



# Experimental Induction and Measurement of Negative Affect Induced by Interacting with In-vehicle Information Systems

Nadja Schömig<sup>1</sup>(✉), Frederik Naujoks<sup>1</sup>, Thomas Hammer<sup>1</sup>, Markus Tomzig<sup>1</sup>, Bastian Hinterleitner<sup>2</sup>, and Stefan Mayer<sup>2</sup>

<sup>1</sup> Würzburger Institut für Verkehrswissenschaften GmbH,  
97209 Veitshöchheim, Germany  
schoemig@wivw.de

<sup>2</sup> Audi Electronics Venture GmbH, 85080 Gaimersheim, Germany

**Abstract.** The goal of this study was to investigate whether it is possible to induce negative affects in the interaction with IVIS in a driving simulator environment and whether this would be reflected in the driver's facial expressions. N = 29 participants completed a 30-min-drive in a high-fidelity driving simulator performing several IVIS tasks using both a visual multi-level menu and a simulated speech recognition navigation system. During the drive, negative affects were induced by increasing the complexity of the menu tasks and decreasing the recognition rate of the speech system. The results show that experiencing difficulties in solving IVIS tasks went along with self-reported negative affect such as frustration and anger. A newly developed observation protocol adapted from the Facial Action Coding System (FACS) used for the ratings of negative emotions based on the video footage of the drivers' face during the simulator trials was established. The method revealed significant correlations with the self-reported measures of negative emotions. However, there were no correlations between the observed facial expressions and the sub-scales of the State-Trait-Anger-Inventory (STAXI). The study shows the importance of a well-designed IVIS in order to prevent negative emotions that might result in non-acceptance of new technologies. The study also shows the potential of facial recognition technology to provide assistance or tutoring functions that could relieve the driver from emotional discomfort.

**Keywords:** In-vehicle information system · Driving simulation  
Anger · Emotion · Facial recognition · HCI design

## 1 Introduction

In-vehicle information systems (IVIS) are available in virtually every car segment from luxury to low price vehicles. They are used for a wide variety of in-vehicle tasks such as adjusting vehicle settings, using navigation systems or in-vehicle entertainment. Providing customers with a positive and pleasurable experience while interaction with IVIS should thus be a primary concern of HCI research. However, increased complexity of IVIS might also cause negative affect such as frustration or anger, especially

when drivers are not familiar with the respective IVIS. Recognizing these negative emotions could be used both to guide the design process (e.g., by re-design of interaction sequences that commonly cause negative affect) or to provide on-line tutoring systems that provide assistance in difficult situations.

## 2 Theoretical Background

In the driving context, driver's emotions are often studied as reaction to certain environmental conditions, such as frustrating situations like traffic jams or as a reaction to the behavior of other road users [1, 2]. Within this context, the intention of investigating driver's negative emotional state lies in the reduction of aggressive driving and therefore crash rates [3].

However, there are only very few studies to date that have tried to induce and measure negative affect by interaction with IVIS. In most cases, a faulty ASR system (Automatic speech recognition system) was investigated. For example, Bachfischer et al. [4] found that users become annoyed and stressed when interacting with a faulty ASR. Malta et al. [5] also tried to measure the frustration level of drivers during interaction with a faulty ASR. The drivers rated their frustration level by watching a video of their face which was recorded during the interaction with the system. For the rating a slider was used, that could be set from neutral to extreme. For predicting the frustration level measured by the participant's responses through his/her physiological state, overall facial expression, and pedal actuation, Bayesian networks were used. Kuhn et al. [6] assessed frustration induced by different recognition rates of an ASR (44% recognition rate vs. 89%) on driving performance. They found a higher steering wheel angle variance with the low recognition rate, which points out that negative affect can indeed influence driving behavior.

According to [7] for the driving context the best measure is to assess driver's emotions by analysing facial expression. Other measures (e.g. questionnaires or EEG) would disturb the driver and would have only a low specificity to differentiate between different emotions.

Eyben et al. [8] summarize different literature with regard to emotions in driving. Among other sources of emotions, they also discuss the effects of IVIS interactions. As a meaningful strategy to reduce these negative emotions they propose a "socially competent human-like car" including an IVIS interaction which behaves like a real human codriver, that would give personalized help and quickly offer alternative solutions. However, the realization of such an approach would require the measurement of drivers' negative affect in the first place, which is exactly what we did in this study.

## 3 Method

$N = 29$  participants completed IVIS-tasks of varying difficulty in a high-fidelity driving simulator. A real-world IVIS was transferred to the simulation environment. The aim of the study was to investigate whether drivers would report negative affect when solving difficult IVIS tasks and whether this would be reflected in the drivers' facial expressions.

Therefore, self-report scales as well as a standardized observation protocol were developed. The observation protocol was used to rate the facial expressions during interactions with the IVIS and was adapted from the Facial Action Coding System (FACS, Ekman et al. [9]). To account for individual differences in the expression of negative affect, the participants completed the State-Trait Anger Expression Inventory (STAXI, Schwenkmezger et al. [10]). During the study, drivers had to complete the IVIS task while following a simulated car. A feedback of whether the task was solved correctly or not was given after each IVIS task.

### 3.1 Setting

The study was conducted at the Wuerzburg Institute for Traffic Sciences (WIVW). A static driving simulator with an Opel Insignia mockup was used (Fig. 1 left). The front of the car was surrounded by a screen illuminated by three projectors providing a 270° field of view. In addition, displays were placed in the back of the car and on the side mirrors to allow the monitoring of surrounding traffic. The driving simulation software SILAB was used.

During the test the experimenter was sitting in an adjacent room separated from the simulator by a window. From that room he was able to monitor the test and see the participant through various cameras being placed in the car. He was also able to communicate with the participant using an intercom system.

A Basler-camera with 50 Hz and a resolution of 2.3 MP placed in front of the driver recorded the driver's face and his/her interaction with the IVIS in a multi-screen video.



**Fig. 1.** Static driving simulator at the Wuerzburg Institute for traffic sciences and the car follow task (described in Sect. 3.2)

### 3.2 Driving Task

The driving task consisted of a variation of the car-follow-task proposed by the NHTSA distraction guidelines [11]. The participant was driving on a mostly straight road with two lanes. The task was to follow a car in front of the participant's car and keep a defined distance as constantly as possible. The car ahead changed its speed in a

sinus-wave profile with slightly varying amplitude so that the driver had to permanently adjust the speed (Fig. 1 right).

To put the driver in the situation of having to maintain a constant headway, a colored enhanced reality strip (ERS) was displayed on the road. Depending on the distance the color of the ERS changed. It appeared yellow when the distance was too large and blue when the distance was too small. Under perfect circumstances the bar was shown in gray (within a range between 1.0 and 1.8 s time distance). Besides the changing color of the bar, driving in the wrong distance had no additional effect.

### 3.3 Interface

For the interaction with the IVIS a simplified version of the current AUDI MMI GUI was simulated. It consisted of a hierarchical menu with different lists showing the different sub-menus available in the system. The driver had to navigate to different end points of the menu. The final selection of the end points did not trigger any actual reaction but resulted in a feedback depending on the task whether the selection was correct.

An AUDI MMI interaction unit was placed on top of the mockup's original console. This unit consisted of a rotary interaction knob and shortcuts for different menu levels (Fig. 2).



Fig. 2. AUDI MMI GUI (left) and interaction unit (right)

### 3.4 IVIS Tasks

During the 30 min drive the participant had to perform a total of 24 IVIS tasks. 20 of them were visual-manual interactions using the rotary interaction knob to navigate to certain entries in the menu system, the remaining 4 required interactions with a speech-based navigation system. When completing the menu tasks, drivers had to search for certain menu entries and to confirm their selection. The tasks were instructed verbally e.g. “please adjust intensity of the heater” at certain defined positions along the route. A computer-generated voice gave the instructions at the defined road section.

The tasks varied in their difficulty to solve. Half of them were very easy to solve while the other tasks were manipulated so that it was very difficult to reach the correct menu entry. This was either reached by a very ambiguous instruction so that it was not clear in which category the menu entry could be found and/or by a very short time buffer available to reach the menu entry before the task was aborted. Some of the tasks were even unsolvable as the instructed menu items were not part of the menu list. The task difficulty was determined in a previous pilot study. During the drive the frequency of tasks with higher complexity increased gradually to induce negative emotions.

The drivers were instructed that by solving the tasks correctly they could gain bonus points. In case of a correct answer, drivers received five bonus points and the feedback “answer was correct” displayed at the lower part of the visual interface. An extra monetary award was promised for that driver who reached a certain target score of 50 points. In case of an incorrect answer the participant got the feedback “task incorrect” in combination with a loud and unpleasant sound. If the driver was not able to reach the entry within a certain time buffer the task was aborted and the same sound was presented.

In addition, four other tasks requiring verbal interactions with the speech system were arranged in between the other menu tasks. They consisted of computer-generated speech outputs of a fictive “navigation system” which informed the driver about an upcoming traffic jam on the road. The driver was asked to either change the original route or to maintain the current route. The driver was instructed to avoid any traffic jam and should therefore answer with “yes” or “no” to the various questions in order to maintain the original route. Two of these tasks were manipulated in the way that the speech recognition system pretended to misinterpret the answer of the driver resulting in an incorrect feedback although it was correct. In one of the speech recognition tasks the answer had to be repeated several times before the system understood it correctly.

In case of wrong answers to the navigation systems 20 points were deducted from driver’s credits. Shortly before the target score of 50 was reached the manipulated navigation task was triggered which resulted in an incorrect feedback and a deduction of the score which should further increase driver’s negative emotions.

Table 1 summarizes the task sequence and the online ratings during the drive.

29 participants took part in the study. To achieve optimal availability of camera data only persons with a minimal height of 1.64 m were invited. 19 men and 10 women participated with an age of 24 to 65 years ( $m = 42$ ,  $sd = 12.8$  years).

**Table 1.** Used tasks and time points for online self-reportings of emotional state. Menu tasks are counted with figures, speech interactions with the navigation system with letters.

Task	Task type	complexity	block
<b>t<sub>0</sub> Online rating of emotional state (Baseline)</b>			
1	Menu task	easy	
2	Menu task	easy	
3	Menu task	easy	I
4	Menu task	easy	
A	Speech Interaction	correct feedback	
<b>t<sub>1</sub> Online rating of emotional state</b>			
5	Menu task	easy	
6	Menu task	easy	
7	Menu task	easy	
8	Menu task	ambiguous	II
9	Menu task	ambiguous+ short	
10	Menu task	not solvable	
<b>t<sub>2</sub> Online rating of emotional state</b>			
<b>t<sub>3</sub> Online rating of emotional state</b>			
11	Menu task	easy	
B	Speech Interaction	wrong feedback	III
<b>t<sub>4</sub> Online rating of emotional state</b>			
12	Menu task	easy	
13	Menu task	easy	IV
<b>t<sub>5</sub> Online rating of emotional state</b>			
C	Speech Interaction	correct feedback	
14	Menu task	not solvable	
15	Menu task	complex	
16	Menu task	complex	
D	Speech Interaction	3x repeat task	V
17	Menu task	ambiguous	
18	Menu task	complex	
19	Menu task	ambiguous+ short	
20	Menu task	complex	
<b>t<sub>6</sub> Online rating of emotional state</b>			

### 3.5 Measures

#### IVIS Task Performance

Driver performance in the IVIS task was coded as dichotomous variable for all of the 24 tasks. The task could either be performed successfully or not. The tasks were summarized into several blocks with each block comprising the tasks between two time points where drivers rated their current emotional state online. As parameter for task performance the relative frequency of successfully solved tasks per block was used. Dependent on the design of the tasks and the study, this frequency per se went down during the course of the drive (as the number of not solvable tasks increased).

#### Self-reported Affects

During the experiment, drivers were repeatedly asked to online rate their current emotions on different dimensions (how joyful, relaxed, angry, frustrated, helpless, irritated are you at the moment?). A 16-point-rating scale with verbal categories (according to Heller [12]) was used. The ratings had to be given after certain blocks of tasks (including an unequal number of tasks, but summarized according to their intention to induce negative affects).

A pre-rating before the drive served as baseline for the later ratings. Altogether, the online ratings were collected at six time points.

#### Observation of Driver's Expression

Drivers' facial expressions were coded by a rater who was blind for the content and the aim of the study. The basis for the coding was the video footage of the participants which showed the driver's face, the road and interaction with the menu system

(see Fig. 3). The sound was muted so that verbal expressions could not be heard. From the video footage, the sections containing IVIS tasks were selected, resulting in 24 sequences per driver that had to be rated. The sequences started with the verbal instructions of the tasks and ended 10 s after the final feedback of the IVIS system had been given.



**Fig. 3.** Video footage used for observer rating

The raters were first trained by a specifically developed training procedure which condensed the principles and instructions of the widely-used FACS (Facial Acting Coding System by Ekman et al. [9] which is a tool for measuring facial expressions. It is an anatomical system for describing all observable facial movements. Each of these movements is labeled as an Action Unit (AU). The combination of AUs then defines a certain emotion (typically the six basic emotions). For the intentions of this study, it would have been too time intensive to use the original material and training. Therefore only the relevant AU combinations for the emotion anger were extracted, extended by other indicators for frustration or irritation (not included in FACS) and explained to the raters in a training session. Afterwards, the raters scored each single facial expression within each video sequence on a specially developed observation protocol. The rating was given on a scale from 0 (no emotion present) to 5 (maximum degree of any kind of negative emotion; positive emotions such as amusement were not coded). As independent measure the expression with the highest value per video sequence was used.

In order to achieve a high reliability and validity of the measure, an interrater-reliability score was calculated in advance from the ratings of three independent persons assessing the video sequences of twelve drivers. It turned out, that it was very difficult to differentiate between specific negative emotions (e.g. anger and irritation) leading to the decision to summarize these emotions to a more global rating in the analysis. The calculation of weighted kappa [13] for each pair of the three raters revealed a medium to high interrater-reliability (.673/.487/.556). The final rating was given by one rater who rated all the remaining video sequences of the 29 drivers.

### **State-Trait-Anger-Inventory**

In order to evaluate whether the degree of self-reported negative emotions and the facial expressions correlate with a certain anger-personality the so called State-Trait-Anger Inventory was used (STAXI, Schwenkmezger et al. [10]). It is a measure for state- and trait-anger with three different forms of anger expression. People differ in the extent to which they express anger overtly and directly (anger out), how often anger feelings are held in or suppressed (anger in) or in the extent to which they are able to control their angry feelings or their overtly expressed anger (anger control).

### **3.6 Study Procedure**

First, the participants had to sign a secrecy agreement and were then presented with general information about the study. All participants were familiar and trained in using a driving simulator. The participants were told that the goal of the study was to evaluate the structure and content of the menu system and how difficult the search for certain menu entries would be. Therefore the tasks would be somewhat difficult but always possible to solve. They were promised a monetary reward if a target score was reached. They were left blind about the real intent of the study.

To become familiar with the system the participants were given six tasks. Each task led to a different section of the menu, so that every section had been seen before the actual test started.

Before the start of the drive, the participants rated their current emotional state on the 6 items using the 16-point-scale. During the 30 min test drive each participant had to perform the 24 IVIS tasks at defined positions on the route. After each of the defined five task blocks the participants again rated their emotional state online. After the drive the participants filled in several questionnaires, including the STAXI. At the end the real intent of the study was resolved. Before participants received the monetary reward (including the promised bonus in case of reaching the target score) a “cool down phase” assured that they could recover from a potentially high arousal level induced by the study. The whole procedure took about 60 min.

## **4 Results**

### **4.1 Task Performance**

According to expectations, task performance decreased during the test drive (see Table 2). The comparison between the five measuring points revealed a highly significant effect  $F(4, 24) = 34.30, p < .001$ .

### **4.2 Self-reported Emotional State**

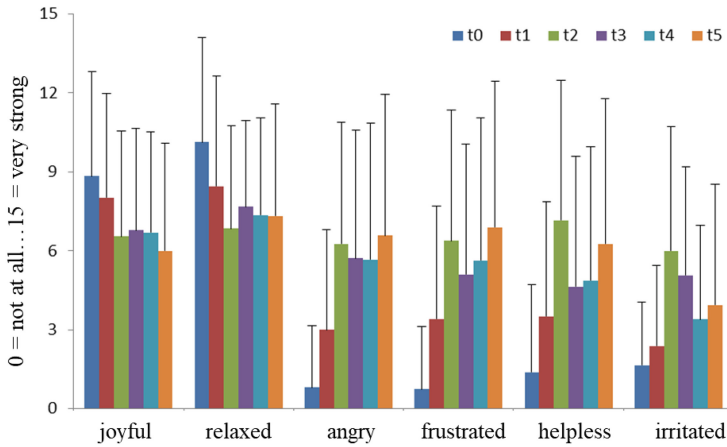
The self-reported emotional state changed gradually towards a more negative direction (see Fig. 4). The emotions anger and frustration heavily increased beginning with the second task block (which included several complex tasks) and then maintained on a constant, moderate level (not higher than 7 on the scale up to 15). However, there was



**Table 2.** Means and standard deviation of the dependent variables.

Measuring point	Task performance		Self-reported anger		Negative facial expression	
	M	SD	M	SD	M	SD
t <sub>0</sub>	–	–	1.29	2.45	–	–
t <sub>1</sub>	0.75	0.20	3.04	3.92	1.26	0.76
t <sub>2</sub>	0.46	0.13	6.32	4.74	1.60	0.67
t <sub>3</sub>	0.48	0.09	5.71	4.88	1.12	0.85
t <sub>4</sub>	0.50	0.44	5.71	5.21	1.74	0.82
t <sub>5</sub>	0.28	0.13	6.68	5.49	1.74	0.64

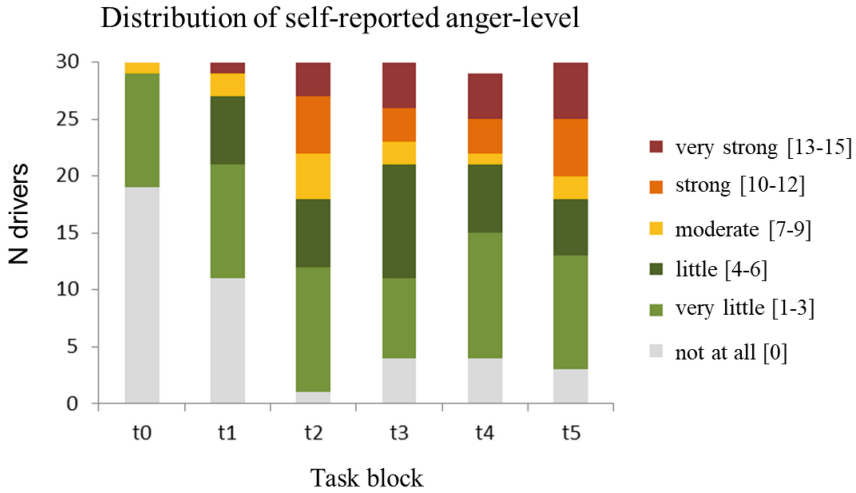
Changes of emotional state across the drive



**Fig. 4.** Self-reported emotional state on six dimensions measured online at six time-points (means and standard deviations are displayed).

also a high variation between the subjects. The self-reported helplessness and irritation also increased beginning with the second block of tasks, however, decreased again towards the end of the drive indicating a kind of resignation if it turned out that the system’s reactions are somewhat independent from the own task performance.

When looking further into detail for the anger ratings, it gets obvious that about one third of the sample (n = 10 drivers) subjectively perceived a strong or very strong feeling of anger (see Fig. 5). The peak lies here at the very end of the drive after block 5. However, there is also a remarkably large group of participants who expressed to experience very little or no anger at all. According to drivers’ comments some of them realized the real intention of the study and therefore did not seem to be influenced by the tasks’ manipulations. This result shows the difficult underpinning of inducing a certain emotional state in such an artificial experimental setup.



**Fig. 5.** Distribution of self-reported anger-level on the 16-point rating scale.

**Table 3.** Correlations between the different measures.

		Self-reported anger	Negative facial expression
Task performance		$r(142) = -.40$	$r(688) = -.28$
Negative facial expression		$r(142) = .37$	–
STAXI	Trait-Anger	$r(27) = .45$	$r(27) = .05$
		$p = .007$	$p = .398$
	Anger-Out	$r(27) = .26$	$r(27) = .15$
		$p = .085$	$p = .222$
	Anger-In	$r(27) = .43$	$r(27) = .20$
		$p = .010$	$p = .147$
	Anger-Control	$r(27) = -.35$	$r(27) = -.29$
		$p = .03$	$p = .061$

### 4.3 Correlations Between the Different Measures

For the description of the relationships between the dependent measures, Pearson coefficients of correlation were calculated (see Table 3), revealing the following relationships:

- Self-reported anger correlated significantly negative with task performance. The lower task performance between two measuring points the higher is the subjectively perceived anger.
- Facial expression correlated significantly positive with self-reported anger. Participants who reported high anger-values during the drive also had a more intensive facial expression.

- In addition, participants showed stronger emotions in their faces when they could not perform a task successfully.

The four subscales of STAXI were correlated with each other and with the mean values of self-reported anger as well as the mean values of negative facial expressions. As expected, participants with higher scores on the Trait-Anger scale had also higher scores on the Anger-Out and Anger-In subscale but lower scores on the Anger-Control subscale. In addition, there was a significant positive correlation with the self-reported anger but no significant correlation with the facial expression. Trait-Anger therefore reflects whether and to what extent a person gets angry but gives no information on the appearance or extent of facial expressions.

Furthermore, participants with higher Anger-In scores expressed higher subjective anger but did not show that mandatorily in their faces. Against expectations, Anger-Out did neither correlate with self-reporting of anger nor with the facial expression although this subscale intended to measure the degree to what a person expresses his/her feelings.

## 5 Discussion and Implications

It could be shown that experiencing difficulties in solving IVIS tasks goes along with self-reported negative affect as well as according facial expressions. The reliability and validity of the newly developed rating tool could also be shown. The study has several implications that are important both for the design of IVIS and advancement of theories of user experience.

Firstly, designers should be aware that cumbersome interaction design will cause negative emotions which are likely to result in non-acceptance of new products. It is well known that negative affect such as anger and frustration can also cause a more risky and ruthless driving style, so that such negative impacts on the driver's emotional state could ultimately become a safety risk. Secondly, the study is a first step into measuring and eventually preventing such disadvantageous developments in the design process. The study also shows the potential of facial recognition technology to provide assistance or tutoring functions that could relieve the driver from emotional discomfort.

From a theoretical point of view, it is somewhat surprising that anger expressions were not correlated with the STAXI subscales, as the STAXI is the most used questionnaire that measures personality traits related to anger expression. From this finding, it appears that significantly more research has to be conducted to fully understand the development and expression of negative affect (such as anger) related to technology use.

## References

1. Hennessy, D.A., Wiesenthal, D.L.: The relationship between traffic congestion, driver stress and direct versus indirect coping behaviours. *Ergonomics* **40**(3), 348–361 (1997)
2. Underwood, G., Chapman, P., Wright, S., Crundall, D.: Anger while driving. *Transp. Res. Part F* **2**(1), 55–68 (1999)

3. Wells-Parker, E., Ceminsky, J., Hallberg, V., Snow, R.W., Dunaway, G., Guiling, S., Williams, M., Anderson, B.: An exploratory study of the relationship between road rage and crash experience in a representative sample of US drivers. *Accid. Anal. Prev.* **34**(3), 271–278 (2002)
4. Bachfischer, K., Henze, S., Wäller, C.: Emotional user interfaces in the car. In: Proceedings of the 1st Workshop on Emotion and Computing—Current Research and Future Impact. In: Reichardt, D., Levi, P., Meyer, J.-J.C. (eds.), pp. 55–59. Bremen, Germany, June 2006
5. Malta, L., Miyajima, C., Kitaoka, N., Takeda, K.: Analysis of real-world driver's frustration. *IEEE Trans. Intell. Transp. Syst.* **12**(1), 109–118 (2010)
6. Kuhn, A., Paek, T., Medenica, Z.: The effect of speech interface accuracy on driving performance. In: *Interspeech* (2007)
7. Eyben, F., Wöllmer, M., Poitschke, T., Schuller, B., Blaschke, C., Färber, B., Thien, N.N.: Emotion on the road: necessity, acceptance, and feasibility of affective computing in the car. *Adv. Hum.-Comput. Interact.* **2010**, 1–17 (2010)
8. Zhang, M., Ihme, K., Jipp, M.: Maschinelle Erkennung der Frustration von Autofahrern aus Videoaufnahmen des Gesichts. In: 5 Interdisziplinärer Workshop Kognitive Systeme: Mensch, Teams, Systeme und Automaten. Bochum, Deutschland (2016)
9. Ekman, P., Friesen, W.V., Hager, J.: *The Investigator's Guide for the Facial Action Coding System. A Human Face*, Salt Lake City (2002)
10. Schwenkmezger, P., Hodapp, V., Spielberger, C.D.: *Das State-Trait-Ärgerausdrucks-Inventar STAXI: Handbuch*. Hans Huber, Bern (1992)
11. National Highway Traffic Safety Administration (NHTSA): Visual-manual NHTSA driver distraction guidelines for in-vehicle electronic devices. Notice in the Federal Register, vol. 77, no. 37, 24 February 2012
12. Heller, O.: Hörfeldaudiometrie mit dem Verfahren der Kategorienunterteilung (KU). *Psychologische Beiträge* **27**, 478–493 (1985)
13. Cohen, J.: Weighted kappa: nominal scale agreement with provision for scaled agreement or partial credit. *Psychol. Bull.* **70**, 213–220 (1968)