Towards Implicit Interaction in Highly Automated Vehicles - A Systematic Literature Review

ANNIKA STAMPF, Institute of Media Informatics, Ulm University, Germany
MARK COLLEY, Institute of Media Informatics, Ulm University, Germany
ENRICO RUKZIO, Institute of Media Informatics, Ulm University, Germany

The inclusion of in-vehicle sensors and increased intention and state recognition capabilities enable implicit in-vehicle interaction. Starting from a systematic literature review (SLR) on implicit in-vehicle interaction, which resulted in 82 publications, we investigated state and intention recognition methods based on (1) their used modalities, (2) their underlying level of automation, and (3) their considered interaction focus. Our SLR revealed a research gap addressing implicit interaction in highly automated vehicles (HAVs). Therefore, we discussed how the requirements for implicit state and intention recognition methods and interaction based on them are changing in HAVs. With this, open questions and opportunities for further research in this area were identified.

CCS Concepts: • General and reference → Surveys and overviews; • Human-centered computing → HCI theory, concepts and models.

Additional Key Words and Phrases: implicit interaction, in-vehicle interaction, systematic literature review

ACM Reference Format:

1 INTRODUCTION

Through the increase of automation, in-vehicle interaction could change fundamentally. A growing number of sensors integrated into the vehicle, major advances in machine learning, and increased processing capabilities allow driver behavior to be captured and interpreted in real-time [73]. This development enables innovative interaction possibilities such as affective in-car voice assistants [15]. This is also reflected through the upcoming trend of research in the fields of Human-Computer Integration [32, 72] or Symbiotic Interaction [59]. However, until now, in-vehicle interaction is mainly based on explicit direct driver inputs such as voice or touch inputs or direct manipulations of knobs or levers [23]. In comparison, human communication is highly based on implicit actions and behaviors, like posture or facial expression. Thus, we are able to make statements about and interact based on another person’s states, even if the person does not explicitly communicate these states [91]. Implicit inputs already play a significant role in advanced driver assistance systems (ADASs). Driver drowsiness, stress, fatigue, or distraction can, for instance, be recognized, e.g., through physiological signals [95, 107]. With a higher level of automation, the vehicle takes over the

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

© 2022 Copyright held by the owner/author(s).
2573-0142/2022/9-ART191
https://doi.org/10.1145/3546726
driving task and the accompanying responsibility. Simultaneously the driver becomes a passenger and is no longer restricted to supervising the vehicle and the environment, enabling a focus on non-driving tasks like reading or sleeping. Therefore, acceptance and trust in the vehicle are indispensable [84, 88]. Additionally, a focus shift of in-vehicle interaction towards user experience takes place [23], which in turn creates new fields of application for implicit interaction. Serim and Jacucci [92] recognized difficulties of a precise definition of implicit interaction in human-computer interaction (HCI). They tried to define the term by differentiating between the various used meanings. A few works already outlined implicit in-vehicle interaction [23, 73]. Some other publications provide surveys on context awareness in intelligent vehicles [33, 79]. However, in the context of in-vehicle interaction, Ohn-Bar and Trivedi [79] firmly focused on driver safety monitoring methods for overtaking situations, not taking into account the shifting interaction focus in HAVs. Other existing literature focused on context awareness and sensing from a rather technical point of view, addressing the realization of distributed human-advanced-vehicle systems [33], or giving a technical overview of in-vehicle sensors [34]. According to that, we are not aware of an overview of sensing methods relevant to implicit in-vehicle interaction, considering the underlying automation level and interaction focus.

In this work, we report the results of a literature review for 23 of the major publication venues within HCI (see Table 1) over the last 10 years (2011-2021) for publications addressing implicit state recognition approaches. Two key research questions (RQs) motivated this research:

- **RQ1**: Which states and intentions can be recognized through implicit input modalities in the current state-of-the-art, and how do the requirements for those machine learning based methods change with increasing vehicle automation?
- **RQ2**: How does implicit in-vehicle interaction change with increasing vehicle automation?

The main contributions of this work are: (1) Results of a systematic literature analysis on implicit interaction, showing what distinct states of a user can be inferred through implicit inputs in the current state-of-the-art and elaborating on the input modalities used. An additional differentiated examination of the state recognition methods on the basis of the underlying considered automation level and interaction focus revealed a research gap addressing state recognition methods in HAVs. (2) Based on these results, we discussed requirements for the translation of implicit interaction from manual to highly automated driving.

## 2 BACKGROUND

We started our analysis with a search for existing definitions of implicit interaction.

### 2.1 Implicit Interaction

In interpersonal conversation, information is communicated via non-verbal cues such as the pitch of the voice, facial expressions, or body language [78]. Those implicit cues improve understanding. In the context of HCI, Schmidt defines implicit interaction as “an action performed by the user that is not primarily aimed to interact with a computerised system but which such a system understands as input” [91, p. 8]. Schmidt further identified the main concepts of implicit interaction as the perception and interpretation of a user’s situational context. Leifer [55] defines implicit interaction as a two-dimensional space, which spans over the variables attentional demand and initiative, showing the diversity of the term’s meaning. Also, Serim and Jacucci [92] defined implicit interaction with different meanings. Thus, they have considered the term implicit differentiated as unintentional (i.e., interaction that is beyond a user’s given intents), attentional background (i.e., interaction without demanding a user’s attention), unawareness (i.e., interaction based on behavior or actions a user is not aware of), unconscious (i.e., interaction as the execution of tasks, which the user is
Towards Implicit Interaction in Highly Automated Vehicles - A Systematic Literature Review

not consciously processing), and *implicature* (i.e., an interaction that is based on meanings a user’s action implies). They summarized their results in the definition of implicit interaction as “user’s attitude towards an input–effect relationship in which the appropriateness of a system response to the user input (i.e., an effect) does not rely on the user having conducted the input to intentionally achieve it” [92, p. 2]. We refer to this definition in the remainder of this paper. Thus, we consider input modalities that are consistent with this definition and are therefore justified by at least one of the differentiated meanings.

2.2 Focus-Shift of In-Vehicle Interaction from Manual to Highly Automated Driving

To reach a shared understanding of the several automation levels, we refer to the common SAE Levels of Driving Automation [96]. By this definition, the automation of vehicles can be divided into five levels, ranging from Level 1, which means no automation, i.e., purely manual driving, to Level 5, which means full automation. With an evolution from current Level 2 to Level 3 vehicles, the system can take over the driving control under limited conditions (e.g., on highways), but the driver is still required to be able to retake control. In Levels 4 and 5, the driver’s role changes towards a supervisory role. Thus, the driver becomes a passenger. The system can handle most driving situations in Level 4 and all driving situations in Level 5.

Looking at the lower SAE Levels, in-vehicle interaction aspects mainly refer to ensuring a safe driving condition (e.g., [89, 99, 100, 123]). This involves focusing on enabling a safe driver state in manual driving and a safe shared control state in conditionally automated driving by focusing, for instance, on achieving successful takeover requests and keeping the driver in the loop. An introduction of Level 4 and 5 vehicles, however, leads to a shift of the driving responsibility from the driver to the vehicle. With this transition, establishing and maintaining trust and acceptance towards the vehicle becomes more important in the design of in-vehicle interaction concepts [84]. Furthermore, as the vehicle takes over the driving task, the driver instead becomes a passenger, able to focus on non-driving-related tasks, such as reading, sleeping, or working. Thus, the interior of future vehicles could more closely resemble the pattern of a working space or living room. In their work, Stevens et al. [97] also draw the comparison of a possible design pattern with a mobile, tiny house, striving for a multi-functional environment. As “trust and use are often associated with users’ experience of the driver-vehicle interfaces and interior design” [35, p. 1], and as the automotive industry realizes it as a unique selling point [23], user experience is another aspect which is gaining high prominence in the design of in-vehicle interaction concepts.

3 SYSTEMATIC LITERATURE REVIEW

To learn about the meaning of implicit input in related work, we conducted a systematic literature survey following the PRISMA Statement methodology [69]. For the search, we selected three digital libraries that include the most relevant conferences and venues on automotive user interfaces (ACM DL, IEEE Xplore and Elsevier Science Direct). We queried the most cited HCI venues [39] plus additional proceedings relevant for automotive research resulting in 23 venues (see Table 1). Our queries each consist of three query parts, which narrow down the results to the thematic focus of the review. (1) The focus is on the automotive domain, which is set by the keywords "vehicle", "car", "driving", or "in-vehicle". (2) The restriction to the context of HCI is constrained by the keywords "input", "interaction", or "interface". (3) The focus on implicit interaction is built out of three sets: The first set is an explicit search for "implicit" interaction. However, most of the literature dealing with implicit input modalities does not directly mention the keyword "implicit". Thus, the second set

---

is composed of gathered possible implicit input modalities. Therefore, we used the presented sensor modalities proposed by Sharma and Pavlovic [93]. In addition, we went through all publications of the 12th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications (2020) and examined them for modalities that could be used implicitly. The goal of processing implicit input is to identify the state/intention of the driver using pattern recognition and machine learning approaches. The processing of data with the goal of state/intention recognition is usually called "prediction", "recognition", "estimation", or "classification" [81, 85]. Thus, the third set is composed of these keywords. This allowed us to include modalities that could not be covered by the preliminary research due to e.g. different wording or synonyms used. Furthermore, this includes literature that obtained implicit input as an implication of explicit input. An example is the processing of explicit pedal input to implicitly understand a driver’s driving style [119].

The resulting queries can be seen in Table 2. In addition, we have narrowed down our search to publications from the last ten years (2011-2021) and only considered titles and abstracts for search. After the removal of duplicates, a total of 1924 publications remained, which we further investigated in an abstract screening.

3.1 Review Process

For an abstract screening, we defined inclusion criteria: (1) The publication focuses on processing implicit human inputs based on our understanding of implicit inputs (see Background) to obtain information about the user’s state or intentions. (2) The publication is related to an automotive context. Further, we defined exclusion criteria: The publication’s main contribution is a taxonomy, design space, literature review, or workshop.

To analyze the selected articles, we used sysrev.com, which enabled us to screen the publications collaboratively. Two authors were involved in the screening process. The first author reviewed each publication and the second one reviewed 805 publications. A calculation of the Cohens Kappa [67] revealed an inter-rater reliability of $\kappa = 73.8 \ (p = 96.6\%)$, indicating a substantial agreement according to Landis and Koch [53]. Afterward, all conflicts were discussed and could be solved. Our Sysrev project is publicly accessible\(^2\). It includes the 1924 screened publications with abstracts and labels attached by the authors.

The abstract screening resulted in a sample of 128 publications. Those publications were further analyzed by a full-text screening. In this step, we excluded an additional 53 papers for which only

\(^2\)https://sysrev.com/u/5332/p/79138

---

the full-text analysis revealed that they do not meet previously defined inclusion and exclusion criteria. We additionally discovered seven works by reference crawling the most relevant results. In total 82 publications were included in the quantitative synthesis. Figure 1 shows the complete PRISMA flow diagram, illustrating our paper selection process.

For the quantitative synthesis, we developed a set of labels to summarize the work and to note the most important aspects for our review. Five labels were extracted: (1) implicit input modalities (string label), (2) vehicle automation (categorical label), (3) contribution (string label), (4) implicit interaction (string label), and (5) notes (string label).

<table>
<thead>
<tr>
<th>Conference / Venue</th>
<th>Number of publications</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACM Conference on Human Factors in Computing Systems (CHI)</td>
<td>119</td>
</tr>
<tr>
<td>ACM Conference on Computer-Supported Cooperative Work &amp; Social Computing (CSCW)</td>
<td>11</td>
</tr>
<tr>
<td>ACM Conference on Pervasive and Ubiquitous Computing (UbiComp)</td>
<td>27</td>
</tr>
<tr>
<td>ACM/IEEE International Conference on Human Robot Interaction (HRI)</td>
<td>52</td>
</tr>
<tr>
<td>ACM Symposium on User Interface Software and Technology (UIST)</td>
<td>9</td>
</tr>
<tr>
<td>ACM Transactions on Computer-Human Interaction (TOCHI)</td>
<td>0</td>
</tr>
<tr>
<td>ACM Multimodal Interfaces and Machine Learning for Multimodal Interaction (ICMI-MLMI)</td>
<td>45</td>
</tr>
<tr>
<td>ACM Designing Interactive Systems (DIS)</td>
<td>14</td>
</tr>
<tr>
<td>ACM Intelligent User Interfaces (IUI)</td>
<td>28</td>
</tr>
<tr>
<td>ACM International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI)</td>
<td>18</td>
</tr>
<tr>
<td>ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutoUI)</td>
<td>192</td>
</tr>
<tr>
<td>IEEE Transactions on Affective Computing</td>
<td>30</td>
</tr>
<tr>
<td>IEEE Transactions on Human-Machine Systems</td>
<td>50</td>
</tr>
<tr>
<td>IEEE Transactions on Haptics</td>
<td>16</td>
</tr>
<tr>
<td>IEEE Transactions on Intelligent Transportation Systems (ITS)</td>
<td>254</td>
</tr>
<tr>
<td>IEEE Transactions on Vehicular Technology</td>
<td>188</td>
</tr>
<tr>
<td>IEEE Transactions on Intelligent Vehicles</td>
<td>49</td>
</tr>
<tr>
<td>IEEE Intelligent Vehicles Symposium (IV)</td>
<td>115</td>
</tr>
<tr>
<td>IEEE International Conference on Intelligent Transportation Systems (ITSC)</td>
<td>114</td>
</tr>
<tr>
<td>IEEE Access</td>
<td>764</td>
</tr>
<tr>
<td>Elsevier International Journal of Human-Computer Studies</td>
<td>13</td>
</tr>
<tr>
<td>Transportation Research Part F: Traffic Psychology and behavior</td>
<td>217</td>
</tr>
<tr>
<td>Elsevier Robotics and Autonomous Systems</td>
<td>54</td>
</tr>
</tbody>
</table>

**Table 1.** Number of publications for the retrieved conferences and venues.

4 RESULTS

In this section, we present the results of the literature review. It includes 75 publications from the previous screening and seven additional works [26, 49, 50, 64, 81, 104, 112] by forward and backward reference searching of relevant results.

4.1 Modalities

The following provides an overview of the modalities and their usage to access a user’s states, also presented in Table 3.

4.1.1 Physiological Signals. Common measured and processed physiological signals are (1) electrical heart activities, like heart rate, heart rate variability (HRV), or other features and cues of
an electrocardiogram (ECG), and (2) electrodermal activity (EDA), which is responsible for the electrical characteristics of the skin [52].

**Electrical Heart Activities.** The activity of the Autonomic Nervous System (ANS) changes with the level of stress, fatigue, or drowsiness [5, 7]. This activity can be sympathetic or parasympathetic. Increased sympathetic activity and decreased parasympathetic activity imply wakefulness states, while inverse activities are characteristic of relaxation states [104]. HRV can be measured to estimate a user’s fatigue level [103]. Balasubramanian and Bhardwaj [7] used a cECG to measure HRV components in an unobtrusive manner through electrodes on the back of the driver’s seat. Du et al. [27] estimated the level of driver fatigue through a measurement of the heart rate and an additional capture of eye and mouth openness level to stabilize possible noisy results of the heart rate measurement.

**Electrodermal Activities.** As the sympathetic nervous system controls sweating by innervating the sweat glands, changes in skin conductivity provide information about the sympathetic arousal of a user [109]. Wickramasuriya and Faghih [109] measured the palm EDA to estimate a user’s state of arousal.

**Thermal Signals.** Our SLR revealed that in multiple publications, the perinasal perspiration is also measured as an unobtrusive alternative to palmar EDA [13, 36, 82] to estimate a user’s state of arousal.

**Respiration.** Grassmann et al. [40] showed that respiratory changes in response to cognitive load. However, we only found publications using those respiratory cues (mainly breathing rate [15, 42]) as part of multimodal input for cognitive load estimation.

**Brain-Activity.** We found six papers that used electroencephalography (EEG; a method of measuring electrical activity on the scalp) to determine a user’s intention with respect to driving decisions, such as the intention to turn right/left or to change speed [6, 11, 12, 31, 45, 58, 61]. The authors focused on paralyzed users or users with movement disorders to enable them to make independent driving decisions, which are then executed by automated maneuvering. As these driving decisions currently require users to focus on external stimuli, this type of interaction is not implicit. Nevertheless, it has been included in the SLR because this input type is a non-negligible opportunity for implicit interaction. The work of Guo et al. [41] even shows in this context that with EEG (without stimuli), it is already possible to detect whether the driver has focused on certain hazards or not. Further, Yu et al. [118] described an approach to detect vigilance.

### 4.1.2 Auditory

In contrast to explicit speech inputs, implicit speech inputs are characterized by verbal utterances that are not directly addressed to the vehicle or paralinguistic cues. Thus, implicit speech inputs are signals of linguistic messages that do not contain direct information. Examples

<table>
<thead>
<tr>
<th>Level</th>
<th>ACM Query</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct</td>
<td>(vehicle? OR car? OR driver? OR &quot;in-vehicle&quot; OR &quot;driving&quot;) AND &quot;implicit&quot; AND (interaction? OR input? OR sensing OR interface?)</td>
</tr>
<tr>
<td>Modalities</td>
<td>(vehicle? OR car? OR driver? OR &quot;in-vehicle&quot; OR &quot;driving&quot;) AND (interaction? OR input? OR interface?) AND (eye? OR gaze OR blink* OR pupil* OR &quot;facial expression&quot; OR &quot;facial behavior&quot; OR mimic OR &quot;visual behavior&quot; OR gesture OR motion OR movement OR posture OR &quot;head pos*&quot; OR &quot;body pos*&quot; OR voice OR speech OR speak* OR verbal OR &quot;physiological signal?&quot; OR &quot;brain activit*&quot; OR &quot;brain-computer&quot; OR biometrics OR emotion OR emotional)</td>
</tr>
<tr>
<td>State</td>
<td>(vehicle? OR car? OR driver? OR &quot;in-vehicle&quot; OR &quot;driving&quot;) AND (interaction? OR input? OR interface?) AND (predict* OR estimat* OR classif* OR recogni*)</td>
</tr>
</tbody>
</table>

Table 2. ACM queries.
<table>
<thead>
<tr>
<th>Modalities</th>
<th>Recognition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physiological</td>
<td>Physiological States</td>
</tr>
<tr>
<td>Heart Activity</td>
<td>Cognitive Load</td>
</tr>
<tr>
<td>Auditory</td>
<td>Fatigue</td>
</tr>
<tr>
<td>Visual</td>
<td>Stressed</td>
</tr>
<tr>
<td>Kinesesthetic</td>
<td>Attention</td>
</tr>
<tr>
<td>Profile</td>
<td>Distraction</td>
</tr>
<tr>
<td></td>
<td>Drowsiness</td>
</tr>
<tr>
<td></td>
<td>Emotional States</td>
</tr>
<tr>
<td>Physiological Parameters</td>
<td>Anxiety</td>
</tr>
<tr>
<td></td>
<td>Valence Arousal State</td>
</tr>
<tr>
<td></td>
<td>Sympathetic Arousal</td>
</tr>
<tr>
<td></td>
<td>Vigilance</td>
</tr>
<tr>
<td></td>
<td>Frustration</td>
</tr>
<tr>
<td>Kinesthetic States</td>
<td>Activity Recognition</td>
</tr>
<tr>
<td>Driving-Related States</td>
<td>Driving Behavior</td>
</tr>
<tr>
<td></td>
<td>Vehicle Control</td>
</tr>
<tr>
<td>Other</td>
<td>User Identification</td>
</tr>
</tbody>
</table>

Table 3. Overview of assessed modalities with reference to recognized user states and intentions.

are pauses in speech, speech rhythm, volume, intonation and pitch [50]. Paralinguistic inputs can provide information about the current emotional state of a user. Work by Jones and Jonsson [50] explored the recognition of the emotional state through paralinguistic inputs as dimensions of valence and arousal. A later publication of Jones and Jonsson [49] obtained paralinguistic cues in speech to recognize emotions in older car drivers. For a deeper understanding, Zepf et al. [120] provided a further overview of studies on the recognition of emotions in the automotive context, including emotion recognition through speech.

4.1.3 Visual. Our SLR revealed that visual signals were mainly used as gaze-based implicit inputs. However, also blink frequency and pupil dilation was considered.

Gaze. Akash et al. [4] used eye gaze to estimate a user’s trust in the automated vehicle and surveillance of the scene. We found multiple publications that showed that gaze direction can be a good indicator of where attention is directed [70], if a driver is distracted [36], or to measure a driver’s cognitive load state [4, 74]. Biswas and Prabhakar [14] estimated workload by exploring saccades, which are fast, erratic movements of the eyes. Wu et al. [113] used saccades to predict a driver’s reaction time.

Blink. Drowsiness can be estimated through the blink frequency [112]. Further, Benedetto et al. [10] considered blink frequency as a way of estimating visual workload and showed that it can serve as an indicator of distraction.

Pupil Dilation. We found several publications which accessed the pupil diameter to estimate a driver’s cognitive load [15, 43, 44]. Niezgoda et al. [74] further accessed the pupil diameter to determine a driver’s distraction.

4.1.4 Kinesthetic. Kinesthetic implicit input includes muscle activities, position and movements of the body, and individual body parts, for example, head, arms, and legs. In addition, we considered facial features.

Body Pose. Our SLR revealed several approaches to recognize user activities by obtaining the body pose or body movements [8, 48, 65, 66, 115, 121]. Activities that can be recognized by their approaches are, for example, driving normally, reaching for center compartment, adjusting the radio, drinking, preparing, talking to a passenger, or talking on a phone. Tran et al. [101] observed the foot position to predict a driver’s foot movement and, thus, could recognize the braking aspect of the driving behavior. Based on posture and sitting position, the tension and, thus, the stress level can be detected [63].

Head Pose. We found work that used a driver’s head pose with head pitch and yaw position to estimate a driver’s focus of attention to predict the lane change intention [22, 71] with additional driving and environmental information. Wang et al. [107] further used the head pose to estimate the driving load and driving capability of a driver.

Facial Features. With the recognition of facial features and understanding of facial expressions, multiple statements can be inferred about a user’s state. Examples are the emotional state [120] or the driving capability through driver fatigue detection [27, 102]. Recognition of facial expressions was additionally used (besides the body and head pose) by Martin et al. [65] and Zhang et al. [121] for the identification of a user’s activity. We further found that Huang et al. [46] used facial features to estimate heart rate (58.60% accuracy), skin conductance (83.78% accuracy), and vehicle speed (59.89% accuracy) as a contactless alternative to multiple bio-sensors in the vehicle.

Driving Activity. Driving is an explicit activity. Direct interactions such as actively moving the steering wheel to determine the vehicle’s direction or using the pedals to consciously control the speed are necessary. However, according to the definition of Serim and Jacucci [92], if the interaction relies on meanings a user’s action implies, such as the driving style, it can be described as implicit.
Therefore, we consider these implications obtained by driving activities to be implicit input. For example, our SLR revealed multiple publications accessing vehicle data to obtain information about a driver’s driving style [29, 82, 119], such as aggressive or normal driving, or a driver’s driving skill level [18]. Gonzales et al. [37] further estimated a driver’s risk-aversion by combining and processing driving activity and body pose. The driving activity can also be used to make statements about the driver’s intention regarding a lane change [22, 38, 71, 108, 114] or driving behavior in general [3, 111, 122].

4.1.5 Profile. In addition to the above-mentioned indirect implicit information about the user’s state, such as driving style, there is also the possibility of implicitly obtaining information about user traits and preferences. The difference is that user traits are stable compared to user states, which are fluctuating [15].

User Traits. A large number of algorithms identifying user traits have been developed in recent years. Ranjan et al. [86], and Aji et al. [2], for example, estimated a person’s gender by fusing facial features. Also, methods for age recognition do exist [87]. However, our SLR did not reveal any publications that make use of implicit gained traits as input for in-vehicle interaction.

Preference. We found a publication that used explicit vehicle data to obtain implicit information about a driver’s action preference to predict driving behavior [37]. Imai et al. [47], and Lassoued et al. [54] recorded route trajectories to gain knowledge about destination and route preferences implicitly and thus predict future destinations and routes.

4.2 Level of Automation and State Transition

<table>
<thead>
<tr>
<th>State</th>
<th>Manual Driving</th>
<th>Shared Control</th>
<th>Autonomous Driving</th>
<th>Independent of Automation Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kinesthetic State</td>
<td>8: [8, 9, 24, 48, 65, 66, 115, 121]</td>
<td>1: [64]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emotional State</td>
<td>4: [15, 49, 50, 109]</td>
<td>1: [26]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>2: [13, 90]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Allocation of the surveyed articles to the considered level of automation.

An essential step toward recognizing user states through implicit input modalities is to look at the above-used methods from the perspective of the underlying level of automation, as the recognition methods and goals may depend on the context in the vehicle, which is changing with SAE Level. While in manual driving the focus is on the primary driving task, at higher levels of automation, the user’s focus shifts to secondary and tertiary tasks due to the elimination of the primary driving task. This transformation also influences possible states of the user concerning the characteristics, i.e., how these states appear or what causes them. Table 4 shows that the majority of the surveyed publications examined are mainly concerned with manual driving (78%). 10% focused on recognizing
states in the condition of shared control, and only 6% put a fully automated vehicle state at the center of their considerations. For the kinesthetic states, the detection of driver activities can be listed as an example of a recognition method that can not be adopted for all SAE Levels without modifications. Deep learning based models for activity recognition usually require data-intensive training. Therefore, most models are limited to the recognition of a small number of pre-classified actions and are directly related to the driving task, such as normal driving, hands on/away from the wheel, or talking on the phone [8, 48, 121] and use manual driving datasets [1, 28]. However, with the transition to automated driving, the condition changes, as both possible activities and the vehicle interior change, e.g., through swiveling chairs that change a user’s observation angle or activities that were not possible while driving. Therefore, to transfer activity recognition models to the automated context, the models need to be re-trained with respect to those evolving conditions. Thus, Martin et al. [64] proposed a new dataset Drive&Act for driver behavior recognition in automated vehicles. Nevertheless, the number of possible activities in automated vehicles is greatly expanding, which is why more datasets are needed to cover these. Our SLR also shows that a lot of work deals with state and intention recognition methods that are in close contact with the manual driving task, such as the recognition of driving behavior, driving performance, crash risk, reaction time, driving uncertainty, and driving style. These will vanish nearly entirely in the automated context. Thus, the shift of tasks towards secondary and tertiary tasks also eliminates driving activity as a contactless input modality, which some methods use to measure distraction [36, 82, 98, 107] or cognitive load [95, 99], for example. This reinforces the need for other (contactless) methods for detecting physiological conditions, such as Huang et al.’s approach [46] of estimating physiological parameters from facial image data. Overall, it is also noticeable that many publications only consider manual driving and refer exclusively to use cases in this context. Still, the methods themselves could be applied regardless of the automation level. For example, detecting physiological and emotional states is mainly considered in the manual driving context and with the ulterior motive of recognizing critical driver states. Although, it could also be useful in an automated condition to detect drowsiness, boredom, frustration, anxiety, or other states, for instance, to enhance the user experience by adapting the driving style, infotainment functions, proactive feedback, etc. Our analysis shows that only a few papers consider state recognition in a highly automated context. Future research should focus on how to detect user states independently of driving behavior and in an unobtrusive manner (contactless, at best) so that the methods apply to drivers as well as passengers, including users of HAVs. Furthermore, researchers should keep the fully automated context in mind when developing state recognition approaches and also consider appropriate use cases in this context and evaluate their methods on these.

4.3 Shift of Interaction Focus

In a further step of our analysis, we extracted the publications that additionally investigated implicit interaction on the basis of implicit input modalities and the physiological, kinesthetic, emotional, and driving-related states and intentions recognized through them. As the interaction focus changes with increasing automation level, we re-analyzed these extracted publications from this point of view. In the previous section, we found that most of the papers examined deal with state recognition in manual driving, which is reflected in the interaction concepts considered (see Table 5): Almost exclusively, concepts with the focus on achieving a safe driving state were implemented. Therefore, implicit inputs are mainly used to recognize the states and intentions of the driver, evaluate them based on driver safety, and intervene if necessary. Taniguchi et al. [99], for instance, used the detected cognitive load to determine the appropriate time for non-driving-related in-vehicle interaction dialogues. An approach to alert the driver if dangerous driving behavior was recognized was presented by Panagopoulos and Pavlidis [82]. Further, Tanveer et al. [100] proposed...
Towards Implicit Interaction in Highly Automated Vehicles - A Systematic Literature Review

Table 5. Focus of implicit interaction based on recognized states.

<table>
<thead>
<tr>
<th></th>
<th>Safe Driving Condition</th>
<th>Acceptance and Trust</th>
<th>User Experience and Comfort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive Load</td>
<td>[15, 99]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Driving Behavior</td>
<td>[25, 82, 123]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drowsiness</td>
<td>[100]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Action Recognition</td>
<td>[48, 57]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distraction</td>
<td>[94]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Driving Uncertainty</td>
<td>[116]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stress</td>
<td>[63]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Driving Style</td>
<td>[29, 82]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive Load and Driving Capability</td>
<td>[107]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Action Recognition</td>
<td>[89]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>User Identification</td>
<td>[89]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a drowsiness warning system. Among all the works, we could identify only two that, starting from implicit inputs, presented an interaction design concept to increase user experience and comfort or acceptance and trust towards the vehicle. The first work is by Akash et al. [4]. They proposed an adaptive augmented reality interface, showing automation transparency by highlighting objects detected in road traffic, which depend on the recognized trust and workload to calibrate the driver’s trust. The second approach was shown by Rivera et al. [89]. They proposed a prototyping framework for adaptive user interfaces and considered the exemplary use case user aware hand-over request, where driver activities are tracked and displayed on a control panel to perform a safe hand-over. Here, the focus is again on maintaining a safe driving condition. In a second use case, they focused on user experience. Thus, a user recognition with automatic adaptation of the infotainment system to the user profile was implemented. However, they did not evaluate the proposed system.

5 DISCUSSION

5.1 Which States and Intentions Can be Recognized Through Implicit Input Modalities in the Current State-Of-the-Art and How Do the Requirements for those Machine Learning Based Methods Change with Increasing Vehicle Automation?

Our SLR shows an overview of current state-of-the-art in-vehicle state recognition methods and goes into more detail about the input modalities used. We found that most methods are implemented with a manual driving condition in mind. For this reason, many methods cannot be used at all for higher levels of automation or only with reconsideration and adaptation.

5.1.1 Translation of Kinesthetic State Recognition Methods. Methods that recognize kinesthetic states are usually based on manual driving data sets. It is, therefore, necessary to re-examine which tasks and associated activities HAVs are likely, such as Wilson et al. [110] did, and generate new data sets based on these tasks, such as the Drive&Act dataset [64]. We suggest that camera-based activity recognition in HAVs should be of further interest. In particular, attention could be paid to recognizing subconscious behavior patterns since they might provide more information about a user’s emotional state enabling an extension of other contactless state recognition methods. As an example, bouncing legs could indicate that a user is nervous.

5.1.2 Translation of Emotional State Recognition Methods. Most emotion recognition methods from the automotive context can be transferred to higher levels of automation, as they are estimated either
by image data and facial expressions or by physiological parameters, which are independent of the factors changing with increasing SAE Levels. Here we see the potential in in-vehicle interaction concepts with the aim of meeting the requirements of the new interaction focus through new use cases and thus having a positive influence on acceptance, trust, user experience, and comfort. Dillen et al. [26], for example, have done just this with the development of a fear and comfort detection approach to individually assess differentiated driving styles. In general the topic of affective interaction is widely discussed for lower SAE Level [15–17, 30, 62]. And even in the context of HAVs, a few works already deal with the design of use cases [76, 77]. Nevertheless, many open questions still need to be explored in future research. Especially for the goal of emotion regulation, some things need to be investigated, such as the recognition of the emotional source (i.e., how to find out of the user is anxious about the driving style or an upcoming appointment) and individual preferences (i.e., how to find out whether the user prefers to listen to music or watch a movie when bored).

5.1.3 Translation of Physiological State Recognition Methods. Considering physiological states, our SLR revealed that most state recognition methods were implemented primarily with the intention of ensuring a safe driving state in manual driving. The best possible target is the contactless recognition of states, which multiple methods already do consider [10, 46, 74]. Nevertheless, our SLR also showed that numerous methods monitored and processed driving-related input modalities to meet this challenge. Future work, therefore, needs to consider which physiological states might be of interest for designing interaction concepts in HAVs and how these states can be detected in a contactless manner and independent of a user’s current activity. Further, as with the emotional state, methods should be found to recognize the source of a state to respond appropriately (e.g., how to find out whether a user is stressed due to high traffic volume or a current phone call).

5.2 How Does Implicit In-Vehicle Interaction Change with Increasing Vehicle Automation?

The dissimilarity between the desired state in manual and autonomous driving results in a new pipeline for implicit input-based interaction. Thus, Figure 2 shows our understanding of the targeting pipeline in implicit interaction in HAVs. First, implicit inputs are processed to recognize a user’s state. For this purpose, identified physiological, emotional, and kinesthetic states are fused. Unlike manual driving, where the driver’s desired state is predefined as the state in which the vehicle can be safely controlled, the desired state is determined by the interpretation of the recognized state with regard to obtained profile information, such as a user’s mood, traits, or preferences and environmental and personal context. Finally, the desired state is translated into action.

5.2.1 Indistinct Desired States Lead to Growing Uncertainties. As we have already seen in our SLR, through this process and the characteristics of current state recognition and intention prediction systems having uncertainties, there is always a residual probability of wrong assumptions leading to inappropriate actions. As a result, it might cause problems at lower levels of automation (e.g., ADAS functions might be triggered erroneously), as well as in high levels of automation, where it might lead, for instance, to inappropriate recommendations or driving adaptions. In the case of HAVs, however, there is the additional factor that the detected condition can also be misinterpreted, resulting in a false desired state being inferred. As an additional explanation: The desired state in manual driving is always the state in which the driver can safely operate the vehicle, meaning, for instance, an attentive, focused, alert, undistracted state. In the context of automated driving, inferring the desired user state from the recognized state is more complex. Thus, for instance, the desired state of a drowsy driver could be awake. However, it could just as well be asleep. This increases the probability of false assumptions. Thus, when designing interaction concepts based on
Towards Implicit Interaction in Highly Automated Vehicles - A Systematic Literature Review

such machine learning methods, it is, especially for HAVs, essential to consider what impacts the user will experience if an action is taken in response to a wrongly recognized and interpreted state. Also, concerning the central importance of acceptance and user experience in HAVs, considering incorrect assumptions should take place, as interactions based on them might influence these aspects negatively. Those negative effects could be described as the cost of failure, which needs to be offset against the expected benefits an action based on correct assumptions would have, both weighted by the accuracy of the machine learning model. Justified by the Halo effect [75], the way a vehicle agent interacts with its users also influences their expectations of the vehicle’s capabilities [35]. For this reason, in the context of implicit in-vehicle interaction, it needs to be explored whether the complexity of user states the vehicle can recognize and react to also influences the perceived driving capabilities. If so, this needs to be taken into account in the design of new interaction concepts as the perceived safety might be positively influenced, which in turn has a direct effect on trust. However, this also means that proactive actions based on misrecognized or misinterpreted implicit inputs could negatively affect perceived safety and trust in the automation. Thus, overall, we consider it necessary to identify influencing factors that quantify the benefits of an implicit input based action and the costs that can arise when this action is caused by incorrect assumptions. This should lead to the ability to perform a cost-benefit analysis based on the accuracy of the machine learning model to create a quality assessment framework for implicit in-vehicle interaction.

6 LIMITATIONS AND FUTURE WORK

Our SLR has mainly shown research gaps in the area of implicit input-based interaction. We see the goal of further work to close the gaps by (1) extending existing use cases and associated implicit input-based interaction concepts, (2) evaluating, adapting, and extending current state recognition methods for a highly automated context to implement designed interaction concepts, and (3) evaluating these concepts in further studies (with consideration of possible false assumptions) based on the identified new requirements. The latter also involves investigating how passengers ultimately interact with such a system. As Serim and Jacucci [92] stated, users might intentionally trigger a system action when they "comprehend the causal relationship between their inputs and the system effects" [92, p. 7]. Thus, in particular, it needs to be investigated if and how the users’ behavior adapts.
7 CONCLUSION

This work’s goals were to examine which states and intentions can be recognized through implicit input modalities in the current state-of-the-art and how the requirements for those methods and for implicit interaction change with increasing SAE Level. Therefore, we conducted a literature review regarding implicit interaction in the in-vehicle context, including 82 publications, which revealed a research gap addressing implicit interaction in HAVs. Findings led to a discussion of requirements for the adoption of implicit interaction to HAVs which opened questions and opportunities for further research in this area.

ACKNOWLEDGMENTS

This work was conducted within the project ‘Self-Supporting Multimodal Interaction’ (Semulin) funded by the Federal Ministry for Economic Affairs and Climate Action (BMWK).

REFERENCES


Towards Implicit Interaction in Highly Automated Vehicles - A Systematic Literature Review


Fei Yan, Mark Eilers, Martin Baumann, and Andreas Luedtke. 2016. Development of a Lane Change Assistance System Adapting to Driver’s Uncertainty During Decision-Making. In *Adjunct Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI ’16 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 93–98. https://doi.org/10.1145/3004323.3004334


Received February 2022; revised May 2022; accepted June 2022