ABSTRACT

In a driving simulator experiment, effects were tested of an integrated support system on driving behaviour, user acceptance and workload. An alternative workload measure was used based upon peripheral vision. Two modes of support (tactile and speech messages) were compared to a control condition without support. Subjects were confronted with critical incidents dealing with lateral and longitudinal control on rural roads and motorways. The experiment showed that speech warnings are better suited for things related to law-enforcement, while tactile warnings are more suitable for situations directly related to driver safety. Although tactile warnings do result in compliance in enforcement type of situations, compliance is less than with speech warnings. For safety-related situations that do not require an immediate response, such as when approaching a sharp curve while speed is too high, speech warnings have the same effects as tactile warnings. However, if the criticality of the situation increases fast, for example when a lead vehicle brakes unexpectedly, tactile warnings result in more favourable effects than speech warnings. The Peripheral Detection Task showed to be sensitive to variations in workload that were caused by different driving situations and to variations caused by short lasting increases in workload as a result of attending to the system messages.

INTRODUCTION

Aim of Project

The research presented in this paper was performed as part of the European IN-ARTE project (Transport Telematics project TR4014 of the EU). IN-ARTE is an acronym for 'Integration of Navigation and Anti-collision for Rural Traffic Environment'. The aim of this project is to develop an in-vehicle system that integrates information from a number of input systems, such as a navigation system, anti-collision radar and CCD camera, and that supports the driver by presenting information in a safe way (IN-ARTE, 1998). The role of TNO Human Factors in this project was to contribute to the development of the warning strategies by means of a driving simulator study.

Visual, Speech or Tactile Warning Strategies

Within the project, warning strategies were developed. In the TNO driving simulator experiment, the effects of presenting these warnings on driver workload, driving behaviour and user acceptance had to be tested. In the project it was suggested that speech warnings may result in a lower workload compared to warnings on a visual display, since the visual input channel is already occupied with processing information related to the driving task. When a driver has to look at an in-vehicle display, he is unable to process some of the visual information present in the outside world. This may result in an increase in the number of errors and late detection of other traffic participants (Rumar, 1990). Presenting warnings in another modality, such as speech messages or tactile messages, would then result in less interference.

For the IN-ARTE system is was suggested that messages or warnings should not be presented exclusively on a visual display, although additional information concerning these messages, such as an explanation why the message is provided, could be presented visually. The idea was to present warnings for situations of immediate danger in a non-visual modality since they have to be presented even if the driver cannot take his eyes off the road. Speech messages combine well with visual information, because they depend on different resources (Wickens, 1984). Despite this, speech messages may induce high levels of mental workload depending on the complexity of the message (Verwey, 1996). Also, depending on how the speech message is designed, the presentation of the speech message may take too long to support the driver effectively. Because of these critical characteristics, the idea of also investigating tactile messages was raised. Especially in critical situations, tactile feedback may have advantages over speech messages. De Vos, Godthelp, Theeuwes and Verwey (1996) found that for a tactile lane keeping support system, drivers drove without immediate visual input for a longer period, which can be interpreted as a reduction in visual workload. In a driving simulator experiment, Janssen and Nilsson (1990) found that tactile feedback in a collision...
avoidance system resulted in a reduction of very short headways, while, in contrast to other types of Human Machine Interface (HMI), there were no negative side effects on driver behaviour. These results suggest that tactile feedback may have a positive effect on both workload and driving behaviour, especially in situations that require a fast reaction of the driver.

Because of these considerations concerning warning modalities, it was decided to include both speech warnings and tactile warnings in the experiment and investigate their specific advantages and drawbacks concerning critical driving situations.

Workload

When measuring the effect of presenting warnings, driving behaviour is not the only important variable. Measuring workload and peaks in workload is also important, since they may endanger traffic safety as well. Driving related workload is added to the workload resulting from, for example, warnings of an in-vehicle system. Especially the sudden increases in workload are potentially dangerous because they often cannot be predicted and therefore cannot be anticipated by the driver. These short-lasting peaks in workload are difficult to measure with the traditional workload methods, since most workload measures are aggregated over time, which means that they are not sensitive to short-lasting peaks in workload. Broadly, the following methods are distinguished:

1) Self-report measures such as the NASA-TLX (Hart & Staveland, 1988) and the RSMI (Rating Scale Mental Effort; Zijlstra & Van Doorn, 1985). The NASA-TLX is a multidimensional scale that integrates six different dimensions of workload into one general workload value. The RSMI is a unidimensional scale to rate the experienced mental effort on a line from 0 to 150.

2) Measures for task performance, either on the primary or a secondary task. Here, poor performance is interpreted as high workload.

3) Physiological measures such as the 0.10 Hz component of heart rate variability, with lower variability being interpreted as higher workload.

Secondary task measures have high potential for detecting variations in workload. The idea behind using secondary tasks is that they measure the ‘capacity’ remaining when performing a main task. If there is not enough remaining capacity for adequate performance on the secondary task, for example, because workload is high and thus most of the capacity is used for the primary task, performance on the secondary task will be poor. The high workload could also affect performance on the primary task, but usually the subject will try to keep performance on the primary task at a certain required level. Verwey and Veltman (1995) found that performance on secondary visual tasks was sensitive to variations in task demand. In a visual detection task, the subject was required to say ‘yes’ when he detected a target on a dashboard mounted display. The percentage of detected targets was found to decrease when a loading task was added to the driving task.

Another approach to workload is that with increasing workload the magnitude of the functional visual field decreases. During a driving experiment on the road, Miura (1986) presented spots of light on the windscreen under different horizontal angles. It was found that with increasing complexity of the driving task (higher traffic density) reaction times to detecting the lights increased. The results were interpreted as reducing the visual field of view with higher complexity of the driving task. Similar results have been reported by Williams (1985, 1995), who found higher reaction times to more peripheral stimuli with increasing workload.

In the driving simulator experiment, it was considered to be especially important to measure workload and the peaks in workload. Therefore, the Peripheral Detection Task (PDT) was developed for the study, based on the idea that the functional visual field decreases with increasing workload. While subjects were driving, a small red square was presented during one second on the simulator projection screen in the visual periphery of the subject. Subjects were required to respond to the appearance of the red square as quickly as possible by pressing a micro-switch that was attached to the index finger of the dominant hand. Reaction time (RT) was measured in ms. If a reaction was not detected within 2s from the onset of the stimulus, it was coded as a missed signal. With random variation between 3 and 5s, a stimulus appeared at a horizontal angle of 11 to 23° to the left of the line between the eyes of the subject and the centre of the screen. Stimuli were presented at a vertical angle between 2 to 4° above the horizon. The task required little conscious attention and could be performed without turning the head to the direction of the stimulus. This meant that subjects could just look at the traffic scene in front of them and still notice the onset of the red square in their visual periphery. A higher RT and a higher fraction of missed signals were interpreted as the result of higher workload.
METHOD

Driving Simulator

The experiment was performed in the driving simulator of TNO Human Factors. It consists of a computer that controls and monitors the experiment, interacts with the experimenter and other subsystems and handles the storage of data, a computer that handles the vehicle model and calculates the position of the simulated vehicle with the dynamic characteristics of a Volvo 240 and a computer generated image system (Evans & Sutherland ESIG 2000) that generates images with a refresh frequency of 60 Hz and an update frequency of 30 Hz. This system computes the visual scene from the position of the driver. The image is projected on a cylindrical screen with a visual angle of $120^\circ$ horizontally and $30^\circ$ vertically in front of the mockup by three high resolution Barcographics 801 projectors. The TNO driving simulator is described in detail by Hogema and Hoekstra (1998), and Hoekstra, van der Horst and Kaptein (1997).

IN-ARTE System Simulation

The idea behind the IN-ARTE system was to warn drivers for critical driving or traffic situations in order to increase traffic safety without increasing workload too much. In all experimental conditions (with speech messages and tactile messages), the IN-ARTE system was simulated and the support messages were stored on file. In the control condition, support messages were not presented to the driver, but they were virtually stored on file in order to analyse how many times a warning message would have been presented compared with the experimental conditions.

Data about the subject's lateral position, speed and position to other vehicles were monitored on-line, together with relevant information about the surrounding road infrastructure, such as local speed limits, curves, distance to the next intersection. These data were sent with a high frequency to a separate PC that applied activation criteria for warnings as they were defined inside the project (Bekiaris & Portouli, 1999). This resulted in warnings about speed and lateral position, taking the specific infrastructure and other traffic into account. Speech messages consisted of a voice telling the driver what to do (in relatively short messages) and tactile feedback was provided on the accelerator pedal (for speed-related warnings) and the steering wheel (for lateral position-related warnings). In addition to these messages, pictograms were presented on a display in the car to give extra information to the driver that he could attend to if desired.

Critical Scenarios

Each subject drove on a two-lane rural road and on a motorway, with curves and straight road segments. The following scenarios were defined as critical scenarios on the rural road:
- Road segments with curves (wide, medium, sharp)
- Intersections with stop signs for which a full stop had to be made
- Preceding vehicle that brakes unexpectedly
- Slowly driving vehicle that has to be overtaken
- Speed limit of 50 km/h
- Default scenario with speed limit of 80 km/h

The motorway database consisted of a 2x2 lane road, designed according to Dutch guidelines. The length of the route was approximately 50 km. The following critical scenarios were defined on the motorway:
- Approaching a traffic cue while driving in fog
- Unexpected braking of lead vehicle
- Vehicle cutting off the subject. This did not result in immediate critical situations because these vehicles were driving faster than the simulator car.
- Package falling off a truck (avoidance action required)
- Speed limit of 100 km/h
- Default scenario of speed limit 120 km/h.

Subjects

Altogether, 54 subjects participated. The number of subjects in the control condition was 18. The number in the tactile support condition 19, while 17 subjects participated in the speech support condition. Ages were between 23 and 56, and both males and females participated. All subjects had their drivers licence for at least 5 years with a minimum annual kilometrage of 5000 km.

Experimental design

Subjects were randomly assigned to either the control or one of the two support modes, resulting in a between-subject design. All subjects encountered the same critical scenarios. In the control condition, no support was given to the driver (although the IN-ARTE warnings that would normally have been provided were stored on file). In the tactile warning condition, tactile feedback was provided on the steering wheel or the gas pedal in case the IN-ARTE activation criteria were exceeded (in case the
situation was critical according to the IN-ARTE system). In the speech warning condition, voice messages were provided in case the same criteria were exceeded.

During the experiment, vehicle speed, lateral position and steering wheel angle were measured together with distance headway, defined as the time gap between the subject’s front bumper to the lead vehicle’s rear bumper. Lateral position was used to compute the standard deviation of lateral position (SDLP) and the fraction of time the centreline and right lane boundary were exceeded. For a number of scenarios, the minimum Time-to-Collision (TTC) was calculated.

Besides driving performance, performance on the PDT was measured, and after completing a ride on one of the road types, NASA-TLX was measures as well as the RSMI. For the rating of the driver acceptance of the IN-ARTE warning system, the scale developed by Van der Laan (1998) was used. This rating scale consists of nine dimensions that are combined into a ‘usefulness score’ and a ‘satisfaction score’. Together with this questionnaire the opinions about the activation criteria were measured.

**RESULTS**

Effects of within-subjects factors are tested with repeated measures analysis of variance, with message mode (Control, Speech and Tactile) as a between-subjects factor. In order to test differences between message modes, Tukey tests are applied where appropriate. Although road type (rural and motorway) was a between-subjects factor, differences as a function of road type were not tested since the scenarios and design of the rural road and motorway differed too much to make a comparison meaningful.

**Frequency of System Activation**

The idea of this dependent variable is that if more messages are provided, the driving performance is less safe. For both the rural road and the motorway, speech messages are most effective in supporting the lane keeping task, since with speech messages, the message ‘stay in lane’ is activated less often compared to the control condition and, for the motorway, the tactile condition. For critical longitudinal support both the speech and the tactile warnings are equally effective. For sharp curves, the speech messages result in fewer warnings, and thus better anticipation by the driver, compared to tactile or the control condition. For rural roads, both the speech and tactile condition are effective in reducing speed limit exceedances.

Based on these results it can be concluded that both speech and tactile warnings are effective in reducing the frequency of criteria activation. This means that drivers respond in a generally safer way. However, stopping for stop signs is an exception. On these scenarios the system did not show any improvement.

All in all, a lot of warnings were generated, which suggests that the criteria for system activation are too strict and must be relaxed in order to be useful and acceptable by drivers.

**Driving Behaviour**

**Speed** On the rural road, the speed limit was more often violated than on the motorway. Figure 1 and 2 show that speech messages were more successful in reducing driving speed to values closer to (or below) the posted speed limit. Figure 1 shows the speed for the rural road and Figure 2 for the motorway.

![Figure 1. Fraction of time that speed is lower than 80 (legal speed limit), between 80-90, between 90-100 and larger than 100 km/h on the rural road.](Image)
These results lead to the conclusion that both tactile and speech messages are successful in increasing the fraction of time that drivers drive below the speed limit. Especially on the rural road, only speech messages reduce the frequency of severe speed violations. Tactile warnings do result in some compliance, but they do not reduce the frequency of severe speed violations.

For curve negotiation, minimum speed in the curve was measured since this is a better indicator for safety than average speed. Figure 3 shows the effect of warning mode on minimum speeds in curves.

Figure 3. Minimum speed as a function of curve radius and feedback mode.

These results indicate that when speed needs to be reduced for negotiating a sharp curve, both speech and tactile messages result in comparable positive effects on driver behaviour.

**Longitudinal Measures** Minimum Time-to-Collision (TTC) was analysed for the scenarios in which a lead vehicle braked unexpectedly. Tactile warnings on the gas pedal resulted in the largest safety margin (largest minimum TTC), which suggests that tactile warnings are most suitable for presenting warnings in safety- and time-critical situations.

Accidents occurred on the motorway only. An accident was defined as an event where the subject’s front bumper touched the rear bumper of another vehicle, in case of driving into the traffic queue, or when the lead vehicle braked unexpectedly, or when the subject drove over the package that fell off the truck. A Chi-square test was performed to test whether the warnings resulted in fewer accidents and to examine any differences between speech and tactile warnings. This analysis showed that both tactile and speech warnings reduced the number of accidents, compared to the non-support condition. This means that the IN-ARTE system, as tested here, clearly had a positive effect in safety-critical situations. However, no differences were found between the speech and the tactile condition.

Standard deviation of lateral position (SDLP) and fraction of time that the right lane boundary or the centreline were exceeded were used as indicators for the quality of lane keeping. Both on the rural road and the motorway, SDLP was not affected by warnings. On the rural road, drivers tended to drive more often over the right lane boundary in the unsupported condition. The narrow lane width together with the oncoming traffic in the other lane may have resulted in driving closer to the right lane boundary in order to enlarge the safety margin to oncoming traffic. Both tactile and speech warnings to stay in lane resulted in a reduction of right lane boundary exceedances. The fraction of time the centreline was exceeded was not affected by warnings for both the rural road and the motorway.

**Driver Workload**

**Peripheral Detection Task** In order to assess the reliability of the PDT and thereby the usefulness for detecting peaks and variations in workload, the average RT and fraction of missed signals were computed for several driving situations. If more
complex driving situations result in larger RTs or higher fractions of missed signals then PDT can be considered as sensitive to variations in workload. Average RT and fractions of missed signals were compared with what was considered to be the easiest situation, normal driving on the rural road and the motorway. Figure 4 shows average RT in the selected situations for the rural road. The data for the misses showed exactly the same effects as the results of the RTs and will therefore not be discussed separately. Figure 5 does the same for the motorway.

Figure 4. Average RT on PDT for selected situations on rural road.

Figure 4 shows that driving in wide or intermediate curves or on stretches with a speed limit of 50 km/h does not result in increased RTs (nor in missed signals) compared to the baseline measurement (normal driving on a rural road). Driving in sharp curves resulted in deteriorated performance on RT (and a higher fraction of misses). Approaching an intersection, and especially if a full stop must be made, resulted in increased RTs (and misses). This also occurred when the driver overtook a lead vehicle or when this lead vehicle braked unexpectedly. The critical incidents on the motorway, especially when a lead vehicle braked unexpectedly or when a package fell off a truck (unexpected obstacle) resulted in a large increase in RTs (and in the fraction of missed signals) as is shown in Figure 5.

Figure 5. Average RT on PDT for selected situations on the motorway.

These results indicate that both RT and fraction of missed signals are sensitive to differences in driving situation. Also, the method is sensitive for measuring sudden peaks in workload since performance deteriorated with situations that require immediate action and that are characterised by sudden and unexpected changes in criticality.

It was tested whether feedback mode (speech versus tactile feedback) and presence of a warning affected workload. For this purpose, average performance on the PDT was measured during moments at which no warnings were given and during moments at which an IN-ARTE warning was presented to the driver. This constitutes the factor MESSAGE. Because speech warnings are longer in time than tactile warnings (because of a sentence that had to be spoken out loud), workload was determined during the 10 s after the moment the message started. This conforms with the presentation duration of 10 s for the messages on the visual display. Figure 6 shows the results of this analysis for RT. The results for the fraction of missed signals were similar.

Performance on the PDT strongly deteriorated when a message was presented compared to when no message was presented. This indicates that the warning messages resulted in increased workload during a short period of time. The effects of feedback mode and the interaction of feedback mode and MESSAGE are also statistically significant for both measures of PDT performance and for both types of roads. Speech messages increased workload compared to the control conditions, but the increases in workload due to the tactile messages are not statistically significant.
NASA-TLX The effects of message mode on dependent variables from the NASA-TLX showed no statistically significant effects. This indicates that the NASA-TLX was not sensitive enough to measure differences in any of the factors of workload between different feedback modes.

RSMI No significant differences were found between different feedback modes on the Rating Scale Mental Effort. On a scale from 0 to 150 these ratings indicate that the task was found to be mildly to rather strenuous. The results show that also the RSMI was not sensitive enough to measure differences in workload between different feedback modes.

Driver Acceptance

There were no statistically significant main effects of feedback mode on any of the items of the driver acceptance scale. However, for both the tactile and the speech conditions together, the system was evaluated as significantly more useful than satisfying. The general impression about the usefulness of the warnings was neutral to mildly positive. The satisfaction score, that measures how satisfied subjects were with the system, was mildly negative. The results show no clear preference for speech or tactile warnings. The results show a neutral opinion about the criteria, probably because the opinions were asked after completion of the experiment.

In a later experiment, questions about the acceptance of the IN-ARTE warnings were asked after every single warning (Martens & van Winsum, 1999). This resulted in the selection of activation criteria that gave good support and were acceptable. Even though subjects indicated they would not brake exactly at that moment in time, the warnings helped raise the attention level. In general, warnings for speed limit violations were not experienced as very supportive or acceptable, since subjects claim they are provided too early.

CONCLUSIONS

Overall, the speech and tactile messages reduced the frequency of speed violations. However, the frequency of severe speed violations was only reduced by speech messages but not by tactile messages. Yet, for speed reductions that were more closely related to driver safety, such as the minimum speed while negotiating a sharp curve, both speech and tactile warnings had the same beneficial effect.

These results can be explained by the different purposes of the messages of the IN-ARTE system. Some warnings are intended to result in higher compliance with traffic rules and are not directly related to traffic safety. Examples are warnings when the speed limit is exceeded and warnings to slow down for a stop sign, even if no other traffic is approaching the intersection. Other warnings are more explicitly related to driver safety, for example warnings for approaching a sharp curve at too high speed and warnings for driving too close to the lead vehicle. Speech warnings are better suited for things related to law-enforcement, while tactile warnings are more suitable for situations directly related to driver safety. Although tactile warnings do result in compliance in enforcement type of situations, compliance is less than with speech warnings. For safety-related situations that do not require an immediate response, such as when approaching a sharp curve while speed is too high, speech warnings have the same effects as tactile warnings. However, if the criticality of the situation increases fast, for example when a lead vehicle brakes unexpectedly, tactile warnings result in more favourable effects than speech warnings. Still, both speech and tactile warnings do have a favourable effect of driver safety in a general way, since they both reduce the accident frequency compared to the unsupported situation. The effect of the warnings on lane keeping performance is about the same for both speech and tactile warnings. It was found that giving support to the driver resulted in compliance with the warnings for both tactile and speech messages in about the same way. The number of warnings was reduced in a comparable way when speech or tactile warnings were presented. In general however, the warning frequency is too high. This means that the criteria are too strict and should be loosened. This would
probably result in a higher perceived usefulness and acceptability.

Performance on the Peripheral Detection Task was both very sensitive to variations in primary (driving) task demand, and to variations that were caused by the warning messages. It appeared that speech messages resulted in a temporary increase of workload that was quite severe. This increase in workload was comparable to the increase in workload that occurred if a lead vehicle braked unexpectedly. Tactile warnings did not result in a statistically significant increase in workload. From this it is concluded that, based on the workload measurements, tactile messages are clearly preferred over speech messages. Furthermore, it is advised that if speech messages are to be used, the length of the message should be as brief as possible. One word messages may result in less severe increases in workload.

The acceptance of the system was not different between tactile of speech messages. Thus, neither of these two modalities of presentation was preferred over the other. The opinions about the system in general or about the specific activation criteria were neutral on average, with the system not being judged as really useful or desirable. In the later study (Martens & van Winsum, 1999), the acceptability of warnings depended on the timing of the activation criteria, and it was possible to find acceptable and supportive warnings by adjusting some of the activation criteria.

In conclusion, it is advised to use tactile messages when possible, especially in situations that require an immediate response of the driver. If speech messages are to be used, it is advised to keep these messages as short as possible. It is advised to think clearly about the type of system that is to be developed. A driver support system for enforcement or improving law adherence has a different functionality and supports different situations than a system that focuses on driver safety. In the first case speech messages may be more appropriate, although tactile warnings have positive effects as well. Also, driver acceptance and perceived usefulness may be different compared to a driver safety oriented system. In a safety oriented system, workload is even more important, since safety-critical situations are often related to high workload induced by the driving task. In these cases, the workload induced by the warning of a support system should be as low as possible. This would favour the use of tactile messages instead of speech messages. Also, tactile messages are more directly related to the required action and, as the results indicate, result in higher safety in critical encounters compared to speech messages.

REFERENCES


