



## **Report on extra seminar on evaluating combinations of functions or systems**

Friday 6 February 2009, Schiphol

All materials are to be found on: <http://www.fot-net.eu/>

Agenda .....	2
Participants.....	3
Work in small groups .....	3
Conclusions.....	4
Appendix A Combinations of systems/functions, hypotheses and performance indicators	5
Appendix B: Conclusions about generating hypotheses for combined functions/systems.	9
Appendix B: Conclusions about baselines.....	10
Appendix D Solutions to problems presented by euroFOT and teleFOT.....	12
Appendix E Recommendations for a methodology of evaluating combinations of systems/functions .....	16
Appendix F Glossary FOT-Net seminar .....	19

## Agenda

MORNING SESSION	10:00 - 10:30	Introduction	<i>Oliver Carsten - FOT-Net</i>
	10:30 - 11:00	In small groups: selection of functions or systems and the expected interaction between them	
	11:00 - 11:30	In small groups: formulation of research questions and hypotheses for a combination of functions or systems	
	11:30 - 12:00	In small groups: define performance indicators	
	12:00 - 12:30	Brainstorm on determining base-lines	
12.30 - 13.15 Lunch			
AFTERNOON SESSION	13:15 - 13:30	Conclusions and summary of morning discussions	<i>Yvonne Barnard - FOT-Net</i>
	13:30 - 14:00	Practical constraints in evaluating combinations of systems and functions	<i>Samantha Jamson - EUROFOT</i> <i>Stig Franzén - TELEFOT</i>
	14:00 - 14:45	In small groups: brainstorm on solutions for practical problems	
	14:45 - 15:15	Presentations from small groups and plenary discussion	
	15:15 - 16:00	In small groups: formulation of recommendations	
	16:00 - 16:30	Plenary discussion and conclusions	<i>Oliver Carsten - FOT-Net</i>

## Participants

Most of the 22 participants were involved in a FOT (euroFOT, SeMiFOT, TSS-FOT, Telefot, AOS, Regional FOT InCar (V21), and Naturalistic Driving Studies). Most of the participants came from research institutes, but also from car manufacturers and public authorities.

## Work in small groups

In small groups of three participants (the composition changing with each exercise) the following exercises were made:

1. Selection of functions or systems and the expected interaction between them  
Select 3 functions or systems that are combined in a FOT. List 3-5 possible interactions between the functions/systems when they are installed in a car and used by the driver.
2. Formulation of research questions and hypotheses for a combination of functions or systems  
Look at the list with research questions:
  - Is it positive for the environment?
  - Is it positive for safety?
  - Is it negative for efficiency?
  - Do drivers accept it?
  - What is the market potential?
  - What are the future costs and benefits?Individually: select a research question for 1 system and formulate 3 hypotheses per research question  
In your group: Discuss the hypotheses and the interactions between them; if possible, formulate common research questions and hypotheses  
Individually: write 1-3 notes with conclusions, stick them to the large sheet
3. Define performance indicators  
Take combination of systems/functions plus research questions and hypotheses from group 11 and generate as many performance indicators as possible
4. On the flip-charts write pros and cons of different baseline solutions:
  - Control group with all systems off
  - Several control groups with permutations of systems on and off
  - No baseline
5. Formulate solutions to the problems presented  
**Experimental Design Questions**  
**Baseline**  
What is an appropriate baseline when dealing with multiple systems and functions?
  1. Are we interested in comparing a package of systems to driving with no package?
    - Does that mean the baseline should be no package?

2. Are we interested in the influence of one part of the package?
  - Does that mean the baseline should be the package without System X?
3. What happens if different systems are targeted at different driving environments?
  - Do we then need multiple baselines?

#### **Exposure**

How do we measure and control for exposure to a system or combination of systems?

Is it in:

1. Time elapsed (calendar months)
2. Distance driven
3. Number of alerts or warnings from a warning system

What happens if a driver never activates a system?

What happens if a driver never receives alerts from a system?

- Do we need to carry on until he/she has “experienced” the system?

#### **Counterbalancing or not**

How do we take care of:

1. Learning effects
2. Seasonal effects

6. Formulate recommendations

**In the appendices the reports from the small groups may be found.**

## **Conclusions**

The day was interesting and highly interactive. Work in the small groups of three participants, in changing combinations, was intensive. Participants rated the sessions as good and useful. Many of them want to apply the results in their own FOT.

Although the topic of evaluating combined systems/functions is a complex one, we succeeded in coming up with ideas for approaches and recommendations to address the challenges. Although, as one participant noted on his or her evaluation form: “One always wants more answers”. The outcomes of this day will be reflected later on in the FOT-Net project in the form of an addition to the FESTA methodology.

# Appendix A Combinations of systems/functions, hypotheses and performance indicators

## Group 1

The following combinations of systems/functions were discussed:

Speed alert/limitation	Navigation	Green driving support
ACC	Forward Collision Warning (FCW)	Emergency braking
Night vision	Pedestrian warning	Driver drowsiness warning
ACC + (FCW)	Lane departure warning (LDW)	Blind spot warning

## Hypotheses

1. system will reduce the number of risky lane changes
2. the system will increase (Lane departure warning) or decrease (blind spot warning) use of turn indicator

Performance indicators:

- Number of lane changes
  - Indicator usage
  - Amount of FCW activation
  - Presence of vehicle in blind spot, number of blind spot activations
  - Distance to preceding vehicle/time headway
  - Number of glances to mirrors
  - Speed
  - Acceleration
  - Steering angle
  - Lane marking distance (distance to centre line)
  - Number of lanes
  - Number of kilometres since first use of system
  - Relative deceleration to preceding vehicle
- 

## Group 2

The following combinations of systems/functions were discussed:

FCW, ACC, LDW, BLIS

Interactions:

- Multiple warnings
- Systems reinforcing

## Hypotheses

Safety

1. ACC will increase the number of lane changes
2. ACC will increase time spent on secondary task
3. ACC will decrease time headway to lead vehicle

4. LDW will increase use of indicators
5. LDW will increase time in centre of lane

Acceptance

6. Acceptance of FCW will increase with decreasing FA rate
7. Acceptance of FCA will increase with intuitive warnings
8. Acceptance of BLIS will be higher on motorways than urban roads

Combined hypotheses

9. Combined ACC & LDW will increase driving in centre of lane compared to LDW only

Performance indicators:

Lateral position  
 Time headway  
 Video recordings of driver  
 Lane changes  
 Use of turning indicator  
 Extra sensors to measure system performance

**Group 3**

The following combinations of systems/functions were discussed:

FCW    Speed alert    Traffic information

Interaction:	FCW	<->	speed alert
Timeline:	short		short
Level:	manoeuvre		manoeuvre
Info channel:	visual, audio, haptic		visual, audio, haptic

Conflict between visual input from outside the vehicle and the warning

With the same reaction required: no conflict

If 2 signals disturb one another: confusion

Interaction:	speed alert	<->	traffic info
Timeline:	short		long
Level:	manoeuvre		planning
Info channel:	?		visual, audio

Same device, information priority: problems

Interaction:	FCW	<->	traffic info
Timeline:	short		long
Level:	manoeuvre		planning
Info channel:	?		visual, audio

**Hypotheses:**

Access to both FCW and traffic info will:

1. Reduce the effect of traffic info by 10%

Performance indicators:

- headway distance
- speed

2. Reduce the number of accidents on motorways and main roads by 50%

- Performance indicators:
    - headway distance
    - speed
    - number of accidents
  - 3. Reduce the number of cars involved in chain rear-end accidents by x%
    - Performance indicators:
      - number of cars involved (defined roads and time period)
      - Increase the driver's workload
  - 4. Increase the driver's reaction time by 10%
    - Performance indicators:
      - reaction time
  - 5. Acceleration/deceleration pattern combination of number of messages received plus number of warnings plus reaction times (braking).
- 

#### Group 4

The following combinations of systems/functions were discussed:  
 Navigation    LDW    FCW

#### Hypotheses:

1. A navigation system will make the driving task easier (if the other systems are not producing warnings)
    - Performance indicators:
      - Workload?
      - Difficult to measure "easy"ness
      - Speed (kms)

Easy has to be defined in order to develop performance indicators  
 Not a good hypothesis!!
  2. For a specific trip, in the case of congestion, the system reduces trip time
    - Performance indicators:
      - Travel time (minutes)
      - Average speed (kms)
      - Traffic situation (normal, congested etc.)
  3. The percentage of kilometres driven on motorways increases
    - Performance indicators:
      - Kms driven
      - Average speed (kms/h)
- 

#### Group 5

The following combinations of systems/functions were discussed:  
 LDW    ACC

**Hypotheses:**

1. LDW increases the use of turning indicators
2. ACC decreases the number of lane changes
3. LDW + ACC result in more time in the middle of the lane

## Performance indicators:

- Turn indicator on/off
  - Number of lane changes
  - Lane position
- 

**Group 6**

The following combinations of systems/functions were discussed:

Navigation      lane departure warning      forward collision warning

**Hypotheses:**

1. The navigation system makes driving task easier (reduced workload) only if the other systems do not provide warnings  
Performance indicators:
  - subjective workload, use of NASA TLX or RSME
  - objective measures : eye glance data, HFS manoeuvres
  - system warning status and on/off status
  - jerk, maximum braking values
2. For a specific driver (daily commute) the system reduces trip time (measured in minutes)
3. The percentage/number of kilometres driven on motorways by drivers previously unfamiliar with motorways increases  
Performance indicators:
  - Number of kilometres driven / road type
  - Percentage of road type
  - Measure of familiarity with road
4. On familiar roads the driver will try to “beat” the navigation system by changing lanes plus close following, therefore triggering alarms

## **Appendix B: Conclusions about generating hypotheses for combined functions/systems**

Defining hypotheses for combined functions/systems is difficult, complex and time-consuming

Hypotheses formulation requires:

- A clear idea of research questions
- System functioning

Probably it should be advisable to follow a 2 step approach:

- 1) identify/list hypotheses per function (isolated)
- 2) and then move to the combination of functions

There may be unexpected interactions between systems despite entirely different functionalities and purposes. Different functions may “hide” each other’s impacts or emphasize them.

Interactions can be divided into:

- user-related interactions
- technical interactions

More hypotheses may emerge when looking at the combination

Categorizing research questions helps to indicate interactions. Research questions should be broad enough to allow the development of both positive and negative hypotheses.

Research questions depend on the viewpoint of the one who asks them.

Hypotheses about combined effects influence the experimental design. Comparison situations may increase/change when more than 1 system is available. Use of long term support may influence use expectations. Long term effects seem to be important, but difficult to specify (and to measure).

## **Appendix B: Conclusions about baselines**

### **No baseline**

#### **The advantages of having no baseline are:**

- It is cheaper
- The FOT can be performed quicker
- The resources of the FOT can be focused on collecting data on usage and behaviour with the system
- It is easier: to implement, to gather and analyse data
- It avoids ethical problems, taking a system away from a driver (after they have paid for it, but also after they have used it)
- Testing user acceptance is easy without baseline, you can compare user acceptance between subjects
- No baseline is needed if you study evolution in time or learning effects
- Instead of a baseline you can use a pre-existing group or bring in baselines from other projects/experiments. A baseline could be extracted from a naturalistic driving study without any functions.

#### **The disadvantages of having no baseline are:**

- It is impossible to assess the effect of systems on driving behaviour, or other impacts such as safety, throughput and acceptance. You will not have a common base to compare your results with.
- You cannot do a cost-benefits analysis
- Trust in results is reduced
- Your study will have poor standing in the research community, unlikely to get results published/disseminated

### **Control group with all systems off**

#### **The advantages of having a control group with all systems off are:**

- Good and easy basis for comparison between groups
- Robust results
- Possibility to realise tests in different contexts
- Tests with different target groups
- Reuse of data
- Easy comparison with and without systems
- Assessment of “normal” driving conditions
- Indications on the effect of introduction of systems
- Shorter duration of FOT when using parallel groups

#### **The disadvantages of having a control group with all systems off are:**

- Large sample needed for the whole FOT
- Expensive
- No comparison is possible of driving behaviour of the same driver (with and without system)

- Matching control and experimental groups is difficult. How to determine what makes up the key characteristics?
- Hard to make sure that the control group is representative
- Difficult to perform similar tests, you need the same context availability (weather, traffic, drivers)
- May be redundant, these results may come also from literature (the “normal” driving)
- You cannot be sure that systems are really off
- Can all systems be easily turned off? (and made to stay off)
- Hard to avoid loss of participants due to lack of interest and motivation

### Several control groups with permutations of systems on and off

The advantages of having several control groups with permutations of systems on and off are:

- More information
- Assesses the different combinations of systems
- You can separate effects of different systems
- Different interactions recorded if systems are on and off
- Identifies (in)compatibilities between systems
- Controlled test design
- Answers more questions, if there is enough data, it is possible to test many (if not all) hypotheses
- Enhances trust in results
- Lots of different analyses are possible

The disadvantages of having several control groups with permutations of systems on and off are:

- Complex design, implementation and data analysis
- Expensive, more resources needed
- Large number of participants needed
- Longer FOT
- Hard to get enough data to do statistical analyses of all the permutations
- A bundle of functions often comes with an integrated HMI which cannot be separated  
-> switching one system off independently is difficult
- How to control for “after effects” of using a system, then turning it off?
- How to control for sequence of exposure, e.g. ACC-LDW versus LDW-ACC?
- Results may be conflicting or misleading
- You may encounter some learning effects which influence the results

Some alternatives:

- Experimental groups: groups with the different combinations compared with one group that has the possibility to put systems on and off
- One group of subjects with the permutations of systems off and on over time

## Appendix D Solutions to problems presented by euroFOT and teleFOT

### Group 1

#### Baseline

What is an appropriate baseline when dealing with multiple systems and functions?

Question 1: Are we interested in comparing a package of systems to driving with no package?  
Does that mean the baseline should be no package?

Answer: Yes, there should be a “no-package” condition, however, it is considered that a most appropriate scheme would be the following:

baselines:  
no package  
S1  
S2  
S3  
-----> S1+S2+S3

Question 2: Are we interested in the influence of one part of the package?

Does that mean the baseline should be the package without System X?

Answer: Yes if there are relevant hypotheses defined. In principle, the influence of a part of the package (i.e. one system) can be assessed by the above scheme. However, if hypotheses require the assessment of partial combinations relevant baselines may be formed.

### Group 2

#### Baseline

What is an appropriate baseline when dealing with multiple systems and functions?

- 1 Individual drivers. New truck or retrofitting.
- 2 Think about the factors of importance.
- 3 Can you match events?

#### Counterbalancing or not:

Fleets determine what is there (huge variety of vehicles)  
Practical problems, even to find 50 trucks  
Matching scenarios events (with ACC, without ACC is same conditions?)

---

### Group 3

How do we measure and control for exposure to a system or combination of systems?

- time elapsed
- distance driven

- number of alerts (number of encounters)
- number of trips must be added

The choice depends on the function! Self-triggered events differ from event-driven events.

speed alerts	fairly frequent	distance driven?
navigation support	fairly infrequent	number of trips?
parking support	fairly frequent	distance driven
traffic info	frequent	time

frequency depends on the particular type of usage  
the choice of users must make sense

You have to choose the solution which provides you with data in the “least frequent exposure”

#### Group 4

**How do we measure and control for exposure to a system or combination of systems?**

Exposure to situation/danger:

It depends on the system

BLIS: all the time, e-Call: seldom

#### Exposure measurement

Number of alerts/warnings per distance driven

+ access to system among the whole population

+ access to system among people who were informed

Exposure to the situation -> environment, condition (night..)

Exposure to combined systems -> Number of warnings during short time/distance driven

#### Different outcomes:

1. Alert never activated
  - > bad design
  - > function not useful
  - > bad information -> subjective study

(What are you testing? Test technical solution? Market acceptance? User acceptance?)

2. Never received alerts
  - > Depends on exposure to danger/situation
  - > Good lesson for usefulness of system
  - > Not answering a problem
  - > Normal-careful driver

#### Group 5

#### Counterbalancing or not

How do we take care of:

- Learning effects
- Seasonal effects

## Seasonal effects

1. Be aware when the baseline trial was conducted. Was it during summer/winter etc. Preferably experimental period should be conducted during same time-frame and the drivers should be matched accordingly.
2. Define what the seasonal effects actually are and their effects on the systems being tested. Different conditions in summer vs winter:
  - Ice + snow
  - Rain
  - Fog
  - heat
3. Be aware that the results may be affected by seasonal differences and be prepared to introduce caveats to the results.
4. If you have the freedom to select a control group running parallel to the test or as near to in time
5. Introduce criteria for the conditions that are selected for the analysis, for example filter out extreme weather conditions (e.g. snowy day).
6. Try to run the experiments over a representative time period when the seasonal effects will be less influential on the overall results
7. Consider also differences in different countries, winter in Spain, Portugal are not like winters in Northern Finland

## Learning effects

1. Be aware the individual effect of learning differs and is difficult to measure
2. Design experimental method so that learning effects are removed as far as possible from the trials
3. However, learning effects are difficult to measure and learning abilities differ from person to person and this may affect the outcomes despite the best intentions
4. If a homogenous group is needed, a certain level of competence may be required and could be tested using a sample trial
5. It could be that learning effects are an independent variable in their own right, for example, we might want to look at how quickly drivers adapt and learn to use a system in which case a group of completely naive system users may be required in order to look at HMI issues and how quickly competency is achieved. So here you are trying to actually evaluate learning effects
6. Finally, need to breakdown learning effects + driving styles according to different types of drivers + driving styles, and evaluate learning effects using appropriate performance indicators.

## Group 6

### Counterbalancing or not

How do we take care of:

- Learning effects
- Seasonal effects

Learning effects;

- Mobile phone use, very common today, assume that all have experience
- Function on mobile phone, can be experienced before “treatment” but by other channel, e.g. on a radio channel for traffic info
- Age, work experience, education etc.
- In the Telefoot case, when will a steady state be reached?  
Some functions (use), speed alert (guide)  
Others (vary)

Seasonal effect:

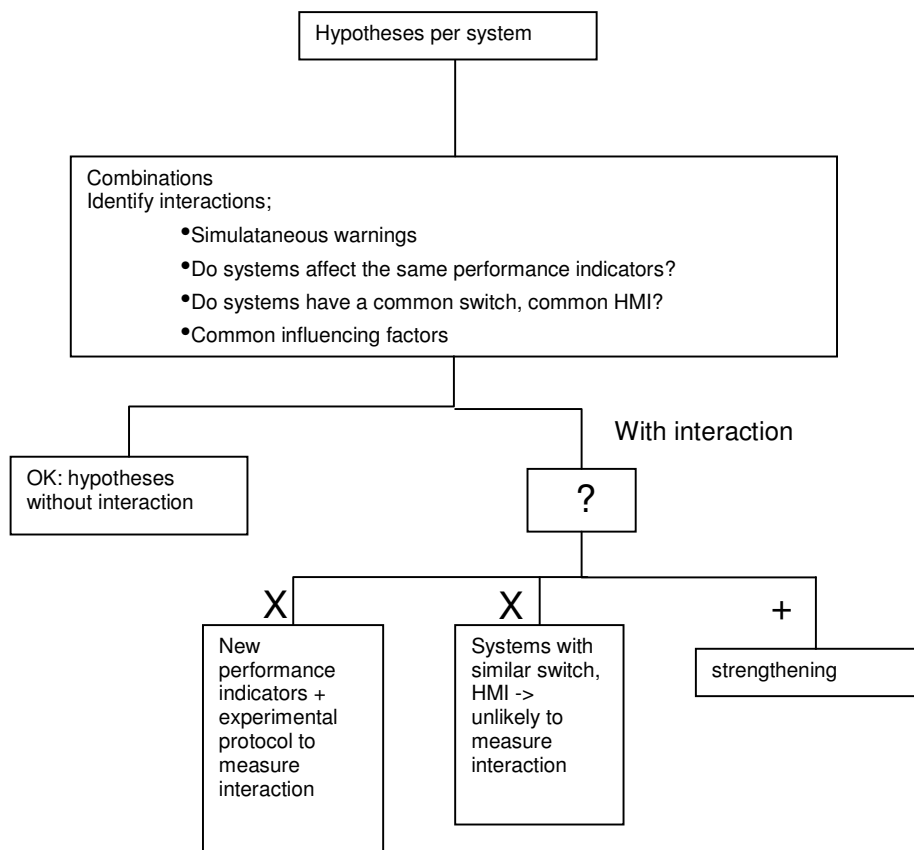
Winter/summer, north/south (weather, light etc)  
Schools, holidays, ...

# Appendix E Recommendations for a methodology of evaluating combinations of systems/functions

## Recommendations 1

When deciding what tests to do in a FOT, what experimental set-up to follow and what hypotheses to test, it is necessary to find the right balance between:

- Costs and price
- Science and practicality
- Research questions and hypotheses and combinations
- Prioritizing of research questions and hypotheses
- Detailed analysis of hypotheses, events and performance indicators
- Combine the hypotheses of the single systems to generate those for the bundle of applications
- Test development of a first “rough” experimental design; iterative process in order to improve the rough design during the next steps (loops) -> FOT-loop



## Recommendations 2

- Provide a structure for dealing with the problem...
- Top-down and bottom-up approach
- Top-down; this applies to FOTs in general just as much as combination of functions
- Bottom-up: these are the functions etc.

Top down:

desired study designs

Bottom-up:

cost benefit approach

*<iterate in a spiral  
between the 2 approaches>*

Interferences -> scenarios

Interactions -> define hypotheses for individual functions

Interactions -> scenarios, use cases are even more important

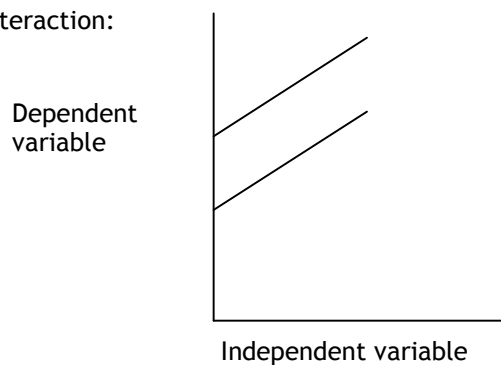
Interactions -> combine hypotheses per function

## Recommendations 3

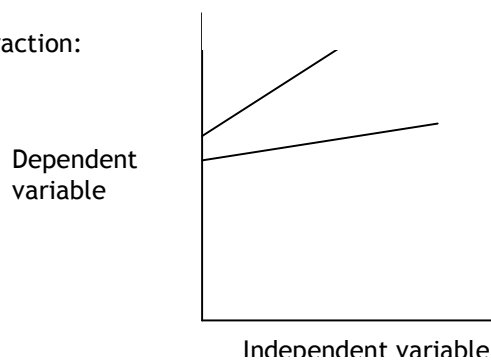
We need to distinguish between:

Dummy interaction and true interaction, where the effects are truly different

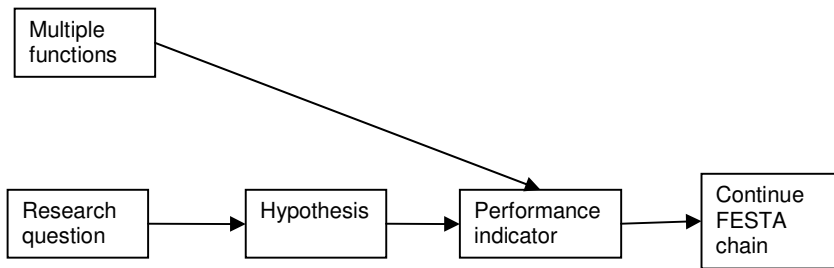
Dummy interaction:



True interaction:



## Recommendations 4



## Appendix F Glossary FOT-Net seminar

Term	Definition	Example	Comments
<b>FOT, Field Operational Test</b>	a study undertaken to evaluate a function, or functions, under normal operating conditions in environments typically encountered by the host vehicle(s) using quasi-experimental methods.	euroFOT, TELEFOT	
<b>system</b>	a combination of hardware and software enabling one or more functions	Headway sensor + controller + HMI	Set of elements (at least sensor, controller, and actuator) in relation with each other according to design. An element of a system can be another system at the same time. Then, it is called subsystem which can be a controlling or controlled system or which can contain hardware, software and manual operations.
<b>function</b>	implementation of a set of rules to achieve a specified goal	FCW	One system may have several functions
<b>research question</b>	general question to be answered by compiling and testing related specific hypotheses	FCW will improve traffic safety	
<b>hypothesis</b>	a specific statement linking a cause to an effect and based on a mechanism linking the two. It is applied to one or more functions and can be tested with statistical means by analysing specific performance indicators in specific scenarios. A hypothesis is expected to predict the direction of the expected change.	FCW will increase TTC	Hypotheses steer the selection of performance indicators, and the criteria against which those should be compared. Hypotheses are seen as questions that can be answered with the help of measurable performance indicators. Criteria can be baseline, different experimental conditions, absolute values, etc.

<b>performance indicator</b>	Performance Indicators are quantitative or qualitative measurements, agreed on beforehand, expressed as a percentage, index, rate or other value, which are monitored at regular or irregular intervals and can be compared to one or more criteria.	mean speed per distance, number of overtaking manoeuvres over time	
<b>measure</b>	A measure can either be direct or pre-processed. A direct measure is logged directly from a sensor, while a pre-processed measure is a combination of different direct or other pre-processed measures. A measure does not have a "denominator" which makes it comparable to other instances of the same measure or to external criteria.	speed, headway, answer to a question in a questionnaire	
<b>baseline period/phase</b>	the part of the data collection during which the <b>function(s)</b> operate in "silent mode", that is, they collect data, but do not give any signals to the driver. From the viewpoint of the driver the <b>function(s)</b> is/are off.		Normally the baseline period is before, but this may cause problems in terms of learning effects and seasonal effects
<b>treatment period/phase</b>	the part of the data collection during which the <b>function(s)</b> are switched on by the experimental leader, such that they are either active all the time, or can be switched on or off by the driver.		
<b>baseline within comparison situation</b>	scenario with system under evaluation "turned off"	car following with speed above 70 km/h sunny day FCW off	