Experimental Evaluation of User Errors at the Skill-Based Level in an Automotive Environment

Frank Althoff, Karla Geiss, Gregor McGlaun, Björn Schuller, Manfred Lang
Institute for Human-Machine-Communication
Technical University of Munich
Arcisstr. 16, 80290 Munich, Germany
{althoff,geiss,mcglaun,schuller,lang}@ei.tum.de

ABSTRACT

Concentrating on the lowest performance level of Reason´s error model, in this work we evaluated the potential of user errors in an automotive environment. Thereby the test subjects had to operate various in-car devices while primarily fulfilling a simulated 3D driving task. The study clearly showed that special error types are related to special distraction effects and that most of the operation errors are not critical and can be resolved by the user himself.

Keywords

Usability study, automotive application, user errors

INTRODUCTION

In the course of time the complexity of information systems has significantly increased, requiring extensive learning periods and adaptation by the user to a high degree. To overcome these limitations the aim of current research projects is to make the interaction more flexible, intuitive and robust. But before dealing with system errors and the appropriate error management strategies, a fundamental task is to understand user behavior.

Classification of User Errors

The theoretical basis for modeling potential error-prone user interaction is given by Reason[1]. Related to the skill-rule-knowledge framework of Rasmussen, 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). 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 (normally known in advance) 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 missclassification 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 one is not prepared for.

EXPERIMENTAL SETUP

Concentrating purely on the SBL, the primary goal of our study is to evaluate user reactions due to various distraction effects while operating audio- and communication devices in an automotive environment. In this context we want to determine the classes of occurred user errors and the frequency of occurrence. As in a real-world scenario, the primary task of the test subjects is to fulfill a driving task and the secondary task is to operate the in-car devices by haptics, natural speech, and both hand and head gestures. Operation errors are provoked by various perturbations in the form of both visual and acoustical effects, massively changing sight conditions and obstacles on the road.

Test Environment

The user study is carried out at a usability laboratory, that has specially been adapted to evaluate user interfaces in automotive environments. To simulate realistic conditions, the lab provides a simple driving simulator of our institute, consisting of a specially prepared BMW 750 with force-feedback steering wheel, a 10"-touchscreen and various haptic instruments. The driving task is projected on a white wall in front of the car. Separated from this area, the control room serves for recording and analyzing user interactions. To carry out identical reproducible test runs, we have developed a special software suit called UsaWiz supporting the wizard in managing the various system parameters, semi-automatically announcing the operation tasks at specified points of time and logging all kind of transactions for a detailed offline analysis. The UsaWiz has already proved its concept in various usability experiments applied in different domains.

Figure 1: Overview of the test scenario

Test Interface

In our 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 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 a central design element, the interface contains a list which shows songs, radio stations or call partners with regard to the current context. The primary input devices are a touchscreen and a separate keypad, providing various buttons. The buttons of this keypad are placed in direct analogy to the design of the buttons on the touchscreen.

Test Methodology

The functionality of the interface is partly realized according to the 'wizard-of-oz' test paradigm[2]. In contrast to haptic interactions, the recognition of semantic higher-level modalities (speech, hand and head gestures) is simulated by a human person supervising the test-subjects via audio- and video-signals. The so-called 'wizard' interprets the user's intention and generates the appropriate system commands. 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.

Test Plan

Before starting the real test run, the test subjects are learning 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 keypad. At the same time, the use of natural speech and gestures as well as potential combinations of the individual modalities is explained. The test consists of two separate runs. Whereas in the first part test subjects have to regulate the speed or to evade potential obstacles on a totally straight road, the speed in the second part is kept constant, but the road is curvy to a high degree. In both parts the test subjects have to handle incoming telephone calls and are distracted by certain visual and acoustical effects. Additional questionnaires after each part help to evaluate subjective user experiences.

RESULTS

A total of 17 persons participated in the usability tests, with three female and 14 male subjects. As mainly graduate students were asked, the average age was 26 years. Most of the subjects were technically highly skilled and nearly two thirds stated to drive cars regularly. Nearly 90% stated to be familiar with the MP3 audio format, two thirds mostly listen to music or radio while driving and only one third stated to make telephone calls while driving. About one third of the test subjects had prior experience with speech or gesture based interaction systems in other domains.

Subjective User Experiences

The test subjects were asked to rate the load superimposed by the driving task, from 1 (low) to 4 (high). Concerning the speed regulation the average rating was 2.65, for evading obstacles 3.06 and for the steering task 1.76. The speed task was rated more difficult because most of the subjects had some problems with the gas pedal. Evading the obstacles on the road was rated most difficult due to the sudden appearance and short reaction time. Test subjects were quite familiar with the steering task as it was closely related to real-world conditions. The influence of the various distraction effects on the system interaction was rated by the participants on the same scale as introduced above (1 to 4). The resulted mean values are as follows: acoustical effects 1.9, visual effects 1.3, incoming telephone calls 2.2, obstacles 3.3.

Causing Operation Errors

Concerning the provocation of SBL errors, the audio and visual effects did not result in any kind of operation errors. Interruptions by incoming calls resulted in forgetting to finish the operation task for 23% of the test persons. Other distraction effects did not cause such errors. Repetitions of similar tasks did not induce errors like strong habit intrusion as intended. The obstacles caused operation errors for 47% of the subjects mainly in the form of wrong buttons or multiple haptic interactions in vain. With regard to the available input modalities, in 80% of the observed operation errors, the test subjects used haptic input by touchscreen or keypad, speech only in 20%. The distribution of errors in the driving task is closely related to these results (79% of the errors while using haptic input).

Although not explicitly tested, errors related to the other two performance levels could be determined, too. Concerning speech, semantically highly complex utterances were used, which additionally applied specific application knowledge. Moreover, implicit mode changes were assumed. In the case of hand gestures, sometimes the meaning of up and down was interchanged.

Error Handling

Concerning the skill-based level, this usability study clearly demonstrates that most operation errors can be resolved by the user himself. Furthermore they can be avoided by a well designed and intuitive system interface. Selectively, the test participants demanded for an adaptive help assistant, that can be activated in case of system or user failures.

REFERENCES

1. Reason, J. *Human Error*. Cambridge Univ. Press, 1990

2. Nielson, J. *Usability Engineering*. Morgan Kaufmann Publishers Inc., San Francisco, 1993