

Review Article

Emotion on the Road—Necessity, Acceptance, and Feasibility of Affective Computing in the Car

Florian Eyben,¹ Martin Wöllmer,¹ Tony Poitschke,¹ Björn Schuller,¹ Christoph Blaschke,² Berthold Färber,² and Nhu Nguyen-Thien³

¹Institute for Human-Machine Communication, Technische Universität München, 80333 München, Germany

²Human Factors Institute, Universität der Bundeswehr München, 85577 Neubiberg, Germany

³Continental Automotive GmbH, Interior BU Infotainment & Connectivity, Advanced Development and Innovation, 93055 Regensburg, Germany

Correspondence should be addressed to Florian Eyben, eyben@tum.de

Received 25 February 2010; Accepted 6 May 2010

Academic Editor: Kostas Karpouzis

Copyright © 2010 Florian Eyben et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Besides reduction of energy consumption, which implies alternate actuation and light construction, the main research domain in automobile development in the near future is dominated by driver assistance and natural driver-car communication. The ability of a car to understand natural speech and provide a human-like driver assistance system can be expected to be a factor decisive for market success on par with automatic driving systems. Emotional factors and affective states are thereby crucial for enhanced safety and comfort. This paper gives an extensive literature overview on work related to influence of emotions on driving safety and comfort, automatic recognition, control of emotions, and improvement of in-car interfaces by affect sensitive technology. Various use-case scenarios are outlined as possible applications for emotion-oriented technology in the vehicle. The possible acceptance of such future technology by drivers is assessed in a Wizard-Of-Oz user study, and feasibility of automatically recognising various driver states is demonstrated by an example system for monitoring driver attentiveness. Thereby an accuracy of 91.3% is reported for classifying in real-time whether the driver is attentive or distracted.

1. Introduction

More than 100 years of history of the automobile are marked by milestones as the combustion engine and mechanical components followed by electrical and electronic device integration, increasing usage of control technique and software. Apart from reduction of fuel consumption, and thus alternative actuation and light weight construction, the main research interest in automobile development in the near future is dominated by driver assistance and natural, intuitive driver-car communication. This statement is supported by various EU-funded research projects such as PREVENT (<http://www.prevent-ip.org/>), SASPENCE (sub-project of PREVENT), and PROSPER [1], which aim at advancing the state-of-the-art in the area of driver assistance systems, and a body of literature on in-car signal processing, for example, [2]. In this respect the ability of a car to talk naturally and provide a virtual companion can be expected

to be a market success decisive factor in future automotive systems as “next big thing” on par with automatic driving systems and intelligent measures to improve driving safety.

Emotional factors are decisive for enhanced safety and comfort while driving a car, as we will show in Section 2.1. It is thus necessary for a car to sense these and by that better understand a driver’s intention and/or state. The aim of in-car emotion recognition should be to support the driver in performing primary, secondary, and tertiary driving tasks. Thereby the primary driving task, which includes steering, accelerating, braking, and choosing the correct lane, speed, route, and distance to other vehicles, as well as the secondary driving task, denoting activities like dimming, operating windscreen wipers, coupling, changing gears, and blinking, can be seen as rather safety-related whereas the tertiary driving task (operating air conditioner, seat heater, radio, and phone) mainly refers to comfort [3].

Constantly increasing provision of speech technology as well as gaze detection and eye/head movement monitoring mark the beginning of more natural ways of human-machine interactions which are based on intuitive communication modalities. Recognition of emotion from vocal and facial expression, physiological measurement, and contextual knowledge will be the next key-factor driving improved naturalness in many fields of Human-Computer Interaction [4]. Next to the modalities speech, facial expression, physiological measurements, and contextual knowledge, driving parameters can be used as an important and reliable modality for driver emotion and state recognition in a car.

This paper will give an introduction to in-car affective computing with an extensive literature overview on studies and existing work in Section 2. The section includes a discussion of the influence of emotions on the driving performance, and lists methods for recognising emotion and control of emotion. Moreover, the concept of a “virtual companion” will be presented. We will outline various use-case examples which illustrate how emotion-oriented technology can be used in a vehicle in Section 3. An open issue is that of user acceptance of emotion aware technology. In order to assess this acceptance, we conduct a pilot study where we interrogate users on their experiences using a Wizard-of-Oz (WoZ) prototype system (see Section 4). To show the feasibility of automatic driver state recognition for safety and infotainment-related tasks we finally and briefly present a system for detecting driver distraction in Section 5, based on measuring various driving style parameters and tracking the driver’s head motion.

2. Literature Review

This section gives an extensive literature overview on the topic of affective computing and the role of emotions and other driver states in the car. We investigate the influence of affective states on the driving performance in Section 2.1. Road rage, fatigue, stress, confusion, nervousness, and sadness are picked as main factors with respect to driving safety. Further, we deal with techniques for recognition and control of various driver states. The two main strategies of countersteering negative emotions, and alternatively adapting car functionalities to the driver’s emotion are the focus of Section 2.2. They are followed by a discussion of the modalities used to actually recognise driver emotion in the car. Next, we discuss the feasibility and benefits of a “socially competent” car in Section 2.3. This type of car is supposed to enhance the driving experience, that is, the driver’s pleasure, as “happy drivers” were shown to be the better drivers in any respect (e.g., [5]).

2.1. The Influence of Affective States on Driving Performance. The role that emotions and other mental states (such as fatigue) play while driving a car becomes evident when considering essential driver abilities and attributes that are affected by emotion: perception and organisation of memory [6, 7], categorisation and preference [8], goal generation, evaluation, decision-making [9], strategic planning [10],

focus and attention [11], motivation and performance [12], intention [13], communication [14–16], and learning [17]. Taking into account the great responsibility a driver has for her or his passengers, other road users, and her- or himself, as well as the fact that steering a car is an activity where even the smallest disturbance potentially has grave repercussions, keeping the driver in an emotional state that is best suited for the driving task is of enormous importance. Of course, similar to a simple control circuit, for an “intelligent” car the first step towards *influencing* or even *controlling* a driver’s emotional state is to *measure* emotion. But what kind of emotion would be ideal to optimally perform primary and secondary driving tasks? Obviously the driver’s emotion should support capabilities like attention, accurate judgement of traffic situations, driving performance, compliance, fast and correct decision making, strategic planning, and appropriate communication with other road users. Literature answers the question of the optimum emotional state with the statement “happy drivers are better drivers” [5, 18, 19]. Research and experience demonstrate that being in a good mood is the best precondition for safe driving and that happy drivers produce fewer accidents [20]. The following sections take a closer look at how other affective states can influence driving performance.

2.1.1. Aggressiveness and Anger. Aggressiveness and anger are emotional states that extremely influence driving behaviour and increase the risk of causing an accident [21]. “Road rage” denotes an extreme case of aggressive driving implying specific incidents of anger intentionally directed at another driver, vehicle or object. Approximately 16 million people in the US might suffer from road rage disorder, as was reported by CNN news in 2006 and is cited in [22]. Extreme forms even involve physical attacks, confrontation with other drivers, “tailgating” (i.e., driving too closely at a distance), and cutting another driver off the road. Apart from these grave misbehaviours, slightly milder forms of road rage like provocation of other road users, obscene gestures, and expressing anger by yelling or honking are part of day-to-day traffic interactions and concern a significantly larger number of traffic participants [23]. Even those comparatively mild levels of aggressiveness disrupt the drivers attention and preclude the driver from concentrating on the traffic, increasing the risk of an accident [24].

On the other hand, a too low level of activation (e.g., resulting from emotional states like sadness or fatigue) also leads to reduced attention as well as prolonged reaction time and therefore lowers driving performance. As stated by Yerkes and Dodson [25], a medium level of activation results in the best performance, whereas the precise optimum level of activation (OLA) varies, depending on the difficulty of the task. This relationship is commonly known as the Yerkes-Dodson Law.

2.1.2. Fatigue. Another example for dangerous driver states is sleepiness, which affects all abilities that are important for driving, in a negative way. The fact that even when people recognise they are tired, they often force themselves

not to take a rest but to go on driving, makes sleepiness a severe problem in today's car traffic [20]. According to [26] up to 3% of all motor vehicle crashes happen due to sleepiness whereas the main risk factors are youth, shift work, alcohol, drugs, and sleep disorders. Since fatigue degrades alertness as well as quick and accurate perception, judgement, and action, tired drivers not only risk accidents from falling asleep while driving, but also from slowing down of reactions or loss of attention during time-critical manoeuvres like abruptly breaking when observing the end of a traffic jam on a highway or avoiding to hit a pedestrian. Various surveys demonstrate that many people experienced excessive sleepiness during driving [27–29]: 55% of 1 000 interviewed individuals had driven while being sleepy during the year preceding the survey [30], 23% had fallen asleep while driving once or more than once in their life, and almost 5% already had an accident due to sleepiness. Those figures indicate that fatigue is a serious problem in traffic and belongs to the most alarming states of a car driver.

2.1.3. Stress. As automobile driving itself can often be a source of stress, it seems obvious that stress is an affective state which is very likely to occur in a car. Driving closely behind other vehicles, changing traffic lanes during rush hour, receiving a phone call, getting to ones destination on time, and paying attention to traffic rules are only some of the tasks which partly have to be fulfilled simultaneously by the driver and therefore cause mental overload. A frequently experienced event is rush hour traffic congestion which is interpreted as stressful by almost every automobile driver and causes many people to use public transport and to dismiss the private car in urban areas. Similar to anger and aggressiveness, stress usually implies a high level of arousal which in turn leads to a lack of focus and attention and therefore lowers driving performance [31]. Excessive workload during driving, for example due to distraction caused by using the cell phone in the car, was shown to downgrade driving style [32]: using a cell phone causes drivers to have higher variations in accelerator pedal position and to drive more slowly with high variation in speed. Obviously such a decrease of driving performance and concentration is also caused by other in-car information and entertainment systems. (cf. <http://www.nuance.com/distracteddriving/>, and Department for Transport Project: Scoping Study of Driver Distraction (2008), Ref. T201T, summary available at: <http://www.dft.gov.uk/rmd/project.asp?intProjectID=12560>) which suggests that with the growth of car functionality the need for monitoring the drivers' stress level increases.

2.1.4. Confusion. Confusion or irritation is a further state which can lead to a loss of self-control and control over the vehicle, increasing the probability of committing a traffic violation or even being involved in an accident [33]. Sources of confusion can on the one hand be nonintuitive user interfaces or defective systems (like e.g., a badly designed navigation system or an error prone speech recogniser; in the latter case detected confusion could be used to increase

the speech recognisers robustness). On the other hand irritating traffic situations like route diversions, mistakable signs or complicated routing of a road can cause confusion. Just like stress, irritation leads to disturbance of driver capabilities such as decision-making, attention, perception, and strategic planning. Particularly older people tend to be confused by the amount of information they have to process simultaneously during driving [34] as today neither car information systems nor all traffic signs or routes are designed for elderly people who often have longer reaction times and slower perception. Supporting irritated drivers through intelligent emotion-sensitive assistance systems will become indispensable for future car generations as confusion potentially increases with the number of car functionalities.

2.1.5. Nervousness. Nervousness is an affective state that implies a level of arousal which is above the degree of activation that is best suited for the driving task. Reasonable decision-making as well as strategic planning and concentration are affected when being nervous. Reasons for nervousness are variable and can be related directly to the driving task (e.g., for novice drivers) or to other personal or physical circumstances. In [35] the nervousness induced by the use of drugs is examined with respect to effects on driving: nervous drivers tend to perform worse as far as driving ability is concerned—mainly due to poorer concentration. Also Li and Ji name nervousness as one of the most dangerous driver states and point out the importance to detect nervousness to provide intelligent assistance and appropriate alerts [36].

2.1.6. Sadness. Also negative emotions with a rather low level of arousal, like sadness or frustration can have perturbing effects on driving ability [37]. An example is shown in [38] where the influence of terror attacks on driving performance was examined: in Israel an increase of traffic accidents by 35% was observed on the third day after terrorist attacks. Sadness seriously affects the level of attention of a driver and therefore endangers the safe operation of a vehicle. As frustration and sadness usually coincide with a certain degree of passiveness or resignation, reaction time in critical situations increases.

Apart from safety aspects, when thinking of the car as a “virtual companion”, the automatic recognition of sadness as an emotional state maybe one day enable the system to cheer up the driver and thus deliver also enhanced driving pleasure besides increased safety.

2.2. Recognition and Control of Driver States. So far, we have pointed out the enormous effect that affective states and emotions have on driving abilities and listed the most dangerous affective states which prevent safe driving. However, the need for automatic in-car emotion recognition and driver state detection only becomes evident when examining adaptation or even “countersteering” strategies that can easily be implemented provided that the drivers emotion is determined accurately. The aim of affect recognition is to provide a kind of “state variable” which serves as input

for subsequent processing in emotion-sensitive accessories, aiming to improve not only driving comfort but also safety [5, 36, 39, 40]. Thereby secure driving can be supported by either attempting to improve the affective state of *the driver*, which would mean making the driver “happy” or at least directing her or him into a neutral emotional state or by adapting *the car* with the emotion of the driver [41, 42]. Both strategies rely on proper emotion recognition and were shown to improve driving performance and reduce the risk of having an accident.

2.2.1. Countersteering Negative Emotions. To reduce the stress level of the driver, dialogue strategies can be adapted to the current workload [43, 44]. Since voice messages potentially distract the driver, an adaptive system would deliver messages only when the driver’s stress level is low. A method to avoid stress caused by traffic jams could be to warn the driver in time as soon as he or she intends to use a route for which congestion had been reported to the system. This, of course, is a desirable feature regardless of affect aware technology, but has the possible benefit of reducing negative emotions.

A possible approach towards making, for example, an angry driver aware of the dangerous driving style, resulting from her or his increased level of arousal, would be to encourage better driving via voice response [45] or to make appropriate alerts [36] (e.g., making the driver aware that the current traffic situation demands more thoughtful actions than emotional actions). Calming down the driver or increasing the consciousness of critical manoeuvres caused by aggressive driving would be a typical method of verbal countersteering that a “virtual companion” might perform in the far future, thereby replacing a reasonable, observant human codriver.

In [5] it is suggested that the car could become more or less talkative depending on the affective state of the driver. Sleepiness is a driver state which requires increased communicativeness of the virtual companion in order to involve the tired driver into a conversation and thereby aiming to make her or him attentive again or even to prevent her or him from falling asleep. This is equivalent to what a good codriver would do to compensate the sleepiness of the driver. However, according to [26], the only safe countermeasure against driving while being sleepy is to stop driving. This advice also counts to the useful alerts an intelligent car could give, provided that the driver state “fatigue” is recognised reliably.

To countersteer detected confusion, an emotion-sensitive system could provide help or more detailed explanations concerning the functionality which the driver is about to use. Complicated information or entertainment systems benefit from automatic guidance through menus. that could be triggered by the detection of irritation. As far as confusion or nervousness due to traffic situations is concerned, it was shown that particularly elderly people profit by the recognition of irritation and subsequent driving support [46] leading to better driving performance and to higher confidence while driving.

2.2.2. Adapting Car Functionalities to Driver Emotion. Apart from trying to influence the driver’s emotion in a positive way, adapting user interfaces to the user’s affective state can also reduce the risk of accidents and potentially leads to higher driving pleasure. Experiments indicate that matching the in-car voice with the driver’s state not only encourages users to communicate with the system, but also improves driving performance [41]. Choice of words, intonation, and tone of voice are important aspects of communication and should be adapted to the emotion of the conversational partner to make the dialogue more natural. Further, it is known that the words which are used to inform the driver of bad driving performance are much more effective when they ascribe the bad performance to the driving environment rather than to the drivers themselves [47]. A voice that matches the driver’s emotion increases the connection between the user and the voice and, like most other adaption strategies, corresponds to what a human codriver would do.

As an important step towards enhanced and reliable speech recognition, adaptation of speech recognition engines to the driver’s current emotion is a technique that has prevailed in increasing the robustness of speech recognition systems [48, 49]. Both the acoustic realisation of a spoken utterance and the choice of words are highly dependent on the speaker’s emotion [50] which makes it necessary to adapt acoustic models (emotionally coloured speech has a different spectral pattern than normal speech, e.g.) as well as the language model of the speech recogniser in order to maintain automatic speech recognition performance for emotionally coloured speech. This again stresses the need for emotion recognition in the car as a major component to guarantee not only safety and comfort but also accuracy and robustness of other functionalities like automatic speech recognition.

In this context the design of emotion dependent speech dialogues for natural interaction with in-car systems will be an upcoming challenge. Besides speech technology improvements, new concepts regarding the interaction design with other input and output modalities are also relevant. Flat menu hierarchy, “one click” solution, user interfaces with seamless multimodality and usage of handwriting recognition (for almost blind text input on a touch display without having to look at buttons) are some examples.

2.2.3. Modalities for In-Car Emotion Recognition. As in many pattern recognition disciplines, the best emotion recognition results are reported for multimodal recognisers [44, 51] that make use of more than one of the four major modalities. These are audio (e.g., [52–54]), video (e.g., [55, 56]), physiology (e.g., [57–59]), and driving style (Section 5). However, not every affective state can be detected equally well from every modality. It will, for example, be hard to reliably recognise sleepiness from the speech signal since a tired driver probably does not talk. Yet, visual information (frequency and duration of twinkling as a trivial example)—if combined with infrared illumination at night time—will be a well-suited indicator for fatigue, and driving style is a good indicator of distraction as we will see in Section 5.

For the recognition of driver states like anger, irritation, or nervousness however, the audio channel was proven to be valuable [19, 40, 60]. This seems little surprising when considering how strongly for example anger is correlated to simple speech features like volume (and energy, resp.) or pitch. The great advantages of the speech modality are low hardware costs, relatively low apparent observation, and high reliability. Furthermore the user is able to control how much emotion is shown, which of course is a disadvantage for constant and reliable driver monitoring; audio information simply is not continuously present if the driver does not constantly speak. Further, not all speech captured by the microphone may be relevant in such an open microphone scenario, which makes recognition tasks more difficult [61].

Recognisers exploiting visual information have been applied for the detection of emotions like anger, sadness, happiness, disgust, fear, irritation, and surprise [44, 62, 63]—mostly from facial expressions. For interest detection, for example, the visual modality also seems to be superior to the audio modality [64]. In contrast to speech, video is omnipresent, however, the use of visual information implies slightly higher hardware costs, and increased observation feeling.

More extensive approaches (at least from the sensory point of view) to measure emotion also include physiology exploiting data from electromyograms, electrocardiograms, respiration, and electrodermal activity [65] measuring quantities like heart rate or skin conductance. These methods are at present mainly used for research and not for Human-Machine Interaction as they require a great amount of noncommercial hardware. Depending on the type of signal, the hardware costs of physiological measurements can also be marginal, however, user acceptance will still be low as the driver—apart from the process of wearing such devices—has a strong feeling of “being watched” and controllability is not granted.

The use of driving style as a modality for emotion recognition is quite self-evident and less costly although not investigated very intensely so far. The fact that driving style and affective state are highly correlated was outlined in Section 2.1 and can be utilised for in-car emotion recognition. A small study on a system using this modality is presented as an example in Section 5.

Motion of the driver in her or his seat is another method to measure nervousness or activity [66], for example. However, we have to be cautious not to confuse nervousness with backache or other problems which could make the driver move more than usual. It has to be carefully researched if backache and nervousness produce different movement patterns. Contextual knowledge is mostly to be seen as additional knowledge source, yet certainly highly reasonable to improve audio- or video-based recognition performance.

2.3. The Socially Competent Human-Like Car. Besides improvements in driving safety by monitoring the driver’s emotional state, the upper class car of tomorrow will also be “socially competent”, that is, more human-like with respect to verbal and nonverbal communication and interaction

skills and, possibly somewhat limited, understanding of nonverbal meaning and contextual information. The car can be expected to be able to interact with driver and passengers in a way quite natural to us as humans. It could be able to serve as a virtual companion or secretary and assist the driver in difficult situations. The car will likely be more like a real human codriver than like the touchscreen based interfaces found in today’s cars [67, 68], however ensuring controllability at all times, that is, the driver will be able to switch off the “talking interface”. One might now ask the question why cars have to be more human-like. The answer is simple: because people want it and a more human-like interface simplifies interaction with advanced technology. The trend is clear, already today people driving upper class cars demand for the latest state-of-the-art technologies in their cars. There exist route guidance systems, finest HiFi entertainment and integrated cell-phones including messaging, calendar and e-mail systems, all synchronisable with your home, work and laptop computer. Still in a development stage are topics like real internet access, just-in-time information about the environment, and real time traffic alerts and warnings [69]—all functions controlled via natural language voice interaction between driver and car. In Section 4, we present the results of a small survey regarding the user acceptance of such technology in the car.

A major problem arising along with the growing number and complexity of in-car entertainment and communication systems is the increased distraction of the driver caused by these systems. When changing your route while driving, for example, the display of your route guidance system will capture your visual and cognitive attention for some time. The same is true for changing the radio station or your music selection in the on-board entertainment system. If these distractions remain few, the driving safety is not affected notably. However, if more tasks are added, especially reading e-mails, retrieving background information about points of interest or communication with other people over the phone, driving safety will certainly suffer, if these systems do not change their way of interfacing with the user [68]. State-of-the-art in-car systems can be controlled via Automatic Speech Recognition (ASR), however robustness is still an issue here. Moreover the user is restricted to a well-defined grammar, and cannot say what he would like the system to do in his own words or even nonverbally (e.g., confirmation or rejection only by tone of voice).

In this section we list reasons and show various sources that indicate a demand for in-car Human-Machine Interfaces to become more human-like in the near future. As mentioned in the previous paragraph, there is primarily the issue of improving driving safety. Section 2.1, shows that this is of major importance and Section 2.2 demonstrates that various ways exist in which technological systems can directly improve safety by incorporating human factors and giving useful feedback. However, it has also been shown that driving safety can be increased indirectly by making driving a more pleasurable experience [19, 41, 42, 70]. Therefore, if the pleasure of the driver is increased and the interfaces are designed more human-like as reported by [68], safety will automatically benefit, even though in-car systems are getting

more complex and are able to offer things not primarily related to driving.

This is an important factor in competitive markets. Users' demands for more technology in the car are obvious. The literature indicates that people experience more driving pleasure in newer, safer, and more comfortable cars, as proven, for example, in [70].

In the following section the demand and feasibility of enhancing users' driving pleasure and enabling the user to be more productive, while still focussing on driving, is discussed in more detail. Section 2.3.1 will summarise existing work that deals with the effects of driving pleasure on driving performance and methods shown to enhance driving pleasure. Section 2.3.2 shows methods that enable the driver to be more productive without affecting the driver's focus on the road.

2.3.1. Enhancing Driving Pleasure. Users of in-car Human-Machine Interaction nowadays most often get frustrated, if a system does not understand their intentions and interface design is complex and nonintuitive. Numerous publications exist that try to improve the interfaces by optimising the amount of time users spend on data input and by restructuring menus in order to access items more quickly, for example, [71]. However, all of these approaches focus only on traditional input and output modalities: haptics/touch and speech for input and displays and speech synthesis for output. The speech-based systems take the first step in making the interaction easier by allowing the driver to operate the system in a hands-free mode. However, the dialogue structure usually follows a fixed set of rules the user has to adapt to [68]. If communication errors occur, due to faulty speech recognition, for example, the user can quickly become annoyed and stressed [68, 72]. Hearing a prerecorded response like "sorry, I did not understand you" over and over again, is likely to provoke anger.

More human-like systems should use multiple modalities (especially adding the visual channel) and thus be able to detect communication problems as quickly as possible, as pointed out by [72]. If the system shows awareness of communication problems and quickly offers alternative solutions and personalised help [73] or uses partial information to re-request specific details that were incorrectly recognised by the speech recognition unit, for example, general perceived reliability and robustness of the system improves. Literature suggests that generally, the more robust a system is, the better the user acceptance is [74]. User studies, such as [73] further suggest that if a driver needs help, the system should be very sensitive to her or his needs and preferences in order to solve the driver's problem as easy and quickly as possible.

A "socially competent" car in the role of a virtual companion can engage the driver into conversation and thus will give the driver the feeling of not being alone [75]. It can further give helpful assistance and driving hints, assuring the driver that there always is somebody to help. Such a form of communication is in strong contrast to the traditional and impersonal way of interfacing via menu structures or well defined dialogue schemes. According to traffic statistics in

the USA published by the US department of Transportation (<http://www.bts.gov/>) for the year 2001, the overall mean occupancy of passenger vehicles varies between 1.1 and 1.3, indicating a large amount of single drivers. Considering this fact, a virtual companion might be well appreciated by a large number of drivers. According to these reports, the mean occupancy is lowest for drives to work (1.14) and work-related trips (1.22). The overall mean occupancy for all observed trips not being much above the number for work-related trips, is an indication that a large amount of traffic is caused by commuters and work-related driving, where car-pooling often is not possible or not appropriate.

Another way to improve the driving experience is personalisation [76]. This trend can be observed in almost all other user interfaces, from full-blown computers to basic mobile phones. Practically all of these devices allow to change settings like the background image, color of the interface or configure favourite menu items. A "socially competent" car should be able to detect who is currently driving, judge emotion and behaviour based on past experiences and automatically adapt a conversation to the preferences of the current driver. This will give the driver the feeling that she or he is in control and the car adapts to her or his wishes and needs and thus also increases driving pleasure.

The car may also offer a good personalised music selection, for example. Music is known to improve mood by directly affecting physical brain processes [77]. Music thus contributes to the overall driving pleasure. However, care has to be exercised when selecting the style of music. Choosing too relaxing music may make the driver tired, choosing music that the driver does not like may annoy her or him, choosing happy music if the driver is sad is also likely to lead to opposite results. The user's personal preferences and the user's current mood have to be considered in such a decision.

Modern upper class vehicles fine tune the engine sound perceived by the driver using a considerable number of hidden speakers throughout the passenger cabin. In some situations the driver might be in a bad mood and bothered by the disturbing sound of her or his engine. An emotionally sensitive car could sense the driver's mood and adjust the engine (the motor of the car) sound (especially as perceived inside the car) based on good guesses or learnt preferences.

2.3.2. Enabling the Driver to be More Productive. Since time is precious, many drivers want to be able to use the time while driving to communicate with other people, access information like news and weather forecast or check reservations and bookings, for example. Today's in-car information systems in combination with mobile phones practically allow drivers to do all these tasks, however, most of these tasks cannot be done safely while driving. Interfaces are still designed in a traditional way like most other Human-Machine Interfaces, using a screen to display information combined with haptic input via buttons, knobs, and touch devices. Some systems use speech input in certain areas such as dialling of phone numbers. Yet, the driver has to spend some cognitive and visual effort on communicating with the system. He or she must learn to interact with the system—it is not the system

that learns how to interact with the user. The latter, however, should be the case in a user-friendly, human-like system [68].

Most users, especially elderly people or people with less practice in interacting with computers, will experience problems properly using an in-car driver interface and thus will require more time to access and use various features [78]. During driving they do not have the time to deal with the system, which leads to poorer acceptance of the system. One could imagine situations where there is a human codriver present in the car. Most of the time, the driver will certainly instruct his or her codriver to enter the new route into the route guidance system, check the weather forecast or call somebody via the cell-phone, for example, instead of doing these tasks on his or her own.

This is exactly where it becomes obvious that a “socially competent” virtual companion would be indeed very helpful for single drivers, especially those unfamiliar with computer interfaces. As many cars are occupied only by one driver, as pointed out in Section 2.3.1 this is an important issue. Communicating with the virtual companion like with a human codriver would increase user acceptance. In the literature the term virtual codriver (VICO) [67] is often used to refer to a human-like virtual companion. Of course a short period of time will be required to get used to the, at first, strange idea of an actually naturally talking car.

Besides being a helpful aid for single drivers, a “socially competent” virtual companion can also be of great help for drivers with passengers requiring special assistance. Children in the back seats can significantly distract the driver if they require too much of her or his attention. Such a virtual companion could detect such a situation and take some load off the driver by engaging the children in conversation or begin telling them stories or showing them cartoons, for example, via a rear seat entertainment system.

At this point it becomes most obvious that for a “socially competent” car it is also necessary to estimate the interest level of the conversation partner. If the entertainment system detects that the children are not interested in, for example, the film currently shown, it probably is time to change to something different in order to keep the children’s attention. Also, the driver should not be bored by noninteresting information.

In Section 3 we summarise use-cases like the above that could be handled by a “socially competent” virtual companion, like reading and writing e-mails and making reservations while driving.

3. Exemplary Use-Cases

In order to design human-machine communication in future upper class cars more naturally and intuitive, the incorporation of innovative applications of pattern recognition and machine learning into in-car dialogue interfaces becomes more and more important. As discussed in the previous sections, emotion recognition is an essential precondition to create a social competent car that can talk to the driver and provide a “virtual companion”. In this section we discuss specific use-cases for emotion related technology in the car

for both fields, namely the safety-related tasks of driver state monitoring and control of driver emotions, and the tasks related to enhancement of driving pleasure and productivity, such as multimodal and affect sensitive interfaces. We start our use-case overview by giving a brief summary of the state-of-the-art in in-car driver assistance and entertainment systems.

3.1. State-of-the-Art. While affect aware technology is missing in today’s automobiles due to the lack of user adaptable and autonomous, reliable technology, speech recognition has started to mature in the automobile market. The most obvious example is navigation systems where the destination selection can be performed via speech input. This speech recognition is based on templates which are stored during a training phase when the user adds a new destination and pronounces its name several times. More advanced systems, are based on subword modelling (phonemes) and include a universal acoustic model. They are thus able to recognise speech input without the need of recording several templates. Some minor voice adaptation might need to be performed in the same way as in modern dictation systems. These systems allow for a voice based command-like interface, where the user can change routes by command (“fast route”, “short route”), change the view, or have traffic information read out aloud. Entertainment systems can be controlled in a similar fashion by commands such as “change station”, “next song”, or even by pronouncing a song title or artist. Yet, these systems are restricted to a set of predefined commands and do not allow for flexible interaction. The user has to know the capabilities of the system, he has to know “what” he can say. Future systems, as proposed in the use-cases in the following sections, must be able to accept all input, filter out information they understand, associate it with available car functions, and “tell” the user what his options are.

3.2. Safety-Related Use Cases. For the safety-related tasks we present three different categories of use-cases, which are countersteering strategies, adaptation strategies, and communicating the driver’s emotional state (e.g., anger/rage, fatigue, and high workload, stress, or uncertainty) to other vehicles.

3.2.1. Countersteering Strategies. This category contains use-cases which aim to “countersteer” negative affective states in order to guide the driver into a happy or neutral state which is known to be best suited for safe driving [5, 18, 19], since most other emotions (anger, fatigue, stress, confusion, nervousness, sadness, etc.) negatively affect driving capabilities like goal generation, evaluation, decision-making, strategic planning, focus, and attention [9–11]. Depending on the context, different voice responses for angry drivers can be given, intending to encourage better driving, make appropriate alerts or calming down the driver. Further, a virtual codriver can react to detected sleepiness—which constitutes another dangerous driver state—by keeping the driver awake or bringing the vehicle to a safe halt in case of danger, if the traffic situation permits (e.g., stopping on a busy highway

is too dangerous, the car has to be directed towards the side lane before it is stopped). Possible measures against stress, confusion, nervousness, and sadness can also be addressed by the virtual assistant through intelligent dialogue strategies. Thereby the responses or actions of the intelligent car always depend on the amount of available contextual and background information regarding the reason for the specific affective state. Especially in situations like stress, the virtual codriver can actively help to reduce the driver's workload, for example, by offering intelligent solutions for tasks related to the on-board entertainment and communication system or temporarily disabling such functions if the traffic situations require the driver's full attention.

3.2.2. Adaptation Strategies. Adapting the personality of an automated in-car assistant to the mood of the driver can also be important. A badly synthesised voice or an overly friendly, notoriously the same voice is likely to annoy the driver which soon will lead to distraction. Therefore, as an important adaptation strategy, matching in-car voice with the driver's emotion is beneficial, as has been found in, for example, [41, 47]. Different parameter settings for the synthesis of emotional speech for different emotions need to be used, as given in [79–82], for example. Other use-cases related to adaptation are emotion dependent spoken language understanding and model adaptation for speech recognition engines. These techniques serve the purpose of improving the accuracy of the in-car speech recogniser, since an inaccurate system is also likely to annoy and distract the user, instead of assisting the driver.

3.2.3. Communicating the Driver's Emotional State. The third category consists of use-cases that describe how a driver's state can be communicated to others. Locating potentially dangerous drivers can aid the driver assistance systems in other vehicles to warn their drivers more timely. Methods of car-to-car communication for preventing road rage are developed by some automobile manufacturers, for example. Further applications include monitoring passengers—especially children—and other road users while driving, to reduce the driver's cognitive workload, logging the driver's emotion to derive statistics for research purposes, and automatically triggering emergency calls in case of accidents, severe pain or dangerous situations.

3.3. Driving Pleasure Related Use-Cases. Similar to the safety-related applications of in-car emotion recognition, the use-cases related to driving pleasure can also be grouped into three different categories: enabling of a mood adequate human-machine dialogue, adaptation of surroundings, and increasing productivity.

3.3.1. Mood Adequate Human-Machine Dialogue. Personalised and “socially competent” small-talk belongs to the first category and is a key feature of a “virtual companion”. Thereby emotion serves as contextual knowledge that indicates how the dialogue system has to interpret the output of the automatic speech recogniser (e.g., the use of irony may

depend on the user's emotional state, also there seems to be a reduced vocabulary in highly emotional speech, such as short angry commands or comments). Such dialogues do not only depend on the current words uttered by the user, but depend also on contextual information like time of day or weather. Similar use-cases are adaptive topic suggestion and switching, dialogue grounding, and reactions to nonlinguistic vocalisations like moaning or sneezing. Further, multimedia content analysis methods enable the car to deliver information from the internet which suits the current interest and affective state of the driver (e.g., love poems if the driver is in love, or only allowing happy news if the driver is in a happy state). Observing the driver's workload also enables the car to adapt the level of entertainment to the current traffic situation. Incoming and outgoing calls can be managed by a “phone guide” who takes into account the affective state of both the driver and the conversational partner. The latter can be determined from speech while the system converses with the caller (i.e., asking for the caller's identification and purpose/importance of his call) before putting him through to the driver.

3.3.2. Adaptation of Surroundings. Depending on the driver's mood, the in-car ambience can be adjusted. This can be done by automatic selection of mood adequate music, for example. Moreover, engine sound, ambient light, and air conditioning can be adapted according to the driver's affect.

3.3.3. Increasing Productivity. Finally, potential use-cases for a virtual codriver can be derived from the goal to increase the driver's productivity. Thereby calendar functions, handling of e-mails, internet access, and automatic translation are relevant as aspects that are likely to be welcomed by car buyers. However, the role affective computing takes in such technological advances is not fully researched, yet. Also increasing productivity on the other hand means higher workload for the driver, and thus reduced focus on the road leading to reduced safety. The aspect of increasing productivity thus should only be addressed if it can be ensured that these tasks do not in any major way keep the driver from his primary task of controlling the vehicle. This would be the case if the virtual codriver had a fully natural speech interface and the capability to robustly understand the driver's intentions from minimal input.

4. User Acceptance

It is important to assess acceptance and success of any new technology as soon as possible to determine whether efforts in developing the technology are well spent. Since it is a well known issue that too much technology might irritate or confuse users or make them feel observed, we address these issues in a user study designed for in-car affective computing. The basic idea is to set up a car with a simulated virtual codriver in a Wizard-of-Oz experiment. Users are asked to perform several tasks in the simulation while being assisted by the virtual codriver. The users' experience with the system is determined via multiple questionnaires which are filled out

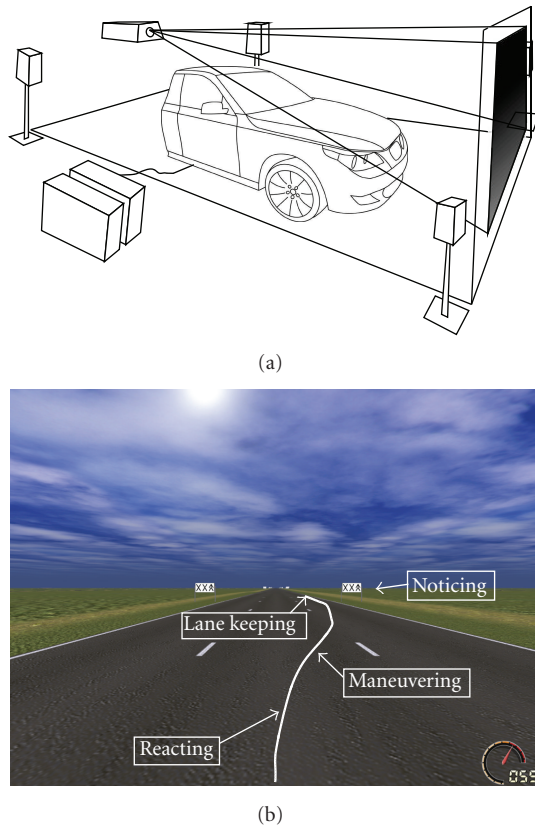


FIGURE 1: Driving simulator and simulation software. (a) Setup of the driving simulator. (b) Lane-Change-Task (the signs indicate that a lane change is to be performed).

after the experiment. The next section describes the setup and procedure of the Wizard-of-Oz (WoZ) experiment. Section 4.2 presents the findings of the survey.

4.1. Wizard-of-Oz Setup. In order to create a realistic driving scenario in a safe and controllable environment, a driving simulator was used. It consists of half the body of a real BMW 5 series vehicle in front of a large screen (see Figure 1(a)). The car's controls (i.e., accelerator and brake) and the steering wheel are used to control the simulation-software, which renders a 3D animation of a road and scenery on to the projected screen. The lab is equipped with two sound systems, one outside of the car to play back the engine and environment sounds, the other inside the car, which is used by the operator to give instructions to the subjects, to play music from the on-board entertainment system, and to output the voice of the virtual in-car assistant. The Lane Change Task [83] was used to simulate a primary driving task. Thereby, a test person has to drive a street of three lanes and switch to the lane, which is signalled by signs on the side of the test track (see Figure 1(b)). Additionally, a simple driver-information system is implemented, which can be controlled with a multifunctional device—the iDrive controller placed in the driving simulator. The following functions are implemented into this system:

- (i) Input of navigation destination.
- (ii) Switching between three alternative options of navigation routes.
- (iii) Number dialling on the car phone.
- (iv) Viewing and editing calendar entries.

In order to simulate as much use-case customised support by the system as possible, the supervisor was able to fully remotely control the driver information system. The virtual in-car assistant's voice is simulated by the Wizard-of-Oz operator. Therefore, the operator's voice is recorded by a microphone in the control room, and after applying on-line effects, simultaneously played back via the car's centre speaker. Instructions are given to the test persons on what task to perform next (it was made clear to the subjects that these instructions did not belong to the virtual driver assistance system). The instructions have been prerecorded to ensure the same test conditions for all subjects. The following instructions were used for the tasks described in the following paragraphs.

- (i) Drive straight with moderate velocity.
- (ii) Drive straight with high velocity.
- (iii) Enter *Schrobenhausen* as destination.
- (iv) Scan the calendar for today's appointments.
- (v) Call your office to inform them of your late arrival.

The experiment was a first-contact situation for the test subjects, and they did not receive instructions on the capabilities of the driver-information system. Subjects were asked to imagine that it was an ordinary Monday morning and they were starting their usual drive to work.

Welcome Dialogue. After taking a seat in the car, the driver is greeted by the car with a short dialogue. Thereby the user is asked whether he or she is driving the usual way to work, or he or she requires navigational assistance, and whether he or she would like to listen to music.

Driving Only. The driver is now asked to start driving a simulated test track with low speed to get used to the primary driving task in the simulator environment. This test track includes the Lane Change Task (see above). Next, the participant is asked to drive the track at a higher velocity, which induces a higher load due to the primary driving task. During this situation the following use-cases are simulated.

Detection and Countersteering High Workload. The operator instructs the subject to enter a navigation destination in parallel. Now, the system (in our WoZ case simulated by the operator) will detect decreased attentiveness in the primary task, ask the driver to pay more attention to his primary driving task via speech output, inactivate the display elements, and offer the user the option of speech-based input of the destination.

Assisting Confused Drivers. Next, a road congestion is simulated. The user is now instructed to inform his office of his delay via his cell phone. The dialling does not work, however, due to a simulated bad network. The reason is not immediately apparent to the user, who only realises that his call is not being connected. The system detects the induced confusion and offers to connect the call once the network is available again. The wizard was instructed to act once he recognised confusion because the user was hesitating or expressing his confusion verbally.

Obtaining Information from the Internet. The subject is now instructed to scan his calendar for appointments. An appointment in a distant city is scheduled for the next day. A comment that a hotel room must be reserved is attached to the appointment. If the subject does not ask the system for available hotels by himself, after a defined timeout the system will ask the user if hotel information is to be obtained from the internet now. The system will guide the user through a hotel reservation process.

Handling Incoming Calls. After finishing the hotel reservation, an incoming phone call is simulated. However, there is no way apparent to the user to answer the call. Again, the system will detect that the user is not answering the call and will ask for the reason while at the same time offering help, that is, to either ignore the call or to accept it.

Smalltalk. Now the system initiates a dialogue, where it comments on the driver’s busy day, and the bad weather, and asks the driver whether a different radio station would be preferred. Finally, an updated traffic report is received with the information that the congestion has not yet cleared. This report is automatically interpreted by the system, and the user is given the option to select three alternative routes from the system display, which will bring him directly to the location of his appointment, instead of his office.

Adapting to the Driver’s Behavior. All the use-cases described so far are fixed, and thus common for all subjects. In addition to these planned scenarios the operator was trained to react individually to the subjects responses and comments, adapt his output voice to the user’s state (thereby changing his tone of voice to match the user’s tone of voice in the current situation), and especially react to nonlinguistic behavior such as laughing, sighing, or hesitation, where it seems appropriate.

4.2. Evaluation and Results. After finishing the experiment, every test subject was asked to fill out a questionnaire, which consists of four parts: The System-Usability-Scale (SUS) [84], and the SEA-Scale (Subjectively Experienced Effort) [85] for rating specific scenarios, and the Attrak-Diff system (<http://www.attrakdiff.de/>), a questionnaire composed of a semantic differential, for rating the complete system. Since Attrak-Diff is a general system for rating product attractiveness, additionally a set of extra questions concerning our specific setup was used.

TABLE 1: Results for the System Usability Scale (SUS) and Subjectively Experienced Effort (SEA) scale for four selected tasks. SUS: maximum score is 100 (best usability), SEA: maximum score 120 (highest workload: worst), except for results marked with *, where the maximum score of 120 indicates the maximum perceived decrease in workload (thus, 120 is best).

Scenario	SUS	SEA
	[0–100]	[0–120]
Intelligent virtual agent support in stress situations	72.2	67.8*
Assisting confused drivers	76.0	66.9*
Smalltalk with virtual agent	76.2	22.2
Adaption of agent speech to the driver’s emotional state	70.0	24.3

4.2.1. Description of Participants. Thirteen subjects (twelve male and one female) took part in the experiment. The average age is 27.8 years with a standard deviation of 3.1 years. All of them had a driver’s license and were interested in new and innovative technical products. The average yearly mileage of each subject is approximately 14 000 kilometers.

4.2.2. System Usability and Subjectively Experienced Effort Scales. The analysis of the System Usability Scale (SUS) was performed with the method proposed in [84]. For each use-case, a total score was determined. This score reflects the user’s impression of the system for the respective scenario. The maximum assignable score is 100, which corresponds to a completely positive evaluation of the system.

Table 1 shows results for the SUS and SEA scales for four selected tasks. Considering the early prototypical stage of the test system (i.e., with respect to look and feel, and range of functionality), the obtained SUS scores are a promising basis for further system improvements, since ratings above 50 are generally in favour of the system in question. The best scores were obtained for assisting confused subjects and smalltalk with the in-car agent. While the first is to be expected, it is not quite obvious that smalltalk does enhance the system’s usability feeling.

The SEA scale describes the subjectively experienced workload for each particular scenario. A high score (maximum value 120) indicates a high perceived workload. Thus, lower values indicate better performance with respect to reducing the driver’s workload and keeping his or her focus on the road. For the first two scenarios, “stress”, and “confusion”, however, a modified scale was used, where a high value (again maximum of 120) indicates the subjectively perceived decrease of workload. The result can also be found in Table 1.

Concluding, every scenario is evaluated positively on average. Both the SUS and the SEA scale show good results regarding the use of the system in spite of the prototypical system setup. The subjectively perceived workload decreased noticeably, if the car gave support to the test person (“stress”, and “confusion” scenarios on the SEA scale). This is a good basis for further development of such driver state-aware functionalities.

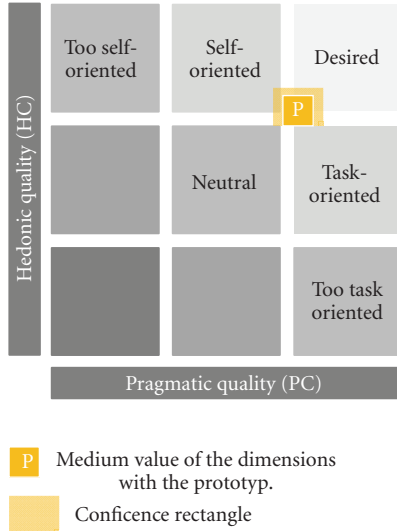


FIGURE 2: Results Attrakdiff-Portfolio Representation.

4.2.3. *Attrak-Diff Rating*. With Attrak-Diff a product is evaluated with respect to the following four dimensions.

- (i) Pragmatic Quality (PQ): usability of the product.
- (ii) Hedonic Quality-Stimulation (HQ-S): support of needs in terms of novel interesting and stimulating functions.
- (iii) Hedonic Quality-Identity (HQ-I): identification with the product.
- (iv) Attractiveness (ATT): global value of the product based on the quality perception.

The so-called portfolio representation (the result in the 2-D space spanned by PQ and HQ) determines in which character-zone the product can be classified. For this study, the Attrak-Diff was evaluated for the entire system with all its new ideas and concepts. Individual features were not evaluated separately. The resulting portfolio presentation is shown in Figure 2.

The system is rated as “rather desired”. However, the classification is neither clearly “pragmatic” nor “hedonic”, because the confidence interval overlaps into other character-zones. So there is room for improvement in terms of usability (PQ and HQ). The small confidence rectangle indicates a high agreement among the test subjects.

4.2.4. *System Specific Questionnaire*. A questionnaire composed of eleven questions using a five point scale was used (“Strongly Agree” (value = 1) to “Strongly Disagree” (value = 5)) for the custom evaluation of the system as a whole. For each question, the mean value of the ratings and the standard deviation (σ) were calculated. The results are summarised in Table 2 and briefly discussed in the following paragraphs.

Nearly every test one thinks that a talking car is reasonable (mean 1.4, $\sigma = 0.4$) and feels rather not observed (mean 3.8, $\sigma = 1.1$) by the car. The question, if the car was disturbing, was evaluated as “moderate” with a light trend to

TABLE 2: Results of system specific questionnaire composed of eleven questions. Mean (μ) and standard deviation (σ) of ratings on a five point scale (“Strongly Agree” (value = 1) to “Strongly Disagree” (value = 5)).

Question	μ	σ
A talking car is reasonable	1.4	0.4
I feel observed by a talking car	3.8	1.1
I feel disturbed by a talking car	3.5	0.3
I would rely on suggestions given by the car	2.7	0.8
I feel the car helps to handle difficult driving situations	1.7	0.7
A car should react to my emotion	3.2	1.1
Starting self-initiated dialogs is desired	2.8	1.3
Automatic evaluation of my stress level and emotional state is desired	2.8	1.6
It is helpful if the car can support me in confusing situations	1.3	0.2
I like the ability to request information from the internet via natural speech input	1.5	0.8
It should be possible to mute the car’s voice	1.3	0.4

“rather not” (mean 3.5, $\sigma = 0.3$). The test persons would rely with a mean value of 2.7 ($\sigma = 0.8$) on suggestions given by the car. The question, whether the users felt the car would help them to handle difficult driving situations more easily, gave a clear positive result, with a mean value of 1.7 ($\sigma = 0.7$).

Unclear are the results of the questions, whether a car should react to the driver’s emotion (mean 3.2, $\sigma = 1.1$) or should determine the stress-state of the driver (mean 2.8, $\sigma = 1.6$), and whether the car should start talking initiated by itself (mean 2.8, $\sigma = 1.3$). The high standard deviations of the answers to these questions indicate that the individual subjects do have quite clear preferences, thus no unifying conclusion for all users can be drawn. Likewise, the recommendation based on this study would be to provide easy ways to disable such functionalities or have them disabled by default and let the users decide to enable them.

The last four questions show all positive results. The test subjects agreed on the fact that it would help, if the car was able to support the driver in confusing situations (mean 1.3, $\sigma = 0.2$), which is in line with the SUS and SEA scale evaluations. Moreover, they liked the car’s ability to request information from the internet via natural speech input (mean 1.5, $\sigma = 0.8$).

Overall, the test persons stated that a talking car—as simulated via the Wizard-of-Oz—makes sense, and they do not feel observed or disturbed. This is an indicator for a good acceptance of such a product.

However, the driver wants to be the master of the situation and makes her or his own decisions, because not all test persons would rely on the car’s suggestions, and high standard deviations were observed for the driver state monitoring questions. Virtually all subjects wish to have a functionality, which allows the user to mute the car’s voice (mean 1.3, $\sigma = 0.4$). From this point of view, the evaluation of the smalltalk-feature gets more comprehensible. Some

subjects commented that this feature would be disturbing, if it happened too many times.

The unclear results regarding the recognition of emotions and stress-states may also relate to this as well as the fact that these functions could not be implemented consistently enough in this experiment (Wizard-of-Oz). Such a consistent evaluation would require ways of reliably and reproducibly inducing emotional and stress-states, which is a very difficult task that can never be performed perfectly. Thus a very large number of subjects is required for these evaluations.

5. Driver Distraction Detection

Driver inattention is one of the major factors in traffic accidents. The US National Highway Traffic Safety Administration estimates that in 25% of all crashes some form of inattention is involved [86]. Distraction (besides drowsiness) as one form of driver inattention may be characterised as: “any activity that takes a driver’s attention away from the task of driving” [87].

In this section we show how reliably driver distraction can be detected using adequate machine learning techniques. The motivation for detecting whether a driver is distracted or not could be adaptive driver assistant systems, for example, lane keeping assistance systems. These systems track the lane markings in front of the vehicle and compute the time until the vehicle will cross the marking. If the driver does not show an intended lane change by using the indicator to signal the change, the systems will use directed steering torques on the steering wheel to guide the car to the middle of the lane.

One problem with lane keeping assistance systems is that they can be annoying in some circumstances [88] since they do not yet respond to the driver’s state or her or his intent but to lane markings and the car’s speed. If it was possible to recognise a driver’s state reliably, the system would give just as much assistance as the driver needed. This would allow for a greater safety margin without annoying the driver with false alarms in normal driving situations.

Our system for online driver distraction detection is based on modeling long-range contextual information in driving and head tracking data. It applies Long Short-Term Memory (LSTM) recurrent neural networks [89, 90] which are able to capture the temporal evolution of low-level data sequences via so-called *memory blocks*. Long Short-Term Memory networks have shown excellent performance in a variety of pattern recognition tasks including emotion recognition from speech [91].

5.1. Database. In order to train and evaluate our system we used data that was recorded during an experiment in which drivers had to fulfil certain “distracting” tasks while driving. The resulting database consists of 32 participants (13 female and 19 male). The car (an Audi A6) was equipped with the “Audi Multimedia System” and an interface to measure CAN-Bus data. Additionally, a head tracking system was installed, which was able to measure head position and head rotation. Head-tracking systems are not common in vehicles today, but the promising research in such cameras for driver

state detection will lead to a higher installation rate in serial cars in the near future. So we decided to use head-tracking information in our approach as well.

Eight typical tasks (performed haptically) on the Multimedia Interface were chosen as distraction conditions:

- (i) adjusting the radio sound settings,
- (ii) skipping to a specific song,
- (iii) searching for a name in the phone book,
- (iv) searching for a nearby gas station,
- (v) dialling a specific phone number,
- (vi) entering a city in the navigation device,
- (vii) switching the TV mode,
- (viii) adjusting the volume of navigation announcements.

The procedure for the experiment was as follows: after a training to become familiar with the car, each participant drove down the same road eight times while performing secondary tasks on the in-vehicle information system. On another two runs the drivers had to drive down the road with full attention on the roadway. In order to account for sequential effects, the order in which the conditions were presented was randomised for each participant. Overall, 53 runs while driving attentively and 314 runs while the drivers were distracted could be measured. The “attentive” runs lasted 3 134.6 seconds altogether, while 9 145.8 seconds of “distracted” driving were logged (see Table 3 for experimental conditions).

An analysis of the influence on lane keeping of the different in-vehicle information system tasks [92] confirmed the tasks to be distracting. Thus, all these tasks were labeled as “distracted” compared to driving down the road with full attention (ground truth: “attentive”). Thereby we labeled runs during which a task had to be completed as *completely* “distracted” since the drivers were engaged with the task during the complete run.

Six signals were chosen for a first analysis:

- (i) steering wheel angle,
- (ii) throttle position,
- (iii) speed,
- (iv) heading angle,
- (v) lateral deviation,
- (vi) head rotation.

Steering wheel angle, throttle position, and speed are direct indicators of the driver behavior. Many studies prove the fact that visually distracted drivers steer their car in a different way than do attentive drivers. The same applies for throttle use and speed (an overview can be found in [93]). The car’s heading angle and its lateral deviation in the lane rely on the amount of attention the driver is allocating to the roadway and may hence give useful information about distraction. Head rotation of the driver is an indicator of the driver’s visual focus. While using the Multimedia Interface, which is located in the middle console just below

TABLE 3: Experimental conditions for driving data collection.

Experimental conditions	
Num. participants	32 (13 f, 19 m)
Age of participants	29 to 59
Driving experience	>10,000 km per year
Car	Audi A6 quattro 2.6 TDI
Road	Ayinger Str.
Num. “attentive” runs	53
Num. “distracted” runs	314

the dashboard, the main rotation of the head is to the right. So the head rotation is the most promising indicator of the head-tracking signals. Note that a trivial way of determining driver distraction due to the operation of the Multimedia Interface would be to simply detect, for example, the touching of the Multimedia Interface buttons. However, we decided to use signals that serve as general indicators of driver distraction in order to be able to also detect distraction which is not caused by the operation of the Multimedia Interface.

5.2. Experiments and Results. The database collected as described above was split into a training, a validation, and a test set. For training we randomly chose 21 drivers. The validation set consists of three randomly chosen drivers, while the system was evaluated on the remaining eight drivers. Thus, our evaluations are completely driver independent, that is, the results indicate the performance of the system for a driver which is not known to the system (the system was not optimised for a specific driver’s style). The training set consists of 35 baseline runs (i.e., runs during which the driver was attentive) and 146 runs during which the driver was distracted. The test set contains 13 baseline and 51 “distracted” runs.

We evaluated the performance for different numbers of memory blocks (70 to 150) in the hidden layer of the LSTM neural network. The number of memory blocks is correlated to the complexity of the network, that is, the number of parameters which are used to describe the relation between inputs and outputs (see, e.g., [94] for a detailed description of the LSTM memory block principle).

Table 4 shows the results for *sample-wise* classification (i.e., quasi-time-continuous prediction every 10 ms) of driver distraction using the two classes “attentive” (baseline runs) and “distracted” (runs during which the driver was involved in a task at the Multimedia System). A total of 286 000 such samples (frames) is contained in the test set. The best F1-measure could be achieved with an LSTM network consisting of 110 memory blocks. Note that due to the imbalance in the class distribution, the F1-measure is a more adequate performance measure than accuracy. Thereby F1-measure is the harmonic mean of unweighted recall and unweighted precision. For the two-class problem, LSTM networks achieve an F1-measure of up to 88.7%. In Table 5 the classification of *complete runs* is evaluated by averaging the sample-wise LSTM predictions over an entire run. With

the best LSTM configuration, an accuracy of 92.9% can be obtained.

By analysing the obtainable classification performance when using only single signals, we can get an impression of the relevance of the individual data streams. The best “single stream” performance can be obtained when using exclusively head rotation, followed by exclusive usage of steering wheel angle, heading angle, throttle position, speed, and lateral deviation, respectively.

Trying to get an impression of the accuracy of distraction detection when driver distraction is not caused by the Multimedia Interface, we tested the system on data that was recorded while the driver had to fulfil tasks like eating a chocolate bar or reading a letter. We found that the obtained F1-measure is only slightly worse for this scenario (83.2%).

Tables 4 and 5 reveal that driver distraction can be detected with relatively high reliability by modeling the temporal evolution of driving and head tracking data. Thus, an adaption of lane-keeping assistance systems which is based on sensor data already available in modern vehicles seems to be a viable and promising approach.

6. Conclusion and Outlook

Summarising all aspects discussed in the past sections, it becomes clear that emotions will be a key issue not only in general oncoming human-computer interaction, but also in the in-car communication.

As we have discussed, emotions affect many cognitive processes, highly relevant to driving, such as categorisation, goal generation, evaluation and decision-making, focus and attention, motivation and performance, intention, communication and learning. There is a need for controlling the driver’s emotional state: the high relevance of an emotionally high valence was documented by a substantial body of literature—“happy drivers are the better drivers”. This control of the emotional state will thus ensure a safer and more pleasant driving experience. At the same time too high arousal may lead to aggressive driving behaviour. For optimal driving performance, a compromise between too high and too low arousal must therefore be found.

Apart from externally induced states of intoxication (alcohol, drugs, medication) or pain, we had found anger, aggressiveness, fatigue, stress, confusion, nervousness, sadness, and boredom as main negative emotions and mental driver states of interest, and happiness as positive factor.

As basic strategies to control emotion, countersteering emotions was found next to adapting car functionalities to driver emotion. The in-car driver interface can thereby influence users’ emotional states in several ways. To provide only few examples, angry drivers could be calmed down and could be made aware of their state, fatigued drivers could be stopped from falling asleep by engagement in a discussion with control of potential boredom for topic-switching, and confused drivers could be offered assistance regarding the current traffic situation.

The growing complexity of in-car electronics demands for new interfaces that do not disturb the drivers’ focus on

TABLE 4: *Sample-wise* classification of driver distraction (attentive vs. distracted) using LSTM networks: accuracy, unweighted recall, unweighted precision, and F1-measure for different numbers of memory blocks.

# Memory blocks	Accuracy (%)	Recall (%)	Precision (%)	F ₁ (%)
70	87.2	78.8	85.6	82.1
90	89.0	82.6	87.0	84.8
100	91.1	88.7	88.0	88.4
110	91.3	89.5	87.9	88.7
130	89.0	83.3	86.4	84.9
150	86.2	77.2	84.2	80.6

TABLE 5: Classification of driver distraction (attentive vs. distracted) for *complete runs* using LSTM networks: correctly classified runs, baseline runs correctly classified as “attentive”, runs with task correctly classified as “distracted” (in percent) for different numbers of memory blocks.

# Memory blocks	Correctly class. runs (%)	Accuracy baseline runs (%)	Accuracy runs with task (%)
70	88.4	62.6	98.0
90	90.9	72.3	97.8
100	92.9	84.5	96.0
110	92.9	86.7	95.2
130	90.7	74.1	96.9
150	87.8	61.8	97.6

the road or annoy the driver because they are so difficult to use. Natural, human-like interfaces that quickly and tolerantly comprehend drivers’ intentions are the key. In Section 4 we evaluated an intelligent driver assistance system, with which users were able to communicate naturally via speech. The evaluation suggests that such a system will generally be accepted by users, as long as they have full control over the system and can mute the system at any time. The driver will expectantly feel more comfortable and safe in such a car because he or she does not need to worry about not knowing how to use the system. The car can also serve as virtual codriver for single drivers, engaging the drivers in conversation and making them feel like having company and not being alone. Further possibilities of increasing driving pleasure are to offer personal settings, personalised conversation (greetings, small talk, etc.) and personalised in-car entertainment and environment customisation. Drivers simply prefer cars where they experience greater pleasure while driving and will therefore likely want to have “socially competent” interfaces in their cars. Further, drivers in the future are expected even more to use the time while driving productively, for example, listen to e-mails (with speech synthesis), make reservations or obtain information about the destination. In order to not interfere with the main task of driving, the driver interface must be operable in hands-free mode and quickly understand the user’s intentions, without the user having to utter predefined commands. In this respect, future cars have to become more “socially competent”, that is, be able to better understand their drivers’ intentions adding the increasingly mandatory intelligent interpretation of multiple modalities such as speech, face

and driving behaviour by incorporation of judgement of emotional and affective states.

As an example for the feasibility of driver state recognition, we presented an automated system for detection of driver distraction, which can be implemented in a car with the technology available today. Using Long Short-Term Memory recurrent neural nets, it is possible to continuously predict the driver’s state based on driving and head tracking data. The strategy is able to detect inattention with an accuracy of up to 91.3% not dependt of the driver, and can be seen as a basis for adaptive lane-keeping assistance.

The presented paper shows the need, the acceptance, the feasibility and doability of intelligent and affective in-car interfaces. Yet, substantially more work is required to develop products which can be manufactured in series and which are robust enough for the end-user market. In this respect, more usability studies with a broader range of users in even more realistic driving situations (e.g., “out in the wild”) are required. Further, implementations of actual prototype systems—instead of the presented Wizard-of-Oz approach—must be built and evaluated by drivers under realistic conditions. Therefore, before implementing such prototypes, more evaluations of, for example, the vocal and visual modalities are required with respect to robustness in the in-car environment and user acceptance.

Naturally, people talk, they talk different from today’s command and control-oriented and in the near future oncoming rudimentary natural language-based in-car interaction, and engineers will have to listen [95]. At the same time, engines might soon observe our affective behaviour patterns—for our safety, comfort, and pleasure.

References

- [1] PROSPER, "Prosper final report, project for research on speed adaptation policies on european roads," Tech. Rep. Project no. GRD2200030217, May 2006.
- [2] K. Takeda, H. Erdogan, J. H. L. Hansen, and H. Abut, Eds., *In-Vehicle Corpus and Signal Processing for Driver Behavior*, Springer, Berlin, Germany, 2009.
- [3] H. Bubb, "Fahrerassistenz primär ein Beitrag zum Komfort oder für die Sicherheit?" *VDI Berichte*, no. 1768, pp. 25–44, 2003.
- [4] A. Batliner, F. Burkhardt, M. van Ballegooy, and E. Nöth, "A taxonomy of applications that utilize emotional awareness," in *Proceedings of the 1st International Language Technologies Conference (IS-LTC '06)*, pp. 246–250, Ljubljana, Slovenia, 2006.
- [5] C. M. Jones and I. M. Jonsson, "Automatic recognition of affective cues in the speech of car drivers to allow appropriate responses," in *Proceedings of the 17th Australia Conference on Computer-Human Interaction: Citizens Online (OZCHI '05)*, vol. 122, pp. 1–10, Canberra, Australia, 2005.
- [6] G. H. Bower, "Mood and memory," *American Psychologist*, vol. 36, no. 2, pp. 129–148, 1981.
- [7] J. A. Singer and P. Salovey, "Mood and memory: evaluating the network theory of affect," *Clinical Psychology Review*, vol. 8, no. 2, pp. 211–251, 1988.
- [8] R. B. Zajonc, "On the primacy of affect," *American Psychologist*, vol. 39, no. 2, pp. 117–123, 1984.
- [9] A. R. Damasio, *Descartes' Error: Emotion, Reason and the Human Brain*, Avon Books, New York, NY, USA, 1994.
- [10] J. E. LeDoux, "Brain mechanisms of emotion and emotional learning," *Current Opinion in Neurobiology*, vol. 2, no. 2, pp. 191–197, 1992.
- [11] D. Derryberry and D. M. Tucker, "Neural mechanisms of emotion," *Journal of Consulting and Clinical Psychology*, vol. 60, no. 3, pp. 329–338, 1992.
- [12] J. A. Colquitt, J. A. LePine, and R. A. Noe, "Toward an integrative theory of training motivation: a meta-analytic path analysis of 20 years of research," *Journal of Applied Psychology*, vol. 85, no. 5, pp. 678–707, 2000.
- [13] N. Frijda, *The Emotions*, Cambridge University Press, New York, NY, USA, 1986.
- [14] R. Birdwhistle, *Kinesics and Context: Essays on Body Motion and Communication*, University of Pennsylvania Press, Philadelphia, Pa, USA, 1970.
- [15] P. Ekman and W. V. Friesen, *Unmasking the Face: A Guide to Recognizing Emotions from Facial Expressions*, Prentice Hall, Englewood Cliffs, NJ, USA, 1975.
- [16] N. Chovil, "Discourse-oriented facial displays in conversation," *Research on Language and Social Interaction*, vol. 25, pp. 163–194, 1991.
- [17] D. Goleman, *Emotional Intelligence*, Bantam Books, New York, NY, USA, 1995.
- [18] J. A. Groeger, *Understanding Driving: Applying Cognitive Psychology to a Complex Everyday Task*, Frontiers of Cognitive Science, Routledge Chapman & Hall, 2000.
- [19] M. Grimm, K. Kroschel, H. Harris, C. Nass, B. Schuller, G. Rigoll, and T. Moosmayr, "On the necessity and feasibility of detecting a driver's emotional state while driving," in *Proceedings of the 2nd International Conference on Affective Computing and Intelligent Interaction*, vol. 4738 of *Lecture Notes in Computer Science*, pp. 126–138, Springer, Lisbon, Portugal, 2007.
- [20] L. James and D. Nahl, *Road Rage and Agressive Driving*, Prometheus Books, Amherst, NY, USA, 2000.
- [21] E. Wells-Parker, J. Ceminsky, V. Hallberg, R. W. Snow, G. Dunaway, S. Guiling, M. Williams, and B. Anderson, "An exploratory study of the relationship between road rage and crash experience in a representative sample of US drivers," *Accident Analysis and Prevention*, vol. 34, no. 3, pp. 271–278, 2002.
- [22] H. Cai, Y. Lin, and R. R. Mourant, "Study on driver emotion in driver-vehicle-environment systems using multiple networked driving simulators," in *Proceedings of the Driving Simulation Conference*, May 2007.
- [23] D. A. Hennessy and D. L. Wiesenthal, "The relationship between traffic congestion, driver stress and direct versus indirect coping behaviours," *Ergonomics*, vol. 40, no. 3, pp. 348–361, 1997.
- [24] J. L. Deffenbacher, E. R. Oetting, and R. S. Lynch, "Development of a driving anger scale," *Psychological Reports*, vol. 74, no. 1, pp. 83–91, 1994.
- [25] R. M. Yerkes and J. D. Dodson, "The relation of strength of stimulus to rapidity of habit-formation," *Journal of Comparative Neurology and Psychology*, vol. 18, pp. 459–482, 1908.
- [26] J. M. Lyznicki, T. C. Doege, R. M. Davis, and M. A. Williams, "Sleepiness, driving, and motor vehicle crashes," *Journal of the American Medical Association*, vol. 279, no. 23, pp. 1908–1913, 1998.
- [27] E. R. Braver, C. W. Preusser, D. F. Preusser, H. M. Baum, R. Beilock, and R. Ulmer, "Long hours and fatigue: a survey of tractor-trailer drivers," *Journal of Public Health Policy*, vol. 13, no. 3, pp. 341–366, 1992.
- [28] M. T. Corfitsen, "Tiredness and visual reaction time among young male nighttime drivers: a roadside survey," *Accident Analysis and Prevention*, vol. 26, no. 5, pp. 617–624, 1994.
- [29] C. L. Marcus and G. M. Loughlin, "Effect of sleep deprivation on driving safety in housestaff," *Sleep*, vol. 19, no. 10, pp. 763–766, 1996.
- [30] A. T. McCartt, S. A. Ribner, A. I. Pack, and M. C. Hammer, "The scope and nature of the drowsy driving problem in New York state," *Accident Analysis and Prevention*, vol. 28, no. 4, pp. 511–517, 1996.
- [31] G. Matthews, L. Dorn, T. W. Hoyes, D. R. Davies, A. I. Glendon, and R. G. Taylor, "Driver stress and performance on a driving simulator," *Human Factors*, vol. 40, no. 1, pp. 136–149, 1998.
- [32] M. E. Rakauskas, L. J. Gugerty, and N. J. Ward, "Effects of naturalistic cell phone conversations on driving performance," *Journal of Safety Research*, vol. 35, no. 4, pp. 453–464, 2004.
- [33] R. Banuls and L. Montoro, "Motivational and emotional aspects involved in driving," in *Traffic Psychology Today*, P.-E. Barjone, Ed., chapter 8, pp. 138–318, Springer, Berlin, Germany, 2001.
- [34] K. Ball and G. Rebok, "Evaluating the driving ability of older adults," *Journal of Applied Gerontology*, vol. 13, no. 1, pp. 20–38, 1994.
- [35] S. MacDonald, R. Mann, M. Chipman, et al., "Driving behaviour under the influence of cannabis and cocaine," *Traffic Injury Prevention*, vol. 9, no. 3, pp. 190–194, 2008.
- [36] X. Li and Q. Ji, "Active affective state detection and user assistance with dynamic bayesian networks," *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, vol. 35, no. 1, pp. 93–105, 2005.
- [37] C. S. Dula and E. S. Geller, "Risky, aggressive, or emotional driving: addressing the need for consistent communication in

- research,” *Journal of Safety Research*, vol. 34, no. 5, pp. 559–566, 2003.
- [38] G. Stecklov and J. R. Goldstein, “Terror attacks influence driving behavior in Israel,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 40, pp. 14551–14556, 2004.
- [39] J. Healey and R. Picard, “Smart car: detecting driver stress,” in *Proceedings of the 15th International Conference on Pattern Recognition (ICPR '00)*, vol. 4, pp. 218–221, 2000.
- [40] M. Grimm, K. Kroschel, B. Schuller, G. Rigoll, and T. Moosmayr, “Acoustic emotion recognition in car environment using a 3d emotion space approach,” in *Proceedings of the 33rd Deutsche Jahrestagung für Akustik (DAGA '07)*, pp. 313–314, DEGA, Stuttgart, Germany, March 2007.
- [41] I. M. Jonsson, C. Nass, H. Harris, and L. Takayama, “Matching in-car voice with drivers state: impact on attitude and driving performance,” in *Proceedings of the 3rd International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, pp. 173–181, Rockport, Me, USA, June 2005.
- [42] C. Nass, I.-M. Jonsson, H. Harris, et al., “Improving automotive safety by pairing driver emotion and car voice emotion,” in *Proceedings of the Conference on Human Factors in Computing Systems (CHI '05)*, pp. 1973–1976, ACM, Portland, Ore, USA, 2005.
- [43] Y. Uchiyama, S. Kojima, T. Hongo, R. Terashima, and T. Wakita, “Voice information system that adapts to driver’s mental workload,” *R&D Review of Toyota CRDL*, vol. 39, no. 1, pp. 16–22, 2004.
- [44] S. Hoch, F. Althoff, G. McGlaun, and G. Rigoll, “Bimodal fusion of emotional data in an automotive environment,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '05)*, vol. 2, pp. 1085–1088, Philadelphia, Pa, USA, March 2005.
- [45] C. M. Jones and I. M. Jonsson, “Performance analysis of acoustic emotion recognition for in-car conversational interfaces,” in *Universal Access in Human-Computer Interaction. Ambient Interaction, Part II*, vol. 4555 of *Lecture Notes in Computer Science*, pp. 411–420, Springer, Berlin, Germany, 2007.
- [46] I. M. Jonsson, M. Zajicek, H. Harris, and C. Nass, “Thank you, i did not see that: in-car speech based information systems for older adults,” in *Proceedings of the Conference on Human Factors in Computing Systems (CHI '05)*, pp. 1953–1956, Portland, Ore, USA, 2005.
- [47] I. M. Jonsson, C. Nass, J. Endo, et al., “Don’t blame me i am only the driver: impact of blame attribution on attitudes and attention to driving task,” in *Proceedings of the Conference on Human Factors in Computing Systems (CHI '04)*, pp. 1219–1222, Vienna, Austria, 2004.
- [48] B. Schuller, J. Stadermann, and G. Rigoll, “Affect-robust speech recognition by dynamic emotional adaptation,” in *Proceedings of the International Conference on Speech Prosody, ISCA, Dresden, Germany, 2006*.
- [49] S. Furui, “Robust methods in automatic speech recognition and understanding,” in *Proceedings of European Conference on Speech Communication and Technology (EUROSPEECH '03)*, pp. 1993–1998, ISCA, Geneva, Switzerland, September 2003.
- [50] T. Athanasis, S. Bakamidis, and I. Dologlou, “Automatic recognition of emotionally coloured speech,” in *Proceedings of the Transactions on Engineering, Computing and Technology (ICCS '06)*, pp. 274–277, Vienna, Austria, March 2006.
- [51] Z. Zeng, J. Tu, M. Liu, T. S. Huang, B. Pianfetti, D. Roth, and S. Levinson, “Audio-visual affect recognition,” *IEEE Transactions on Multimedia*, vol. 9, no. 2, pp. 424–428, 2007.
- [52] R. Fernandez, *A computational model for the automatic recognition of affect in speech*, Ph.D. thesis, MIT Media Arts and Science, 2004.
- [53] B. Schuller, D. Seppi, A. Batliner, A. Maier, and S. Steidl, “Towards more reality in the recognition of emotional speech,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '07)*, vol. 4, pp. 941–944, Honolulu, Hawaii, USA, 2007.
- [54] A. Batliner, S. Steidl, C. Hacker, and E. Nöth, “Private emotions versus social interaction—a data-driven approach towards analysing emotion in speech,” *User Modelling and User-Adapted Interaction*, vol. 18, no. 1-2, pp. 175–206, 2008.
- [55] S. D. Kollias and K. Karpouzis, “Multimodal emotion recognition and expressivity analysis,” in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '05)*, pp. 779–783, Amsterdam, The Netherlands.
- [56] G. Caridakis, G. Castellano, L. Kessous, A. Raouzaoui, L. Malatesta, S. Asteriadis, and K. Karpouzis, “Multimodal emotion recognition from expressive faces, body gestures and speech,” in *Proceedings the 4th IFIP International Conference on Artificial Intelligence Applications and Innovations (AIAI '07)*, vol. 247 of *IFIP International Federation for Information Processing*, pp. 375–388, 2007.
- [57] J. Kim and E. André, “Emotion recognition using physiological and speech signal in short-term observation,” in *Proceedings of the International Tutorial and Research Workshop on Perception and Interactive Technologies (PIT '06)*, vol. 4021 of *Lecture Notes in Computer Science*, pp. 53–64, Kloster Irsee, Germany, June 2006.
- [58] F. Hönig, J. Wagner, A. Batliner, and E. Nöth, “Classification of user states with physiological signals: on-line generic features vs. specialized,” in *Proceedings of the 17th European Signal Processing Conference (EUSIPCO '09)*, B. Stewart and S. Weiss, Eds., vol. 15, pp. 2357–2316, Glasgow, UK, 2009.
- [59] E. André and J. Y. Chai, “Workshop: eye gaze in intelligent human machine interaction,” in *Proceedings of the International Conference on Intelligent User Interfaces*, pp. 431–432, Hong Kong, 2010.
- [60] F. Eyben, M. Wöllmer, A. Graves, B. Schuller, E. Douglas-Cowie, and R. Cowie, “On-line emotion recognition in a 3-D activation-valence-time continuum using acoustic and linguistic cues,” *Journal on Multimodal User Interfaces*, vol. 3, no. 1, pp. 7–19, 2010.
- [61] S. Steidl, B. Schuller, A. Batliner, and D. Seppi, “The hinterland of emotions: facing the open-microphone challenge,” in *Proceedings of the 4th International HUMAINE Association Conference on Affective Computing and Intelligent Interaction (ACII '09)*, vol. 1, pp. 690–697, IEEE, Amsterdam, The Netherlands, 2009.
- [62] C. Busso, Z. Deng, S. Yildirim, M. Bulut, C. M. Lee, A. Kazemzadeh, S. Lee, U. Neumann, and S. Narayanan, “Analysis of emotion recognition using facial expressions, speech and multimodal information,” in *Proceedings of the 6th International Conference on Multimodal Interfaces*, pp. 205–211, State College, Pa, USA, October 2004.
- [63] I. Cohen, G. Ashutosh, and T. S. Huang, “Emotion recognition from facial expressions using multilevel hmm,” in *Neural Information Processing Systems*, 2000.
- [64] B. Schuller, R. Müller, F. Eyben, et al., “Being bored? Recognising natural interest by extensive audiovisual integration for real-life application,” *Image and Vision Computing*, vol. 27, no. 12, pp. 1760–1774, 2009.
- [65] C. D. Katsis, N. Katertsidis, G. Ganiatsas, and D. I. Fotiadis, “Toward emotion recognition in car-racing drivers: a biosignal

- processing approach,” *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, vol. 38, no. 3, pp. 502–512, 2008.
- [66] T. Paul-Stueve, Ed., *Driver Activity Recognition from Sitting Postures*, Bauhaus-Universität Weimar, Weimar, Germany, 2007.
- [67] N. O. Bernsen and L. Dybkjær, “A multimodal virtual codriver’s problems with the driver,” in *Proceedings of the ISCA Tutorial and Research Workshop on Spoken Dialogue in Mobile Environments*, p. 15, ISCA, Bonn, Germany, 2002.
- [68] K. Bachfischer, S. Henze, and C. Wäller, “Emotional user interfaces in the car,” in *Proceedings of the 1st Workshop on Emotion and Computing—Current Research and Future Impact*, D. Reichardt, P. Levi, and J.-J. C. Meyer, Eds., pp. 55–59, Bremen, Germany, June 2006.
- [69] T. Sasayuki, K. Shin, and Y. Shin, “A driver-adaptive driver assistance system. A case of human-centered its view aid system,” *Nippon Kikai Gakkai Robotikusu, Mekatoronikusu Koenkai Koen Ronbunshu*, pp. 2A1–L2-4, 2004, (CD-ROM).
- [70] M. A. Tischler, C. Perter, A. Batliner, M. Wimmer, and J. Voskamp, “Application of emotion recognition methods in automotive research,” in *Proceedings of the 2nd Workshop on Emotion and Computing—Current Research and Future Impact*, vol. 1, pp. 55–60, Osnabrück, Germany, 2007.
- [71] M. Ablassmeier, T. Poitschke, and G. Rigoll, “A new approach of a context-adaptive search agent for automotive environments,” in *Extended Abstracts on Human Factors in Computing Systems (CHI ’06)*, pp. 1613–1618, ACM Press, New York, NY, USA, April 2006.
- [72] S. B. Wang, D. Demirdjian, T. Darrell, and H. Kjellström, “Multimodal communication error detection for driver-car interaction,” in *Proceedings of the 4th International Conference on Informatics in Control, Automation and Robotics (ICINCO ’07)*, pp. 365–371, Angers, France, May 2007.
- [73] F. Althoff, G. McGlaun, P. Schuller, M. Lang, and G. Rigoll, “Evaluating misinterpretations during human-machine communication in automotive environments,” in *Proceedings of the 6th World Multiconference on Systemics, Cybernetics and Informatics (SCI ’02)*, N. Callaos, J. H. Porter, and N. Rishe, Eds., vol. 7, pp. 13–17, Orlando, Fla, USA, 2002.
- [74] A. Amditis, A. Polychronopoulos, E. Bekiaris, and P. C. Antonello, “System architecture of a driver’s monitoring and hypovigilance warning system,” in *Proceedings of the Intelligent Vehicle Symposium*, vol. 2, pp. 527–532, IEEE, Athens, Greece, March 2003.
- [75] C. L. Lisetti and F. Nasoz, “Affective intelligent car interfaces with emotion recognition,” in *Proceedings of the of the 11th International Conference on Human Computer Interaction (HCI ’00)*, vol. 2, Las Vegas, Nev, USA, July 2005.
- [76] V. Kostov and S. Fukuda, “Emotion in user interface, voice interaction system,” in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, vol. 2, pp. 798–803, Nashville, Tenn, USA, 2000.
- [77] F. Angelucci, E. Ricci, L. Padua, A. Sabino, and P. A. Tonali, “Music exposure differentially alters the levels of brain-derived neurotrophic factor and nerve growth factor in the mouse hypothalamus,” *Neuroscience Letters*, vol. 429, no. 2-3, pp. 152–155, 2007.
- [78] N. Sheehy, “User-interface requirements for elderly people,” Tech. Rep., School of Psychology, The Queen’s University of Belfast, 2000.
- [79] J. E. Cahn, “The generation of affect in synthesized speech,” *Journal of the American Voice I/O Society*, vol. 8, no. 5, pp. 1–19, 1990.
- [80] I. R. Murray and J. L. Arnott, “Implementation and testing of a system for producing emotion-by-rule in synthetic speech,” *Speech Communication*, vol. 16, no. 4, pp. 369–390, 1995.
- [81] F. Burkhardt and W. F. Sendlmeier, “Verification of acoustical correlates of emotional speech using formant synthesis,” in *Proceedings of the ICSA Workshop on Speech and Emotion*, pp. 151–156, ISCA, Belfast, Northern Ireland, 2000.
- [82] S. J. L. Mozziconacci, *Speech variability and emotion: production and perception*, Ph.D. thesis, echnical University Eindhoven, 1998.
- [83] F. Kuhn, “Methode zur Bewertung der Fahrerablenkung durch Fahrerinformationssysteme,” April 2010, http://www.gui-design.de/download/wud_LCT_2005-11_Stuttgart.pdf.
- [84] J. Brooke, “SUS—a quick and dirty usability scale,” Tech. Rep., Redhatch Consulting, 1986.
- [85] K. Eilers, F. Nachreiner, and K. Hänecke, “Entwicklung und Überprüfung einer Skala zur Erfassung subjektiv erlebter Anstrengung,” *Zeitschrift für Arbeitswissenschaft*, vol. 40, pp. 215–224, 1986.
- [86] J. Wang, R. Knipling, and M. Goodman, “The role of driver inattention in crashes; new statistics from the 1995 crashworthiness data system (CDS),” in *Proceedings of the Annual Conference of the Association for the Advancement of Automotive Medicine*, Des Plaines, Ill, USA, 1996.
- [87] T. Ranney, E. Mazzae, R. Garrott, and M. Goodman, “NHTSA driver distraction research: past, present and future,” Tech. Rep., National Highway Traffic Safety Administration, Washington, DC, USA, 2000.
- [88] T. P. Alkim, G. Bootsma, and S. P. Hoogendoorn, “Field operational test “the assisted driver”,” in *Proceedings of the IEEE Intelligent Vehicles Symposium*, pp. 1198–1203, Istanbul, Turkey, 2007.
- [89] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [90] A. Graves and J. Schmidhuber, “Frame-wise phoneme classification with bidirectional LSTM and other neural network architectures,” *Neural Networks*, vol. 18, no. 5-6, pp. 602–610, 2005.
- [91] M. Wöllmer, F. Eyben, B. Schuller, E. Douglas-Cowie, and R. Cowie, “Data-driven clustering in emotional space for affect recognition using discriminatively trained LSTM networks,” in *Proceedings of the 10th Annual Conference of the International Speech Communication Association (INTERSPEECH ’09)*, pp. 1595–1598, Brighton, UK, September 2009.
- [92] C. Blaschke, F. Breyer, B. Färber, J. Freyer, and R. Limbacher, “Driver distraction based lane-keeping assistance,” *Transportation Research Part F*, vol. 12, no. 4, pp. 288–299, 2009.
- [93] K. Young, M. Regan, and M. Hammer, “Driver distraction: a review of literature,” Tech. Rep., Monash University Accident Research Center, 2003.
- [94] A. Graves, *Supervised sequence labelling with recurrent neural network*, Ph.D. thesis, Technische Universität München, 2008.
- [95] E. E. Shriberg, “Spontaneous speech: how people really talk and why engineers should care,” in *Proceedings of the 9th European Conference on Speech Communication and Technology (EUROSPEECH ’05)*, pp. 1781–1784, ISCA, 2005.