INTERNATIONAL STANDARD

# ISO 15007

First edition
2020-08

# Road vehicles — Measurement and analysis of driver visual behaviour with respect to transport information and control systems

*Véhicules routiers — Mesurage et analyse du comportement visuel du conducteur en relation avec les systèmes de commande et d'information du transport*



Reference number
ISO 15007:2020(E)

# Contents

Page





# Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/TC 22, *Road vehicles*, Subcommittee SC 39, *Ergonomics*.

This edition cancels and replaces ISO 15007-1:2014 and ISO/TS 15007-2:2014, which have been technically revised.

The main changes compared to the previous editions are as follows:

— integration of ISO 15007-1 (*Part 1: Definitions and parameters*) and ISO/TS 15007-2 (*Part 2: Equipment and procedures*) into one document;

— detailed description of different data reduction procedures;

— detailed description of procedures and criteria for quality assurance.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

# Introduction

This document supports the quantification and description of visual behaviour while using TICS (transport information and control systems) and driving vehicles in different driving levels of automation. It supports the quantification of information acquisition related to internal and vehicle external environment/objects (e.g. vehicles, billboards, information displays, variable message signs).

It provides assistance in the assessment of driver state considering visual attention. This document does not address fatigue and drowsiness.

This document describes the phases of visual behaviour assessment including the following steps:

— calibration setup and calibration verification (piloting phase);

— data collection;

— data reduction;

— quality assessment;

— data presentation.

Each of these steps should be handled with care, documented and checked for quality before moving to the next step.



# Road vehicles — Measurement and analysis of driver visual behaviour with respect to transport information and control systems

## 1 Scope

This document defines key terms and parameters applied in the analysis of driver visual behaviour focused on glance and glance-related measures. It provides guidelines and minimum requirements on equipment and procedures for analysing driver visual behaviour including assessment of TICS to:

— plan evaluation trials;

— specify (and install) data capture equipment; and

— validate, analyse, interpret and report visual-behaviour metrics (standards of measurement).

The parameters and definitions described below provide a common source of reference for driver visual behaviour data.

It is applicable to on-road trials (e.g. field operational tests or naturalistic studies), and laboratory-based driving studies. The procedures described in this document can also apply to more general assessments of driver visual behaviour. Data collected and analysed according to this document will allow comparisons to be performed across different TICS applications and experimental scenarios.

## 2 Normative references

There are no normative references in this document.

## 3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— IEC Electropedia: available at http://www.electropedia.org/

— ISO Online browsing platform: available at https://www.iso.org/obp

### 3.1 Basic terms

**3.1.1**
**area of interest**
**AOI**
pre-determined area within the visual scene

Note 1 to entry: Region of interest (ROI) is used as a synonym.

Note 2 to entry: An AOI will be no smaller than the normal resolution of the eye-measurement system being used (e.g. no smaller than 0,5 ° for typical eye tracking systems). See F.1.

EXAMPLE A rear-view mirror.



### 3.1.2
**blink**
short period of time in which the eye is closed by the eyelid

Note 1 to entry: The blink starts when the eyelid starts moving downwards and ends when the eye is fully opened again.

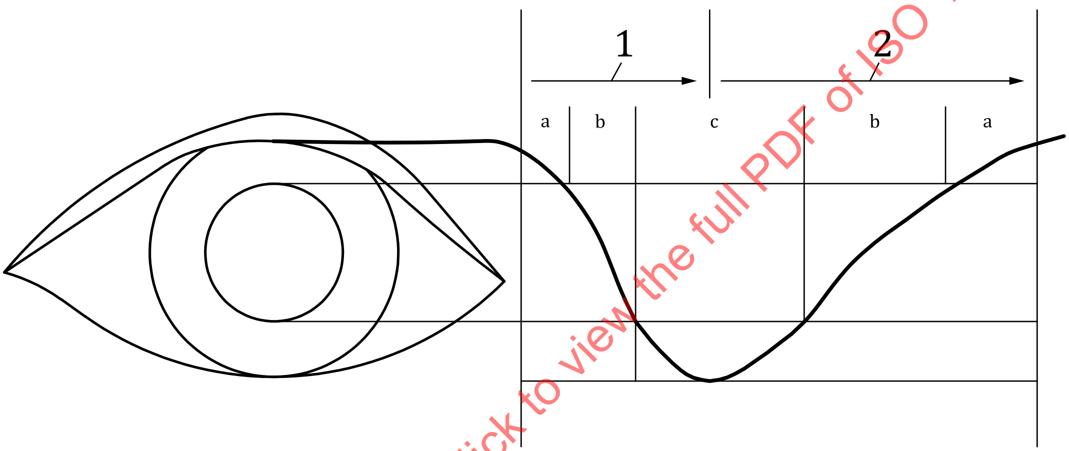
Note 2 to entry: According to the duration for which the eye is closed the following classification applies (see Reference [13]):

— normal blinks: ≤300 ms (mean duration 257 ms; standard deviation 11 ms);

— long closed durations: 300 ms to 500 ms;

— eye-lid closures: ≥500 ms (indicating microsleeps).

Note 3 to entry: Depending on participant state and/or *glance* (3.1.6) direction of a participant the eyelid may not fully open, although the blink might end. Especially in these cases the visibility of the pupil or major part of the pupil is important for correct gaze recognition.




**Key**

1 closing phase

2 opening phase

x vertical positioning of upper eyelid

a Open.

b Partially closed.

c Closed.

**Figure 1 — Phases of a blink (see Reference [17])**



Note 4 to entry: A blink is performed by the down movements the upper eye lid while lower eyelid remains static. Upper eyelid movement is considered in defining a blink.

### 3.1.3
**direction of gaze**
orientation of the eye to the *area of interest* (3.1.1) to which the eyes are directed to

**3.1.4**
**fixation**
short temporal holds of movements that keep alignment of the eyes to a particular point within an *AOI* (3.1.1) which falls on the fovea (the middle of the retina responsible for our central, sharpest vision) for a given time period

Note 1 to entry: Typically, individual fixations last from 100 ms to 2 000 ms (see Reference [5]). Fixations are the briefest of pauses during which visual information extraction is done by the eyes-and-brain from spatial areas that fall on the fovea of the eye (and hence are quite small). During fixation, there are believed to be at least three processes taking place (see Reference [16]) 1) analysis of the image falling on the fovea, 2) selection of a new *saccade* (3.1.7) target and 3) programming of the saccade to-be-made-next. It is not yet known how these processes are synchronized by the brain, nor how precisely they are synchronized – since fixation durations are not always long enough to comprehend completion of all the processes. (Sometimes the eyes move before information extraction from the site of fixation has been completed, as evidenced by frequent corrective return fixations to a site under some conditions that was fixated too briefly.) There is evidence that the brain both pre-programs fixation duration, and also does "process-monitoring" during a fixation to determine if analysis of the foveal image is complete within the fixation's duration before moving on. Thus, fixation time is dependent on both the immediate stimulus and the history of prior fixations. The contribution of both components suggests that fixation time may depend on the task and the amount of useful information in the fixated display (or viewed information) (see Reference [7]).

Note 2 to entry: See F.1 to F.4.

**3.1.5**
**fly through**
small 'snapshot' of a *saccade* (3.1.7) (<120 ms) that may be an artefact captured when the eye is moving from one *area of interest* (3.1.1) to another area of interest, and passing through one or more intermediate Areas of Interest in between

EXAMPLE The eye moves from the road scene ahead to the instrument cluster and passes the head-up display.

Note 1 to entry: Sometimes a small 'snapshot' of such a saccade may appear to be a *short fixation* (3.1.4), when it is really still part of the saccade. Such fly throughs (<120 ms) are not treated as fixations. Fly throughs may be grouped with the saccade they are part of, if saccades are being measured.

Note 2 to entry: Research shows that fixations cannot be shorter than 100 ms (see Reference [10]).

**3.1.6**
**glance**
temporal maintaining of visual gaze within an *AOI* (3.1.1), bounded by the perimeter of the AOI which can be comprised of more than one *fixation* (3.1.4) and *saccades* (3.1.7) within the AOI and its duration is measured as *glance duration* (3.2.1.3)

Note 1 to entry: A glance is a scientific *construct* that sums over one or more fixations that are made contiguously within a given area of interest (but one that is larger than the area corresponding to the eye's foveal region – an area that usually requires more than one fixation to view). The construct of a glance, therefore, typically comprehends more than a single fixation and is a coarser unit of analysis than a single fixation (since it is summing over fixations that are contiguous in time and spatially proximal within an area of interest. The construct of a "glance" is needed because often the salient questions in a study relate to the amount of contiguous time spent gazing at a particular area of interest (before the eyes move away from it). Of course, in some instances, the "glance" construct is also necessary because some measurement approaches are not capable of the fine discriminations needed to identify individual fixations (spatially and temporally) – and can only discriminate at the spatial/temporal granularity of glances. Thus, "glances" are a coarser measure of visual information extraction by the eyes/brain from a continuously viewed but somewhat larger spatial region. Typical glance lengths vary by stimulus and task but might (for example) range from 500 ms to 3 s for a task like "tuning the radio" (see Reference [11]).

Note 2 to entry: See F.1 to F.4.

**3.1.7**
**saccade**
brief, fast movement of the eyes that changes the point of *fixation* (3.1.4), within an *AOI* (3.1.1) or between different AOIs

Note 1 to entry: Saccades reach velocities as high as 500°/s (see Reference [10]), whereby the mean saccade ranges between 1° (text reading) to 5° (scene perception) (see Reference [14]).

Note 2 to entry: See F.1 to F.4.

**3.1.8**
**smooth pursuit movement**
smooth continuous movement of the eyes made to closely follow/pursue a moving object or signal

Note 1 to entry: Humans generally perform smooth pursuit movements better in the horizontal than vertical dimension, and better in the downward than upward direction. Smooth pursuit movements can have a velocity as high as 90°/s (see Reference [9]).

**3.1.9**
**sample interval**
the epoch of time over which measurements are taken

EXAMPLE The duration of an in-vehicle task on entering a destination into a route guidance system, where the evaluation in interest is to evaluate the driver behaviour when the driver performs a task of entering the destination.

Note 1 to entry: Usually, this will be the contiguous epoch of time that is associated with an event or task that is of interest in the study. The sample interval is the period of time (from start to end) during which data are extracted.

**3.1.10**
**transition**
change in eye *fixation* (3.1.4) location from one defined *area of interest* (3.1.1) location to a different defined area of interest

Note 1 to entry: A transition could be composed of a large *saccade* (3.1.7) with further head movements to compensate the required total amount of viewing deviation when changing from one AOI to another. E.g. visual task to observe mirror, diverted from the main task while observing the road ahead.

Note 2 to entry: See F.1 and F.2.

**3.1.11**
**visual angle**
angle subtended at the eye by a viewed object or separation between viewed objects

Note 1 to entry: The figure below shows the visual angle $\alpha$.

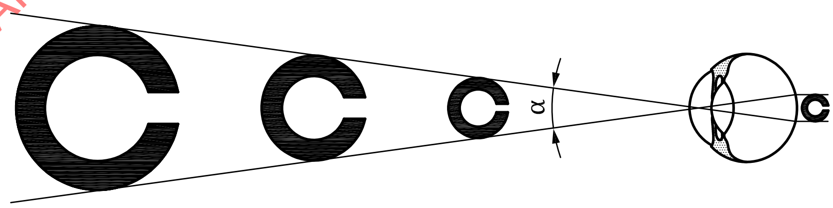



**Figure 2 — Visual angle $\alpha$**

**3.1.12**
**visual demand**
degree or quantity of visual activity required to extract information from an object to perform a specific task

Note 1 to entry: This can be determined by, e.g. *number of glances* (3.2.2.6), *total glance time* (3.2.2.11). See F.1.

**3.1.13**
**visual display**
device used to present visual information

Note 1 to entry: See F.1.

## 3.2 Terms for metrics

### 3.2.1 Basic direct metrics

**3.2.1.1**
**duration of diversion**
period of *glance duration(s)* (3.2.1.3) associated with directions of gaze away from one *area of interest* (3.1.1) to another area of interest during a defined period of interest

Note 1 to entry: See F.3.

**3.2.1.2**
**dwell time**
sum of consecutive individual *fixation* (3.1.4) and *saccade* (3.1.7) times to an *AOI* (3.1.1) in a single *glance* (3.1.6)

Note 1 to entry: See F.4.

**3.2.1.3**
**glance duration**
time from the moment at which the *direction of gaze* (3.1.3) moves towards an *AOI* (3.1.1) to the moment it moves away from it

EXAMPLE A look at the inside rear-view mirror.

Note 1 to entry: Under certain specific conditions, this may be operationalized differently. See F.2.

**3.2.1.4**
**minimum glance duration**
shortest possible duration for a *fixation* (3.1.4) to an *AOI* (3.1.1)

Note 1 to entry: The “small fixations” observed to an area of interest of ≤120 ms are classified as *fly through* (3.1.5).

Note 2 to entry: Fixations to an area of interest ≤120 ms are physically not possible. If an eye tracker is in use and records such duration it can be classified as part of a *transition* (3.1.10) between areas of interest and it is not taken into a *glance* (3.1.6) period.

**3.2.1.5**
**transition time**
duration between the end of the last *fixation* (3.1.4) on an *AOI* (3.1.1) and the start of the first fixation on another AOI

EXAMPLE Visual task to observe mirror, diverted from the main task while observing the road ahead.

Note 1 to entry: A *transition* (3.1.10) could be composed of a large *saccade* (3.1.7) with further head movement to compensate the required total amount of viewing deviation when changing from one AOI to another.

### 3.2.2 Glance derived metrics

**3.2.2.1**
**glance location probability**
probability that the eyes are fixated at an *AOI* (3.1.1) (or set of related AOIs) during a condition, task, subtask or sub-subtask

EXAMPLE 7,85 %.



Note 1 to entry: Glance location probability = *number of glances* (3.2.2.6) to an AOI during a condition, task, subtask or sub-subtask) / (number of glances to all AOIs during a condition, task, subtask or sub-subtask) × 100 %; unit [%].

**3.2.2.2**
**glance rate**
number of glances per unit of time
*number of glances* (3.2.2.6) divided by the duration of condition, task, subtask or sub-subtask

EXAMPLE 0,53 glances/s.

Note 1 to entry: Unit [number of glances/s].

Note 2 to entry: To obtain the glance rate of a task, only glances to the task are counted and divided by the duration of condition, task, subtask or sub-subtask. To obtain the glance rate to all locations during a test scenario, all glances to all locations are counted and the total divided by the total length of the condition, task, subtask or sub-subtask.

**3.2.2.3**
**link value probability**
probability of a *glance* (3.1.6) *transition* (3.1.10) between two different *AOIs* (3.1.1) during a condition, task, subtask or sub-subtask

EXAMPLE 17,39 %.

Note 1 to entry: Link value probability between AOI A and AOI B = (number of glance transitions from A to B + the number of glance transitions from B to A) / (number of glance transitions between all AOIs) × 100 %; unit [%].

**3.2.2.4**
**maximum glance duration**
longest *glance duration* (3.2.1.3) to an *AOI* (3.1.1) (or set of related AOIs) during a condition, task, subtask or sub-subtask

EXAMPLE 2,12 s.

Note 1 to entry: Maximum glance duration = max [glance duration1, glance duration2, ..., glance duration *n*]; unit [s].

**3.2.2.5**
**mean glance duration**
mean duration of all *glance durations* (3.2.1.3) to an *AOI* (3.1.1) (or set of related AOIs) during a condition, task, subtask or sub-subtask

EXAMPLE 1,28 s.

Note 1 to entry: Mean glance duration = [*total glance time* (3.2.2.11)] / [*number of glances* (3.2.2.6)] during a condition, task or subtask; unit [s].

**3.2.2.6**
**number of glances**
count of *glances* (3.1.6) to an *AOI* (3.1.1) (or set of related AOIs) during a condition, task, subtask or sub-subtask

EXAMPLE 9 glances.

Note 1 to entry: Unit [count].

**3.2.2.7**
**percentage time on AOI**
ratio representing the percentage of time *glances* (3.1.6) are within an *AOI* (3.1.1) (or set of related AOIs) during a condition, task, subtask or sub-subtask

EXAMPLE 53,47 %.



Note 1 to entry: Percentage time on AOI = Σ (glance duration 1, glance duration 2, ..., glance duration *n*)/(duration of condition, task, subtask or sub-subtask) × 100 %; unit [%].

**3.2.2.8**
**percentage of eyes off road time**
**PEORT**
percentage of time during a condition, task, subtask or sub-subtask [i.e. during a *sample interval* (3.1.9) of interest] that *glances* (3.1.6) are not on the road scene ahead

EXAMPLE 41,29 %.

Note 1 to entry: Percentage of eyes off road time = [*TEORT* (3.2.2.10)] / (duration of condition, task, subtask or sub-subtask) × 100 %; unit [%].

**3.2.2.9**
**percentage transition time**
percentage of time during a condition, task, subtask or sub-subtask (i.e. during a *sample interval* (3.1.9) of interest) during which the eyes are in *transition* (3.1.10) (or are in movement between areas-of-interest)

EXAMPLE 3,23 %.

Note 1 to entry: Percentage transition time = Σ[*transition time 1* (3.2.1.5), transition time 2, ..., transition time *n*) / (duration of condition, task, subtask or sub-subtask] × 100 %; unit [%].

**3.2.2.10**
**total eyes off road time**
**TEORT**
summation of all *glance durations* (3.2.1.3) to *AOIs* (3.1.1) other than the road scene ahead during a condition, task, subtask or sub-subtask

EXAMPLE 103,32 s.

Note 1 to entry: TEORT = Σ(glance durations to AOIs defined as off road-scene-ahead); unit [s].

Note 2 to entry: The road scene ahead excludes driver's rear-view mirror *glances* (3.1.6).

Note 3 to entry: This metric should be operationalized in an appropriate way for each study – since the concept of "not on the road" may comprehend different AOIs depending upon study objectives. When TEORT is utilized, clearly specify which AOIs are counted on the road-scene ahead and which AOIs are defined as off the road scene ahead.

**3.2.2.11**
**total glance time**
summation of all *glance durations* (3.2.1.3) to an *AOI* (3.1.1) (or set of related AOIs) during a condition, task, subtask or sub-subtask

EXAMPLE 17,88 s.

Note 1 to entry: Total glance time= Σ(glance duration 1, glance duration 2, ..., glance duration *n*) ; unit [s].

# 4 Trial planning and evaluation

## 4.1 General

Assessment of driver visual demand can be carried out in relation to many forms of TICS (transport information and control systems) applications and road environments. Consideration should be given to the factors/aspects influencing driver visual behaviour.

## 4.2 Trial planning

### 4.2.1 General

The following aspect given in 4.2.2 to 4.2.6 should be considered.

### 4.2.2 Roadway/traffic specification

An appropriate operational environment for the specific TICS application under evaluation should be chosen. The type of roadway and likely traffic conditions to be encountered should be defined within the trial (or study). This may entail defining and documenting the roadway geometry, signals, and surroundings, as well as describing the driving scenarios participants will experience (including speeds of travel, manoeuvers, traffic densities, movement of traffic, weather conditions and so forth). For naturalistic driving studies or long-term data collection refer to Annex H.

### 4.2.3 Vehicle specification

Experimental apparatus used to represent the driving task should be described as fully as practicable.

EXAMPLE Documenting the make and model of the road vehicle employed or the driving simulator characteristics employed (including key parameters of the vehicle dynamics model, whether the simulator has a fixed- or moving-base, the breadth of its field of view, etc.).

### 4.2.4 TICS specification

The characteristics of the TICS should be reported.

EXAMPLE Type, position and image quality of a visual display (resolution, contrast, colour-rendition, reflectivity/glare).

### 4.2.5 Participant selection

Evaluation trials of TICS applications should use a representative sample from the target population for the specific TICS. This driver sample should be categorized by age, gender, visual ability (including colour vision deficiencies, as well as whether and what type of corrective lenses are required to drive) and driving experience.

### 4.2.6 Participant training

Trial objectives will determine the need for participant training in the use of the TICS. Assuming that some form of training is required, participants should receive clear and consistent guidance. The tasks and subtasks associated with the TICS should be fully explained to the participant and the limitations of responsibility and pacing in time of these between the driver and experimenter should be specified. Each participant's familiarity with the TICS prior to the trial should be reported. When determining the usability of the TICS device, consideration should be given to the level and assessment of training required.

### 4.2.7 Data exclusion

Control procedures for individual evaluation trials within an experimental programme should include guidelines for the conditions under which the trial is to be terminated. The procedures for verification of calibration and quality assurance in Annexes B and E also give guidance if data have to be excluded.

EXAMPLE Trial aborted for failure to complete a task or subtask: document how this is to be recorded or how the trial is to be re-scheduled.



## 4.3 Steps for data acquisition and data processing

The acquisition and analysis of visual behaviour data is a process including the following process steps as described in Figure 3.

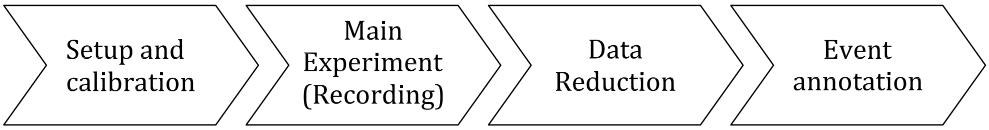



**Figure 3 — Process steps of acquisition and analysis of visual behaviour**

Each of these steps can suffer from different factors leading to artefacts, insufficient data quality or missing data.

Depending on the data acquisition technology and technical setup this clause describes procedures and metrics to check the quality of data and engage countermeasures.

The procedures can be grouped into the following categories:

— preparation of participant individually;

— check of technical availability of tracking equipment;

— check of tracking accuracy of tracking setup;

— check for quality of manual reduction;

— check for quality of automated reduction.

While a representative sample of participants is necessary to generalize results, some preparations can have a significant benefit when collecting data with an eye tracker by minimizing extraneous factors not related to the study. In some cases, the appearance of the participants might negatively affect the quality of the eye tracking data (e.g. makeup, hair position, eye glasses, amongst others). For example, it might be necessary to ask participants beforehand to wear correction contact lenses, avoid use of fake colour contact lenses or have them tie back long hair to improve the eye tracking quality.

## 4.4 Experimental conditions, tasks, subtasks, sub-subtasks, and relationship

### 4.4.1 Experimental condition

This is considered to encompass all visual behaviour of the driver during an experimental session.

EXAMPLE The distributions of visual scanning to all specified AOIs of the visual scene (including the TICS), from the specified start of a test route to its specified purpose.

Experimenters will need the flexibility to define experimental conditions that are relevant for their research goals. However, when studies involve examining glance patterns for secondary tasks while driving, the following experimental conditions may be useful for planning and for performing the research. The following terms are introduced because they define intervals of time and behaviour that may be of particular interest when evaluating a TICS – and, hence, in analysing the glance data associated with a TICS.

### 4.4.2 Task

Refers to a sequence of interactions undertaken to achieve a goal. Glance behaviour may be measured over the duration of a task.

Definitions of start and end of a task from the Visual-Manual NHTSA Driver Distraction Guidelines (see Reference [15]): “Start state for a task means the pre-defined device state from which testing of a task always begins. This is frequently the "home" screen, default visual display state, or other default driver interface state from which a driver initiates performance of the task (IV.2.j). Start of data collection means the time at which the experimenter tells a test participant "begin" (or, by some means, issues a non-verbal command indicating the same thing). Test participant eye glances and vehicle driving performance are examined only after the start of data collection (IV.2.i). End State for a Testable Task means the pre-defined device state sought by a test participant to achieve the goal of that testable task (IV.2.d). End of Data Collection means the time at which a test participant tells the experimenter "done" (or, by some means, indicates non-verbally the same thing) (IV.2.c)."

EXAMPLE All visual behaviour occurring during the procedure (task) of entering a destination into a route guidance system.

### 4.4.3 Subtask

Refers to a sequence of interactions undertaken to achieve a subgoal of the task (often one specific interaction). Glance behaviour may be measured over this (shorter) duration of the subtask.

EXAMPLE When entering a destination into the route guidance system, all visual behaviour associated with entering the "city name" portion of the destination.

### 4.4.4 Sub-subtask

Refers to operations or interactions with lower-level subtask elements (e.g. individual controls or screens).

EXAMPLE 1 On the route guidance system, glance behaviour associated with the sub-subtask of "verify the city name appeared" on destination entry screen.

EXAMPLE 2 On the route guidance system, entering a single letter of the destination by aligning the finger on to the touch screen of keyboard.

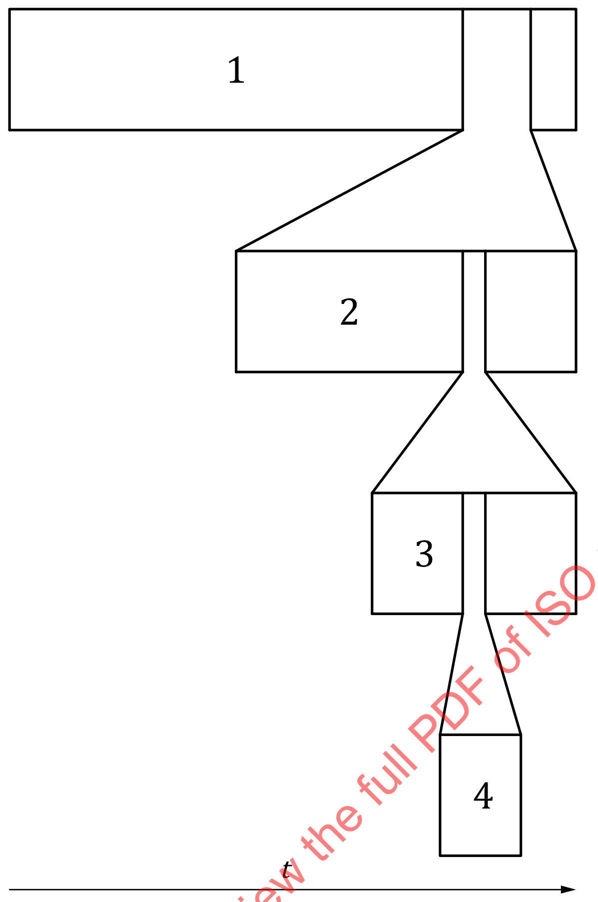
### 4.4.5 Relationship

Refers to the relation between an experimental condition, a task, a subtask and a lower level subtask element; and is graphically represented in Figure 4.

NOTE Users of this document can consult references on hierarchical task analysis for guidance on how to decompose a task (e.g. Reference [12]).






**Key**

1 experimental condition
2 task
3 subtask
4 sub-subtask
*t* time

**Figure 4 — Experimental condition, task and subtask — Relationship**



# 5 Recording equipment

## 5.1 General

Clause 5 provides practical advice on the use of data recording equipment to monitor driver visual demand.

Figure 5 shows basic technical setups which can be used for data acquisition and data processing in a visual behaviour study.

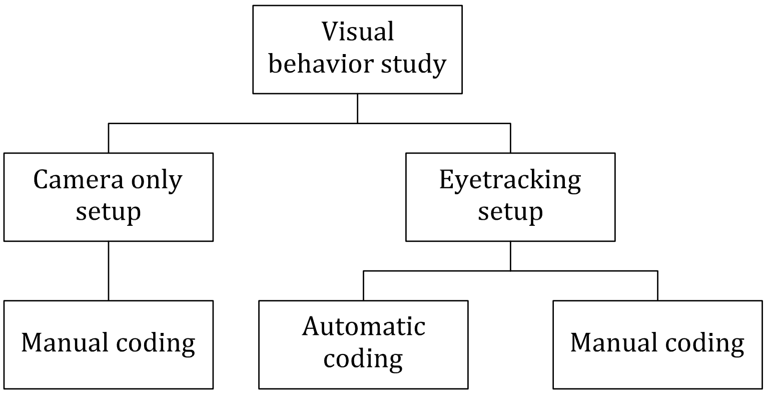



**Figure 5 — Basic technical setups**

## 5.2 Eye tracking equipment

### 5.2.1 General

In general, there are three different methodologies for recording eye tracking data:

— head-mounted eye tracking systems (5.2.2);

— remote eye tracking systems (5.2.3);

— camera only systems (5.3.4).

(See also Reference [4].)

Additional recording equipment is described in Annex I.

### 5.2.2 Head-mounted eye tracking systems

With head-mounted eye tracking systems, the participant wears components of the eye tracking system directly on the head. The components necessary for the eye tracking are mounted on a helmet, a cap or on a device similar to glasses. Head-mounted eye trackers may consist of the following components.

— Scene camera: this camera records what the participant can see.

— Eye camera: this camera records at least one eye.

NOTE 1 The eye can be recorded directly or via an infrared (IR) reflective mirror.

NOTE 2 Calibration of eye camera to scene camera is necessary to transform the x and y coordinates from the eye camera coordinate system to the scene camera coordinate system. Thereby the head-mounted eye tracking system is able to correlate in the scene camera view where the participant is gazing at.

— Infrared LED: the infrared LED typically makes the eye visible in the infrared spectrum. Thereby the system becomes more independent and robust from the surrounding lighting conditions, without disturbing the driver view.

Infrared light: the infrared light is typically not visible to human eye visual spectrum. If improperly used it can be hazardous to the human eye. Follow the safety precautions, for example given in IEC 60825-1 and IEC 62471.



### 5.2.3 Remote eye tracking systems

With remote eye tracking systems, the glance behaviour is recorded by at least one camera mounted in some location that can record the driver's eyes, such as on the dashboard.

— Remote eye trackers (and/or image recording systems) may consist of the following components.

  — Eye camera: at least one camera is directed to the driver's face to capture where the participant is looking.

  — Scene camera: the scene camera records the road scene ahead.

  Additional to the road scene camera, a second in cabin driver TICS operation camera may be installed if the study target to track the driver task. Eye glance observed to align the finger onto the touch panel prompt for example. Or finding the toggle switch positioning to achieve targeted operation. Care shall be taken that no interferences (ligthing, frequencies) between the recording systems occur.

  NOTE 1 Calibration of eye-camera to scene-camera is necessary to transform the x and y coordinates from the eye-camera coordinate system to the scene-camera coordinate system. Thereby the remote eye tracking system is able to indicate in the scene-camera view where the participant is looking.

  — Infrared LED: the infrared LED typically makes the eye visible in the infrared spectrum. Thereby the system becomes more independent and robust from the surrounding lighting conditions, without disturbing the driver view.

  Infrared light: the infrared light is typically not visible to human eye visual spectrum. If improperly used it can be hazardous to the human eye. Follow the safety precautions, for example given in IEC 60825-1 and IEC 62471.

  — Annotation: remote systems include conventional methods of capturing glance behaviour on video.

## 5.3 Setup and verification of calibration of eye tracking systems

### 5.3.1 General

Users of eye tracking systems may check the correct recording of an eye tracking system once if an eye tracking system is integrated in a new experimental setup. If they decide to check the correct recording, they shall use the eye tracker validation task (EVT) procedure, follow Annex C. The application of the EVT shows whether the combination of eye tracking system and experimental environment leads to reducible data sets free or nearly free from tracking artefacts.

### 5.3.2 Recording with eye tracking systems

For each participant and each experimental condition, the setup of the eye tracking system in the experimental setting and its calibration shall be checked using the verification of calibration procedure (follow Annex B) prior to the experiment. The accuracy and quality of calibration shall be checked. The verification criteria are given in Annex B. The Cohen’s kappa shall be checked (kappa prior).

This procedure shall be applied for each participant and experimental condition after the experiment (post verification of calibration) leading to kappa post.

If initial calibration is not acceptable the experimenter shall improve the setting and perform the verification of calibration again until the quality is sufficient for the intended experiment needs.

If Kappa post is not acceptable the experimenter shall document the recalibration and potential countermeasures shall be taken during data reduction if an automatic recalibration of the data is possible. Otherwise the data set cannot be included into the sample.

### 5.3.3 Recording with remote eye tracking systems

In principle, follow the same procedure as described in (5.3.1) but in this setup the scene camera of road ahead is fixed.

### 5.3.4 Camera only systems for manual eye-glance analysis

Cameras for manual eye-glance annotation require at least one camera directed towards the driver's face. The camera directed at the face should be placed in such a way that the view of the camera is not being blocked by driver's hands, sun visor, etc. When multiple face cameras can be used, the observation to eye glance becomes easier if a camera is positioned at nearby the AOI or the AOI of most importance. Ideally a scene camera(s) that captures what the driver could possibly be looking is suggested to be installed. (For example, a camera that captures the hands helps to differentiate a potential glance to a cell phone in the driver's hand instead of a glance to the centre stack.)

The video should be collected at a minimum of 10 Hz.

An infrared sensitive camera combined with infrared light source can be used to supplement ambient illumination. When considering the camera type, field of view, position and infrared lighting it is important to consider whether the data may also be used for automated eye-glance coding.

## 5.4 Setup and check of recording

**If data will be reduced manually** users of eye tracking systems **may** check the correct recording of an eye tracking system. If they decide to check the recording, they shall use the EVT procedure (refer to Annex C) once an eye tracking system is integrated in a new experimental setup. The application of the EVT helps to clarify whether the combination of eye tracking system and experimental environment leads to manually reducible data sets free from tracking artefacts, in case the eye tracking system produces video material.

**If data will be reduced automatically** users of eye tracking systems **should** check the correct recording of an eye tracking system using the EVT procedure (refer to Annex C) for each experimental condition.

## 5.5 Additional equipments

The following additional equipments are also typically required:

— computer for storage and control; and

— eye tracking software: the eye tracking software records, processes and stores the data.

## 5.6 Installation

Although the specific conditions of an experiment may vary, the following general principles should be applied.

The installed data collection system and employed procedures should not obscure the driver's view of the roadway or any in-vehicle driving relevant equipment and should not cause the driver any unnecessary distraction.

These criteria also apply to experimenters who may be present within the test vehicle to avoid failure of judgement during experiments.



# 6 Data reduction

## 6.1 General

The amount of raw recorded data is enormous and the recorded sequences of data which are not relevant to the study may be deleted and relevant data converted to quantitative metrics to reduce the amount of recorded data.

A reduction of experimental records of visual behaviour data can be accomplished by extracting and calculating quantitative metrics of visual behaviour. Clause 6 provides guidance on the assumptions to be made and the procedure to be performed in data analysis to achieve the data reduction appropriately.



## 6.2 Sample interval

Two regimes may be adopted in the reduction of visual behaviour records:

— reduction of the entire experimental session, for all identified AOIs of the visual scene;

— reduction of the forward view and AOIs relevant for the study (e.g. TICS display).

## 6.3 Manual reduction by raters/data analysts

For manual reduction, the data analyst shall be trained to carry out the following steps.

a) Advance the record of visual behaviour to the start of a sample interval (experimental condition, task or subtask) of interest.

b) Examine and identify the first video frame of the information on gaze direction to determine the glance location, then record this as the AOI applicable and the starting time for that specific glance.

c) Advance the record of visual behaviour frame-by-frame until the driver's eyes start to move to another specified AOI. When this occurs, record the glance duration for the previous AOI, record the new AOI and the time code for that frame.

d) Repeat the previous steps frame-by-frame until the sample interval has been fully reduced.

Annex A gives an overview of the details for the training of the data analyst.

## 6.4 Manual reduction by raters/data analysts of data from a camera only setup

Before a data analyst/rater starts with data analysis his/her ability shall be checked using an AOI specific reference template (follow Annex A). A Kappa of 0,7 shall be reached. If not the training of this rater shall be continued.

If only one rater reduces a data set, the 5 % post hoc check data verification shall be applied (follow Annex D).

It shall be documented which rater did the data reduction.

## 6.5 Manual reduction by raters/data analysts of data from an eye tracking system

In case of an eye tracking setup in combination with manual data reduction, the following steps and quality measures are specified.

The eye tracking system delivers a video recording of the participants' visual behaviour including a visual representation of the glance direction that is used as the basis for a manual reduction.

The reduction of data resulting from an eye tracking system application shall be checked for its quality by the 5 % post hoc data verification procedure (follow Annex D).

It shall be documented which rater did the data reduction.

### 6.6 Data Protocol for manual reduction

An example of a reduction record is given in Table 1.

**Table 1 — Example reduction record (glance durations are in [s])**

| Clock time | Driver mirror | Right region | Left region | Into car | Notes |
|---|---|---|---|---|---|
| 54:51:31 | 0,6 | | | | |
| 54:52:44 | | 0,5 | | | |
| 54:56:22 | 0,8 | | | | |
| etc. | | | | 1,5 | Looks at the instruments. |

### 6.7 Summary data

Data that summarizes the trial, encompassing the information as given in Tables 2 to 5, should be reported.

**Table 2 — Participant summary information**

| Parameter | Information required |
|---|---|
| Age | Range, mean and standard deviation |
| Gender | Number of each gender |
| Distance (kilometres or miles/year during the previous five years) | Range and mean |
| Years of driving | Range, mean and standard deviation (if absolute values are reported) |
| Visual legal compliance | Statement that all participants comply with relevant legal requirements for minimum driving visual ability |
| Visual ability | Definition of range of participants visual ability relevant to the experimental design |
| Exclusion criterion | Description and frequency of exclusions |

**Table 3 — Experimental design summary information**

| Parameter | Information required |
|---|---|
| Experimental conditions | Number and description |
| Duration of condition | Range, mean and standard deviation |
| Independent variables | Number and description |
| Dependent variables | Number and description |
| Vehicle environment | Public road, test track or simulator |
| Type of roadway | Arterials, collectors, locals (see FHWA functional classification system) |
| Traffic density | Use level of service (A-F) |
| Exclusion criterion | Description and frequency of exclusions |

**Table 4 — TICS and control condition summary information**

| Parameter | Information required |
|---|---|
| System | Description of system including functions, controls and displays |
| Tasks | Number and description |
| Subtasks per task | Number and description |
| Task, subtask and sub-subtask pacing | Frequency (number of tasks, subtasks and sub-subtasks/ per time unit) and description |
| Participant experience of TICS | Categorization of experience |
| Exclusion criterion | Description and frequency of exclusions |

**Table 5 — Visual data classification summary information**

| Parameter | Information required |
|---|---|
| Number of regions | Number and boundaries (forward view, driver mirror etc.) |
| Calibration of AOIs with respect to driver's glances | Statement that all participants instructed to fixate on each AOIs prior to experimental condition, including relevant participant instructions |
| Start of experimental conditions, tasks and subtasks | Time (and definition of environmental cue if any) |
| Stop of experimental conditions, tasks and subtasks | Time (and definition of environmental cue if any) |
| Basic unit of observation for data reduction | Data recording resolution |
| Exclusion criterion | Description and frequency of exclusions |

# 7 Data reduction using automated gaze analysis of eye tracking system

## 7.1 General

Some eye tracking systems enable an automated gaze detection and data analysis by using a software-based automated AOI detection. By using such function, participants' gaze behaviour does not need to be analysed manually frame by frame, saving a lot of time and effort. However, such automated detection of AOIs may induce a new source of errors. Clause 7 provides guidance and procedures to be performed when using automated gaze analysis.

## 7.2 Data quality verification using 5 % of entire collected data

### 7.2.1 Positional/orientation errors

The automated software-based AOI classification may contain deviations of the detected position, size and shape from the real position, size and shape of the actual physical object (e.g. a display).

EXAMPLE A participant glances onto the TICS (= AOI) but the gaze is not correctly counted because the AOI detection algorithm classifies the gaze as a gaze onto the passenger side mirror at a shifted position or not at all (because of technical problems with the AOI-related marker detection).

### 7.2.2 Detection time errors

Another kind of errors may occur if the software does not classify glances on an AOI from the very first frame appearance and classifies them with some delay causing deviation of the time-related variables.

EXAMPLE A participant's gaze turns, e.g. to a display (= AOI) but the marker detection algorithm classifies the AOI just a few moments later which leads to a wrongly calculated glance duration.

### 7.2.3 Verification of Cohen's kappa to secure accuracy of automated analysis

In view of the possible influence of errors described in 7.2.2 and 7.2.3 when analysing eye gaze behaviour using an automated AOI detection, the performance of the detection shall be checked to verify that the automated AOI detection works with sufficient accuracy.

A subsample of 5 % of the complete data set shall be analysed and compared in the following two ways:

(1) coding the glance locations by a rater manually; and

(2) automated AOI detection and coding.

Then the agreement between both methods shall be calculated using Cohen's kappa (follow Annex F).

The verification procedure is the same as for calculating the interrater agreement between two raters who judge eye gaze behaviour based on videos filming the participant's face as described in Annex E.

The quality criterion for accuracy is also the same, as described in Annex E: if the comparison of manual and automated coding between the analysed subsample shows that all Cohen's kappas have a value of 0,70 or higher, the remaining 95 % of the study may be analysed by the automated AOI detection.

However, if one or more Cohen's kappas are lower than 0,70, readjust the parameter set of the software detection algorithm or replace the software detection algorithm to analyse and achieve Cohen's kappas that are higher than 0,70 with the 5 % of the data that was manually coded.

If Cohen's kappas equal or above 0,70 cannot be achieved, then the study data cannot be analysed fully automatically.

It is not allowed to draw another 5 % subsample.

Each of these steps shall be handled with care, documented and checked for quality before moving to the next step.

It is recommended to apply this rationale and procedure described in 7.2.3 to an even bigger subsample of the data set.

NOTE Different parameter sets or software detection algorithms for the automated reduction that correspond to heterogeneous experimental conditions (e.g. nighttime vs. daytime) in one data-set can help to improve the quality of the automated reduction approach.



## 7.3 Availability of the eye tracker data

As eye gaze behaviour recording is based on a recording process there might be missing data during an experiment due to the following reasons:

— missing video signal;

— participant is moving out of expected measurement range;

— inappropriate lighting conditions to rely on automated gaze analysis (e.g. low light);

— eye tracking algorithm loss of detection signal due to glaring noise source.

This may lead to the situation where a manual rater or an automated reduction algorithm has to code data sets including missing data.

In general, the availability of data should be as high as possible and be equally distributed over experimental condition, tasks and participants. Annex G provides recommendations and examples for how to treat situations with reduced availability of eye gaze data.

# 8 Data analysis and presentation

## 8.1 General

Glance AOIs may fall into four high level categories which can be further subdivided if necessary:

1) road scene ahead;

2) other traffic related AOIs (left road scene, right road scene, left side mirror, right side mirror, inside rear-view mirror, speedometer);

3) displays and control devices of interest;

   NOTE Special care is required when using transparently overlapping AOIs such as head-up display and symbols superimposed on the inside rear-view mirror, the left side mirror or the right side mirror (e.g. warning symbols of a blind spot warning system).

4) other non-traffic related AOIs (e.g. billboards, sky, and so on).

The fundamental glance metrics should be considered for visual demand assessment (as defined in Clause 3) and shall be calculated in relation to the above listed key AOIs. From these glance metrics, a number of derived metrics of visual behaviour have been defined and interpreted from the standpoint of visual demand. To gather valid results, one should care about the quality of eye tracking data and potential artefacts in the recorded data. For more information about these issues see G.1 and G.2.

## 8.2 Interpretation of findings from analyses of glance metrics

Example interpretations of some of the commonly derived glance metrics follow.

NOTE The driver accommodation and adaptation to the TICS can differ by age, influencing the time required for the test participant to perceive and extract the necessary information.

— A high **number of glances** to an AOI may indicate a high importance of the AOI, such that multiple glances are needed to extract information.

— The **total glance time** associated with an AOI (e.g. in-vehicle device) provides a measure of the visual demand posed by that AOI or the visual information in that particular location. As visual demand increases, total glance time should increase.

— The **mean glance duration** describes how long a participant has to look at a certain AOI (e.g. a TICS display) to perceive and extract necessary information from that specific AOI to perform an intended task. Shorter mean glance durations are an indicator that information can be perceived and extracted fast from an AOI and longer mean glance durations indicate the opposite.

— A high **glance rate** to an AOI may indicate a high importance of the AOI. This measure is related to event detection. Research has indicated that the largest decrements in event detection tend to occur when glance rate is very high (the eyes are moving a lot, in other words, making many transitions). During transitions, vision is suppressed – and perhaps the continuity of situation awareness is also disrupted as the number of glances increases (and, hence, the number of transitions increases). This is the only glance metric that correlates significantly with event detection, besides task-related glance metrics (percentage and durations for glances to task) (see Reference [1]). Together, these metrics account for between 66 % and 82 % of the variance in specific aspects of on-road event detection. It may be important to include this metric for distraction assessment whenever there are conditions that could lead to high levels of scanning (such as with some types of visual-manual tasks), or very low levels of scanning (which are sometimes associated with states of cognitive load).

— The **percentage time on AOI** describes the visual demand of an AOI (especially when operating a TICS). A high percentage time on AOI combined with a long mean glance duration while operating a TICS may indicate that the task's design does not allow it to be visually interrupted and resumed easily without loss of information. Design improvements to this aspect of the task may reduce its visual demand.

The above-mentioned glance metrics should be interpreted in conjunction for the TICS evaluation and not isolated from each other, because this may lead to wrong conclusions (see 8.3).

— Long **maximum glance durations** to a TICS display while operating may be a sign of high visual and mental demand caused by the TICS. This may be due to human factors, problems such as unexpected or delayed reaction from the TICS.

— **Glance location probability** on a given AOI reflects the relative visual demand associated with that AOI. Across a mutually exclusive and exhaustive set of AOIs, fixation probabilities capture where the eyes were fixated throughout a sample interval. Given such a distribution, visual demand assessment might statistically compare two such distributions (under two experimental conditions or tasks).

  EXAMPLE If device use were to induce a relative decrease in the fixation probabilities associated with the driving scene, such as road scene or rear-view mirrors, this would be considered indicative of the visual demand associated with the device.

— **Link value probabilities** represent the strength of the relationship between one AOI and another. The greater the link value probability, the stronger the need to time share attention between the two locations. In visual demand assessment, the link value probabilities may be analysed to assess how visual attention has been affected by TICS use or the driving conditions.

— Increasing **total eyes off road time (TEORT) and percentage of eyes off road time (PEORT)** indicate that the participant may be distracted by TICS. It can also be a sign for low primary task workload which may have the effect that the driver starts operating TICS in the car (which can in turn also lead to increased TEORT and PEORT).

— A **transition time** is roughly a linear function of the distance from one AOI to another. During the transition time, there is relatively little new visual information acquired by the driver. Thus, increased transition times reflect reduced availability for driver information-gathering.

## 8.3 Interpretation of multiple glance metrics

When drawing conclusions about visual behaviour and driver workload, users should examine multiple visual metrics.

EXAMPLE 1 Glance frequency and mean glance duration can be traded off within a fixed sample interval. That is, very long glance durations (indicative of high workload demand) can be associated with fewer rather than more glances. Thus, it is important to consider the two measures together, especially if the sample interval is fixed rather than allowed to reflect task completion time.

EXAMPLE 2 When comparing the speed gauge in a head up display (HUD) with a conventional speedometer in the instrumental cluster one can find that participants have a higher number of glances and a higher glance frequency to the speed gauge in the HUD than to the speedometer in the instrumental cluster. When only taking into account those two metrics this might lead to the conclusion that the HUD is more distracting. When also taking into account that the mean glance duration to the HUD is shorter than to the instrumental cluster and that the total glance time to both AOIs for the same time interval is the same, one will draw the conclusion that participants control their speed more often with the HUD without being more distracted which leads to an increase in safety (see Reference [3]).

EXAMPLE 3 The HUD also offers an opportunity to highlight another example of where multiple metrics can be examined. Because the HUD is typically located high in the field of view and closer to the driver's line of sight to the forward roadway, transitions to nearby AOIs on the roadway are shorter – and the probability of noticing events in nearby regions (such as unexpected roadway events like pedestrians or braking vehicles) is increased, with response times to them facilitated (see Reference [6]).



NOTE 1 Another consideration in interpreting glance metrics for new technologies (such as HUDs) is that they can have a novelty effect, which often results in increasing the number of glances to the AOI containing the new technology or display. This was the case for HUDs, and it was found that the novelty effect of increased glances to the HUD declined across 4-days of use (1 session of use per day). Thus, it is important to disentangle 'novelty effects' on glance metrics from effects that are more stable and representative of long-term use of a display once novelty has worn off.

Offering standard interpretations is a challenge because the context can have a strong effect on their meaning. For example, short mean glance duration might indicate efficient uptake of information, but it might also indicate the presence of many short, useless glances and consequently an inefficient uptake of the visual information.

NOTE 2 Regarding the issues on assessing HUD with visual measures there are other concerns that can be even more prominent. The HUD AOI can overlap the road ahead which makes some glance-based measures inappropriate.

# Annex A
(normative)

# Manual reduction procedures

## A.1 Introduction

This annex provides information focusing on manual analyses procedure targeting the classification of the participant eye movement down to inter AOI glance granularity. And in principle, a driver eye movement to observe an AOI is classified as a single glance.

As the duration of the raw recorded data becomes huge, recorded data are analysed to extract the glance relevant time information. This subclause describes the procedure and recommendation, especially when the procedure is done manually by rater, as manual glance reduction procedure.

When manually analysing visual behaviour data, special equipment, procedures and care shall be applied to ensure a high data quality. These are described in the following clauses.

## A.2 Procedure

### A.2.1 Video reduction software

For manual glance reduction it is recommended that computer software be provided to the rater to view and navigate the collected multiple recorded video in a synchronized way. The rater should have several camera recorded views available, for example including the driver's face, the driver's hands, and the forward roadway. Additional cameras that provide increased context for the driving environment are ideal and the software should allow the rater to navigate through the video with synchronization of different views, frame-by-frame and provide a video playback rate of 10:1 (10 frames per second) or higher for precise frame identification and registration.

It is also recommended that the computer software provides an editing function to enter glance information in sync with the video (e.g. one data point per frame of video). Alternatively, one can manually write down the glance information on paper or using Excel with time stamp information, for example. The image below illustrates an example software interface. Note that 4 camera views are available (face, hands, forward, rear), as well as a time series speed variable to assist with video navigation.

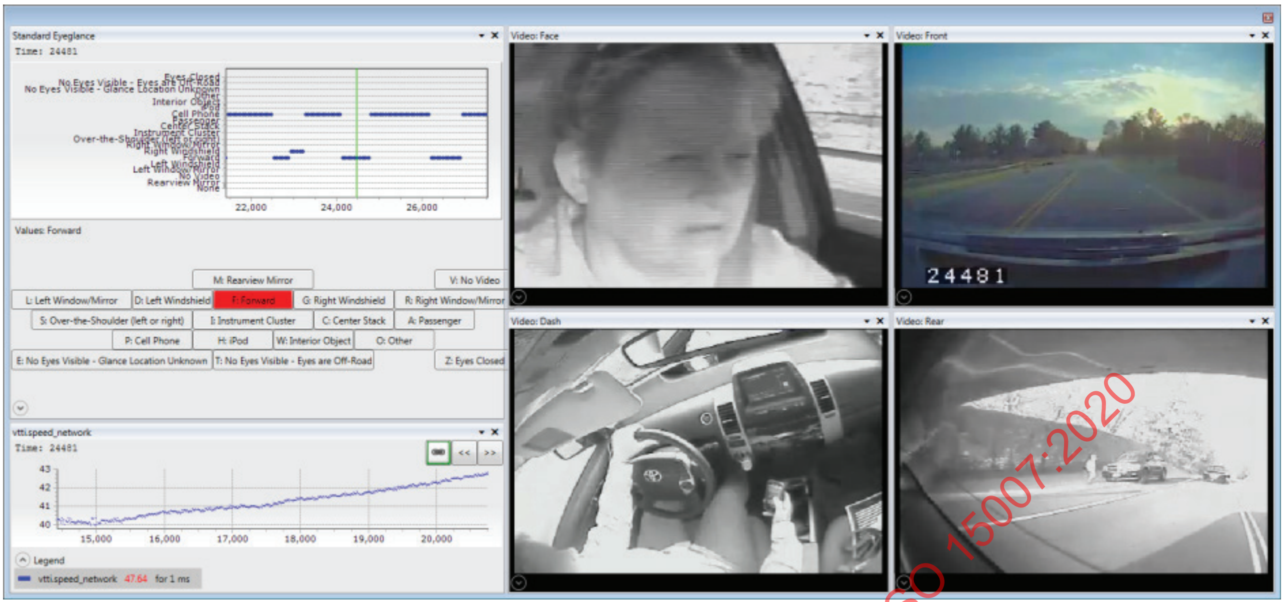



Figure A.1 — Video reduction software

### A.2.2 Coding of the glance in manual data

When manually coding glances, coding methods shall be determined according to the research needs and available resources. Following are different strategies in determining the coding method.

— **Fixation by fixation method**

In manual data reduction, a glance to an AOI is determined as the segmented time intervals starting either from last fixation on the last AOI to end of the fixation of the next AOI or from the first fixation on one AOI to the start of the fixation on the next AOI.

These intervals should be chosen as small as possible. The smallest unit is a frame in the video. A glance might be composed of several fixations and small saccade as shown in Figure F.4. For the glance to the road scene ahead, this AOI is in general composed of multiple driving relevant points of interest and the driver might exhibit several large saccadic eye movements. But for the purpose of the study, the transition observed in these points of interest and the saccadic eye movement does not need to be extracted during the analysis. They are grouped into a single AOI for a glance to the road scene ahead.

— **AOI dwell time method**

Saccades between fixations are generally not coded during manual glance reduction. All fixations and associated saccades are instead coded as one AOI. Thus, dwell time is coded rather than fixation time in manual glance reduction.

Multiple video cues can often help the rater to determine the difference between a glance to the radio and a glance to a cell phone, for instance, or between a glance out the right window and a glance to an adjacent passenger.

A glance might be composed of several fixations and small saccade as shown in Figure F.4. For the glance to the road scene ahead, this AOI is in general composed of multiple driving relevant points of interest and the driver might exhibit several large saccadic eye movements. For the purpose of the study, the transition observed in these points of interest and the saccadic eye movement does not need to be extracted during the analysis, they are grouped into a single AOI for a glance to the road scene ahead.

In manual glance reduction procedure, transitions between AOIs may be coded inclusive to a glance or as separate entities depending on the purpose of the experiment. There are several ways to handle transitions: code them as separate entities, code them with the origin glance, or code them with the destination glance. McClafferty et al (see Reference [8]) discussed the pros and cons of each method and presented a comparison of the resulting data and key glance metrics. Key points from that report are as follows.

1) Coding transitions as separate entities requires more time and resources than including them with either the origin or destination glance because two shifts shall be identified for each transition rather than one.

2) Coding transitions with the destination glance require more time and resources than including them with the origin glance because the destination glance shall be determined before the transition can be coded. This effect may be minimized if the reduction software allows video to be played and coded in reverse.

3) For naturalistic driving data, there are no significant differences in calculated glance durations, other key glance metrics, or rater reliability between the origin and destination methods or when separately coded transitions are included with either the origin or destination glance.

Due to the above, any of the three options for transition coding are considered equivalent, and the method selected should be determined according to the research need and available resources. However, if transition metrics (e.g. transition time, percentage transition time) are required by the experimenter, then transitions shall be coded separately. The way of coding should be consistent for different raters rating one data set.

### A.2.3 Quality assurance

Just as eye tracker data requires validation, manual glance reduction should always be accompanied by a documented quality assurance protocol that depends on specific institutional or laboratory specific requirements. The designated quality assurance staff should be well trained with extensive experience in video eye glance annotation. A minimum of 6 months coding experience and exceptional performance on numerous projects and proficiency tests is recommended. Quality assurance staff should also possess good communication skills such that errors can be both shown and explained to data reductionists; a good quality assurance staff member will simultaneously help improve the quality of the current dataset and further develop the skills of the data reductionists to improve future work. Suggested quality assurance procedures are listed below.

— Limit the time spent on reduction by each rater each day. For example, four hours per shift.

— Require raters to take frequent breaks during reduction hours. For example, ten minutes per hour of work.

— Establish feedback loops between each rater and the quality assurance staff. Quality assurance staff should periodically review some or all of each rater's work and request that the original rater re-review and make corrections to potential coding errors.

— Revise reduction protocols as needed. In any reduction process, there will be new scenarios that may not have been explicitly defined in the protocol. As these situations arise, or anytime that the research need shifts, reduction protocols should be updated and raters apprised of and/or re-trained on the changes.

— All project staff should sign off on both new and updated protocols to document and confirm effective communication and comprehension.

### A.2.4 How to produce the video material for the exemplar standard/template

The exemplar standard videos and data sets should be developed in a way that is consistent with the glance reduction task that raters will be required to perform after training is complete. This requires consideration of epoch length, AOIs to be coded, environmental conditions of the collected data, participant characteristics, and task types.

— The length of epochs in the example, practice, and test epochs should be similar in length to the epochs to be coded after proficiency is achieved. (i.e. if the epochs to be coded range from 30 s – 90 s, then the training and testing epochs should be similar). It is recommended to use training and testing epochs that are ≥30 s long in order to allow raters to learn to use context and follow driver glances through a natural progression.

— The AOIs coded during the training/testing phase should be the same AOIs that raters will be expected to code once proficiency is achieved.

— The environmental conditions present in the training and testing video epochs should include a similar range of the conditions that will be present in the epochs to be coded after proficiency is achieved. If the epochs to be coded include specific camera angles, face size and distance, different lighting scenarios (e.g. day vs night vs dawn/dusk), different sun angles (e.g. in driver's face vs casting a shadow), different driving environments (e.g. heavy vs. light traffic, intersection traversals vs merging vs straight road segments), etc. then the training and testing epochs should include those same conditions.

— The characteristics of the participants to be coded should also be reflected as much as possible in the training and testing epochs. This primarily is concerned with the ensuring that multiple drivers of heterogeneous characteristics are included in the training and testing epochs and providing for the presence or absence of different types of glasses. It can also include consideration for different eye shapes and colourings.

— Finally, the tasks to be performed by research participants should be the same tasks that are included in the training and testing epochs. For example, if the research question relates specifically to a cell phone texting task or a lane change task, then the training and testing epochs should include those same types of tasks to ensure that raters are proficient in identifying the glance patterns likely to be present in the research data.

The pre-coded, practice, and test epoch sets will ideally contain 10 – 15 video epochs each, or 10 – 15 total minutes of video to be reviewed or coded. Having multiple sets of each type of event (pre-coded, practice, and test), while not always feasible, does allow for re-training and re-testing if initial proficiency scores are unsatisfactory.

The training provided to raters should include a demonstration of the tools to be used (e.g. specific software and reduction interface), a video-coding demonstration of different driver tasks and glances, and a discussion of different context clues that can be used to identify glances to each AOI. This initial training should take place in person whenever possible, with a written protocol provided for later reference. The rater-in-training can then work alone (with access to the trainer for questions) to review the pre-coded epochs and code the practice epochs. Finally, the test epochs should be coded by the rater without input from the trainer.

— All glance raters shall go through an initial training phase before coding any data. Prior to doing manual eye-glance reduction, an experienced glance rater should train each rater specific to the project in the reduction protocol and in the use of the software interface. The training should include both pre-coded epochs (i.e. defined period of times) where the rater can see the correct codes and a set of "practice" epochs where the rater can practice the reduction to become comfortable with the software and glance definitions. The training should be concluded with a set of "test" epochs where the rater completes the coding and the resulting glance data is compared to that of an exemplar standard. The exemplar standard should have been coded by a reduction trainer, supervisor, or manager. Ideally multiple expert raters should jointly develop the exemplar standard. A rater's test results (e.g. %-agreement with exemplar standard, types of glances missed, transition timing off) should be shared with that rater to familiarize with the possible erroneous cases, and if necessary, a second round of training and/or testing should be completed. The experienced rater or another quality assurance staff will determine if a rater has achieved an accepted level of reduction proficiency compared to exemplar standard.

— New, inexperienced raters that have recently completed the training process and are newly assigned to a research project should initially receive quality assurance checks on all of their work (i.e. actual data). These checks should be performed by an experienced rater with a documented high level of accuracy, and raters should be required to review and correct their mistakes before coding additional data. The experienced rater or another quality assurance staff will determine if the new rater has achieved an accepted level of reduction proficiency.

— Once a rater has demonstrated consistent accuracy that meets pre-defined criteria, then that rater's data should be checked periodically. The amount of data checked will depend of several factors, such as the number of AOIs, size of the AOI, and relative angle between AOIs. If the established accuracy threshold is not met by a given rater, the amount of data checked should again be increased for that rater.

— Expert-rater reliability should be tested and reported for manual glance reduction efforts. This requires all of the raters assigned to a given reduction effort to code the same small sample of epochs independently, and the results to be compared to an exemplar standard for agreement. These tests should occur at regular intervals during a reduction effort (monthly is recommended). Each test should include a similar number of epochs (for example, n = 10 or n = 20, depending on the length of the epochs). Ideally raters would be "blind" to this testing and would not know they were being tested on a given event. If expert rater test scores drop below a preset threshold (e.g. 90 %), the quality assurance staff should respond by increasing the percentage of epochs that are reviewed for that rater and/or the rater may require retraining.

— Intra-rater reliability should also be checked by requiring each rater to code a small sample of epochs multiple times over the course of the reduction task. In this case, the rater's data from each subsequent completion would be compared to the previous completion for that epoch and/or to the first completion. Intra-rater testing should also occur at regular intervals during a reduction effort with the raters remaining "blind" to the testing process. Low levels of intra-rater agreement (e.g. <90 %) should be followed up with an examination of trends in a given rater's coded data and test scores to verify that any changes in coding practices that have evolved for that rater are beneficial rather than detrimental to the results.

— Upon completion of the reduction task, and prior to performing statistical analyses, the reduction manager should conduct a thorough validation process. Specific checks to be done include (but may not be limited to):

  — confirm that all video frames within the desired epochs are coded;

  — check for glances coded to AOIs that seem unlikely or rare given the experimental conditions;

  — check for glances that may not fit defined criteria. (For example, if glances may not be less than 120 ms in duration, then confirm that is the case in the coded data.)

— Check for unusually long glances. (For example, it is recommended to check any non-forward glance that is more than 3 s long.)

— If transitions are coded separately, it is recommended to check any transitions that are longer than 0,3 to 0,4 s.

— Check any epochs that are coded exclusively or almost exclusively to one AOI, including the forward roadway.

— Check any glance sequences that are unexpected. (For example, a non-forward glance immediately followed by another non-forward glance is unusual, though certainly possible. These should be checked for accuracy.)



# Annex B
(normative)

# Verification of calibration — Check of availability and calibration accuracy of tracking equipment before recording data using a verification of calibration procedure



## B.1 Introduction

The procedure is as follows:

— at the beginning of each participant session a verification of calibration should be done;

— at the end of the experiment the calibration should be checked using the same verification procedure.

The verification of calibration ensures that the eye tracking system has calculated the point of fixation correctly after the calibration and that there is no drift in calibration during the experiment.

If the verification of the calibration shows that a shift of calibration occurred, there are two options to be taken:

— re-calibrate at the moment of drift;

— discard the data of the participant and replace with a new participant.

The verification of calibration method shall be applied to each participant before and after an experimental session.

The verification/proof of quality of calibration can be performed after the measurement campaign. It is recommended to perform the verification/proof of quality of calibration within the experimental session with this participant.

For the verification of calibration there are two possibilities:

1) iterative instructed glances onto specific locations in different AOIs;

2) glances onto specific locations in reference grids ("calibrations boards") in front of an AOI or within an AOI.

The experimenter should take into account the size of the AOI and research question for the selection of the specific locations and the number of specific locations.

## B.2 Method — Iterative instructed glances onto natural AOIs ("bounding boxes")

This method is particularly suitable for AOIs of different sizes and distances to the participant which results in different visual angles. The AOI should be an area which is specific. If an abstract AOI is used for the experiment (i.e. one that does not have clear instructions or is difficult to perceive; e.g. the right area of the windscreen), calibration shall use a concrete object. For example, this concrete object can be a marker or a piece of paper with a marked midpoint to glance at.

For checking the accuracy of the calibration by iterative instructed glances onto natural AOIs the procedure is as follows.

— At the beginning of the measurement the participant is instructed to glance at a certain point within an AOI, e.g. the midpoint of a display or the midpoint of the speedometer, or multiple locations depending on the size and/or shape of the AOI.

— All AOIs and all locations of each AOI that shall be analysed in an experiment have to be included into the verification of calibration.

— The participant is instructed to gaze alternately in about 1-second intervals at the instructed location within the AOI.

  An example of a sequence: road scene – dashboard – display – road scene – dashboard – display – road scene.

  It is recommended that the instructed sequence covers both long (maximum) and short transitions between different locations.

  An example for a maximum transition: left mirror – right mirror.

— Each AOI location shall be glanced 5 to 10 times. When using 3 AOIs this results in 30 trials (3 × 10). The iteration is an important aspect for statistical reasons because the redundancy increases the reliability.

  If in the calibration verification prior to the experiment the first 1-2 fixations already show deviations from the AOI locations, the experimenter could abort the process, recalibrate the system and restart the check of calibration. The calibration verification shall not be interrupted or repeated after the experiment.

— The experimenter scores (correct/incorrect) every trial if the fixation point displayed by the eye tracking system is at the AOI location.

— The percentage of correct scored trials ($P_{ct}$) is calculated by the formula below:

$$P_{ct} = \frac{N_{ct}}{N_{tt} \times 100}$$

where

$N_{ct}$ is the number of correct scored trials;

$N_{tt}$ is the total number of trials.

— If in at least 85 % of all cases (e.g. 26 of 30 cases = 87 %, when using 3 AOIs) the displayed fixation point is within the instructed AOI, the accuracy is precise enough. If not, the verification of calibration procedure shall be redone. If the verification of calibration is redone the product specific eye tracking procedures shall be checked.

— Recommendation: If an AOI or a specific location within that AOI exhibits low score the AOI or location should be reconsidered in regard to the research question of the experiment.

Potential markers and grids which are needed only for the verification of calibration shall be removed before the start of the experiment.

## B.3 Method — Glances onto reference grids in front of each AOI

The validity of the calibration also can be tested with a standardized reference grid like Figure B.1.



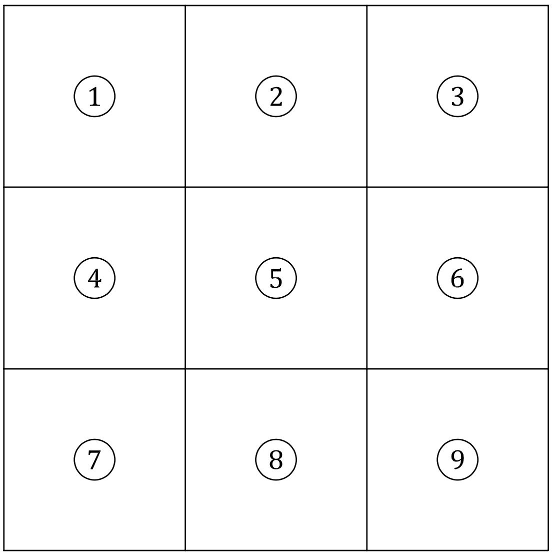



**Figure B.1 — The reference grid**

The grid should be printed out in a size of 15 cm × 15 cm. Each single square has a size of 5 cm × 5 cm (This figure should be printed 1:1 15 cm × 15 cm).

The reference grid is adapted to measuring eye tracking behaviour within the car or driving simulator. Due to the usual size of AOIs in an automotive context (displays and mirrors) and their usual distance to the participant (about one arm length) a minimum local resolution of 5 cm shall be revealed after the verification of calibration procedure.

The eye tracking systems shall have an accuracy of pupil tracking so that the accuracy of the fixation point is at least enough to discriminate all different target points set as AOI, and different points of its sub-task according the purpose of the study.

The target accuracy of the eye tracking system may be estimated by separation of distance in interest, distance of two fixated points and distance of an AOI to the participant by calculating the visual separation angle (alpha).

NOTE The accuracy depends on participant characteristics, calibration, gaze direction, illumination, the eye-tracking system and potential additional factors.

The visual separation angle ($A$) is calculated as follows:

$$A = 2 \times \tan - 1 \left( \frac{d_{fp} \times 0{,}5}{d_{AOI,p}} \right)$$

where

$A$ is the visual separation angle;

$d_{fp}$ is the distance of two fixed points;

$d_{AOI,p}$ is the distance of an AOI to the participant.

Table B.1 provides some example of these visual separation angle for reference.

For measuring eye gaze behaviour in practice, the accuracy of the fixation point is crucial. For example, an accuracy of 1 °means that two points with a distance of 5 cm to each other can be distinguished by the eye tracker in a distance of 286 cm.



**Table B.1 — The relation between the distance of two fixated points, observation distance and visual angle**

| Distance of two fixated points (cm) | Distance to participant (cm) | Visual angle (degree) |
|---|---|---|
| 5 | 50 | 5,7 |
| 5 | 75 | 3,8 |
| 5 | 100 | 2,9 |
| 5 | 125 | 2,3 |
| 5 | 150 | 1,9 |
| 5 | 175 | 1,6 |
| 5 | 200 | 1,4 |
| 5 | 225 | 1,3 |
| 5 | 250 | 1,1 |
| 5 | 275 | 1,0 |
| 5 | 300 | 0,96 |
| 5 | 325 | 0,88 |
| 5 | 350 | 0,82 |
| 5 | 375 | 0,76 |
| 5 | 400 | 0,72 |
| 5 | 425 | 0,67 |
| 5 | 450 | 0,64 |
| 5 | 475 | 0,60 |
| 5 | 500 | 0,57 |
| 5 | 525 | 0,55 |
| 5 | 550 | 0,52 |
| 5 | 575 | 0,50 |
| 5 | 600 | 0,48 |



For checking the accuracy of the calibration, the procedure is as follows.

— The reference grid is printed out in a size of 15 cm × 15 cm and placed in front of each AOI within the car or driving simulator which is relevant for analysing eye gaze behaviour (e.g. dashboard, displays, rear-view mirror, outside mirror).

— At the beginning of the measurement the participant is instructed to glance at the numbers 1 to 9 successively in climbing sequence. The experimenter checks if the fixation point displayed by the eye tracking system is within the square of the particular number. If in at least 85 % of all cases (e.g. 8 of 9 cases = 89 %; or in at least 31 of 36 cases when using 4 AOIs) the displayed fixation point is in the correct square the accuracy is precise enough. If not, the verification of calibration procedure shall be redone.

The most AOIs within a car are located in a distance nearer than 100 cm to the participant. Therefore, a grid space of 5 cm is appropriate. However, for large distances the 5-cm grid space can be too small for the accuracy of an eye tracker (e.g. the right outside mirror, objects on the screen of a driving simulator, or objects in the real traffic environment). Hence for viewing distances larger than 100 cm the calibration board may be used with an 8-cm grid space. This conforms to a 24 cm × 24 cm square which can be printed out onto a DIN A3 paper.

Independent of the used method for checking the validity of the calibration before data recording, the following points shall be respected.

— At the end of the measurement the verification of calibration is repeated to verify that each glance is still properly displayed/calculated. If it is not, it is a sign that there is a drift in calibration that occurred (e.g. by shifting down the eye tracking glasses during the session). At the end of the experiment the percentage of correctly displayed fixation points should be still >85 %. If it is <50 % the recording is not reliable and shall be discarded. If the used eye tracking system enables a post-hoc recalibration (which demands to save the video of the eye camera and not only X- and Y-coordinate values) discarding the recording can be avoided, if after this recalibration the accuracy is sufficient. If the percentage of correctly displayed fixation points is between 50 % and 85 % a recalibration is recommended.

— During longer experimental sessions it is recommended to repeat the verification of calibration at appropriate times (e.g. between different tasks or during breaks) to realize problems with low accuracy early and to intervene by recalibrating.

— The verification of calibration and their result have to be recorded to documenting that these checks were done, and the eye tracking accuracy was sufficient.



# Annex C
(normative)

# Eye tracker validation task (EVT)

## C.1 Introduction

The application of a given eye tracking system in a given experimental environment (e.g. experimental vehicle, driving simulator, lab) might lead to specific artefacts in tracking and glance segmentation. The EVT method is able to check for these artefacts and to uncover some systematic biases produced by an eye tracking system.

For example, one long gaze onto an AOI may be separated erroneously into two short glances or a gaze onto an AOI may be erroneously not detected by the eye tracking system. These are example questions of validity.

The EVT procedure is an independent study (performed by a minimum of 24 test participants) which can be done separately from the actual experimental study. The validation of the test setup procedure described in this annex shall be performed in combination of a given experimental environment and application purpose, and it shall be repeated if set up or test condition is modified.

The basic idea of the EVT is to instruct via a sound cue the target AOI the participant should glance at. For that, the same AOIs should be used which will also be used in the actual driving study. Since the order and points in time of all instructed AOIs are pre-defined by the experimenter, the glance behaviour measured by the eye tracking system can be compared with the instructed glance behaviour (target-performance comparison).

The EVT focuses on statistical values of three relevant parameters:

1) mean glance duration;

2) total glance time;

3) glance frequency.

These values are calculated by aggregating data over all trials and all participants of the sample. In C.6 the pass-fail-criteria are defined.

If the statistical value of one or more of the three parameters lies outside of the defined interval of confidence, then it means that the eye tracking system measures gaze behaviour systematically biased and its quality is not sufficient to measure participants' gaze behaviour in this driving environment.

The primary use cases of the EVT are:

1) testing the general quality of an eye tracking system, e.g. before deciding to use this system;

2) comparing two eye tracking systems for deciding which one to use for a driving study (benchmark test).

The EVT is not applicable to perform a verification of calibration before recording data or to judge the quality of a single recording of the actual study.



## C.2 Procedure

Three different sound cues are used for instructing the participant's viewing direction (e.g. road scene, central information display (CID), instruments). Every sound cue consists of: tone_1 (100 ms) + interval (100 ms) + tone_2 (100 ms).

— cue_1 (e.g. road): 1 200 Hz, 2 000 Hz;

— cue_2 (e.g. CID): 1 200 Hz, 1 200 Hz;

— cue_3 (e.g. instruments): 1 200 Hz, 500 Hz.

The advantage of sinusoidal tone is the clearly defined onset and the possibility to present the same short durations. In contrast, instructions via words would be more difficult for the predefinition of short glance durations, because words have longer and more variable durations. Sound cues with the discriminative stimulus (specifying the target location) being the second sound should be used. These sound combinations prevent a too early execution of the saccade which could occur when using a precued response, i.e. the discriminative stimulus being the first sound.

The sound cues shall be presented with a level of 55 dB (which corresponds to room volume). It should be presented via loudspeakers. Headphones can be used but shall not interfere with the eye tracking system equipment. The sound shall be recorded and synchronized with the eye tracking video.

The overall test procedure per participant is composed of 6 sequences and each sequence consists of 18 sound cues (see vertical column in Table C.1, where a set of AOI and interval times for each step of a sequence is shown in vertical order). The time intervals between each sound combination shall be randomly distributed in a balanced order to avoid predictive behaviour by the participant which may affect the test results.

Each sequence may be started by a long “start” tone (1 200 Hz for 300 ms) to prepare the participant for the start of a sequence, followed by the AOI specifying sound cue with a random inter stimulus interval of 1,5 s, 2,0 s or 2,5 s. In the beginning of each sequence, 5 cues are presented to familiarize the participant with the task (dummy trials). Then, 12 pairs of tones are presented (each pair four times in a balanced order). Only these 12 trials should be included into the analysis. Also, the last trial serves as a dummy trial, because the resulting last glance differs from the others due to the fact that it is not followed by a glance to another AOI. The sequence ends with a long “end” tone. Before the analysis, data from the first 5 dummy trials and the last dummy trial shall be deleted.

See Table C.1 as an example for the detailed order of cues (i = instruments, r = road, c = central information display) and time intervals (in seconds). The intervals used here vary between 1,5 s, 2 s and 2,5 s according to the critical glance durations during driving. The criterion of 2 s as maximum single glance duration was based on a distribution of single glance durations described by Rockwell (1988)[11].

Naturally, time intervals can be adapted according to the purpose of the eye tracking recording. If the time intervals are adapted the intervals should be same as the time to be detected in the study. For driver distraction testing use these time intervals. Other time intervals may be chosen in relation to the research question and glance durations to be measured. For example, in automated driving studies with prolonged glance durations to the CID or to other devices longer time intervals can be presented.

**Table C.1 — Example for the order of sound cues and time intervals**

| | Cue sequences | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Sequence 1 | | Sequence 2 | | Sequence 3 | | Sequence 4 | | Sequence 5 | | Sequence 6 | |
| | AOI | sec | AOI | sec | AOI | sec | AOI | sec | AOI | sec | AOI | sec |
| dummy trials | start | 2 | start | 2 | start | 2 | start | 2 | start | 2 | start | 2 |
| | i | 2,5 | c | 2,5 | r | 2,5 | c | 2,5 | i | 2,5 | r | 2,5 |
| | r | 2 | i | 2 | c | 2 | r | 2 | c | 2 | i | 2 |
| | i | 1,5 | r | 1,5 | i | 1,5 | c | 1,5 | r | 1,5 | c | 1,5 |
| | c | 2,5 | i | 2 | r | 1,5 | i | 2 | c | 2,5 | r | 1,5 |
| | r | 1,5 | c | 1,5 | i | 2 | r | 1,5 | i | 2 | c | 2,5 |
| 1 | c | 2 | r | 2,5 | c | 2,5 | i | 2,5 | r | 1,5 | i | 2 |
| 2 | i | 1,5 | c | 2 | r | 2 | c | 2 | i | 2 | r | 2 |
| 3 | r | 2 | i | 2 | c | 2 | r | 2 | c | 2 | i | 1,5 |
| 4 | i | 2 | r | 2 | i | 2 | c | 2,5 | r | 2 | c | 1,5 |
| 5 | c | 1,5 | i | 1,5 | r | 2,5 | i | 2 | c | 1,5 | r | 1,5 |
| 6 | r | 2,5 | c | 2,5 | i | 2,5 | r | 1,5 | i | 2,5 | c | 2 |
| 7 | c | 2 | r | 1,5 | c | 1,5 | i | 1,5 | r | 2,5 | i | 2,5 |
| 8 | i | 2,5 | c | 1,5 | r | 2 | c | 2 | i | 2 | r | 2 |
| 9 | r | 1,5 | i | 2 | c | 2 | r | 2,5 | c | 2,5 | i | 2 |
| 10 | i | 2 | r | 2 | i | 1,5 | c | 1,5 | r | 2 | c | 2,5 |
| 11 | c | 2,5 | i | 2,5 | r | 1,5 | i | 2 | c | 2 | r | 2,5 |
| 12 | r | 2 | c | 2 | i | 2 | r | 2 | i | 1,5 | c | 2 |
| dummy trials | c | 2 | r | 2 | c | 2 | i | 2 | r | 2 | i | 2 |
| | end | | end | | end | | end | | end | | end | |

In this example the mean values within a sequence and mean values over all sequences are 2 s per single glance duration, add up to 8 s total glance duration and have a frequency of 4 glances for each AOI.

Each sequence should be tested only once. During testing the examiner should check if participants look at the correct AOI (e.g. by using a video camera which films the face of the participant). If the participant makes a mistake, the sequence should be stopped and the recording of the eye tracking data for this sequence should be repeated. Note, if an error occurs during the first four sound presentations, the sequence shall not be aborted, because these trials are excluded from analysis.

The whole test setup, including lighting conditions, should be chosen according to the later test setup the eye tracker is normally used for (e.g. a driving simulator test setup, if the eye tracker is used for the investigation of eye movements during driving in a driving simulator).

Equivalent, AOIs as well as AOI size should be chosen according to the AOIs normally examined by eye tracking data (e.g. road scene, central information display and instruments, if normally the eye tracker is used for the investigation of eye movements away from the road scene ahead during driver`s interaction with a non-driving related task).

AOI definition shall be standardized for the test, because the size of AOI will have an influence on the results. Additionally, the position of the fixation point within the AOI will have an influence. Hence, the fixation point should be located in the middle of the AOI.

AOI size should be transferred into visual angle ($v_a$)according to the formula:

$$v_a = 2 \times \arctan\left(\frac{w}{2 \times d}\right)$$

where

*w* is the width of the AOI, in mm;

*d* is the distance between eye and AOI, in mm.

This is done in order to generate AOIs of approximately the same size as on the retina regardless of the viewer distance.

For example, for an AOI of 7 ° × 7 ° visual angle:

— instruments: 8 × 8 cm square, distance: approximately 65 cm;

— CID: 10 × 10 cm square, distance: approximately 80 cm;

— road: 36 × 36 cm square, distance: approximately 300 cm.

The three fixation points as physical markers should be crosses within a circle of 1,5 cm diameter for the 8 × 8 cm square. Fixations points of the other AOIs should be adapted according to the formula to a visual angle of 1,3°. See Table C.2 for possible combinations of distances and sizes and Figure C.1 for picture of an AOI. AOIs with fixation points can be printed on paper in order to define it precisely later in the video.

**Table C.2 — The relation between observation, AOI size and size of the fixation cross**

| Distance to participant (cm) | Size AOI size (cm) for visual angle (degree) = 7 | size circle/cross (cm) for visual angle (degree) = 1,3 |
|---|---|---|
| 50 | 6,1 | 1,1 |
| 55 | 6,7 | 1,2 |
| 60 | 7,3 | 1,4 |
| 65 | 8,0 | 1,5 |
| 70 | 8,6 | 1,6 |
| 75 | 9,2 | 1,7 |
| 80 | 9,8 | 1,8 |
| 85 | 10,4 | 1,9 |
| 90 | 11,0 | 2,0 |
| 95 | 11,6 | 2,2 |
| 100 | 12,2 | 2,3 |
| 110 | 13,5 | 2,5 |
| 120 | 14,7 | 2,7 |
| 130 | 15,9 | 2,9 |
| 140 | 17,1 | 3,2 |
| 150 | 18,3 | 3,4 |
| 160 | 19,6 | 3,6 |
| 170 | 20,8 | 3,9 |
| 180 | 22,0 | 4,1 |
| 190 | 23,2 | 4,3 |
| 200 | 24,5 | 45 |
| 210 | 25,7 | 4,8 |
| 220 | 26,9 | 5,0 |
| 230 | 28,1 | 5,2 |
| 240 | 29,4 | 5,4 |

**Table C.2** *(continued)*

| Distance to participant (cm) | Size AOI size (cm) for visual angle (degree) = 7 | size circle/cross (cm) for visual angle (degree) = 1,3 |
|---|---|---|
| 250 | 30,6 | 5,7 |
| 260 | 34,8 | 5,9 |
| 270 | 33,0 | 6,1 |
| 280 | 34,3 | 6,4 |
| 290 | 35,5 | 6,6 |
| 300 | 36,7 | 6,8 |

Dimensions in centimetres

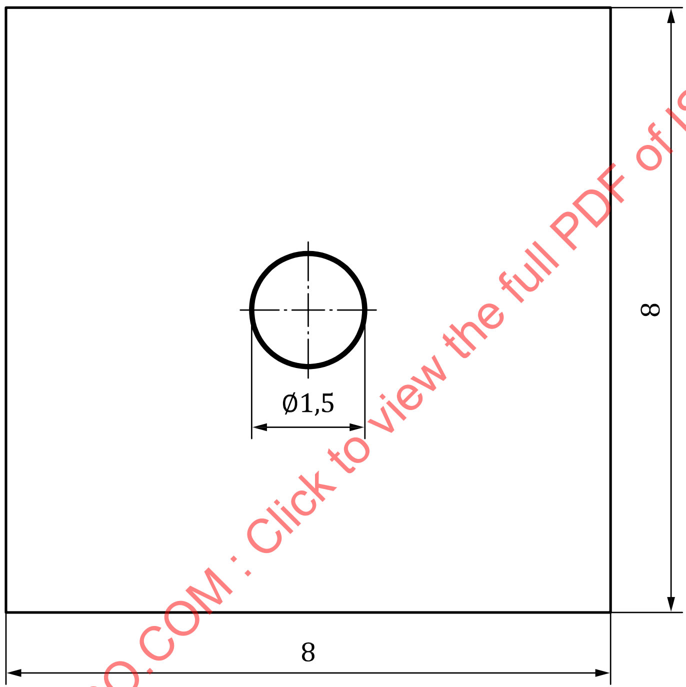



**Figure C.1 — AOI and fixation cross (This figure should be printed 1:1 8 × 8 cm)**



Additionally, if available a scene video including a fixation overlay can be manually coded following the procedures described in Annex C. Doing this shall take into account that this material includes potential artefacts of the eye tracking software already.

## C.3 Participants for EVT

The EVT should be performed by a minimum of 24 test participants [$n$ = 24] if the EVT is used for the qualification of an eye tracking system within a specific experimental environment.

For preparation of participants see 4.3.

## C.4 Participant instruction for EVT

The following verbal/written instructions for EVT should be given to the participant:

"In the following experiment you will be asked to look at specific locations based on three different sound cues. Each cue will consist of two tones, one immediately following the other. The first tone in the pair will always be the same. The second tone will either be the same, higher, or lower in pitch as the first. This second tone will indicate which target you should look at. There will be a series of 18 pairs of tones for each trial, with a total of six trials. Please continue to look at each target until you hear the next sound cue. Before we start with the experiment, I will show you which sound represents which target and give you as much time to practice as you might need. If you make a mistake, we will redo that trial, so please try to be as accurate as you can."

## C.5 Practice trials

The glance sequences should be practiced until the participants are familiar with the task and one sequence can be performed without an error.

## C.6 Test metrics

After exclusion of the first five and last dummy trial of each sequence the mean values of single glance duration, total glance time and glance frequency over all sequences for each AOI with automatic reduction should be computed and compared with the chosen values.

In our example, following the criterion of 2 s for single glance durations derived from the AAM guidelines for driver distractions, the mean values should be:

— mean glance duration (MGD) [s]: 2;

— total glance time (TGT) [s]: 8;

— glance frequency (GF): 4.

The deviation of measurement shall not exceed one or more of the three criterions: ±15 % for MGD or ±1,5 s TGT or ±1 glance for GF. If the combination of the given eye tracking system and experimental environment does not meet one or more of these three criteria, it is not suitable to use it for a study on eye glance behaviour. The measurement configuration shall be improved (e.g. lighting conditions, data availability, AOI size, data segmentation) and the EVT shall be repeated.

Besides the mean values also the standard deviations of MGD, TGT and GF should be considered and be as low as possible and should be documented.

The experimenter may fix a more stringent criterion and should document this.

In addition to the analysis of above values (MGD, TGT, GF), also single glance durations can be compared with the pre-defined time intervals of the sequence in detail. For example, data acuity can be tested also while comparing the 1,5 S, 2 s, 2,5 s intervals with recorded single glance durations.

# Annex D
(normative)

# 5 % data verification — Check of reliability after data recording and before statistical data analysis

To analysing eye gaze behaviour an eye tracker is not always used, but one or more video cameras filming the driver's face (= manual reduction) may be used as well. Glances are in this case analysed by a rater analysing the video. The rater judges based on the head movements and eye movements what AOIs are gazed at each point of time and codes the eye gaze behaviour directly using as many categories as AOIs exist. Since this method is principally less precise and less objective than using an eye tracking system, before analysing the data statistically it shall be proved that the reliability is sufficiently high.

To verify the quality of the data produced by an eye tracking system or the coding of a manual rater it is recommended to manually code a subsample of 5 % of the whole data set in accordance with Annex A and compare the result with the data produced by automatic reduction or manual coding by the original rater.

An interrater agreement using Cohen's kappa shall be calculated. Cohen's kappa is the most used method for calculating the agreement between different raters with categorial data. Here a Cohen's kappa is always calculated for the raters' consensus regarding the gaze behaviour of each analysed participant and shall not be lower than 0,70.

How to select the data:

— Procedure A: random sample of 5 % over 100 % of the data set;

— Procedure B: select a random of 50 % of the participants and analyse 10 % of their data. It is recommended to balance this subset of data equally distributed over the experimental conditions and tasks.

Annex E explains how to calculate Kappa. Tabulate agree and disagree glances between the original coding and the check coding synchronized using the experimental time stamp.

If Kappa has a value of 0,7 higher, the analysis of the study should be done by one of the two raters. If Kappa is less than 0,7, the whole dataset should be analysed by 2 raters following the instructions in (see Annex A).



# Annex E
(normative)

# Calculating a Cohen's kappa for one participant

1) The video of the participant is segmented into time intervals. These intervals should be chosen as small as possible. The smallest unit is a frame in the video.

2) In each time interval the rater judges independently from each other what AOI the participant is looking at and code the glanced AOI with a number, e.g. 1 = road scene ahead, 2 = rear mirror, 3 = display. The AOIs are the categories to which the gazes are assigned by the raters.

3) By coding glances on AOIs in this way there results a data matrix consisting of 2 columns (for rater A and B) and as many rows as there are time intervals. The cells contain the gazed AOIs coded by each rater within each time intervals (numbers 1, 2 or 3).

4) A Cohen's kappa is calculated. It quantifies the consensus between both raters for the data set of this participant. The agreements and disagreements between both raters can for this participant also be summarized in a frequency table (see Figure E.1).

5) The same procedure is done with the data sets of the next participants until the rater agreement regarding 10 % of all participants are checked in this way.

| | | Rater A | | | |
|---|---|---|---|---|---|
| | | Gazes on road scene | Gazes on rear mirror | Gazes on display | Total |
| Rater B | Gazes on road scene | 530 | 50 | 20 | 600 |
| | Gazes on rear mirror | 110 | 140 | 50 | 300 |
| | Gazes on display | 10 | 60 | 30 | 100 |
| | Total | 650 | 250 | 100 | 1 000 |

NOTE Altogether the gazes onto three AOIs within 1 000 time intervals were judged by the raters.

**Figure E.1 — A frequency table showing the agreements (dark grey cells) and disagreements (light grey cells) between rater A and rater B in the data set of one participant**

NOTE 1 A simple calculation of the percentage of agreement (= sum of cases in which both raters came to the same judgment divided the number of all cases) is not a suitable measurement, because the agreement by chance is not respected. However, Cohen's kappa takes into account the agreement occurring by chance.

Cohen's kappa is calculated as follows (see Reference [2]):

kappa = (relative observed agreement among raters − probability of chance agreement)/(1 − probability of chance agreement) relative observed agreement among raters = sum of agreements/number of observations

probability of chance agreement = ((rater A total number of gazes to AOI 1 + rater B total number of gazes to AOI 1/number of observations) + (rater A total number of gazes to AOI 2 + rater B total number of gazes to AOI 2/number of observations))/number of observations

**Dealing with missing values:** When recording gaze data, it may happen that there are sequences where data are not available (e.g. for technical reasons) or where a rater is not able to judge which AOI was glanced. Therefore, an additional coding is necessary to catch also these time intervals and to handle them as missing data. This event is added as an additional category to the categories of AOIs (and coded with an own number, e.g. "99"). Thus, if rater A judges that a participant is glancing, e.g. rear mirror, whereas rater B in the same time interval cannot deliver a judgment then this is a kind of disagreement which is considered by calculating Cohen's kappa.

To prevent the statistical artefact of an overestimated rater agreement caused by a high percentage of cases where both raters agree in their decision that they cannot judge where the participant is looking at, missing data should be coded with different numbers. For example, rater A always codes missing data with "99" and rater B always with "98".

**Minimum criterions for interrater agreement**

— A subsample of 5 % of the entire data set shall be checked according to interrater agreement using Cohen's kappa.

— To prevent selection biases by systematically analysing participants whose eye gaze behaviour is easier to analyse for some reasons, always those 5 % of data sets shall be analysed which were recorded first. For example, in a study with 100 participants for the first 5 participants the interrater agreement shall be calculated.

— If within this analysed subsample all Cohen's kappas have a value of 0,70 or higher, the remaining rest of the study may be analysed by only one of these both raters. However, if one or more Cohen's kappas are lower than 0,70, all data sets of the study shall be analysed by two raters, a Cohen's kappa shall be calculated for each data set, and all data sets with a Cohen's kappa of less than 0,70 shall be discarded.

NOTE 2 In this context the significance value which is a standard output of most statistic software programs does not matter because significance depends also from the number of cases. Since here the *N* is the number of time intervals which can be chosen very small (till down to the level of frames in the video) the *N* can be inflated artificially and is extremely large in most cases. However, the crucial point is only the Cohens's kappa itself as a measurement of effect size which quantifies the amount of agreement between two raters.

