EVALUATION OF MANUAL VS. SPEECH INPUT WHEN USING A DRIVER INFORMATION SYSTEM IN REAL TRAFFIC

Ulrich Gärtner[1], Winfried König[2], Thomas Wittig[3]
Robert Bosch GmbH
[1]Hildesheim, Germany
[2,3]Schwieberdingen, Germany
E-mail: Ulrich.Gaertner@de.bosch.com

Summary: The executed study evaluated the influence of manual and speech input on driving quality, stress and strain situation and user acceptance when using a Driver Information System (DIS). The study is part of the EU-project SENECA. 16 subjects took part in the investigations. A car was equipped with a modified DIS to carry out the evaluation in real traffic situations. The used DIS is a standard product with manual input control elements. This DIS was extended by a speech input system with a speaker independent speech recogniser. For the use of the different DIS devices (radio, CD player, telephone, navigation) 12 different representative tasks were given to the subjects. Independently the type of task speech input needs longer operation times than manual input. In case of complex tasks a distinct improvement of the driving quality can be observed with speech instead of manual input. The subjective safety feeling is stronger with speech than with manual input. With speech input the number of glances at the mirrors and aside is clearly higher than with manual input. The most frequent user errors can be explained by problems when spelling and by the selection of wrong speech commands. The rate of speech recognition errors amounts on the average to 20.6 –% what makes it necessary to increase the recognition performance of the examined speech system. This improvement of system performance is the task of the development for the system demonstrator in the 2nd half of the SENECA project.

OBJECTIVES

The study was part of the EU-project SENECA (Speech control modules for Entertainment, Navigation and communication Equipment in Cars) in the 4th ESPRIT programme and under the Human Language Technology (HLT) [1, 2]. In detail the following questions were investigated:

- Are there differences between speech and manual input regarding road safety or driving quality?
- How good is the recognition performance of the speech input system?
- Does the driver have problems with the human machine interface of the speech input system (dialog structure, speech syntax)?
- How is the acceptance of the speech input system and is it possible to observe learning effects?

METHODS

A car was equipped with a DIS and an additional speech input system. The speech input system used a speaker independent continuos word recogniser which was developed in the SENECA project. To investigate the mentioned objective a multitude of dependent variables were analysed which can be assigned to the following topics: operating times, driving quality, mental workload, glance behaviour, speech recognition errors and user errors when using speech input system. Data were collected with a protocol file on a laptop containing all manual and speech inter-actions with the DIS, an audio/video recording with three cameras, notes of a driving instructor and interviews with the subjects. The trials took place on a route nearby a middle size town. The course was composed of express roadways, highways, driveways and streets across villages. The subjects drove the course twice to cover both experimental conditions (speech and manual mode). Reference segments without any operation tasks were mixed with the test segments and served as basis for comparison. For the main experiment 16 subjects were recruited, 13 were male and three female. They were experienced in driving luxury cars and in using common infotainment techniques. They drove at least 10.000 km/year and were safe drivers. In addition, three subjects were recruited for the pre-test.

For the use of the different DIS devices (radio, CD player, telephone, navigation) 12 different simple and complex tasks were given to the subjects. They contained representative operation actions: Activating a main or sub function, tuning parameters, partial spelling, choosing out of a list and entering connected digits. In case of a simple task, i.e. audio functions, 2 or 3 input steps are needed; complex tasks, especially when using telephone and navigation, need 6 to 8 input steps. The experimental design considered in addition to the independent variable "input mode (manual, speech)" the independent variables "route complexity (low, high)" and "task complexity (low, high)" which were varied systematically. So the experimental design consisted of 2 modalities, 2 route complexities and 2 task complexities, each combination with 3 tasks. The route complexity is a function of needed driving manoeuvres. The reference mode was mixed with the test mode, i.e. particular 12 parts of the course were defined as test segments with manual or speech input operations and 5 parts were defined as reference segments without operation tasks. The complete trial comprised written and verbal instructions, a pre-experimental questionnaire, a training of manual and speech input operations and driving, the test trial with intermixed reference segments, interviews and a post-experimental questionnaire. A complete experiment took about 3 hours. For speech input the relevant vocabulary was reviewed with the subject. The subjects trained system operations mainly in the stationary car. The main intention of the training trial was to become familiar with the car and the experimental conditions. The subjects were not assisted during a task or reference segment.

RESULTS

The evaluation in the SENECA project was carried out in three countries (Germany, Italy, France). The used methods for evaluation were the same in all three countries. The presented results are based on the German evaluation [3] but equivalent with the Italian and French ones.

Recognition ~~errors~~Errors

The overall recognition error rate was 20.6 % (figure 1~~Fehler! Verweisquelle konnte nicht gefunden werden.~~). Every 5th input was rejected, substituted etc. This conflicts with the subjectively tolerated error rate of mostly 10 – 15 %. I.e. in each 3rd of the simple tasks one recognition error occurred. In complex tasks 1,5 - 2 recognition errors occurred. The relatively low recognition error rates of navigation and dialling do not sufficiently reflect the problems of those complex tasks. While most of the control commands were recognised correctly, the cities and streets were often not recognised

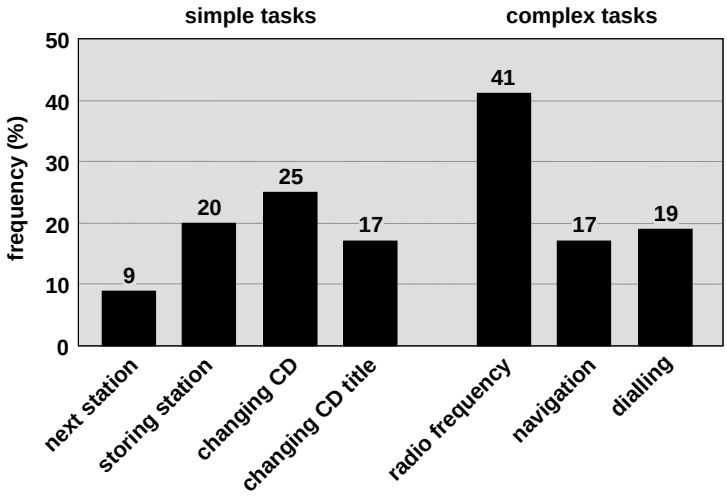

Figure 1: Recognition errors of speech input system

~~were often not recognised~~ correctly. The actual weather situations during the trials had a very important effect on recognition errors: While there were 12.5 % errors without rain (sunny or cloudy), the error rate amounted to 36.6% with rain, both for task 2 (navigation). During the operation, the wind-screen wipers mostly squeaked on the glass and caused insertion and substitution errors. There is a correlation between recognition errors and driving quality, i.e. more recognition errors resulted in more driving errors and in a worse driving mastery. Recognition errors distracted the subjects from their driving task (looking to the display to identify the recognition error and to verify). Subjectively, the most disturbing errors were the substitutions and omissions, while rejections were tolerated to a certain degree.

Operation ~~times~~Times

The speech inputs took considerably more time than the manual inputs. I.e. the speech input times are underestimated. Most of the complex speech inputs were followed by time consuming speech output, and verifications needed additional time. Those factors dominated the time consuming sequential operations and observations with manual input. Complex tasks lasted considerably longer than simple ones. For speech input this reflects relatively exactly the relation between complex tasks (6-8 steps) and simple tasks (2-3) ~~concerning~~ concerning input steps, if no verifications are considered. For speech input needed longer than manual input for simple as well as for complex tasks (figure 2~~**Fehler! Verweisquelle konnte nicht gefunden werden.**~~).

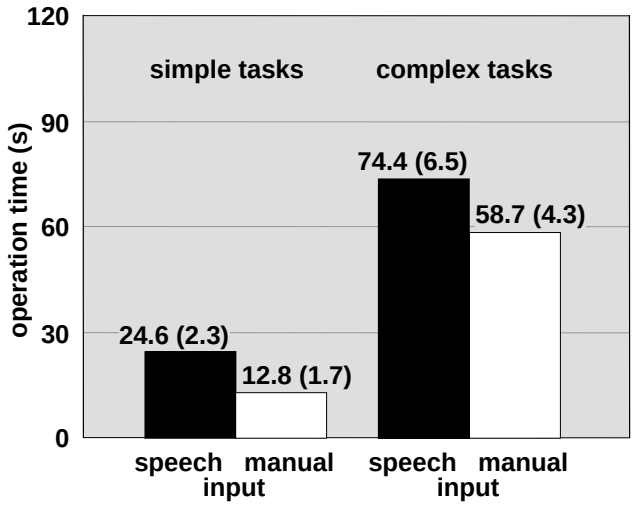

Figure 2: Operation time for speech and manual input

User Eerrors

On average one user error occurred in each task. The user error rate depends on task complexity (figure 3~~Fehler! Verweisquelle konnte nicht gefunden werden.~~). Most of the user errors were made by the older subjects. Concerning the complex tasks, most errors were within the navigation and dialling task. The problem in the navigation task was to follow the long sequence of 8 actions. The most frequent errors were spelling problems, entering or choosing the wrong street and using wrong commands. The most frequent user errors during the dialling tasks were stops within digit sequences, which led to recognition errors. From time to time a misrecognition was not detected by the subjects, which is a hint that a pure spoken feedback is prone to attention and remembrance problems. Concerning the kind of user errors (independent on task complexity), most of the user errors were vocabulary errors, then orientation/breaking off errors and "Push To Talk (PTT)" errors. Words from another context were used, which did not work in the actual context. Other errors were activating a device explicitly, which was not always possible. Two old subjects repeated human expressions frequently. Several errors were commands, which are used as synonyms in the everyday language, but had not been implemented in the SENECA concept demonstrator.

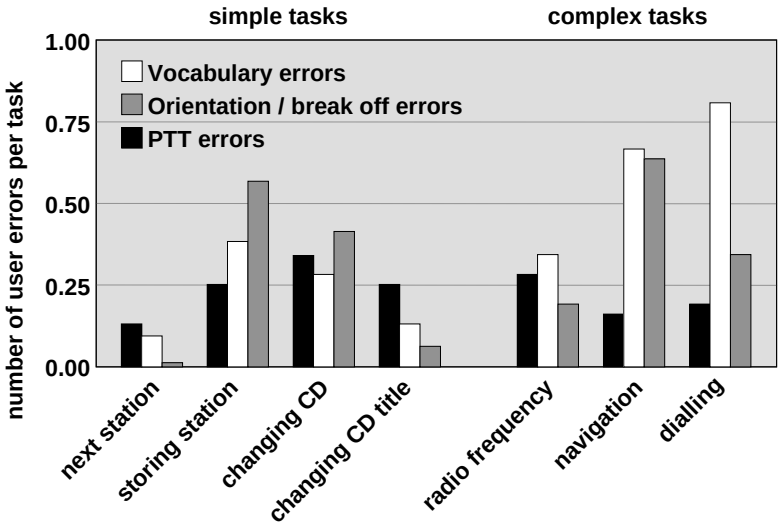

Figure 3: User errors for different simple and complex tasks

Driving ~~quality~~Quality

31 different driving errors were classified into 8 main criteria (criteria A-H, see figure 4). For two of the eight driving error classes there are significantly less errors with speech input in comparison with manual input (figure 4). "Poor lane keeping" is the most frequent kind of error for both modalities, i.e. the mental, visual and partly manual diversion by the additional operation tasks is reflected mainly in a reduced visual-motor control of the car. With manual input there were more errors of "poor lane keeping", less traffic observations and more often "speed too low". Thus, drivers try to compensate their additional mental load by reducing their speed, which is more marked with manual than with speech input.

If relating the driving errors to input time, there is a much more pronounced difference between modalities: During the (shorter) manual inputs there were twice as many driving errors (3,8/min) as during the (longer) speech inputs (1,9/min). While there is no marked difference of driving errors between speech input and the equivalent reference segments, there is a considerable difference with manual input (1,7 more driving errors per min during test). I.e. the usual driving quality level can be kept up additional operations are done by speech, not, however, by hand.

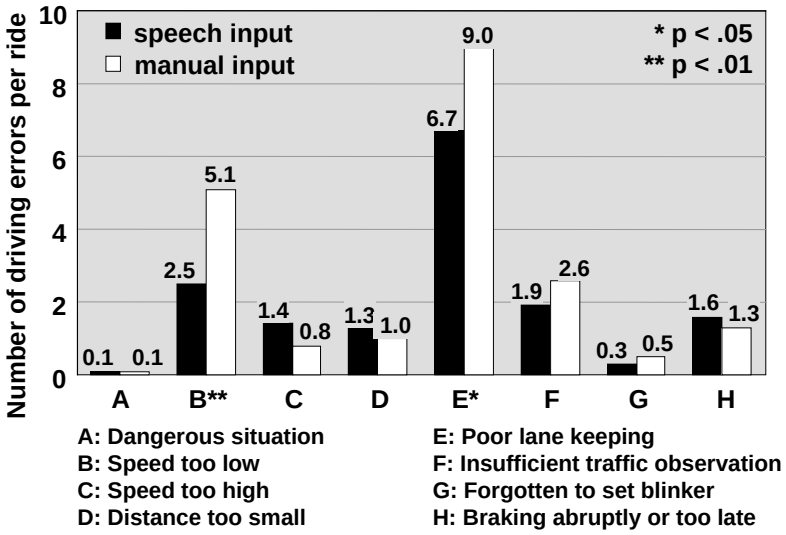

Figure 4: Driving errors for different error classes

Subjectively ~~felt~~ Felt ~~strain~~Strain

The subjectively felt strain was on the average nearly identical for both modalities. This result indicates too, that the strain posed by the additional operation of the system was usually felt by the subjects as relatively low - independent of the input mode. Most of the tasks, even with some amount of user and system errors and in different traffic situations, can be performed without a too high subjectively felt strain.

Glances

From the video data glances to the display, speedometer/steering wheel, rear mirror, aside (including outside mirrors) were extracted for 4 subjects. Short glances (below 1 s) and long glances (more than 1 s) were separated, because 1 s is often considered as a critical visual absence time during traffic observation. For simple tasks the total numbers of display glances per task for speech and manual input are equally about 9 (figure 5). These are surprisingly many glances, though simple tasks needed not more than 3 inputs (if entered and recognised perfectly). For complex tasks speech input needed a less number of shorter glances and longer glances than manual input (figure 5). This result is due to the fact that speech input required just a few integrated inputs while manual input required a lot of single actions. The navigation task (without any errors) needed typically 1 manual action and 7 utterances, most of them were feed back acoustically. Manual input for the navigation task, however, required about 20 manual actions (without user errors). The dialling task had an exclusively spoken dialogue. Therefore, the glances for complex tasks reflect the inherent quality of speech input as well as the spoken feedback within the speech dialogue. The advantage of speech input for complex tasks as to glances onto the display is marked in spite of the speech recognition errors, which needed extra glances for orientation.

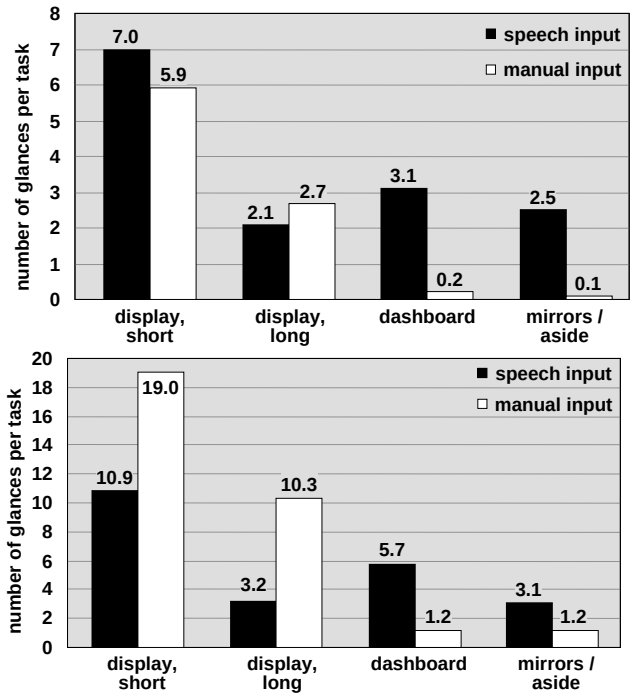

Figure 5: Glances during simple (left side) and complex tasks (right side)

CONCLUSIONS

Based on the results of the investigation it is possible to point out the following conclusions:

- The input with a DIS should be possible both manually and with speech.
- With speech input there is a chance to improve road safety, especially in case of complex tasks.
- The feeling of being distracted from driving is smaller with speech input than with manual input.
- The recognition performance of the examined system has to be increased for complex tasks.
- The most frequent user errors can be explained by problems when spelling and by the selection of wrong speech commands. Future development has to improve the speech dialog and speech syntax. This leads to the conclusion that the main focus for speech recognition within the car has to be on the design of the human machine interface. [3, 4]

ACKNOWLEDGEMENT

The SENECA project is supported by the European Commission within the 4th framework of ESPRIT in the so called program "System Integration and Applications" and under Human Language Technology under the Contract number ESPRIT 26 981. The authors wants to thank all the partners involved in the project. [5]

REFERENCES

[1] Project Program SENECA, Technical Annex for "Speech control modules for Entertainment, Navigation and communication Equipment in Cars Project number ESPRIT 26981, 1998

[2] Gärtner, U.: The SENECA Project: Speech Recognition within the Car for Entertainment and Communication Systems, Detroit Auto Interior Show 2001

[3] Mutschler, H; Baum, W; Waschulewski, H.: Report of evaluation results in German, SENECA D22-3.1, 2000

[4] Mangold, H.: Usability und Robustheit von Spracherkennungssystemen als Voraussetzung für eine breite Vielfalt von Anwendungen, KONVENS 2000

[5] Reference: http://www.seneca-project.de/