

A New Icon Selection Test for the Automotive Domain Based on Relevant ISO Usability Criteria

Verena Johann
Automotive Group, DFKI

German Research Center for Artificial Intelligence
Saarbrücken, Germany

verena.johann@dfki.de

Angela Mahr
Automotive Group, DFKI

German Research Center for Artificial Intelligence
Saarbrücken, Germany

angela.mahr@dfki.de

ABSTRACT

In the automotive domain icons are popular and widespread, but it is not absolutely clear how to select icons according to a number of usability requirements. In this paper, we introduce a procedure for the evaluation of icons. Besides taking into account former icon test approaches, this new test is the first to explicitly consider various relevant criteria from current ISO standards in the field of Human-Computer-Interaction. We provide a test description including four diverse subtasks, and provide exemplary results for the icon drafts belonging to two features.

Categories and Subject Descriptors

H.5.2 [Information Interfaces and Presentation]: User Interfaces – Evaluation/methodology, User centered design

General Terms

Measurement, Performance, Design, Human Factors

Keywords

Icon, selection, test, understandability, learnability, memorability

1. INTRODUCTION

User-friendly designed icons can influence a human machine interface positively in several ways: First, they can be found and recognized quicker, they need much less space than text, they are better memorable, and are not bound to a specific language [1]. According to ISO/TR 16352 [2] their use can also be seen critically: If the meaning of the respective icons is not obvious and captured entirely, an increasing error rate could result. Thus, icons have to fulfill various criteria, such as manifestness (i.e. most people associate the same, intended concept [3]). Similarly, ISO norm 9241 [4] has defined certain criteria that interactive systems have to meet. For icons as part of an in-vehicle driver information or assistance system, the most important criteria are task adequacy, self-descriptiveness, conformity to expectations, and learning supportiveness. In literature, there are few attempts to create icon-based tests. In [5] the subjects' task was to freely name the meaning of presented icons. Afterwards, raters evaluated the answers with the help of a nine-step scale. Another possibility is to present several icon drafts together with the written meaning and let subjects rate, which one captures the intended meaning best [6]. In order to test predesigned icons according to the various design criteria for human-machine-interaction, we developed a new method, which exceeds previously applied techniques.

2. ICON SET

As an evaluation corpus for this new procedure, we used 18 diverse functionalities from the automotive domain, which were extracted from the car-to-x communication research project sim^{TD}. By pressing the respective icon, more detailed information about

the feature can be accessed on a larger screen area. For each content (warning or information) three icons were created, resulting in 54 icons overall. Warning icons had a red triangular frame according to German hazard traffic signs, whereas information icons displayed a rectangular, blue frame (cf. Figure 1).

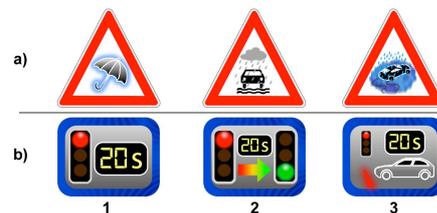


Figure 1. Items for a) warning icons “Attention, heavy rain!” and for b) information icons “Remaining red light time”.

Notably, the information of most warning icons was quite precise, not too complex, and easy to picture (e.g. icy conditions, a braking car). In contrast, the contents of information icons were mostly abstract and rather hard to design (e.g. dynamic route planning, street foresight). Accordingly, we expect an icon test being able to differentiate according to ISO criteria, to reveal better results for warning icons than for information icons.

3. ICON TEST

With reference to the aforementioned, relevant ISO criteria, we developed four different tasks. The entire test was conducted with 24 participants (mean age = 30.7, range = 20-57; 12 males and 12 females).

3.1 Task 1

The first task was designed to test the icons' manifestness, understandability, and task adequacy. When looking for specific information, a user should be able to choose the corresponding icon. Moreover, when encountering a new icon, the intended meaning should immediately be evident. In the first task all icons were presented sequentially in their original size (1.1”x 0.8”) on a 7” display with a viewing distance of about 20”. The subject should assess the respective meaning by ranking four possible descriptions available. In order to achieve additional, reasonable interpretations, two independent subjects had generated these alternatives via free associations prior to the actual test. The statement fitting best from the subjects' point of view, should be labeled with a ‘1’, the second best with a ‘2’, and so on. If the correct meaning was selected in first place, the icon was later evaluated with three points; in second place it received two points and so on. In order to control for possible sequence and motivation effects, the 54 icons were divided into three equal blocks containing one icon for each feature. Block order was counterbalanced between subjects. After having finished the whole first task, the intended “correct” meaning of each icon was revealed.

Copyright held by author(s)

AutomotiveUT11, November 29-December 2, 2011, Salzburg, Austria
Adjunct Proceedings

3.2 Task 2

In addition, icons should be concordant with the users' expectation: After selecting an icon, the expected information appears. Therefore, subjects should assign each icon to a more detailed illustration of the information, which it represents. For this test, printouts were prepared of all 54 icons and of the 18 larger, detailed feature information screens. Subjects assigned the small icons to the larger information screens manually. Encoded numbers on the backside of the printouts helped the experimenter to take down the results. Particularly with regard to information and warning icons, it can be assessed, whether a graphical distinction is reasonable and prevents category mix-ups. Icons that were assigned correctly, gathered one point in this test.

3.3 Task 3

Following [6], in the third task subjects rated which of the three alternative icons captures the intended meaning the best. In an evaluation booklet, one feature after the other was shown together with a short description of its intended content and its three respective icon drafts. The best icon should be labeled with '1', the middle one with '2', and the worst with '3'. In case of labeling an icon with '1' the icon received two points, whereas '2' gained one, and '3' obtained no points.

3.4 Task 4

The last task was designed as a free recall task. The purpose was to figure out, how well learnable and memorable the respective icons are. To achieve this goal, three days after completion of task one to three, we sent a list containing all 54 icons to the subjects via email. The icons were again divided into three blocks, and their sequence was balanced between subjects like in the first task, but the order was changed. In addition, the email message contained instructions as well as an example. The task was to describe the correct meaning of each icon shortly and precisely. Furthermore, the subjects were advised to respond within 24 hours. The answers were evaluated by two independent raters with the help of a four level rating scale ranging from „The description corresponds the meaning completely” to “The description does not correspond with the meaning at all”. At best an icon could achieve three points in this task, at worst it earned no points.

3.5 Aggregation Procedure

As the four tasks have equal priority in standard evaluation cases, differences between the tasks' maximum scores had to be outweighed, before adding up the single scores. In the second and third task, only one or respectively two points could be achieved. These scores were multiplied by the factors 3 and 1.5. Another possible way is to assign 3 points to correctly assigned icons in task 2, and accordingly 3 or 1.5 points in task 3. This allows skipping the multiplication. In any case, the maximum score for each task is three points, resulting in a total value from zero up to twelve points for each icon. Eight points were defined as a threshold for using an icon draft without major revision.

4. RESULTS

Below, we present exemplary results for the warning content „Attention, heavy rain!” and for the more abstract information feature “remaining red light time” (cf. Figure 1). Scores for each icon in tasks one to four and their total value can be taken from Table 1. Pairwise comparisons (bonferroni corrected) revealed, that the second warning icon was significantly better than icon number 1 ($p < .05$ and icon 3 ($p < .001$). Information icon number 2 earned a significantly higher score than icon 1 ($p < .001$) and icon 3 ($p < .001$).

Table 1. Scores for two exemplary, car-to-x based features.

Feature	'Attention, heavy rain!'			'Remaining red light time'		
	1.	2.	3.	1.	2.	3.
Task 1	2.4	2.7	2.8	2.2	2.8	2.2
Task 2 ($\times 3$)	1	1	1	1	0.9	1
Task 3 ($\times 1.5$)	0.5	1.9	0.5	1.1	1.4	0.4
Task 4	2.5	2.9	2.9	2.6	2.9	2.8
Total	8.9	11.5	9.5	9.6	10.8	8.7

This demonstrates, that our new icon test is able to differentiate between different illustrations by the combination of tests, as we had intended. Also, subjects overall rated warning icons significantly higher than information icons (mean = 9.6 vs. 9.0; $t(46) = 5.2$, $p < .001$) as we had expected from the first. By the way, for each of the 18 features, we identified at least one usable icon (score > 8) and got insights for which of those improvement might still be worthwhile (score < 10).

5. SUMMARY & CONCLUSION

The aim of this study was to develop a test that verifies icons with regard to several ISO criteria. By this icon test, pictograms for different warning and information features were evaluated. First of all, the results show that this icon test can differentiate between diverse forms of illustration and fundamentally support the selection process. Moreover, compared with former techniques, it is a more extensive method to review icons and especially fits the automotive domain. Furthermore, if needed, the test could (carefully) be adapted to specific requirements by changing the weighting of tasks or selecting just a subset of tasks. For the future, it is still necessary to look closer into the quality criteria of this test (validity, reliability), as well as an additional measure for consistency regarding the icons of a final set could be introduced as a next step.

6. ACKNOWLEDGMENTS

This work was funded within the project sim^{TD} by the German Federal Ministries of Economics and Technology as well as Education and Research, and supported by the Federal Ministry of Transport, Building and Urban Development.

7. REFERENCES

- [1] B. Färber, M.M. Popp, J. Ripper, H. Waschulewski-Floruß, B. Wittenbrink, and T. Wolfmaier, “Ergonomische Gestaltung von Fahrer-Informationssystemen“, *ErgoFis*, 1990.
- [2] ISO/TR 16352: *Road vehicles – Ergonomic aspects of in-vehicle presentation for transport information and control systems – warning systems*, 2005.
- [3] G.R. Marshal, and C.N. Cofer, “Associative indices as measures of word relatedness: A summary and comparison of ten methods” *Journal of Verbal Learning and Verbal Behavior*, 1963.
- [4] ISO 9241: *Ergonomie der Mensch-System-Interaktion – Teil 110: Grundsätze der Dialoggestaltung*, 2006.
- [5] J.L. Campbell, D.H. Hoffmeister, R.J. Kiefer, D.J. Selke, P. Green, and J.B. Richman, “Comprehension Testing of Active Safety Symbols”, *SAE International*, 2004.
- [6] P. Green, “Development of pictographic symbols for vehicle controls and displays”, *SAE Paper 790383, Warrendale, PA, Society of automotive engineers*, 1979