Evaluating Misinterpretations during Human-Machine Communication in Automotive Environments

Frank Althoff, Gregor McGlaun, Björn Schuller, Manfred Lang, Gerhard Rigoll

Institute for Human-Machine-Communication
Technical University of Munich
Arcisstr. 16, 80290 Munich, Germany
phone: +49 89 289 28538

{althoff,mcglaun,schuller,lang,rigoll}@ei.tum.de

ABSTRACT

In this work, we present the results of a user study conducted to evaluate the potential of misinterpretations while interacting with a multimodal user interface in an automotive environment. Using classical haptic input combined with natural speech and dynamic gestures, the user has to control various in-car comfort facilities and simultaneously performing in a driving task. Thereby, the test subjects are confused by various perturbations and ambiguous task descriptions. Multiple features give rise to an increased potential of misinterpretations. Amongst others, untypical system commands are used, the general interaction behavior changes and the use of various modalities massively increases. Supervising the individual voice characteristics and the display attention frequency proved to be important resources. Concerning help dialogs, the test subjects expect individually adapted strategies taking into account the current system context, driving situation and personal preferences. In most cases, a brief description of possible alternatives is preferred.

Keywords: Human errors, misinterpretations, automotive environment, multimodal user interface, error handling.

1. INTRODUCTION

The user workload in an automotive environment is influenced by various factors. First and foremost, the primary task of the driver is to be in entire control of the car at any point of time. Second, he must be prepared to quickly react to external events like changing sight conditions, obstacles on the road and the behavior of the passengers. Third, the handling of various in-car comfort facilities like audio- and telecommunication applications requires additional attention. With regard to a save driving performance, the basic requirement in the design of such in-car devices is to induce as little additional distraction potential as possible. In the course of time, the complexity of automotive information systems has significantly increased, but the majority of the currently available applications integrated in automobiles are still mostly restricted to haptic interaction i.e. in the form of buttons and sometimes touchscreens. These interfaces can be worked with very effectively, but, on the downside, they require extensive learning periods and adaptation by the user to a high degree. More advanced interaction styles, such as the use of speech and gesture recognition as well as combinations of various input modalities, can only be seen in dedicated applications thus far. To overcome these limitations the aim of current projects is to make the interaction more flexible, intuitive and robust with regard to occurring errors. In our overall research work, we concentrate on the development of various error handling strategies to cope with both system and user errors. An important issue in this context is devoted to detecting and managing misinterpretations that might occur during the human-machine communication.

2. FORMAL DESCRIPTION

Before developing any appropriate error management strategies, a fundamental task is to understand the nature of both human and system errors and clearly identify the difference between an error and a misinterpretation.

Theoretical Background

Strictly following an absolute philosophical point of view, Festinger[1] has developed a theory of cognitive dissonance for describing user errors. In his model, a human error is always an expression of certain habits that cannot automatically be used in specific situations and thus result in an error during the operation. Rigby[2] differentiates between sporadic, accidental and systematic

Levels (Rasmussen)	Skill-based (SBL)	Rule-based (RBL)	Knowledge-based (KBL)
Description	routine actions	productions	analytical processes
Error - type (J. Reason)	slips, lapses	planning-failure (stored rules)	planning-failure (novel situation)
Causation	deviation from a trained routine	misclassification of situations	unpredictable changes

Table 1: Classification of errors [3]

errors. In his phenomenological approach, sporadic errors are singular events and are often considered as outliers. Accidental errors have a high mean variation with regard to the intended target status, but, in contrast to systematic errors, they do not show any unambiguous tendency towards a special direction. Unfortunately, these two approaches cannot be used in a practical application since they suffer from a significant drawback. As the flow of interactions is assumed to be controlled by the system exclusively, the user is not involved sufficiently.

The theoretical basis for modeling potential error-prone user interactions has been developed by Reason[3]. Related to the skill-rule-knowledge framework of Rasmussen[4], he differentiates between errors on three different performance levels: the skill-based level (SBL), the rule-based level (RBL) and finally the knowledge-based level (KBL). As shown in table 1, user interactions at the SBL comprise actions, which have already become routine by multiple execution. Errors at this level, which are either execution failures (slips) or failures of memory (lapses), imply a deviation from a well-trained routine. At the RBL, human performance is determined by stored rules (productions). Errors at this level are planning failures (mistakes) and typically related to the misclassification of situations. Finally, at the KBL, in novel situations, problems are solved by applying conscious analytical processes and stored knowledge. Here, mistakes arise from unpredictable changes in the environment that one is not prepared for.

Interaction Errors

Based on the formal description and classification of human errors discussed above, we will derive a practicable definition of an interaction error that additionally handles system failures and faults. In the following, we briefly list some prototypical error-prone situations in human-machine communication that have to be covered by the definition to be of any practical use. In the first case, the user gives a command, which is interpreted by the system in a certain context that does not match the primary intention of the user. Second, if feedback information is interpreted in the wrong by the user who then reacts surprised or in an unexpected way. Finally, if in the course of multiple interactions the mental model of the user (more precisely: the task model combined with the system model) and the user model of the system differ and, thus, bilateral information is interpreted in the wrong context. Thereby, the significance of the errors becomes higher, the later the existing divergence of the two models is detected.

Taking into account both the theoretical background and the individual mentioned error cases, we will define an interaction error as follows: *An error in human machine communication is a consequent result, if the requirements and the intention of the acting part are not covered in a sufficient way by the reacting part.* Thereby, the acting part can be both the system and the user. Finding a measure that estimates a threshold for fulfilling the requirements is the central part of this work.

Misinterpretations

In inter-human communication, misinterpretations can emerge in various situations, i.e. if the concentration of one partner is reduced by external factors like stress, weariness or abrupt distractions. Moreover, the risk can be increased by giving ambiguous propositions or while communicating via different modalities, i.e. a certain topic is described both orally and in form of a sketch by one partner, but the other person does not look at it and, thus perhaps misses important context information.

Concerning misclassifications during human-machine communication, the difference with regard to a normal error situation has to be worked out. Both errors and misclassifications result in an unintended behaviour. The unexpected occurrence of an interaction error at least presupposes an error-prone system or situation. Otherwise, a potential malfunction neither can occur nor be detected. On the other side, a misinterpretation is a result of a certain system context that might be influenced by various both internal and external factors and events. As especially these events are responsible for the misinterpretation, in general, we have to consider the underlying system as being functional correct and, additionally, the user as acting absolute correctly. Thus we simply have to extend the former definition of a interaction error by this assumption: *Provided that both the system and the user are acting absolute correctly, a misinterpretation in human-machine communication is a consequent result, if the requirements and the intention of the acting part are not covered in a sufficient way by the reacting part.* Same as for the definition of an interaction error, the acting part can be both the system and the user.

3. EXPERIMENTAL SETUP

The primary goal of our user study is to evaluate the potential of misinterpretations while operating a multimodal user interface for controlling various audio- and communication devices in an automotive environment. With regard to this application scenario, the

Figure 1: Working environment

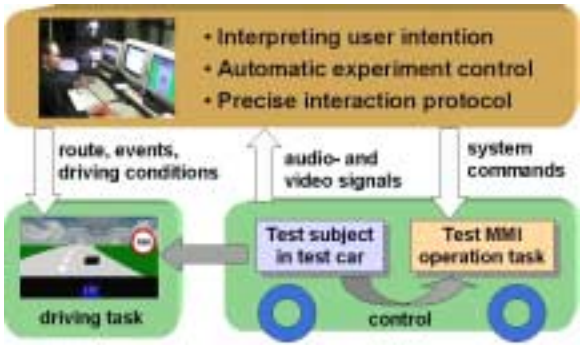

Figure 2: Experimental setup

first task is to identify measurable parameters that might indicate upcoming misinterpretations. Moreover, we want to find out by which specific strategies misinterpretations can be resolved or, even better, totally prevented. In this context, we look for specific strategies in initialising help dialogs. As in a real-world scenario, the primary task of the test subjects is to perform in a driving task and the secondary task is to operate the in-car devices either by classical haptic interaction or by semantically higher-level modalities like speech and hand or head gestures. Diverse operation errors are provoked by increasing the cognitive workload of the test subjects, i.e. they are confronted with highly complex and often ambiguous operation tasks, various perturbations in the form of both visual and acoustical effects, changing sight conditions and obstacles on the road, which they have to evade. Based on these events, misinterpretations are provoked by strongly confusing and mentally stressing the user.

Test Interface

In our experimental setup, the test application is a multimodal interface for controlling audio devices (MP3-player and radio) and standard telecommunication tasks. The player module provides well-known CD-player functionalities (*play, pause, stop, skip, random/repeat*, etc.). In radio mode, the user can switch between 25 different predefined radio stations. The telephone functions are restricted to basic call handling (*call, end, accept, deny, hold*, etc.) of predefined address-book entries. Moreover the volume of the audio signal can be controlled in a separate mode.

As shown in figure 1, the interface can be operated by a touchscreen and a special key console box. Thereby, the interface itself is organized in four separated horizontal areas. The top line is composed of four buttons representing the individual modes of the application (*mp3, radio, telephone*, and *control*). Directly beneath this button line, as the central design element, the interface provides a list containing individual items that can vertically be scrolled through by the two buttons on the right. The area in the in the lower part contains context specific buttons varying from five buttons in mp3 mode, three in radio and control and two in telephone mode. The appropriate functionality is given by the system and the interaction context. In addition, the last line of the interface contains a feedback line continuously informing the user of the current volume and the status of the interface, e.g. indicating the name of an incoming call connection or additional information for the current radio station or the mp3 song that is currently played. The key console box, which is located near to the gear stick of the test car, provides various buttons that are organized in direct analogy to the layout of the buttons on the touchscreen. In addition, the console provides two special knobs: one for adjusting the volume and the other for browsing in the list display. By pressing these knobs, the volume is muted or the current list item is selected, respectively.

Test Environment

The user study is carried out at the usability laboratory of our institute, that has specially been adapted to evaluate multimodal user interfaces in automotive environments. To simulate realistic conditions in non-field studies, the lab provides a simple driving simulator, consisting of a specially prepared BMW limousine with force-feedback steering wheel, gas and break pedals, as well as a gear stick. The probands have to use these devices to control a 3D driving task, which is projected on a white wall in front of the car. Thus, they can experience the driving scenario from a natural “in-car” perspective and better anticipate the roadway. The individual parameters of the simulation can fully be controlled, e.g. the degree of the curves, day- or night sight conditions, speed regulations, obstacles or passing cars. For interacting with the interface, the test car contains a 10”-touchscreen, a special key console and additional buttons on the steering wheel. Furthermore, the car is equipped with a number of microphones and cameras to supervise the test subjects.

The audio- and video signals from inside the car are transferred to a separated control room that serves for recording and analyzing the user interactions with the test interface and the driving performance. To carry out identical reproducible test runs, we have developed a special software suit[5], simplifying the management of the various system parameters, semi-automatically announcing the operation tasks at specified points of time and logging all kind of transactions. The concept has successfully been applied in various experiments[7].

Test Methodology

As shown in figure 2, the functionality of the interface is partly realized according to the so-called *Wizard-of-Oz* test paradigm[9]. In contrast to the haptic user interactions by the touchscreen and the key console that are directly transcribed by the system, the recognition of the semantic higher-level modalities (speech, hand and head gestures) is simulated by a human person supervising the test-subjects via audio- and video-signals in the test room. The so-called 'wizard' interprets the users intention and generates the appropriate system commands, which are sent back to the interface in the car to trigger the intended functionality. Thereby the wizard is instructed to be extremely cooperative. In the case of ambiguous user actions, the interaction is to be interpreted at best in the current system context. By following this approach, we can absolutely guarantee that no additional error potential is introduced by randomly distributed malfunctions of the recognition modules. Moreover, as the driving simulation demands each test subject in a different way, we have developed a baseline measurement that, when applied to the driving task, makes sure that each subject is exposed to the same cognitive load concerning the operation of the interface that is independent of personal precognitions[6].

Test Plan

Before starting the real test run, the test subjects are familiarized with the functionality of both the driving task and the audio and telephone interface in an extensive, interactive training period together with the wizard, mainly by using the haptic input devices touchscreen and key console. At the same time, the use of natural speech and gestures as well as potential combinations of the individual modalities are explained. Although, in general, the subjects are free to use their own speech and gesture vocabulary, certain possibilities are explained, that have already proven to be meaningful[8]. For example, making a wiping move with the right hand stands for skipping in the playlist or shaking the head can be used to deny an incoming telephone call. To ensure that the probands take the driving task seriously, they are told that their reward at the end of the test depends on their driving performance.

The real test session consists of three different stages. In the first stage (*learning phase*), the test subjects learn to operate the interface under realistic test conditions. On the background of a very simple driving task, they can devote most of their attention to fulfill 15 different tasks. This phase normally last for eight to ten minutes.

In the second stage (*reference phase*), the test subjects have to accomplish 27 different operation tasks while confronted with a much more complex driving simulation (obstacles on the road, speed checks, changing boundary conditions, etc.). Additionally, the drivers are distracted by incoming telephone calls and by certain visual and acoustical effects. This phase normally last twice as long, typically from 14 to 17 minutes. Its main purpose is to analyze user behavior with respect to a perfect interface.

In the third stage (*test phase*), which last as long as the reference phase, the test subjects have to deal with the same operation tasks, but the difficulty-level of the driving simulation has noticeably been increased once more. Moreover, misinterpretations are explicitly provoked and various strategies are tested to resolve these problematic situations. Additional questionnaires after each part help to evaluate subjective user experiences.

4. RESULTS

A total of 19 persons participated in the test trials with 15 male and four female subjects. The average age was 32 years with a majority of people being younger than 30 years. Besides a couple of engineers, people of different education and profession took part in the tests. Most of the subjects were technically highly skilled and stated to drive cars on a regular basis. All subjects except for one were familiar with the MP3 audio format. When selecting the test subjects, great emphasis was put on the fact that nobody had any prior experiences with the interface resulting from similar tests in the environment.

Video Analysis

To evaluate user behavior shortly before, during and after a provoked misinterpretation, the video material was both quantitative and qualitative analyzed. Thereby, the focus was put two aspects. First, we wanted to know, if under the extreme stress conditions in the test phase, the test subjects still apply the same interaction style as learned in the reference phase and second, if certain patterns can be observed directly after a misinterpretation.

Multiple features can give rise to an increased potential of misinterpretations. Most and foremost, the visual attention towards the interface massively advances. Thereby, it could be observed, that, in general, people do not use one long check, but instead use multiple glances in short periods of time. Compared to the reference phase, the average frequency of glances was nearly three times higher in the test phase. Moreover, the test subjects showed a strong tendency to use emotional interjections or astonished question constructions. When repeating a command, the voice characteristics changed significantly, i.e. the rate of speech slowed down or the voice became louder. Additionally, alternating command phrases have been observed with a tendency towards using shorter constructs. If the system did not react in case of a second oral command repetition, nearly 70% of the test subjects change the modality and made use of the touchscreen.

In general, the test subjects behaved strongly uncertain and hectic after a system-induced misinterpretation. Amongst other features, untypical system commands were used, the interaction behavior changed and the use of various modalities massively increased.

Help dialogs

When a misinterpretation is provoked, the test subjects are exposed to an increased cognitive workload, which lead to uncertain and hectic user behavior and, as a result, to a poor driving performance. Introducing error-handling dialogs helped the subjects to calm down and to better concentrate on the primary task of driving the car again.

Two different dialog forms have been evaluated. The first approach consists of a long help dialog that guides the user through the complete command hierarchy of the system. By using context information, the second approach simply lists the currently available options and thus facilitates shortcuts in the command hierarchy.

The time elapsing from the end of the task announcement to the help dialog was 25.2 seconds(s) on the average. The test persons looked at the display for about 6.8s, resulting in a relative display attention (RDA) of nearly 27% of the whole time. The long help dialog itself lasted about 72.6s and the RDA during this time was only 5%. This clearly indicates, that in the case of the dialogs, the test subject could devote much more attention to the driving task since they were not looking at the display that frequently and thus also made less driving mistakes. Concerning the short help dialog, this observation could be confirmed. With an average of 26s for the dialog and 3.5s average display time, the RDA was 13.5%, still significantly less compared to the average value without any help (average task completion time of 37s with 14.1s display attention resulting in a RDA of nearly 40%).

As an important result we found out that, on the average, the same amount of total time is needed if, first the user explicitly demands for a help dialog and then solves the task or second, the user directly gives a command and the system assists him in case of an ambiguous interaction with a context specific list of potential alternatives. Omitting the help dialogs resulted in task completion times that were about 10 seconds longer on the average and, as already stated, with a worse driving performance.

Subjective user experience

Concerning the available input modalities, most of the test subjects preferred speech compared to gesture and haptic input, which confirmed the results of various prior studies[7]. Only for selected functionalities like adjusting the volume, the majority preferred the key console.

Concerning the help dialogs, the test subjects expected individually adapted strategies taking into account the current system context, the driving situation and personal preferences. In 74% (14 out of 19 test persons), a brief description of possible alternatives was preferred with an option for further explanations if requested by the user. When asked, at which point of time a help dialog should be initiated by the system, all subjects were of the same opinion. The system should mainly react in cases of ambiguous commands or when several options are possible, e.g. when multiple entries in the telephone list match the given user utterance.

5. CONCLUSIONS AND FURTHER WORK

We have conducted a series of usability studies to evaluate user behavior when operating various in-car comfort devices and simultaneously performing in a 3D driving task. To summarize the results, with regard to the detection of a potential misinterpretation, the voice of the test subjects proved to be the most important knowledge resource. By applying context specific help dialogs, the task completion time could significantly be reduced.

Currently we are working on the design of a dedicated model, measuring the divergence of the users mental model and the current functional system status. Fusing the individual parameters identified in test into a multidimensional threshold vector, the point in time for initiating help dialogs can better be estimated. For the future, we plan to run more usability studies, especially evaluating user preferences and modality changes when operating with a multimodal interface.

6. ACKNOWLEDGMENTS

The presented work has been supported by the FERMUS project, which is a cooperation of the BMW Group, DaimlerChyrsler, SiemensVDO and the Institute for Human Machine Communication at the Technical University of Munich. The project name FERMUS stands for error-robust multimodal speech dialogs.

7. REFERENCES

[1] L. Festinger, “A theory of cognitive dissonance”, Stanford University Press, 1957

[2] L. Rigby, “The nature of human error”, In Annual technical conference transactions of the ASQC, 1970

[3] J. Reason, “Modeling the basic error tendencies of human operators”, Reliability Engineering and System Safety, 22, pp 137-153, 1988

[4] J. Rasmussen, “Skills, rules, knowledge: signals, signs and symbols and other distinctions in human performance models”, IEEE Transactions: Systems, Man & Cybernetics, SMC-13, pp 257-267, 1983

[5] R. Nieschulz, B. Schuller et al, “Aspects of efficient Usability Engineering”, Journal IT+TI Vol 44, 2002

[6] G. McGlaun, F. Althoff et al, “A new technique for adjusting distraction moments in multitasking non-field usability tests”, Proceedings of CHI 2002

[7] F. Althoff, K. Geiss, G. McGlaun, “Experimental evaluation of user errors at the skill-based level in automotive environments”, Proc. of CHI 2002

[8] M. Zobl et al, “A Usability Study on hand-gesture controlled operation of in-car devices”, Proceedings of Human-Computer-Interaction (HCI), 2001

[9] Jacob Nielson, “Usability Engineering”, Morgan Kaufmann Publishers Inc., San Francisco, 1993