UR:BAN

Technical Report

Suitability for Use while Driving
Introduction for (App) Developers

Michael Krause, Klaus Bengler
Technische Universität München
Institute of Ergonomics

UR:BAN MV
Urbaner Raum:
Benutzergerechte Assistenz-
Systeme und Netzmanagement

Gefördert durch:
Bundesministerium
für Wirtschaft
und Technologie
aufgrund eines Beschlusses
des Deutschen Bundestages

Document Informationen

Project	MV – Mensch im Verkehr
Subproject	MMI – Stadtgerechte Mensch Maschine Interaktion
Category/Topic	Mobile Devices
Title	Suitability for Use while Driving — Introduction for (App) Developers
Authors	Michael Krause, krause@tum.de Klaus Bengler, bengler@tum.de Institute of Ergonomics, Technische Universität München
Preparation Date	September 2015
Status	final
Dissemination Level	public

History			
Version	Date	Author	Change
1.0	September 1st .2015	Krause, Bengler	Initial

Index

About the document

This document should help and guide a software engineer (e.g. an app developer) on the first steps in the field of driver distraction. We tried to keep the document short and nevertheless cover most of the topics. In adaption of the Pareto logic, we can hopefully provide you 80% of the important topics in 20% of the words that others would need. Therefore, this document is definitely not a replacement for reading specific guidelines and standards, but should support you in finding the right documents for your purpose, understand the connections, ideas and philosophies and give some background for understanding; e.g. when talking to psychologists about driver distraction and user tests.

First we explain some terms (***General and Wording***); so you have a chance to follow the content of other documents, studies or personal discussions. See it as a kind of crash course in human factors engineering.

The central documents related to the reduction of driver distraction are guidelines. The relationship between these guidelines and some background are presented in ***Guidelines and Documents***.

Important Concepts present some important concepts when designing interfaces for use while driving.

In ***Assessment Methods*** we give an overview of measurement techniques, which are currently used to assess driver distraction.

Hints, Caveats, Anomalies and Controversies presents our view to some ongoing actions and discussions.

Some sections are marked with ℹ Informational. You can skip these if you are in hurry or on a first reading.

General and Wording

The terms **nomadic device**, **mobile device** or **portable device** are used interchangeably; they stand for devices broad into the car and could be also taken easily away, but are more complex than a simple garage door opener. Typically the terms refer to satnavs or phones (consumer electronics). The term **after-market device** is closely connected, but this could be also a radio that is permanently installed into your OEM car, after you bought the car.

In the field of **ergonomics** (= **human factors**) often the interchangeable terms **Human Machine Interface (HMI)** or **Man Machine Interface (MMI)** show up. Sometimes in vehicles this is also named **Driver Vehicle Interface (DVI)**. The driver and the car have many interfaces (e.g., steering wheel, pedals, speedometer, etc.). For this document especially information provided to the driver is important. Often In the automobile domain the marketing word **infotainment** (information/entertainment) is used. But entertainment is mainly restricted to audio for the driver; visual entertainment is not allowed while driving. Systems providing such information are referred to as **In-Vehicle Information Systems (IVIS).** In Germany we would call it *Fahrerinformationssystem* (Driver Information System, DIS). The information is presented by a **User Interface (UI)** which is typically a **Graphical User Interface (GUI)**. IVIS are separated from the **Advanced Driver Assistance Systems (ADAS)**. An ADAS normally has sensors and actuators. The word 'advanced' is an indication, that it more fancy than an automatic windscreen wiper; which also definitely supports the driver with its sensors and actuators on a rainy day. A typical example for an IVIS is a navigation system. An example for an ADAS is Adaptive Cruise Control (ACC), which keeps your time gap to the car in front of you and combines it with a speed control.

ADAS are sometimes split into (cf. definition in [1]): information ADAS (inform the driver), warning ADAS (warn the driver), control ADAS (control the car/drivetrain). The information ADAS would heavily overlap with IVIS and would make the definitions fuzzy. We therefore prefer no information-ADAS. These are IVIS instead and so we can use the terms IVIS and ADAS separately, as the ESoP [2] also does. The ISO group *ISO TC 22 SC13 WG8* circumvents this problem and uses the term **Traffic Information and Control Systems (TICS)** which covers both (IVIS/ADAS). Afterwards in the scope section of an ISO document, they can decide which systems is meant in a specific document.

The term **suitability** is in the title of ISO 17287 (*Road vehicles -- Ergonomic aspects of transport information and control systems -- Procedure for assessing suitability for use while driving*). It refers to the usability of a product/function while driving, and emphasizes the importance of the driving task, while weakening the classic (subjective) usability aspect of user satisfaction. The classic definition of **usability** from (ISO 9241) consists of context-dependent: effectiveness, efficiency, and satisfaction. Some people find the classical term usability could not cover all hedonic aspects of an interaction, so they come up with **user experience (UX)**. The problem nowadays is that UX experts and UX departments in companies are handling just plain old usability. In this case, UX is used as a marketing buzzword and not covering hedonic aspects which extend usability.

When researchers are talking about the evaluation of TICS , they often refer to the driving task as the **primary task** and e.g., the (potentially) distraction information system as the **secondary task**. Sometimes these secondary tasks are further categorized in non-driving related (e.g., a phone call) and driving-related, but not directly needed to control the car (e.g., switch the wiper on). Non-driving related secondary tasks are then called **tertiary tasks**. When setting up an experiment, (secondary/tertiary) **tasks** are defined, where the **subject**s should do during the examination (e.g. 'Navigate from the home screen to the navigation option and enter the address XY'). Typically the subjects perform the defined task more than once (different **trials**). So that an individual mean/average or median value can be calculated from the trials afterwards.

Informational

To circumvent the negative term **driver distraction,** euphemistic the wording **driver workload** is sometimes (mis)used. Driver workload can be one reason for driver distraction. There are many definitions about driver distraction. Some categorize e.g. it, into vehicle internal or external distraction. A good concept in our opinion is to separate **distraction** and **diversion**. If an experienced and responsible driver diverts his attention from the driving task, it is intrinsic/intentional and the diversion is prepared, e.g. the road is scanned, if it is ok to read the time from a clock afterwards. Distraction is often extrinsic and more unintentional (e.g., a blinking light attracts attention) or can evolve from diversion: E.g., the clock is so badly designed that intrinsic motivated reading with prepared diversion requires an undesired long and continuous glance and results in distraction. There are plenty of terms and psychological constructs in this field which we cannot handle here. If you are more interested, also look for: daydreaming, looked-but-failed-to-see, sleepiness, arousal, vigilance or cognitive workload.

When setting up an experiment, there are two main design concepts (within-subject-design, between-subject-design). The experimental question and context is important to choose or to combine these concepts. You will see the typical setup when your app is tested, it will be a **within-subject-design**, which means that every test subject performs every defined task. In some experiments a **treatment** (e.g., using your app) is compared to a **baseline** (e.g., just driving) or the treatment is compared to a **reference** treatment (e.g. distraction potential of a reference radio tuning task), which then could be also referred to as baseline.

There are many psychological effects that can disturb an experiment and make it different from a physical measurement. If a subject is for some reason (highly) motivated in an experimental condition, he/she can **over-compensate** potential negative effects to some extent. An example would be a person who (in a non-blind setup) with a little bit alcohol has a better driving performance, than in the baseline (no-alcohol) condition.

A related effect in behavior is the **Hawthorne effect**. If subjects feel they are part of an experiment and/or their performance is measured, they can change their normal behavior. Therefore, typical instructions for a test subject include a phrase like: *"We don't want to test you or your performance, we want to test the application"*

An often cited concept in human factors is an (inverted) U-shaped arousal curve (**Yerkes-Dodson law**). The idea is, to get a good performance in humans, you need a certain amount of arousal. If there is less arousal, it is boring (under-load) with reduced performance, but too much will result in overload and the performance also decreases.

In an experiment, a subject typically performs many different tasks, this can lead to other problems. The subjects can **carry-over** from one of the task **strategies** that were learned or influenced from other tasks. Imagine your app was tested in one experiment, together with apps from developer X and failed. But maybe it would have passed, if it would have been tested alone, or together with apps from developer Z.

In typical driver distraction studies the tasks are trained by the subjects before measurement trials. If they have severe problems in this training phase, there is something wrong with your app. Because time is an issue when performing experiments, this training has to be somehow limited, but nevertheless standardized. There are often statements, e.g.: The subject should perform at least two error-free trials until feeling comfortable. These procedures often seem odd to new people in the field. Nevertheless, it is definitely state of art: The subjects are trained to a task prior to measuring.

The order of the tasks could also be an issue (**order effects**). If the tasks are performed by all test subjects in the same order, the subjects will perhaps become more familiar with the situations after some time. Therefore, the last tasks maybe benefit by learning effects. On the other hand these experiments are also exhausting; the last task could also be influenced by fatigue effects. Therefore the task order is mixed for the participants. Some prefer **permutations** for this mixing, others like **randomizations** (sometimes with some boundary restrictions); or both are combined. To make a long story short: Permutations are intentional 'designed' orders (combinatory). If some data sets / subjects are excluded afterwards (e.g., due to technical problems) or an error occurred in the permutation generation and application, randomization could have been the better choice. When fighting unwanted effects (like order), often the term **counter balancing** is used.

The people who carry out the experiment could also have an influence. If self-designed questionnaires are used, **suggestive and leading questions** could be worst-case scenarios. The person (examiner) could also have a more subtle effect on the subject's behavior and results with his opinion (**Rosenthal Effect: self-fulfilling prophecy; Greenspoon Effect: positive/negative feedback**). Even if the experiment and communication is standardized as strictly as possible, situations are always coming up where the examiner has to make a decision (e.g., repeat the last measurement trial, dismiss this test subject, etc.). It's like being the referee in sports.

Some concepts which should also be mentioned:

Statistics. The experimenters typically have a strong background in statistics. This field cannot be covered here. If you need a search starting point for: standard error, confidence interval, percentiles/quantiles, average versus median, effect size, statistical power, sample size, t-test, ANOVA, odd ratios, non-inferiority testing.

Hick's Law. When the number of choices in a reaction time experiment is increased, the reaction time increases logarithmic with the number of choices.

Fitt's Law. How to calculate prediction about the hand movement time; if you know the distance to a target (e.g. a button) and the width of the target.

Model Human Processor (MHP). To calculate a model of human processing as to how long an action will approximately take. MHP is basis for other modeling techniques like GOMS, GOMS-KLM and so on.

7 +/-2 Chunks (Miller's Law) stems from a paper "*The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information*" which shows that your short-term memory is limited. **Chunks**: Information units/items. Don't take it literally, an interface is better which needs less chunks or nothing to remember in short term memory.

Wickens' Multiple Resource Theory. Concept that a human has different resources along different stages of information processing; which can explain that some tasks can be performed simultaneously without much interference (i.e. deterioration in task performance).

Rasmussen's Skill-Rules-Knowledge-Model (SRK). Allocates human actions on three levels, depending on the individuals' experience and training for a specific action. E.g., for an experienced driver to shift gears is an automated skill. A novice has to follow rules in mind (first press clutch, then shift gear, finally release clutch). For the lowest level (skill), the lowest amount of cognitive resource is needed (automated, highly-trained). Deriving the actions by knowledge would need the most cognitive resources.

Psychological terms and concepts that are also connected to driving, are **risk perception** and **risk homeostasis**. Risk perception covers theories how individuals are affected by situations and come to their personal conclusions and actions. E.g., an often seen strategy of test subjects is to drive slower or increasing the time gap to the car in front, when tasks appear demanding. Risk homeostasis is a theory that safety measures can be counterbalanced or even overcompensated by riskier behavior (e.g., if someone with a seemingly safe car would drive more reckless).

Human Centered Design (HCD) or User Centered Design (UCD). HCD and UX are closely connected buzzwords. HCD is an iterative process focused on user needs; see also ISO 9241-210:2010 *Ergonomics of human-system interaction -- Part 210: Human-centred design for interactive systems. (former ISO 13407)*
Similar: *Plan-Do-Check-Act* or the *spiral model.*

A recommendable 'cookbook' to evaluate within UCD an IVIS/ADAS can be found in the public chapter 5 (known as the 'AIDE cookbook') of:

Janssen, W., Keinath, A., Nodari, E., Alonso, M., Rimini-Doering, M., Brouwer, R., Portuoli, V., Horst, D., Marberger, C.,Vega, H., & Plaza, J.(2007), Specification of AIDE methodology EU project AIDE, Deliverable 2.1.4., http://www.aide-eu.org/pdf/sp2_deliv_new/aide_d2_1_4_summary.pdf

The current state of the art to assess ADAS is the:
Response 3. Code of Practice for the Design and Evaluation of ADAS (2009)
http://www.acea.be/uploads/publications/20090831_Code_of_Practice_ADAS.pdf

Guidelines and Documents

In Europe there is a recommendation from the European Commission about IVIS-HMIs, which was worked out by field experts from different countries and backgrounds (research/industry). The official number is *2008/653/EC* [2]. In practice it is called *European Statement of Principles (ESoP)*. The guideline has about 40 pages and depending how you count with 34 to 43 principles. The main idea of the document is to support the development process to get good interfaces, but not to restrict development with criteria. This freedom often results in surprising displeasure from developers. They read the ESoP and are perplexed afterwards: No criterion is mentioned for testing, how could this be helpful?

Strong input to the ESoP (2006) came from the Transportation Research Laboratories (TRL, UK), which before had a checklist, to assess the topic of driver distraction. They later (2011) updated the checklist; now you can use this updated TRL-checklist as supporting material for the ESoP. The ESoP makes a lot of references to ISO standards and uses a trick: It states that the newest version of the standards should always be used. This is how the ESoP keeps itself up to date. ISO standards have the advantage, they are internationally agreed and have the disadvantage they are not available for free. ISO standards also typically refer to other ISO standards, which refer to other ISO standards (and so on); but even the direct connection from ESoP to ISO standards is hard enough to handle.

The Alliance of Automobile Manufacturers (AAM), the industry coalition of mainly American automobile manufacturers worked on a self-commitment (2006); the final title is: *Statement of Principles, Criteria and Verification Procedures on Driver Interactions with Advanced In-Vehicle Information and Communication Systems. Driver Focus-Telematics Working Group [2].* The title is long and is normally referred to as *AAM* in short. The term *DFT* (*Driver Focus-Telematics*) is also rarely used. The guideline has about 90 pages. It proposes two test procedures, typically referred to as *2.1.A* and *2.1.B*. If a car manufacturer decides to use procedure 2.1.A to assess an interface the glance data is investigated in an experiment with test subjects. For this purpose two methods can be used, eye-tracking or occlusion. The results are then compared to criteria stated in the guideline.

In procedure 2.1.B the driving performance data is explored in an experiment with participants. The test subjects use the IVIS under investigation and a reference task. The reference task is a strictly standardized radio tuning task. Afterwards the results of the IVIS and radio tuning are compared. The idea is, that radio tuning is performed in the car for decades and the associated risk seems accepted by society. Also the glance criteria from 2.1.A are based on radio tuning (from literature).

A self-commitment of the Japan Automobile Manufacturers Association (JAMA) is the *Guideline for In-vehicle Display Systems – Version 3 (2004).* It is typically referred to as *JAMA.* It relies on eye-tracking or occlusion and specifies the strictest criteria. With about 15 pages it is also the densest guideline.

In February 2012 the National Highway Traffic Safety Administration (NHTSA, USA) released a draft of the *Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices* (Docket No. NHTSA-2010-0053) which was finalized in April 2013 and clarified in September 2014 (official document citation 79 FR 55530). It is a voluntary, nonbinding federal guideline. Therefore, the legal state is about comparable to the ESoP. The NHTSA follows a plan and wants to release three guidelines: Phase 1 for visual manual interfaces and equipment, Phase 2 for portable and aftermarket devices and Phase 3 for voice-based interfaces. The current guideline is Phase 1. This guideline completely relies on glance data (eye-tracking or occlusion) that are compared to criteria.

You can see different philosophies: While AAM, JAMA and NHTSA specify criteria for the total glance time (thus the task length), this is not the case for the ESoP. This reflects different regional opinions regarding the importance. While the ESoP refers to ISO 15005, which includes (at the moment) a dwell duration criterion for single glances of 1.5s and several times also emphasizes the interruptibility, the JAMA has no single glance criterion and AAM/NHTSA use 2s.

The ISO working group *ISO TC 22 SC13 WG8* addresses driver distraction topics with their standards. If your company/organization has a long-term interest in this field, participation in the group could be an option. Also the Society of Automotive Engineers (SAE) must be mentioned. The SAE is well-known for international standards in the automobile domain.

Many car manufacturers, suppliers and phone producers, allied in the CarConnectivityConsortium (CCC). This is the association that develops the industry standard MirrorLink to seamlessly connect phone and car. Part of this effort are guidelines and test procedures for driver distraction, based on the established guidelines (ESoP, JAMA, AAM). The CCC provides information for app developers. Due to the fact that many software developers are involved in the CCC processes, the driver distraction documents could be as you desire.

Most documents are guidelines/recommendations, so it may look like it is very easygoing and voluntary. On the other hand these documents are partly summaries of knowledge (e.g., include reviews and studies) and represent a state of the art. Therefore, a lawyer for product safety might see or use them differently. Also the NHTSA became famous for the NCAP test; perhaps later, they use their guideline for testing and improved consumer information, developers/manufacturers could encounter an undesired presence in the media with unfavorable test results.

To read the guidelines is recommended but could be tedious without some further help. There are three documents that provide an overview of different times and with different foci. They could be helpful prior to, or while reading the guidelines:

The paper *Fighting driver distraction: worldwide approaches* (Christian Heinrich, 2013 ESV Korea) http://www-esv.nhtsa.dot.gov/Proceedings/23/files/23ESV-000290.PDF compares directly ESoP, AAM and JAMA and provides background information.

Organisation Internationale des Constructeurs d'Automobiles (OICA) provides a position paper *Recommended OICA Worldwide Distraction Guideline Policy Position* (http://www.oica.net/wp-content/uploads//OICA-Position-Paper-Driver-Distraction-Final-2015-03-03.pdf)

From the European Project *Adaptive Integrated Driver-vehicle InterfacE* (AIDE, 2004-2008) the deliverable *Report on the review of the available guidelines and standards*, Deliverable No. D4.3.1 http://www.aide-eu.org/pdf/sp4_deliv_new/aide_d4_3_1.pdf provides a general and broad overview.

Further Reading:

M. A. Regan , J. D. Lee and K. L. Young Driver Distraction: Theory, Effects, and Mitigation, 2009, CRC Press ISBN-13: 978-0849374265

Important Concepts

Interruptibility

Interruptibility means the driver is every time in full control of the interaction with the IVIS at. An adjective closely connected to Interruptibility is: *user-paced.* The opposite concept would be *system-paced*. A typical system-paced component in automobiles is scrolling texts automatically (e.g. for long music titles); which is bad and prohibited. User-paced means, the user can interrupt the interaction at an arbitrary point and has no drawback when resuming the interaction in the future. This is a central concept, when designing for use while driving.

To achieve *user-paced*-ness it is important that your interface/app includes no time-outs. If you start a timer to delete user input after x seconds, or your application automatically jumps back to the home screen after x seconds, these are no-goes. If your user enters now '*sup'* and continuous in 3 minutes with '*ermarket'*, the result is 'supermarket'. The no-timeouts-rule includes all states and screens; even on-screen keyboards do not automatically disappear, time-based. In general: If you stop a time for use in if-statements or use other timers, alarms or delayed events and could be an indication that your implementations turns in the wrong direction.

It is slightly different for some traffic and car related information. E.g., the navigation announcement (e.g. turn left) of a satnav must to some extend be *system-paced (*related to the current position).

Interruptibility enables the driver to turn the eyes back to the road at any point in time, without feeling any pressure to look longer to the interface.

Feedback / Visible system status

Concepts that also support interruptibility are a timely feedback and a visible (perceivable) system status. Both are similar, but it is a good idea to separate them; because they are working on different levels and in different time domains.

Feedback means, the user directly receives a confirmation that the system recognized the input. E.g., if a button is pressed from an on-screen keyboard, the button changes color, optional to above the touch point letter is displayed enlarged, the letter is added to a text field, maybe even a short click or beep is sounded off or some devices even yet give a haptic push.

For this, some guidelines recommend a maximum time value of 250ms. We would recommend to give the feedback as fast as possible and recommend a value below 100ms. 100ms are a long time for modern processors and the period were users feel events as instantaneous. As a developer you maybe run into two problems. First: The user interface must be responsive at any time, so it's maybe your job to organize threads or manage the user interface architecture. A modern operating system typically keeps track of this and e.g. does not allow a long blocking download activity in a user interface thread.

Second: If you are responsible for the whole on-board IVIS you must probably convince the business department that even a luxurious car will have a cheap user experience, if the on-board infotainment uses the lowest possible processor.

System status works on the dialog level or for the system in general. A general system status could be e.g.: the system is muted. An example on the dialog level is the calculation of a route: The driver has entered an address and pressed 'start'; now your app typically needs time for calculating/downloading. The AAM recommends to show an indication after 2s (e.g., please wait). For implementations it is far easier to directly show e.g. a progress bar or message. This will also support the feedback level. It would be good, if this waiting message and the end of the waiting period can easily be grasped by the driver. E.g. if the whole screen is greyed out (excluding the waiting indication). We would even add another optional requirement: Your app should nevertheless be able to respond to user input. When a progress bar is displayed, add a cancel button.

A clearly visible (perceivable) system status also lowers cognitive load and confusion. E.g., if the user cannot see the system is muted, but it is silent, many causes would be possible (it is muted, no music title selected, fault in the audio hardware, streaming connection failed, etc.).

You can check your app/system with a short thought experiment: If your system needs no information from (human) memory and has a clearly perceivable system state, one user (Alice) can interrupt the interaction with the app/system at any point and another user (Bob) can later resume working with the app/system, without asking Alice what she has done before.

Accessibility / Legibility

If developers think of suitability while driving, their first thought probably is, to make the fonts and buttons bigger. This is not wrong, but as you have seen from the other concepts, also with big fonts and big buttons you can design inappropriate interfaces. If the content is of such design, that it can be easily grasped with short single glances (e.g. no cluttered interfaces with a lot of details and overlapping/transparent content), again interruptibility is supported.

The newest- and for developers easiest way to get information about e.g., font size, button size, button placement, contrasts, handling of color gradients and recommendations how to handle screen sizes / resolutions should be the developer documents from CCC (MirrorLink). If the involved car manufacturers and suppliers would carry on updating these documents regularly with their new knowledge and results from studies, they will stay as a valuable resource of information.

A trivial concept that nevertheless should be also mentioned here: While interacting with the interface, one hand must be at the steering wheel. Some recommendations also forbid that simultaneously actions with both hands are needed (e.g. with buttons on the steering wheel)

Prioritization and Integration

When the system provides information from different sources or internally generates messages that can interfere with messages from other ADAS or IVIS, prioritization can be a problem. Not all messages can be presented to the driver simultaneously. The ISO/TS 16951 proposes two methods, how to determine the priority. In one method all messages are compared amongst themselves (matrix) and for each combination it is decided beforehand which message is more important. The second method assigns two factors to each message: Urgency (time pressure, how fast must the message be delivered), criticality; (what happens if the delivery is delayed, potential damage). With this priority is determined. If you design a complete IVIS you must be aware of the priority issue and implement appropriate methods or APIs so other developers can provide information to your system about their message priority (e.g. urgency and criticality).

Nowadays, navigation apps on the smartphone are prominent examples for retrofitting in cars. The prototypes of KOLIBRI[1] and SignalGuru[2] (traffic light assistants) or the students' project from TU Clausthal[3] (wrong-way driver information) demonstrate that a smartphone could provide useful driver information. Due to the widespread smartphones even older cars could be equipped. Some applications try to use the camera of the phone to offer surrogate lane departure or forward collision warning systems. Due to the focus on critical warnings, based on local driving related real-time data processing (computer vision), these functions and implementations should be classified as (warning) ADAS. The ESoP [2] states: *ADAS are fundamentally different and require additional considerations in terms of human-machine interface.* Or in other words: When implementing an IVIS you must be careful; when implementing an ADAS you must be extremely careful.

Another problem of *retrofitting by smartphone* is: often phone and car are not coupled. E.g., if you want to mute the sound, you maybe have to mute the phone's navigation app and the car radio separately. New systems (e.g., MirrorLink, Google Android Auto, Apple CarPlay) will integrate phone and car.

Safe mounting

The interface components must be mounted safely. This means they don't injure the passengers in case of a crash and that during normal operation they don't disturb or interfere with the primary driving task. Provide information for mobile devices, that mounting them should be safe and reliable. The suction cups often get loose after some time; mounting with form closure e.g., clamps in the ventilation slots are far better. If someone really likes the suction cup mounting, he should be aware to never try to catch the device when it is falling off, while driving. He has to mentally anchor the decision beforehand: If it gets loose, I let it fall. Also the mobile device must not occlude **any** small part from the road scene (also think of children); not even at the lower part of the windshield or close to the A-pillar. If you move your head around, make sure the borders of the device don't occlude more than part of the dashboard or engine hood. Some guidelines propose to mount displays as close as possible to the line of sight and not lower than 30° below horizontal view. To give you a feeling: Most classical speedometers are around 20° below horizontal. So, the ventilation slots that are typically at that height could be an acceptable option; also with respect to reachability. A good possibility for OEMs to set the display higher, is to separate display and input device. Which means e.g. rotary knob instead of touch screen. If the rotary knob provides appropriate haptic feedback (intends) it also has advantages regarding interruptibility and blind-handling. With this solution you can also better solve an age effect/problem. For older people your car package can run into a conflict of objectives: For presbyopia the display should be at a distance, but the arm length is a constraint.

[1] TUM press release 2013. *Cooperative optimization of traffic lights outside of cities improves traffic flow: Progressive traffic signal systems save time and fuel* https://www.tum.de/en/about-tum/news/press-releases/short/article/30925/

[2] E. Koukoumidis, L.-S. Peh, and M. R. Martonosi, "SignalGuru: leveraging mobile phones for collaborative traffic signal schedule advisory," in Proceedings of the 9th international conference on Mobile systems, applications, and services, New York, NY, USA: ACM, 2011, pp. 127-140.

[3] automotiveIT April 2014. *App prototype provides wrong-direction alert* http://www.automotiveit.com/app-prototype-provides-wrong-direction-alert/news/id-009173

Assessment Methods

Occlusion

Occlusion is anchored in every one of the guidelines mentioned in the guideline section. So it as an internationally agreed and accepted method. While there is an ISO 16673 which standardizes the occlusion method, most guidelines (AAM, JAMA, NHTSA) apply regionally specific protocols, task training, sample size and criteria.

The basic idea behind occlusion is, that while driving and handling a secondary task, the view of the driver continuously switches from the road to the HMI and back to the road. This interruption of the view to the task is mimicked with shutter glasses that continuously switch system paced between opaque and transparent. Therefore you can see the occlusion method as a kind of driving surrogate.

Figure 1 Shutter glasses (PLATO, Translucent Technologies, CA) open typically 1.5 seconds and closed typically 1.5s (ISO 16673)

If the shutter is closed (opaque/blurred) you would look to the road. If the shutter is open you can see the HMI. Important metrics of an occlusion experiment are:

- the Total Task Time (TTT) that is needed to perform a task without occlusion
- the summed up time the shutter is open when the task is performed with occlusion glasses (Total Shutter Open Time, TSOT)
- the relationship between TSOT and TTT: The R-ratio (TSOT/TTT). A low R-ratio is an indication the task is interruptible and/or can be handled to some extend without vision.

We prepared an open source Android app, so you can experience the general idea behind the method on a smartphone screen: http://www.lfe.mw.tum.de/android-occlusion/

Occlusion is probable the cheapest and easiest assessment method.

Informational

Sometimes there are ideas to repurpose 3D shutter glasses instead of dedicated occlusion spectacles. If you try it, you will probably run into some problems: Many 3D glasses rely on polarization filters built into PC/TV screens; light from other sources is not affected. Depending on the purpose, you first have to find 'real' shutter glasses. Even when you manage to 'close' these shutter glasses, you encounter they are just like dark sun glasses and ISO 16673 especially recommends the illumination between open and close state should be similar. Therefore no continuous light adaption by the eyes is needed. For the same reason it is not advisable to repurpose the visor of automated welding helmets.

Occlusion can be also seen as a (slow) derivate of tachistoscopy. A technique from the field of experimental psychology. Or the other way around: If the occlusion shutters glasses can be used for precise and fast switching, they are probably suitable for tachistoscopic experiments.

Eye tracking

In eye tracking you will encounter two different concepts: Remote eye-tracking and head mounted eye tracking. In remote eye-tracking one or more cameras are observing the test subject frontal. Sometimes a single camera is used, other setups use a stereo camera configuration. In expensive setups even several stereo camera setups can be coupled. The main advantage of remote eye tracking is, that setups are not intrusive, nothing is hooked onto the test subjects.

The head mounted systems are fixed on the head of the participant. They need at least one camera to observe the pupil of one eye, some observe both eyes, each with a cam. Due to this closer and fixed position to the eye, they can achieve better results in some situations (e.g., if test subjects make expansive head movements). Head mounted systems are typically be equipped with a head-mounted cam that records what the subjects can see from the ego-perspective. In this ego-view the gaze direction which is detected by the pupil-cams is usually visualized. Depending on the manufacturer and system these eye trackers are either frames with cameras or the cameras are integrated into special eyeglasses (e.g. with combiners/mirrors).

Eye trackers sometimes rely on special light reflections in the eye, or try to get better results in the infrared domain; so it is common that infrared light sources are incorporated.

ISO 15007 is the central document for eye tracking. The standard specifies the eye tracking process and defines metrics and wording. It is also worth noting, that despite of the fact that modern automated eye trackers all usually cost several 10,000 dollars, it is still valid and allowed to record the glance behavior with an ordinary camera and code the result manually. Also some low cost open source projects for automated eye trackers are coming up.

There are different communities that use eye tracking for different purposes and have different technical requirements. Automobile engineering is only one of them, further applications are in general HMI and workplace assessment (e.g., website usability) or reading.

In the automobile domain, the two most important metrics are the *single glance duration* and the *total glance time*. The *total glance time* is the sum of the *single glance durations* needed by a person to perform a task. Nevertheless, the regional guidelines have a slightly different understanding and definition, what their criteria address is (e.g. eyes-off-road time).

If a subject needs 4 single glances for a task each 1s, the total glance time would be 4s. It is important to understand that 4 glances with 1s are far better, than 1 glance with 4s, which would also give a total glance time of 4s. If you now can optimize your app to only need 3 glances with each 1s, it would be the next step of improvement.
A cheap but reliable alternative for an eye-tracking study can be an occlusion experiment.

Informational

There are different other methods to assess gaze behavior, but not often used. You can search for oculography to find more. An intrusive method are contact lenses with coils in a magnetic field to detect the gaze direction. Other methods rely on the muscular activity (electromyography) and are not designed to detect the accurate position, but the general gaze activity.

Lane Change Test (LCT)

The lane change test is a standardized method (ISO 26022). The official title is Lane Change T**est**, but is also often named Lane Change T**ask.** Typically the test subjects drive with a game steering wheel at a normal desktop PC on a visualized road with three lanes (LCT V1.2 Stefan Mattes Daimler AG 2011). Supporting material for ISO 26022 (the Lane Change Test itself and an analysis program) can be downloaded from ISO: http://isotc.iso.org/livelink/livelink?func=ll&objId=11560806&objAction=browse&viewType=1

The participants have two tasks: To stay in the middle of the lane (tracking) and change the lane as fast as possible, when signs show up in the visualization (reaction). The visualized car constantly drives at 60km/h.

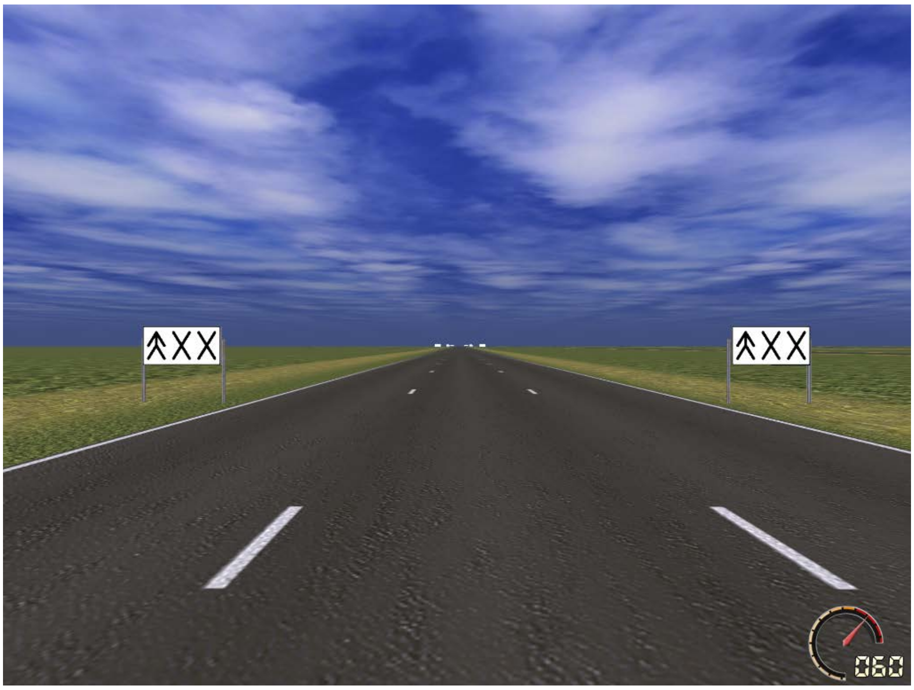

Figure 2 Lane Change Test. Subject should change from middle lane to left lane

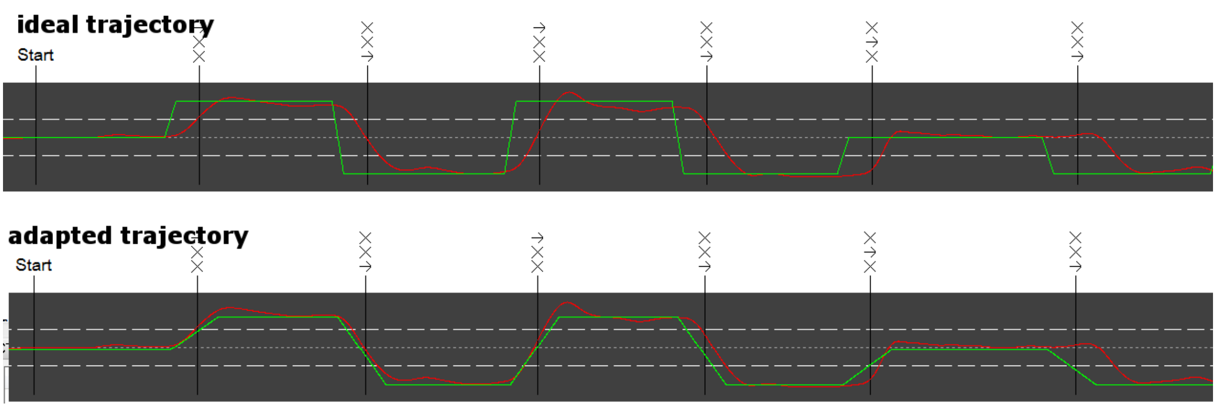

Figure 3 Lane Change Test Analysis with ideal/adapted trajectory (green) and test subject performance (red). The difference between the curves is the basis for the MDEV metric.

The LCT is driven by the participants without any IVIS (baseline) and with the IVIS. The performance in the two tasks (tracking & reaction) is combined into one metric, the Mean Deviation (MDEV). It is the error integral from a standardized 'ideal' trajectory, normalized by the driven distance. The contributors to the standard noticed the lane middle position and steering reaction is individual, therefore the standard also includes a proposed Adaptive Mean Deviation (Adaptive MDEV). The adaption uses the baseline driving of each individual test subject to adjust (adapt) some parameters in the 'ideal' trajectory for individual behavior. E.g. when the subject has the tendency to drive on the right side of a specific lane (tracking). The lane change trajectory (reaction) is also adapted to some extend to the individual behavior.

The LCT was calibrated in a multi-laboratory study [5], these results can help to range your results and should be used to validate the experimental setup itself. For validating an experimental setup, two reference tasks and a comparison procedure were specified in a supporting standard (ISO TS 14198) based on [5].

The LCT was finalized after the ESoP was edited. Therefore, the LCT at the moment has no link from a guideline. Due to the fact this is an internationally agreed standard (ISO 26022) it is only a marginal and hopefully temporary fact.

Informational

New analysis software from Daimler (V3.03 Stefan Mattes Daimler AG 2011) even includes procedures that take care of missed lane changes and therefore does deepened analysis of lane change behavior. If a test subject misses a lane change from the left to the lane on the right, the MDEV 'penalty' is nearly double, to if he would have just had missed a lane change from the left to the middle lane. This could result in problems during analysis. A miss is a strong indication the IVIS is challenging (distracting), but nevertheless misses should result in the same penalty.

Detection Response Tasks (DRT)

Most methods, principles and guidelines address visual/manual distraction. But how could cognitive distraction be assessed? This question led to the development of the Detection Response Tasks (DRT) ISO/DIS 17488. While driving and using e.g., a voice based IVIS an eye tracker maybe detects that eyes are on the road. The experimenter can see that hands are on the steering wheel. But how could be assessed how cognitive distracting is voice interaction with the system? To answer a question like this, the participants randomly receive a stimulus every 3 to 5 seconds and have to respond as fast as possible to this stimulus by pressing a button. If a task is cognitive, distracting the reaction time for the DRT is increased, or the subjects even miss to respond to some stimuli. At the moment two types of DRT are favored: The Tactile DRT (TDRT) with a vibration motor for the stimulus or a visual stimulus that is mounted to the head (Head-mounted DRT, HDRT) or alternatively on the dashboard (Remote DRT, RDRT). If you are looking for further information it could be helpful to know the TDRT was formerly known as Tactile Detection Task (TDT) and had a slightly different protocol. Also some researchers preferred to use multiple visual stimuli in the peripheral view (Peripheral Detection Task PDT). Some scientists like a modification of the HDRT: The stimulus can show up in two colors (red/green) and the subject is instructed to respond to the red color (oddball experiments).

The simple setup (one stimulus, one button) is very sensitive. We prepared a free Android App for you: http://www.lfe.mw.tum.de/mdt/ in order for you to experience the idea. Just hold the smartphone in your hand and as one condition, try to count up in steps of one (+1) for about 30-60 seconds. Than write down your reaction time. Afterwards count back in steps of seven (-7) starting at 150 for 30-60 seconds. Typically your reaction time should increase.

At the moment the DRT standard is under development (but nearly finished).
If you need a DRT device, you can have a look at our open source Arduino Detection Response Task: http://www.lfe.mw.tum.de/arduino-drt/

Informational

Hints, Caveats, Anomalies and Controversies

Simple straight road

When people are new to the field and experience the AAM driving track (straight simple road, constantly leading vehicle), their first impression is: This is too simple. It is important to understand that AAM, NHTSA and some researchers use these simple tracks intentionally. If you follow the AAM specifications, you can calculate the appropriate metrics, compare your results to the literature values and check guideline criteria. Therefore, it is not advisable e.g., to insert curves or roundabouts into the track, or to modulate the speed of the leading vehicle.

Reference tasks

For one measurement protocol the AAM uses a reference radio tuning task. The Lane Change Test has its own additional standard (ISO TS 14198) which specifies two reference tasks. If you implement a process to assess driver distraction (e.g., in your company) and your assessment method does not require a reference task, include one nevertheless. This way you can compare from experiment to experiment, if there are artifacts and problems. At this point we would also advertise our modified radio tuning task for Android: http://www.lfe.mw.tum.de/radiotask/

Lane Change Test (LCT) is not a driving simulation

The original software for the LCT (Daimler AG) looks like a driving simulation, but it is not a driving simulator. The LCT is a tracking task (stay in the middle of the lane) combined with a reaction test (change the lane as fast as possible). The results of both tasks are typically measured in one performance metric. The creators of this test decided that it should look like a driving task and programmed it into a nice visualization, to overcome acceptance problems. This sometimes leads to the problem, that LCT is misused as driving simulator.
In the LCT standard (ISO 26022) the technical details of the task are specified. Therefore it is even possible to program the LCT into a driving simulator, which is a lot of work. But it is not possible the other way around: The Lane Change Task (LCT) is not a driving simulation.

15-seconds-rule

There is a 15 seconds-rule around (SAE J 2364). It states, that a task could be acceptable for use while driving, if the task is finished within 15 seconds while the car is standing still. [6] has shown that unsuitable tasks that could be easily handled while standing still (e.g. reading scrolling texts) are not identified by the 15-seconds-rule. An appropriate alternative to the 15-seconds-rule is the occlusion technique.

Glance Durations and Time Headway

You can approach the question about glance duration criteria from different directions [7]. The AAM looked at a glance distributions from a radio tuning task as a social accepted distraction. Other researchers instructed participants to look inside the vehicle as long as they feel comfortable. Others occluded the view to the road and measured the driving deterioration. [7] finally concluded "*At this point, the authors have not be able to draw firms conclusions about what each approach concludes are maximum times for various glance statistics, how long too long to look away from the road actually is. There resources were insufficient.* […]" [sic]

Normally, glance duration away from the road is self-limited by the participants. So, if your application really exceeds a single glance criteria, there is definitely no room for discussions. And shorter is better: The ESoP holds an example which a good designed interface could be grasp within 1s ([Good Example:]: "*Easily legible and well-structured graphics on a well-positioned visual display which allows* *identification of the relevant menu item with a single one-second glance.*"). And ISO 15005:2002 mentions that glances should be shorter than 1.5s.

Sometimes there is a misunderstanding/wording: If you look away from the road, you are not blind. You can still have peripheral view. Nevertheless, eyes on the road is a superior concept to drive a car.

[8] looked at a time window 5s before and 1s after an event that led to crashes or near-crashes (*precipitating factor*, e.g. leading vehicle decelerating) in a naturalistic driving study. In this 6 second time window they found: "When the *total time eyes were off the forward roadway* was greater than 2seconds, regardless of where the driver was looking, an increased risk of crash or near-crash involvement (OR = 2.3) was observed."
To look 2s away from the road in this 6s time window is a proportion of 1/3 (33%) of the time.

[9] used the proportional perspective for a 2s time windows before a crash or near-crash and examined which time slot is the best predictor for an incident. The proportion of eyes-off-the-road in the slot 3s to 1s before a crash or near-crash, gave a good predictive model (they advanced this model, when they incorporated two other metrics).

Another ambiguity is the treatment of Total Glance Times (TGT). Most guidelines (except the ESoP) restrict the TGT for a task. If you think about it, TGT restrictions cannot be transferred to continuous tasks; e.g., route guidance (navigation), which is allowed. If you would drive 5 hours from Munich to Berlin and look to the satnav every minute for one second your TGT would be 5 minutes. Recommendations, e.g., regarding acceptable and typical glance frequencies and visual sampling behavior would be helpful.

Also a general change in behavior of the drivers may help. [9] mentions: "It can also be noted that crashes with initial headways above 2 seconds were very rare." German car drivers are taught rules of thumb: The right distance is half of the velocity (km/h). So, if you drive 100km/h the right distance would be 50m (=time headway 1.8s); 50m is the distance between reflector posts in Germany. Another trick is: To look when the car in front is passing a landmark/waypoint and mentally count up two seconds.
All this cognitive patchwork indicates that human factors engineering is required for solutions. Just one thought: The dominant instruments in most vehicles at the moment are the speedometer and revolution counter. For normal automatic drivers the second one is not that important and for manual gear shifting the auditory component is likely the primary information. The only useful situation probable is driving downhill. For electric cars the revolution information is superfluous in any situation. Why not simply change from dominantly visualizing speed & RPM to speed & time headway (when sensors are available)?

Median and 85th percentile

Some of the measurements rely on the median and the 85th percentile. The median is the 50th percentile. A percentile means that x% of the measured values are below this value. Therefore, the median (50th percentile) splits the results of an experiment into two halves; half of the results are below the median, half are above. Traditionally the 85th percentile value is used by traffic engineers. e.g., when they have to state a speed limit on a road. They measure the speed of free driving vehicles and use the 85th percentile. This percentile handling made its way to driver distraction assessment. The 85th percentile includes some information about the tail of typical distributions. If the tail is longer, the 85th percentile is higher. The 85th percentile is applied at different points. For example, for the AAM, when the average single glance duration for each individual is measured and the 85th percentile is taken from these individual average values. This is interesting, because the AAM derived at its criterion from an overall histogram taken from literature, where many glances from different subjects were combined. Therefore, a different calculation approach.

The NHTSA guideline applies the 85th percentile nested twice: First the 85th percentile for each individual is calculated and then checks if 85% of the test subjects are below the criterion. You may wonder what is the 85th percentile, if a test subject needed only two glances for a task (e.g., 1s and 4s)? NHTSA explains the 85th percentile of the number of glances is rounded up to the next full number and provides a table. The complementary event of 85% is 15%. 15% of two glances is 0.3; rounded up to 1. One glance can be above 2s. Therefore, our example (1s ; 4s) would be in line with the 85th percentile criterion. But they also state in another criterion the average also needs to be below 2s. In our example the average (1s ; 4s) is 2.5s; therefore above the criterion for this single test subject.

You will sometimes see researchers are not aware of 85th percentiles, typically mean (M) and standard deviations (SD) are published. As a rule of thumb, for a normal distribution the 85th percentile is about the average plus of the standard deviation: M+SD (with an error of about 4%). Glance metrics often look skewed, but M+SD is just a rule of thumb.

Sample Size

Novices are often surprised that not 1000 or 2000 people are typically tested in human factors studies, but often only 10-30. Statistics has formulas and theory how to determine the right number of test subjects, depending on the question, study design and previous knowledge. One of the main influence is the effect size. If you looking for a big effect you only need a few people. Or the other way around, with a few people you can detect big effects. This is also important to have in mind, when designing usability. Would you test your prototype application with 1000 people, just to know that 872 had a specific problem afterward? However, never test your prototypes directly while driving in real traffic. Make sure to employ other means before (see *Assessment Methods* e.g., occlusion, LCT, etc.).

Tools

At the moment we are investigating how we can support developers, when they are developing applications for use while driving. This document is a first step. Beside the driver, we assume that you (the developer) is an essential stakeholder. You typically have a lot of freedom to design the interaction with your application. Hopefully, we could have shown you that suitability while driving is more than just making buttons and fonts bigger.

Our ideas to further help you, are to provide information about the distraction potential of typical subtasks (e.g. select an item from a list, adjust a slider) in the near future. We also test, if an artificial driving surrogate task (tracking task) on Android could support the development process. The occlusion app for Android is also a step in this direction.

Acknowledgement

This document was written to address the topic mobile devices for the group Man Machine Interaction (MMI) within the German project Urban Space: User oriented assistance systems and network management (UR:BAN 2013-2016); funded by the Federal Ministry for Economic Affairs and Energy. The authors would like to thank the group MMI for commenting a first version of the document.

References

[1] United Nations ECE/TRANS/WP.29/2011/90 Guidelines on establishing requirements for high-priority warning signals http://www.unece.org/fileadmin/DAM/trans/doc/2011/wp29/ECE-TRANS-WP29-2011-90e.pdf

[2] ESoP – European Statement of Principles (2006) 2008/653/EC COMMISSION RECOMMENDATION of 26 May 2008 on safe and efficient in-vehicle information and communication systems: update of the European Statement of Principles on human-machine interface
(notified under document number C(2008) 1742)
http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2008.216.01.0001.01.ENG

[3] AAM – Alliance of Automobile Manufacturers (2006) Statement of Principles, Criteria and Verification Procedures on Driver Interactions with Advanced In-Vehicle Information and Communication Systems. Driver Focus-Telematics Working Group
http://www.autoalliance.org/index.cfm?objectid=D6819130-B985-11E1-9E4C000C296BA163

[4] JAMA – Japan Automobile Manufacturers Association (2004) Guideline for In-vehicle Display Systems – Version 3

[5] Bengler, K., Mattes, S., Hamm, O., Hensel, M. (2010). Lane Change Test: Preliminary Results of a Multi-Laboratory Calibration Study In: Rupp, G.L. (Ed.). Performance Metrics for Assessing Driver Distraction: The Quest for Improved Road Safety. Chpt. 14 (pg. 243-253). SAE International. Warrendale, Pennsylvania.

[6] Baumann, M., Keinath, A., Krems, J.F. & Bengler, K. (2004). Evaluation of in-vehicle HMI using occlusion techniques: experimental results and practical implications. Applied Ergonomics, 35, 197-205.

[7] Green, P. Tsang-Wei Lin, B. (2014), How Long Can Drivers Look Away from the Road: Some Key Documents. UMTRI Technical Report UMTRI-2014-xx April, 2014 draft final report

[8] Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J., Ramsey, (2006) The Impact of Driver Inattention on Near-Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data DOT HS 810 594

[9] Victor, T., Dozza, M., Bärgman, J., Boda,C.-N., Engström, J., Flannagan, C., Lee, J.D., Markkula, G. (2015) Analysis of Naturalistic Driving Study Data: Safer Glances, Driver Inattention, and Crash Risk. TRB REPORT S2-S08A-RW-1 http://onlinepubs.trb.org/onlinepubs/shrp2/SHRP2_S2-S08A-RW-1.pdf

Related ISO standards

Specifications:

ISO 15005:2002 *Road vehicles -- Ergonomic aspects of transport information and control systems -- **Dialogue management principles** and compliance procedures*

ISO 15006:2011 *Road vehicles -- Ergonomic aspects of transport information and control systems -- Specifications for in-vehicle **auditory presentation***

ISO 15008:2009 *Road vehicles -- Ergonomic aspects of transport information and control systems -- Specifications and test procedures for in-vehicle **visual presentation***

ISO 11429:1996 *Ergonomics -- System of auditory and visual **danger and information signals***

Measurement:

ISO 15007-1:2014 *Road vehicles -- Measurement of driver **visual behaviour** with respect to transport information and control systems -- Part 1: Definitions and parameters*

ISO/TS 15007-2:2014 *Road vehicles -- Measurement of driver **visual behaviour** with respect to transport information and control systems -- Part 2: Equipment and procedures*

ISO 16673:2007 *Road vehicles -- Ergonomic aspects of transport information and control systems -- **Occlusion method** to assess visual demand due to the use of in-vehicle system*

ISO 26022:2010 *Road vehicles -- Ergonomic aspects of transport information and control systems -- Simulated **lane change test** to assess in-vehicle secondary task demand*

ISO/TS 14198:2012 *Road vehicles -- Ergonomic aspects of transport information and control systems -- **Calibration tasks** for methods which assess driver demand due to the use of in-vehicle systems*

ISO/DIS 17488 *Road vehicles -- Transport information and control systems -- **Detection-Response Task** (DRT) for assessing attentional effects of cognitive load in driving*

Process standards:

ISO/TS 16951:2004 *Road vehicles -- Ergonomic aspects of transport information and control systems (TICS) -- Procedures for **determining priority** of on-board messages presented to driver*

ISO 17287:2003 *Road vehicles -- Ergonomic aspects of transport information and control systems -- Procedure for assessing **suitability for use while driving***

Reports:

ISO/TR 16352:2005 Road vehicles -- Ergonomic aspects of in-vehicle presentation for transport information and control systems -- **Warning systems**

ISO/TR 12204:2012 Road vehicles -- Ergonomic aspects of transport information and control systems -- Introduction to integrating safety critical and time **critical warning signals**

The ISO TC 204 (Intelligent transport systems) WG14 (Vehicle/roadway warning and control systems) works on related topics, often ADAS:
http://www.iso.org/iso/home/store/catalogue_tc/catalogue_tc_browse.htm?commid=54706&published=on&includesc=true