A new technique for adjusting distraction moments in multitasking non-field usability tests

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

ABSTRACT
Evaluating errors that result from user interactions with in-car applications, it has to be considered that the user is permanently involved with driving the car. Reproducing this driving workload in non-field usability tests, it showed that the driving simulation demanded each test subject in a different way because of individual precognitions and properties. To ensure an identical driving workload for each test subject, it becomes necessary to individually adapt the degree of difficulty (DOD) of the driving task. We present a new technique in which, concerning the driving performance, each test participant is pre-classified in a baseline investigation before the main trial. In this context, a special measurement for objectively validating the driving performance of the subjects is being introduced.

Keywords
Usability Engineering, distraction task, driving task, operation task, workload, human error, pre-classification

INTRODUCTION
With regard to the evaluation of errors in the human-machine interaction in the automotive environment, a substantial problem is the permanent superposition of several user acts. Implicated by the domain, the user has to execute all operations concerning the in-car applications simultaneously with driving the car. To meet these facts in non-field usability studies, the participants have to perform in a 3D-driving task in addition to the operation task to be evaluated. Yet, it often showed that a set of participants had nearly no difficulty with the driving task, whereas others were demanded so much that they were hardly able to operate the in-car application. Subjects of the latter group showed the tendency to no longer take the driving task seriously, and solely concentrated on the operation of the in-car application. To avoid these unrealistic scenarios, the design of this kind of trials requires a rather sensitive taxonomy of the specific influence factors. The participant with her or his personal properties and individual precognitions, the DOD of the driving task simulation, and

the scope of the operation task form a barycentrical triangle that has to be kept constant over the whole progression of the trial. In terms of the operation task, the driving thread represents a distraction moment. To reproduce the test scenarios in equal terms, it is necessary to ensure a distraction moment that is not too powerful and, above all, identical for each test subject.

METHODOLOGY
Implementation in a Test Concept
Before the main test, each test person participates in a so-called baseline investigation in which the subject is pre-classified with regard to his driving performance. Within the baseline investigation, which runs over a period of 180s, the participant is not yet confronted with any kind of operation task. The subject has to follow a predefined course of a road that is unknown to her or him. For reproducibility reasons, in this ascertainment, the participant cannot manipulate the velocity, but by a predefined run chart, the speed of the car is varied automatically between five DODs. The degrees are characterized by the frequency and the narrowness of the curves and the speed. In general, the time slots in the sequence of the DODs are of different length, and the levels are not monotonously de- or increasing to eliminate anticipation effects by the test persons. The concrete driving task is to always keep the car on the right lane of the street. This means, the vertex of the triangle T (representing the current position of the car) must be kept within the limits (distance q) given by the broken white line A in the middle of the road and the right road side B (c.f. 1).
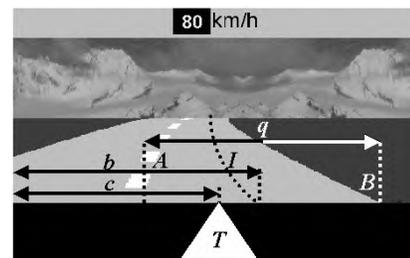

Figure 1: screenshot of the 3D-driving simulation

During the whole study, the absciss c (c.f. 1) of the position of the vertex of T in relation to the absciss b of the ideal line I (c.f. 1) and the correspondent timestamps t are

recorded in a log file (at a sample rate of 13Hz. Directly after the baseline investigation, the test subject is pre-classified, and, based on the result, the DOD of the driving task in the main trial is determined. If a participant had a bad performance in the baseline investigation, the DOD will be respectively low in the main part. To objectively rate the driving performance of each test subject objectively, a special numerical validation measurement based on the driving data is introduced.

Computation of the Error Score

If the driver leaves the right lane of the road, i.e. in case the condition $| c - b | - 0.5q < 0$ is violated, a so-called *deviation event* D_j is generated. Then, $d = | c - b | - 0.5q$ represents the distance of the car from the boundary of the lane in the simulation. Let j be a counting index for the deviation events, i for the distance samples, $m(j) = |D_j|$, and $d_{1,1}$ denote the first recorded distance of the first deviation event D_1. Then, a deviation event D_j can be characterized by a set of distance samples, as follows: $D_j = \{d_{1,j}, d_{2,j}, \ldots, d_{m(j),j}\}$. Each deviation event is validated with a numerical value E_j. This measurement is dependent on the individual performance factors:

i) the current distance d_{ij} at the timestamp t_{ij} and

ii) the DOD s_{ij}, at the timestamp t_{ij}.

The j-th contribution E_j of one deviation event D_j is computed as the linear approximation of the area between the edge of the lane (A, and B, respectively,) and the driven line. Assumed the car has exceeded the left border A, in figure 2, an exemplary visualization for the computation of E_j is given (bird view, scene rotated by -90°):
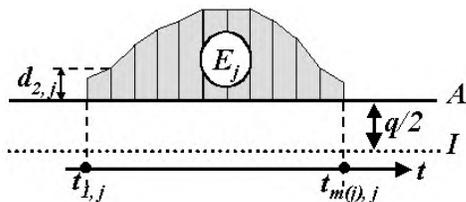

Figure 2: visualization of the computation of E_j

Thus, E_j can be computed via the following formula:

$$E_j = \Re_{i=2}^{m(j)} \frac{(t_{i,j} - t_{i-1,j})(d_{i,j} + d_{i-1,j})}{2}.$$

Assumed n deviation events, the total error score E^* is calculated as the weighted sum of all n score contributions. The weight of each single deviation is determined by $s_{1,j}$ which is the DOD having been effective at the time $t_{1,j}$:

$$E^* = \Re_{j=1}^{n} \frac{E_j}{s_{1,j}} = \Re_{j=1}^{n} \Re_{i=2}^{m(j)} \frac{(t_{i,j} - t_{i-1,j})(d_{i,j} + d_{i-1,j})}{2s_{1,j}}.$$

With E^* being a monotonously increasing function of n, it is a direct measurement for the user performance in the driving task, as the contributions of the deviation areas are all non-negative).

Classification Scheme

For the classification of the test subjects, the score E^* is compared with values of a look-up table (LUT). The LUT consists of several score intervals. To each interval, a correspondent DOD of the driving task to be used in the main task is assigned. The single score intervals of the LUT depend on peripheral constants which are

i) the determination of the course of the street,

ii) the determination of the particular speed levels,

iii) the width r of the car, and

iv) the width q of the right lane

Based on the results In a preliminary study, which will be described elsewhere, the LUT (see table 1) was determined on the following conditions: resolution 1024x768 pixels, $r = 1$pixel (car represented by vertex of a triangle), $q = 380$ pixels, DOD 1 = easy,…, DOD 5 = hard:

Error score (in ms*pixels)	0-3000	3001-6000	5001-7000	7001-9000	9001-
Level	5	4	3	2	1

Table 1: LUT for the determination of the DOD

RESULTS AND CONCLUSIONS

The presented technique proved its efficiency in several tests series (see also [1]). In the latest one, 15 subjects (from the age of 20 to 65, average age 32) participated. The frequency of the particular DODs is given in brackets after the DOD itself 5(5), 4(3), 3(2), 2(4), 1(1). The average DOD on all subjects determined was 3.5. The large mean variation showed that a pre-classification is necessary and not negligible in this kind of investigations. Exclusively regarding the results of the subjects in the eldest group (age over 50), which consisted of 5 participants, we got the distribution as follows: 4(1), 2(3), and 1(1), with gives an average level of 2.2. On the other hand, the youngest group (5 subjects) had 5(3) and 4(2), which is 4.6 on average. This shows that especially elderly people are demanded more by the driving task. In further works, the logged driving data will be explicitly evaluated with regard to the driven curve of the participant in relation to the ideal line. Thus, we hope to be able to make special statements about the influence of the operation task on the driving performance of the subjects.

REFERENCES

1. G. McGlaun et al.: "A generic operation concept for an ergonomic speech MMI under fixed constraints in the automotive environment," 9th Int. Conference on HCI 2001, New Orleans