in the far infrared sensor image. It detects and enhances specific features, such as the outlines of warm objects (e.g., Bertozzi, Broggi, Pascoli, Graf, & Meinecke, 2004). In a knowledge-based categorization process, these features are used by a computer program to identify, for example, a pedestrian (e.g., Bertozzi, Broggi, Grisleri, Graf, & Meinecke, 2003). Caird, Horrey, and Edwards (2001) suggested that the main application of night vision enhancement should be aimed not at improving general visibility but at actively supporting the detection of critical targets. Rumar (2002) also envisioned an NVES that would present only selected information in order to prevent drivers from completely focusing on the system instead of looking through the windshield. Advanced systems with automatic categorization of image contents could inform the driver in a different way than by showing an analogue video image. Event-related warnings could be generated as optical or acoustical signals to shift the driver's attention to the object identified, thereby supporting the perceptive and cognitive processing in hazard prevention.

## Objectives

Besides a few exceptions (e.g., Tsimhoni et al., 2004), previous empirical investigations on NVESs have evaluated only single systems (e.g., Barham et al., 1998; Gish et al., 1999). Advantages and

drawbacks exist for all these systems, but no comparison of usability has been made for different systems under real traffic conditions in a single study (see also Rumar, 2003; and Tijerina et al., 1995). Therefore, we started a detailed investigation of six different NVES prototypes under real traffic conditions with the aim of deducing recommendations for the further development of such systems. Prototypes differed in infrared sensor, display, and image-processing technologies. Two studies were conducted. In the first, a qualitative study, experts in ergonomics, perception, and traffic psychology assessed the six NVESs in a heuristic evaluation (Nielsen, 1993, 1994). The goal was to collect substantial valid comments about different usability aspects of the prototypes. These reported aspects provided the basis for the second experimental study, in which additional quantitative data were collected from experienced drivers. The usability dimensions of effectiveness, efficiency, and user satisfaction were of particular interest (International Organization for Standardization, 1998; Pataki, Schulze-Kissing, Mahlke, & Thüring, 2005). Effectiveness can be described as the degree of accuracy and completeness with which the user's goals are satisfied. Efficiency can be characterized as the effectiveness of system usage in relation to its costs in terms of effort or time. Satisfaction means the comfort and the acceptability of working with the system.

### Heuristic Evaluation

Eight experts participated in the heuristic evaluation. Two methods for data collection were employed in each case: (a) free verbal reports (Ericsson & Simon, 1993, 1998) during the use of the six different NVES prototypes under real traffic conditions at night, lasting about 3 to 4 hr, and (b) semistructured interviews with heuristic guidelines the day after the night drives. The items for the interview were derived from previous research (mainly by Mariani et al., 2002; Mariani, Andreone, Palmieri, & Tango, 2003; Rumar, 2002; and Tsimhoni & Green, 2002). In the following analysis, the most important comments, which will be enhanced by quantitative data, are summarized.

*Applicability of NVESs.* Experts regarded night vision enhancement as a very promising and worthwhile approach to driver assistance. All of them endorsed further development. They rec-

ommended that the main focus should be on the active, software-supported detection of critical events (see also Rumar, 2002). Such event-based systems were considered as more effective than displays with analogue video presentations, which require additional attention and regular scanning.

*Event-based systems with automatic object detection.* An event-based system with automatic object detection has the advantage that only potential hazards are signaled. Drivers can keep their eyes on the road all the time. Conspicuous signals should be used to provide warnings of critical obstacles in a fast and direct way. Positioning the display at an unobtrusive location at the edge of the visual field was judged as important to make it peripherally perceivable. Such a system does not cause a risk of unwanted distraction, nor does it create additional visual demand. Therefore, it is unlikely to impair driving performance. Experts could imagine an event-based system being used regularly for several hours per night, as it would reside passively in the background and require attention only when a hazardous event is detected.

*Information systems with analogue video presentation.* Because systems with analogue information presentation need considerable visual attention and enhanced visual glance times for display scanning, these systems were judged as fatiguing and likely to decrease traffic safety. Nevertheless, many features of the image of a near infrared sensor system, such as contrast, quality, and the comparability to real-world scenes, were assessed positively. However, the potential of these systems to warn about a hazard was judged as unsatisfactory. Different objects in the scene (e.g., a reflective post and a pedestrian) were presented too similarly, making it hard to draw the driver's attention to critical targets. In addition, drivers can overlook critical events because of glare from oncoming headlights that might preclude seeing. Far infrared sensor systems, on the other hand, hold the potential to show additional information that the driver cannot see under normal light conditions. They provide useful warning hints because of the accentuation of warm objects, which appear as very bright in contrast to a colder surrounding. Grave problems were anticipated by the experts with respect to learnability because far infrared images differ considerably from usual video images. Moreover,

**TABLE 1:** Evaluated Combinations of Sensor, Image Processing, and Display Types

NVES Prototype	Abbreviation	Test Car
Near infrared sensor with head-down display	NI-HD	Audi A8
Far infrared sensor with head-down display	FI-HD	Audi A8
Far infrared sensor with virtual head-up display	FI-VHU	Audi A8
Near infrared sensor with mirror-based head-up display	NI-MHU	VW Touareg
Far infrared sensor with automatic pedestrian recognition and an analogue mirror-based head-up display with highlighting of recognized objects in the video picture	APR-MHU	VW Touareg
Far infrared sensor with automatic pedestrian recognition and an event-based LED display	APR-LED	VW Touareg

temperature-dependent variance in the images, resulting from the surroundings cooling off overnight, makes it even harder to interpret them in a constant and consistent manner during longer nighttime driving (see also Parks et al., 1995). Experts regarded the absence of an active component that tells the driver when to look at the display as a major problem for all video display systems. All these issues limit the efficiency of such systems to a considerable degree. Experts could not imagine analogue systems being used on a regular basis because observing a continuous video presentation would require too much effort.

## METHOD

To amplify the heuristic evaluation, an experimental study was conducted in which the six NVES prototypes were tested with ordinary drivers. The goal was to collect data about the impact of the differently designed systems on drivers' workload and performance in real-life situations.

### Participants

Fifteen drivers (7 women, 8 men) participated in the study. They were between 40 and 65 years old. This group was chosen because night vision tends to become impaired with age (Rumar, 2002). Hence, older persons in particular should benefit from NVESs. All participants were private drivers who rated their driving experience as high or very high, especially with respect to driving at night. Drivers were paid for their participation.

### Apparatus

*NVES prototypes.* Two test cars, a VW Touareg and an Audi A8, were equipped with a far and a near infrared sensor and different displays. An automatic pedestrian recognition system (Bertozzi et al., 2004) was implemented in the VW Touareg.

Because of space limitations, it was impossible to integrate all display types in each test car, but both cars were chosen from the same vehicle class. The output of the infrared sensors was presented on different types of displays: (a) a head-down display integrated in the instrument panel behind the steering wheel, in front of the driver; (b) a virtual head-up display in which the image was projected into the windshield; (c) a real head-up display in which the image was projected on a narrow mirror at the bottom of the windshield; and (d) an abstract LED display, in combination with the active pedestrian recognition system, that signaled the presence of an automatically detected pedestrian. The combinations of sensor, image processing, and display types that were evaluated are summarized in Table 1 (see also examples in Figure 2). The first five systems in Table 1 presented the information as an analogue video image. Intelligent image processing was used in two systems: In system APR-MHU, the pedestrian was highlighted in the analogue video image; in system APR-LED, the pedestrian's location was indicated by the abstract LED display. A flashing LED symbolized the event "pedestrian detected" and indicated the direction of the pedestrian's location.

*Data recording.* Both cars were equipped with a head-mounted eye-tracking system to measure the driver's gaze (iView, SensoMotoric Instruments), a video-recording system with a video splitter to synchronize all camera outputs (near and far infrared sensors as well as gaze), a controller area network (CAN) bus to log steering activity as well as velocity, and an audio recorder to document all utterances by the drivers.

*Test route.* For each test drive, a circuit course of about 14 km on public roads in Berlin, Germany, was used. Half of this route was on rural



Figure 2. Examples for two different display types: APR-LED (left) and APR-MHU (right). (See Table 1 for explanations of abbreviations.)

streets leading through a forest without any streetlights; the other half consisted of small streets in a suburban area with sparse illumination. Traffic on the course was regular – that is, the roads were not closed for the test drives. There were a number of curves on the course as well as nine branching points. On average, drivers needed 20 min per trip. At varying locations along the course, four pedestrians were positioned to ensure a minimum of four critical events to which the drivers had to react. Pedestrians did not stand in curves and had to keep a safety distance of about 1 m from the lane. They did not wear any safety clothing to improve visibility. Half of them stood on the left side of the road, the other half on the right. All of them were in clear sight of the drivers.

### Experimental Design

All drivers drove the predefined route with each of the six systems and twice without any system (baseline drives). One baseline drive was conducted with each car. Thus, a one-factorial within-subjects design was realized, in which system was a single factor consisting of eight treatments.

### Dependent Variables

A variety of measures were used to examine driving behavior as well as the usability criteria of effectiveness, efficiency, and user satisfaction.

**Effectiveness.** Usually, effectiveness is defined as the degree of accuracy and completeness with which the user's goals are satisfied by using the system. With respect to using an NVES, the central goal is to detect any hazardous object and to react to it as quickly as possible. In our experiment, drivers had to verbally indicate the occur-

rence of any pedestrian, biker, or wildlife as fast as they could. Three measures served as indicators for effectiveness:

- The *recognition rate* is defined as the percentage of correct detections. It represents the degree of accuracy and completeness with respect to identifying potential hazards.
- The *recognition time* is defined as the time required for a correct detection. It represents how fast drivers can respond to a critical event indicated by an NVES. To compare the different systems, we calculated recognition times from the time when the first pixel of the pedestrian appeared in video recording until recognition. For the APR-LED system, the first flashing LED related to the given event and visible to the driver was used as the starting point. The comparison of recognition times between systems aims to show which constellation of sensor, display type, and data processing is most effective for triggering fast responses.
- The *recognition distance* is defined as the distance between car and pedestrian at the time of recognition. It was calculated from the speed of the car and the time between recognizing and passing the pedestrian. The comparison of recognition distances between baseline drives and drives supported by a NVES aims to show if any of the systems increases the distance between car and pedestrian at the time of detection.

In addition to these three measures, false alarm rates could serve as an additional indicator for effectiveness. Two kinds of false alarms can be distinguished. First of all, false alarms may be caused by the system (i.e., the automatic pedestrian recognition implemented in the APR-MHU and the APR-LED may respond erroneously). None of the other systems can produce a false alarm because all of them present only a video image without any machine-based interpretation. Second, false alarms may be caused by the drivers.

In this case, high alarm rates might indicate that an NVES is more distracting and irritating than supportive. We will return to this issue when reporting our results.

*Efficiency.* The cost aspect of efficiency was measured by two dependent variables: subjective effort and time spent on watching the NVES. The subjectively perceived effort was assessed by drivers using two rating scales: (a) the Subjective Mental Effort Questionnaire Scale (SMEQ-Scale; Arnold, 1999; Eilers, Nachreiner, & Hänecke, 1986) and (b) a slightly enhanced version of the NASA Task Load Index (NASA-TLX; Hart & Staveland, 1988), which was used to assess different dimensions of demand (e.g., mental demand, temporal demand, and frustration). For this study, we added one item to the NASA-TLX to measure visual demand. As second cost factor, we captured the time of attending to the NVES by recording the frequency of eye movements and the duration of fixations. The overall time spent on a system was computed from these values.

*Satisfaction.* To measure users' satisfaction with the systems, two questionnaires were employed in addition to an overall ranking: (a) an acceptance questionnaire based on the acceptance theory of Kollmann (2004) for obtaining information on attitudes to use and purchase a system, and (b) a questionnaire based on the user experience model by Mahlke (2002) for analyzing the users' perception of relevant aspects such as utility, ease of use, hedonic quality, and visual attractiveness.

*Driving behavior.* Several driving parameters were logged via the CAN bus of the car to investigate whether the use of an NVES influenced central driving activities, such as steering, as compared with driving without a NVES.

## Procedure

For each driver, the experiment was split into two sessions, which were a week apart and took place at night between 10:00 p.m. and 1:00 a.m. in October. Eight drivers used the systems of the VW Touareg in the first session; the other 7 started with the systems of the Audi A8. Each session consisted of a briefing, a test phase, and a debriefing.

*Briefing.* Drivers were informed about the goals of the study and the duration of the sessions. The car as well as the NVESs were explained and demonstrated. Drivers familiarized themselves with the car and its equipment by driving a short training course. At the end of the

briefing, drivers received detailed information about the route and their tasks. They were instructed to verbally indicate the detection of any pedestrians, bicyclists, or wildlife as quickly as possible. In addition, they were asked to report what they had seen and whether or not their perception was based on the NVES.

*Test phase.* Drivers completed four test drives: one drive without system support and three drives with different NVESs. The order of the test drives varied for each driver. The four pedestrians were positioned at different locations for each test drive. Two supervisors supported the driver in navigating on the route, surveyed the data recording, and gave assistance in case of an unforeseen event. After each test drive, there was a 15-min break during which drivers filled in the questionnaires.

*Debriefing.* Drivers were thanked, paid for their participation, and given the time and date of the next session. The second session was identical to the first one, except that the three remaining systems were tested with the other car.

## RESULTS

### Effectiveness

To evaluate the effectiveness of system usage, we analyzed recognition rates, recognition times, and recognition distances for all planned pedestrian incidents.

*Recognition rate.* The rate obtained with the APR-LED (57.7%) was closest to that of the baseline drives (58.2%). Using one of the other five systems, drivers detected between 46.6% and 56.9% of the pedestrians (Figure 3). All these differences were too small to be statistically significant – that is an ANOVA showed no significant effect of the system factor,  $F(7, 98) = 1.15, p = .34$ .

*Recognition time.* The analysis of recognition times focused on potential differences between systems (Figure 4). Usage of the APR-LED system and of the two near infrared-based systems produced reactions between 4.9 s and 6.4 s, whereas the three other systems led to much longer recognition times (8.6–9.5 s). An ANOVA showed a significant effect of the system factor on recognition times,  $F(5, 70) = 10.41, p < .01, \eta^2 = .43$ . Pairwise comparisons (Bonferroni-corrected  $t$  tests) indicated significant differences between both APR-LED and NI-MHU, as compared with FI-VHU, APR-MHU, and FI-HD.

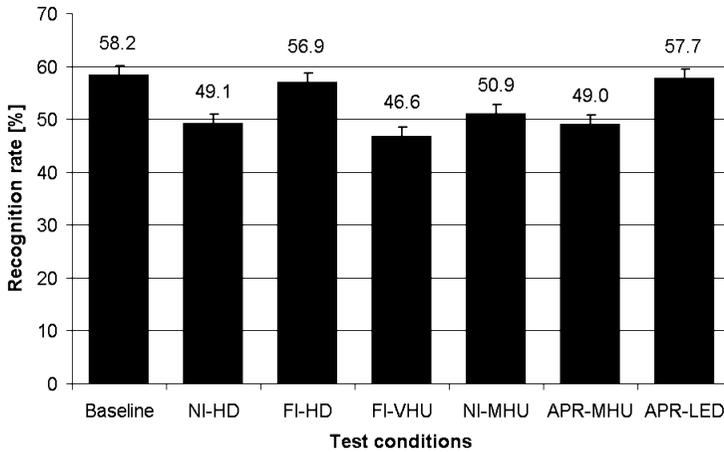


Figure 3. Recognition rates for all test conditions (the two baseline drives are grouped). (See Table 1 for explanations of abbreviations.)

The recognition times of the NI-HD system were significantly smaller than those of the FI-VHU system.

*Recognition distance.* The recognition distance for three systems (NI-MHU: 49 m; NI-HD: 47 m; APR-LED: 36 m) was slightly higher than for the baseline drives (34 m). For the other systems, the average distance lay between 30 and 32 m (Figure 5). None of these differences was statistically significant, and an ANOVA revealed no effect of the system factor,  $F(7, 98) = 1.08, p = .38$ .

In addition to these three measures, we also investigated false alarm rates. Both systems that were equipped with automatic pedestrian detection (APR-MHU and APR-LED) occasionally showed false alarms. These erroneous system responses were rather infrequent and varied with

weather conditions. Future studies at a more technical level might be conducted to analyze the false alarms of the two systems in more detail, because our weather recordings were not precise enough for that purpose. A screening of the other type of false alarms – those produced by the drivers – showed that they were extremely rare. Even at a purely descriptive level, it was apparent that there were neither differences between baseline drives and drives supported by a NVES nor differences among the six prototypes under investigation. Therefore, no statistical analysis was conducted.

**Efficiency**

To learn more about efficiency, we took a closer look at two cost factors associated with system usage: mental workload and time spent on a system.

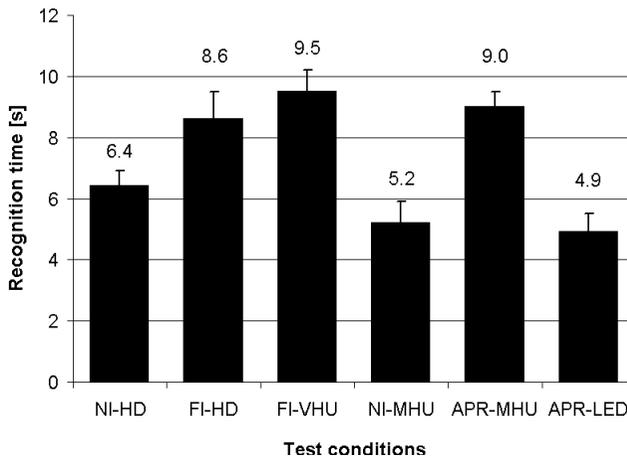


Figure 4. Average recognition times for all system variants. (See Table 1 for explanations of abbreviations.)

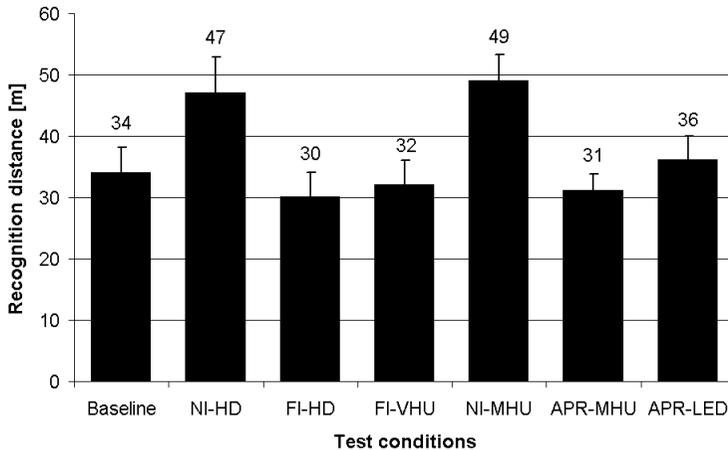


Figure 5. Average recognition distances for all test conditions (the two baseline drives are grouped). (See Table 1 for explanations of abbreviations.)

**Mental workload.** We measured mental workload by the NASA-TLX and the SMEQ-Scale. On both scales, drives with the NI-HD and APR-LED systems were judged as essentially equivalent to the baseline drives. Ratings for the other systems showed higher workloads. The ANOVA of the NASA-TLX data indicated a significant effect of the systems factor,  $F(7, 98) = 3.92, p < .01, \eta^2 = .22$ . Similar results were found for the SMEQ-Scale,  $F(7, 98) = 5.79, p < .01, \eta^2 = .29$ . Regarding both scales, pairwise comparisons (Bonferroni-corrected  $t$  tests) revealed significant differences between the baseline drives and all systems except APR-LED and NI-HD. Workload measured for APR-LED was significantly lower than that for all systems except NI-HD. The workload for the NI-HD system was significantly lower than that for the FI-VHU system. An in-depth analysis of the subscales of the NASA-TLX revealed that these differences arose from different ratings of visual demand,  $F(7, 98) = 4.54, p < .01, \eta^2 = 0.24$  – that is, higher mental workload resulted mainly from greater visual efforts.

**Eye-tracking data.** The analysis of eye-tracking data focused on the six conditions with system usage. The results showed that the frequency of fixation hardly varied by system but that the total duration of fixations differed between the systems. Systems displaying analogue video images were fixated longer (11%–18% of driving time) than was the APR-LED (6%). The ANOVA of the total duration of fixations showed a significant effect of the systems factor,  $F(5, 70) = 2.65, p < .05, \eta^2 = .16$ . The analysis of the other three fixation

categories of windshield/road, instruments, and mirrors indicated that the main part of the additional time that drivers spent fixating the NVEs with analogue video displays was taken from the time looking at the road.

In a second step, we focused on selected eye-tracking data to learn more about the relation between correct detections and the time spent on a system. Only those fixation times were included that were related to the detection of pedestrians obtained by explicitly using the system. The selection criteria for this subset of data were twofold: (a) The fixations had to occur between pedestrian appearance and recognition, and (b) the driver had to verbally indicate that the detection was accomplished with the system. All other fixation times (e.g., those associated with correct identifications without system usage) were excluded. On average, about 25% of the pedestrians were recognized by using one of the systems. The APR-LED was used most often, (i.e., in almost 40% of the cases). Successful usage of the video-based systems required between 22% (NI-MHU) and 38% (FI-HU) of the time between pedestrian appearance and detection. In contrast, the APR-LED system required only 6% of this time interval to produce a recognition event. The results are shown in Figure 6. The ANOVA of the proportion of fixation time regarding 60 detections with system usage indicated a significant effect of the system factor,  $F(5, 54) = 5.53, p < .01, \eta^2 = 0.34$ . Pairwise comparisons (Bonferroni-corrected  $t$  tests) showed significant differences between APR-LED as compared with FI-VHU, FI-HD, and NI-HD.

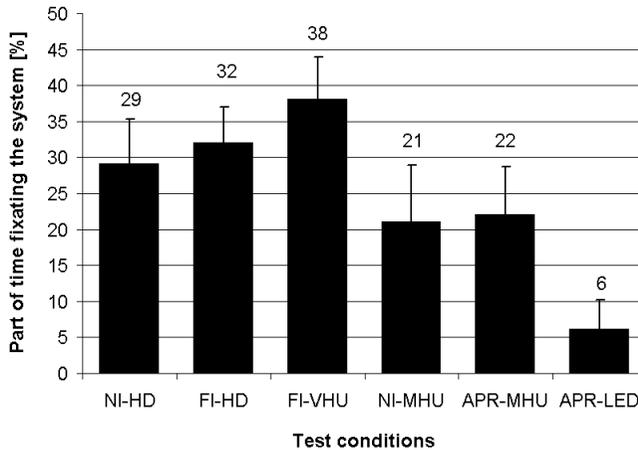


Figure 6. Average times that systems were fixated between pedestrian appearances and detection if an object was recognized using the system. (See Table 1 for explanations of abbreviations.)

### Satisfaction

Three approaches served to assess users' satisfaction with the NVEs: a system ranking, the acceptance questionnaire, and the user experience questionnaire. Again, data were analyzed for the six conditions with system usage. In the system ranking, the NI-HD system was ranked best (average rank 2.5), followed by FI-HD (2.9) and APR-LED (3.3), but a Friedman test revealed no significant differences between systems,  $\chi^2(5, 15) = 9.09, p = .11$ . Clearer results were gained from the acceptance questionnaire. Here, an ANOVA of the ratings showed a significant effect of the system factor,  $F(5, 70) = 3.17, p < .05, \eta^2 = .19$ . The NI-HD system got the highest acceptance value, and APR-LED was ranked second. In the experience questionnaire, an ANOVA showed a significant effect of the system factor only on the utility dimension,  $F(5, 70) = 2.85, p < .05, \eta^2 = .17$ . Again, perceived utility was rated highest for the NI-HD, with the second best being the APR-LED.

### Driving Behavior

In addition to effectiveness, efficiency, and satisfaction, selected parameters of driving behavior were evaluated. No differences were found between the baseline drives and the drives with one of the systems regarding average speed and steering angle speed. To examine the driving behavior at the moment of detection, we analyzed speed, steering angle, and steering angle speed before detection, at the moment of detec-

tion, and after detection. Again, no differences were found.

## DISCUSSION

As a number of studies have documented (e.g., Mariani et al., 2002), detecting pedestrians at night is a demanding task. The task in our study might have been additionally difficult because, for safety reasons, pedestrians had to keep a distance of about 1 m from the lane. Sometimes blinded by the headlights of approaching traffic, drivers recognized less than 60% of our pedestrians during the baseline drives. This rate was not improved by any of the NVEs under investigation. Although there was no significant impairment of recognition, either, it should be noted that only drivers using the APR-LED reached about the same level of performance as drivers without any support. Hence, none of the systems increased drivers' effectiveness with respect to the number of detected pedestrians.

Another aspect of effectiveness concerns the speed of correct detections. Two issues are particularly relevant in this context:

- Recognition time: Do the systems differ with respect to the time required by the drivers for recognizing the signalization of a pedestrian and for responding to it?
- Recognition distance: Do any of the investigated NVEs lead to an increase of distance between car and pedestrian at the time of detection as compared with baseline drives?

Of course, both issues are closely related. The sooner an unequivocal signal is generated by an

NVES, the sooner the driver can react and the greater the gain of distance between car and pedestrian (see Figure 7).

With respect to the first issue, we found that recognition times were significantly shorter when the APR-LED system or the near infrared-based systems were used. The reason for this effect probably lies in one feature shared by these rather different systems: All three systems can be easily understood and require little effort to be learned. As already mentioned by the experts, the APR-LED is highly intelligible because the driver has only to establish an association between the flashing of a diode and the presence of a person close to the road. The near infrared-based systems convey images with a close resemblance to the usual visual image of the environment and therefore are fairly self-explanatory. In contrast, it is more difficult to interpret information from the three far infrared video display systems (FI-HD, FI-VHU, and APR-MHU) because the images they provide deviate considerably from what drivers are used to seeing. Therefore, drivers probably need more time to become accustomed to far infrared images before they can benefit from them. Even the highlighting of automatically detected objects in the far infrared video pictures of the APR-MHU system did not improve recognition performance.

With respect to the second issue however, we found no advantage of any NVES in comparison with the baseline drives. None of the systems provided a significant increase in distance between car and pedestrian at the time of detection. Although the APR-LED and the two near infrared-based systems (NI-HD and NI-MHU) led to slightly better results than the baseline drives, the

effect was not statistically significant. Compared with the distance other studies have assumed to be necessary for choosing an appropriate action, the increase was too minor to be of any real help (e.g., see Rumar, 2003).

Summing up the effectiveness results, it seems that the three far infrared video display systems neither improve nor impair recognition performance. The two near infrared-based systems and the APR-LED system increase the effectiveness of detection in at least some respects. The APR-LED system in particular lends valuable support. Its usage entails recognition rates close to those found in the baseline drives and leads to the shortest recognition times. Hence, the APR-LED probably has the most potential to be developed into a system that also increases recognition distance. In particular, a faster algorithm to identify pedestrians would lead to an early signaling and thus help to lengthen the stopping distance.

The second factor we investigated was the efficiency of NVESs: How much effort must be invested to make system usage successful? For only the NI-HD and the APR-LED systems, the subjectively perceived effort was as low as that for the baseline drives. Ratings for the other systems were significantly higher, thus indicating a perceived increase in workload experienced by our drivers. This increase resulted from higher visual demands for the far infrared video display systems and the NI-MHU system. As noted before, far infrared video display systems are likely to be more demanding because their pictures are difficult to interpret. The NI-MHU is more intelligible but may cause another kind of strain: Because its head-up display permanently resides in the visual field, it must be continuously processed to be reconciled with the usual visual information. If both the intelligibility and the location of the display increase visual demand, then the FI-VHU system should lead to the highest mental workload, and this is what we found.

Consistent with the experts' predictions, evaluation of the eye movements also showed that all video-based systems required more time in comparison with the APR-LED system. This difference probably occurs because these systems suggest a kind of usage behavior different from that with the APR-LED system. Video-based systems require intentional eye movements and some fixation time to process the continuous flow of information from the video image. Therefore,

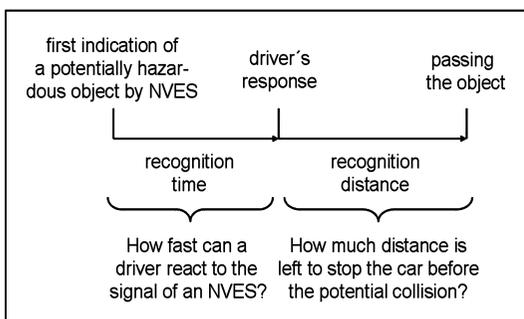


Figure 7. Two issues concerning the speed of correct detections.

drivers must deliberately fixate on the system, shift their attention to the relevant parts of the image, identify any objects presented, compare them with the outside world, and then decide whether or not these are hazardous. Using the APR-LED system is obviously quite different. Because the system is event based and designed to automatically indicate living objects, it supports drivers in two respects. First, there is no need to look at the system when it is inactive. Second, when the system reacts, the meaning of its response is unequivocal – that is, it warns the driver of a living being close to the path of the car. The behavioral pattern suggested by the APR-LED system is to concentrate on the road, to keep the system in the periphery of the visual field, to shift gaze and attention to the APR-LED only when it responds, to use the information provided by the system as a cue for the position of a person, and then to look through the windshield to find the indicated object. As our data suggest, this strategy is not supported by the other system with automatic pedestrian recognition and analogue video presentation, the APR-MHU.

The third factor investigated in our study was the users' satisfaction with the systems. The analysis of our acceptance data showed that the NI-HD system was judged as best, followed by the APR-LED system. The high ranking of the NI-HD system may result from the good quality of its display, which conveyed images of high resolution and the fairly self-explanatory near infrared image. Another reason may be that it did not block out part of the driver's visual field and, therefore, did not obstruct the view of the road. This feature also applied to the APR-LED system, which was located in the periphery of the driver's visual perception. The good ranking of this system deserves particular emphasis because it was a very early prototype. Although it had no impressive appeal and occasionally produced a false alarm, drivers' satisfaction with the APR-LED was surprisingly high.

### IMPLICATIONS FOR DESIGN

To summarize, the APR-LED system showed the best results regarding effectiveness and efficiency and also received good satisfaction ratings. The central features of this system are probably responsible for this outcome and should

be carefully considered for the development of future NVEs. Three characteristics are particularly important in this context:

- The system is very easy to learn. Because it presents information in a highly intelligible way, neither extensive instructions nor training are required for interpreting its output.
- Hazardous events are detected and signaled automatically. The event-based character of system responses relieves drivers from the burden of continuously scanning a display in addition to driving.
- The unobtrusive location of the system avoids blocking central parts of the visual field. Its LED display can be perceived peripherally and draws attention only when necessary.

In accordance with the results of our user study, experts' recommendations derived from the heuristic evaluation focused on these central features – that is, experts favored a system with automatic pedestrian recognition and event-based warnings. Future versions of such a system could be improved by indicating more than the direction of a detected hazard. Enhancements could be the indication of distance, direction of movement, speed, and type of object. If such features are implemented, it must be ensured that intelligibility and ease of interpretation are not hampered.

Further potential for improving NVEs lies in the development of precise and fast algorithms for the automatic detection of pedestrians, bicyclists, and wildlife. As shown in Figure 7, the earlier a warning is generated by the system, the sooner drivers can react, and the more distance can be gained from a hazardous object on the road. Moreover, any decrease of false alarms caused by an algorithm for automatic detection would certainly deepen drivers' trust in the system and increase their satisfaction. Improved that way, NVEs of the future promise to facilitate driving at night significantly and may thus help to reduce the number of severe and fatal crashes.

### ACKNOWLEDGMENTS

This research was funded by Volkswagen AG Group Research. We would like to thank Matthias Rötting and Frank E. Ritter for many clarifying comments on an earlier draft of this paper. The comments of two anonymous reviewers and of Patricia R. DeLucia and John D. Lee greatly enhanced the quality of the manuscript.

## REFERENCES

- Arnold, A. G. (1999). Mental effort and evaluation of user interfaces: A questionnaire approach. In H.-J. Bullinger & J. Ziegler (Eds.), *Human-computer interaction: Ergonomics and user interfaces* (pp. 1003–1007). Mahwah, NJ: Erlbaum.
- Barham, P., Oxley, P., & Ayala, B. (1998). Evaluation of the human factors implications of Jaguar's first prototype near infrared night vision system. In A. Gale, I. D. Brown, C. M. Haslegrave, & S. P. Taylor (Eds.), *Vision in vehicles VI* (pp. 203–211). Oxford, UK: Elsevier.
- Barham, P., Oxley, P., Thompson, C., Fish, D., & Rio, A. (1999). Jaguar cars' near infrared night vision system – Overview of human factors research to date. In A. Gale, I. D. Brown, C. M. Haslegrave, & S. P. Taylor (Eds.), *Vision in vehicles VII* (pp. 177–185). Oxford, UK: Elsevier.
- Bertozzi, M., Broggi, A., Grisleri, P., Graf, T., & Meinecke, M. (2003). Pedestrian detection in infrared images. In *Proceedings of the IEEE Intelligent Vehicles Symposium 2003* (pp. 662–667). Washington, DC: IEEE Computer Society.
- Bertozzi, M., Broggi, A., Pascoli, A., Graf, T. & Meinecke, M. (2004). Pedestrian detection for driver assistance using multi-resolution infrared vision. *IEEE Transactions on Vehicular Technology*, 53, 1666–1678.
- Brickner, M. S. (1989). Apparent limitations of head-up displays and thermal imaging systems. In R. S. Jensen (Ed.), *Proceedings of the Fifth International Symposium on Aviation Psychology* (pp. 703–707). Columbus: Ohio State University.
- Caird, J. K., Horrey, W. J., & Edwards, C. J. (2001). Effects of conformal and non-conformal vision enhancement systems on older driver performance. *Transportation Research Record*, 1759, 38–45.
- Eilers, K., Nachreiner, F., & Hänecke, K. (1986). Entwicklung und Überprüfung einer Skala zur Erfassung subjektiv erlebter Anstrengung. [Development and test of a scale to measure subjectively perceived demand]. *Zeitschrift für Arbeitswissenschaft*, 40, 215–224.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (Rev. ed.). Cambridge, MA: Bradford Books/MIT Press.
- Ericsson, K. A., & Simon, H. A. (1998). How to study thinking in everyday life: Contrasting think-aloud protocols with descriptions and explanations of thinking. *Mind, Culture, and Activity*, 5, 178–186.
- Gish, K. W., & Staplin, L. (1995). *Human factors aspects of using head up displays in automobiles: A review of the literature* (DOT-HS-808-320). Washington, DC: U.S. Department of Transportation.
- Gish, K. W., Staplin, L., & Perel, M. (1999). Human factors issues related to use of vision enhancement systems. *Transportation Research Record*, 1694, 1–9.
- Hart, S.G., & Staveland, L.E. (1988). Development of a multi-dimensional workload rating scale: Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139–183). Amsterdam: Elsevier.
- International Organization for Standardization. (1998). *Ergonomic requirements for office work with visual display terminals (VDTs): Part 11. Guidance on usability* (ISO 9241-11:1998). Bern, Switzerland: ISO.
- Johnson, C. W. (2004). The role of night vision equipment in military incidents and accidents. In C. W. Johnson & P. Palanque (Eds.), *Human error, safety and systems development* (pp. 1–16). Boston: Kluwer Academic Press.
- Kiefer, R. J. (1995). Human factors issues surrounding an automotive vision enhancement system. In *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting* (pp. 1097–1101). Santa Monica, CA: Human Factors and Ergonomics Society.
- Kollmann, T. (2004). Attitude, adoption or acceptance? Measuring the market success of telecommunication and multimedia technology. *International Journal of Business Performance Management*, 6, 133–152.
- Leibowitz, H. W., & Owens, D. A. (1977). Nighttime accidents and selective visual degradation. *Science*, 197, 422–423.
- Leibowitz, H. W., & Owens, D. A. (1986). We drive by night. *Psychology Today*, 20, 54–58.
- Levine, M. W., & Shefner, J. M. (2000). *Fundamentals of sensation and perception* (3rd ed.). New York: Oxford University Press.
- Mahlke, S. (2002). Factors influencing the experience of website usage. In *CHI '02 extended abstracts on human factors in computing systems* (pp. 846–847). New York: Association for Computing Machinery.
- Mariani, M., Andreone, L., Palmieri, S., & Tango, F. (2003). User requirements and customer benefit analysis in the design of a novel driver support system for night vision. In C. Stephanidis & J. Jacko (Eds.), *Human-computer interaction: Theory and practice* (Part 2, Vol. 2, pp. 163–167). Mahwah, NJ: Erlbaum.
- Mariani, M., Palmieri, S., Foschi, P., Pisani, T., Shelton, R., Andreone, L., et al. (2002). *Analysis of customer benefit and system performances specifications*. (IST-2001-34076; Enhanced Driver's Performance in Poor Visibility [EDEL] Tech. Deliverable 2.1). Siena, Italy: University of Siena.
- Mourant, R. R., & Rockwell, T. H. (1972). Strategies of visual search by novice and experienced drivers. *Human Factors*, 14, 325–335.
- Nielsen, J. (1993). *Usability engineering*. Boston: Academic Press.
- Nielsen, J. (1994). Heuristic evaluation. In J. Nielsen & R. L. Mack (Eds.), *Usability inspection method* (pp. 25–62). New York: Wiley.
- Owens, D. A., & Sivak, M. (1996). Differentiation of visibility and alcohol as contributors to twilight road fatalities. *Human Factors*, 38, 680–689.
- Padmos, P., & van Erp, J. B. F. (1996). Driving with camera view. In A. Gale, I. D. Brown, C. M. Haslegrave, & S. P. Taylor (Eds.), *Vision in vehicles V* (pp. 219–228). Amsterdam: Elsevier Science.
- Parks, A. M., Ward, N. J., & Bossi, L. L. M. (1995). The potential to improve driver safety. *Le Travail Humain*, 58, 151–169.
- Pataki, K., Schulze-Kissing, D., Mahlke, S., & Thüring, M. (2005). Anwendung von Usability-Maßen zur Nutzeneinschätzung von Fahrerassistenzsystemen. [Application of usability measures to evaluate the benefit of driver assistance systems]. In K. Karrer, B. Gaus, & C. Steffens (Eds.), *Beiträge zur Mensch-Maschine-Systemtechnik aus Forschung und Praxis* (pp. 211–228). Düsseldorf, Germany: Symposium.
- Rumar, K. (2002). *Night vision enhancement systems: What should they do and what more do we need to know?* (UMTRI-2002-12). Ann Arbor: University of Michigan Transportation Research Institute.
- Rumar, K. (2003). Night vision enhancement systems (NVES) – Research and requirements. In H. J. Schmidt-Clausen (Ed.), *Progress in automobile lighting – 5th International Symposium* (pp. 895–908). Munich, Germany: Herbert Utz Verlag.
- Schlesinger, M. (1995). *Infrared technology fundamentals*. New York: Marcel Dekker.
- Stahl, A., Oxley, P., Berntman, M., & Lind, L. (1994). The use of vision enhancements to assist elderly drivers. In *Proceedings of the First World Congress on Applications of Transport Telematics and Intelligent Vehicle-Highway Systems* (pp. 1999–2007). Boston: Artech House.
- Sullivan, J. M., & Flannagan, M. J. (2001). *Characteristics of pedestrian risk in darkness* (UMTRI-2001-33). Ann Arbor: University of Michigan Transportation Research Institute.
- Summala, H., Nieminen, T., & Punto, M. (1996). Maintaining lane position with peripheral vision during in-vehicle tasks. *Human Factors*, 38, 442–451.
- Tijerina, L., Browning, N., Mangold, S. J., Madigan, E. F., & Pierowicz, J. A. (1995). *Examination of reduced visibility crashes and potential IVHS countermeasures* (DOT-VNTSC-NHTSA-94-6). Washington, DC: U.S. Department of Transportation.
- Tsimhoni, O., Bärgerman, J., Minoda, T., & Flannagan, M. J. (2004). *Pedestrian detection with near and far infrared night vision enhancement* (UMTRI-2004-38). Ann Arbor: University of Michigan Transportation Research Institute.
- Tsimhoni, O., & Green, P. (2002). *Night vision enhancement systems for ground vehicles: The human factors literature* (UMTRI-2002-05). Ann Arbor: University of Michigan Transportation Research Institute.
- Tyrell, R. A., Patton, C. W., & Brooks, J. O. (2004). Educational interventions successfully reduce pedestrians' overestimates of their own nighttime visibility. *Human Factors*, 46, 170–182.

Sascha Mahlke is a research scientist at the Center of Human-Machine Systems at the Berlin University of Technology, where he received his M.S. in psychology in 2002.

Diana Rösler is a research scientist in the Department of Cognitive and Industrial Psychology at the Chemnitz University of Technology, where she received her M.S. in psychology in 2003.

Katharina Seifert is a project manager at Volkswagen AG Group Research, Wolfsburg, Germany. She received her Ph.D. in mechanical engineering in 2002 from the Berlin University of Technology.

Josef F. Krems is a full professor and the head of the Department of Cognitive and Industrial Psychology at the Chemnitz University of Technology. He received his Ph.D. in language understanding from the University of Regensburg in 1984.

Manfred Thüring is a full professor and head of the Department of General and Cognitive Psychology and a member of the scientific board of the Center of Human-Machine Systems at the Berlin University of Technology. He received his Ph.D. in causal models from Berlin University of Technology in 1990.

*Date received: May 3, 2005*

*Date accepted: April 1, 2006*