When Do Drivers Interact with In-Vehicle Well-being Interventions? An Exploratory Analysis of a Longitudinal Study on Public Roads

KEVIN KOCH*, University of St. Gallen, Switzerland

VARUN MISHRA*, Dartmouth College, USA

SHU LIU, ETH Zürich, Switzerland

THOMAS BERGER, University of Bern, Switzerland

ELGAR FLEISCH, ETH Zürich and University of St. Gallen, Switzerland

DAVID KOTZ, Dartmouth College, USA

FELIX WORTMANN, University of St. Gallen, Switzerland

Recent developments of novel in-vehicle interventions show the potential to transform the otherwise routine and mundane task of commuting into opportunities to improve the drivers' health and well-being. Prior research has explored the effectiveness of various in-vehicle interventions and has identified moments in which drivers could be interruptible to interventions. All the previous studies, however, were conducted in either simulated or constrained real-world driving scenarios on a pre-determined route. In this paper, we take a step forward and evaluate when drivers interact with in-vehicle interventions in unconstrained free-living conditions.

To this end, we conducted a two-month longitudinal study with 10 participants, in which each participant was provided with a study car for their daily driving needs. We delivered two in-vehicle interventions – each aimed at improving affective well-being – and simultaneously recorded the participants' driving behavior. In our analysis, we found that several pre-trip characteristics (like trip length, traffic flow, and vehicle occupancy) and the pre-trip affective state of the participants had significant associations with whether the participants started an intervention or canceled a started intervention. Next, we found that several in-the-moment driving characteristics (like current road type, past average speed, and future brake behavior) showed significant associations with drivers' responsiveness to the intervention. Further, we identified several driving behaviors that "negated" the effectiveness of interventions and highlight the potential of using such "negative" driving characteristics to better inform intervention delivery. Finally, we compared trips with and without intervention and found that both interventions employed in our study did not have a negative effect on driving behavior. Based on our analyses, we provide solid recommendations on how to deliver interventions to maximize responsiveness and effectiveness and minimize the burden on the drivers.

CCS Concepts: • **Human-centered computing** → **Empirical studies in HCI**; *Ubiquitous and mobile computing systems and tools*; • **Applied computing** → *Consumer health.*

---

*Contributed equally to this work

Proc. ACM Interact. Mob. Wearable Ubiquitous Technol., Vol. 5, No. 1, Article 19. Publication date: March 2021.

19

1 INTRODUCTION

Commuting by car is a routine, albeit integral, part of life for many people. In the United States, the average person spends just under an hour driving every day [65]. Daily driving is rarely pleasant and is often considered a burden and a source of stress. Drivers regularly face a lack of control over their trips, caused by traffic jams, congestion, and unpredictability [10, 40]. Consequently, people generally experience a worse mood while driving a car than during other daily activities [10, 32]. Further, prior research has found that longer commute times to and from work adversely affect drivers' well-being [62], which, by extension, could lead to negative effects on the drivers' health [14]. Beyond long-term health consequences, low well-being due to certain emotions could have an immediate negative impact on the driver's driving and safety. Angry driving or driving under strong negative emotions impairs cognitive abilities, which is reflected in driving behavior, through increased risk-taking and reduced lateral control [9, 28].

Given that people spend a significant amount of time in a dedicated space, the context of a car provides a unique opportunity to deliver interventions to improve drivers' well-being and long-term health. In recent years, there has been a variety of research aimed at in-vehicle interventions; these prior works show the potential of such interventions to improve drivers' well-being. Paredes et al. designed and evaluated the effectiveness of a guided slow-breathing intervention to reduce drivers' physiological stress [53, 54]. Balters et al. further evaluated the slow-breathing intervention in a controlled, real-world driving condition on a test track, and found that the intervention led to a reduction in breathing rate and physiological arousal for drivers who engaged with the intervention [4]. Dimitijs et al. designed an odor-based intervention to improve well-being while driving [17]. In related work, Braun et al. conducted a controlled, real-world study to evaluate the effectiveness of a personalized voice assistant on positive emotions experienced by a driver [8].

Although interest in such in-vehicle interventions is rising, with almost 3,000 deaths due to distracted driving [50], care must be taken when delivering such interventions to avoid distracting the driver from their primary task – driving. Recent works in the UbiComp and HCI community have explored these questions. Kim et al. conducted a simulator-based, and a controlled, real-world driving study to predict opportune moments for in-vehicle proactive auditory-verbal tasks (*n*-back tests) [33]. Building on their prior work, Kim et al. conducted another controlled, real-world driving study to understand how varying demands of proactive voice tasks affect driver interruptibility, and how drivers adapt to concurrent driving and voice tasks [34]. Semmens et al. conducted another controlled, real-world driving study to evaluate when participants are willing to engage in voice-based information from the car, by frequently asking the participants, "Is now a good time?" [60]. In other work, Kim et al. conducted a controlled, real-world study to detect interruptibility based on peripheral interaction [35].

These existing works, however, pose certain limitations. First, none of the works modeled or evaluated in-vehicle interruptibility delivered actual interventions; they either prompted questions to the drivers [60], asked them to do some task [33, 34], or used a proxy to determine interruptibility [35]. Second, the data in these prior works was either collected in a simulated environment or in a controlled, real-world driving scenario, where the participants were asked to drive in a predetermined route for a certain duration. Such controlled driving conditions do not account for the variables in a person's daily drive. In unconstrained scenarios, drivers experience several conditions and situations: routine commute (e.g., home → work and work → home), non-routine drives (e.g., weekend outing or a holiday road-trip), challenging driving conditions (e.g., bad weather or driving at night),

and driving with passengers are some examples. Hence, it is important to understand when drivers interact with interventions in unconstrained driving conditions.

To this end, we conducted a two-month longitudinal study with 10 participants. We equipped a fleet of ten cars with a data collection system that we handed over to ten participants for their daily drives and recorded their driving behavior. In 66% of the drives, we randomly prompted the participants with one of two interventions (a *mindfulness* or a *music* intervention) to be completed while driving; these interventions aim to improve affective well-being of the drivers [36]. The goal of this work is to conduct an exploratory analysis of when participants interacted with these interventions. Specifically, we make the following contributions.

- We evaluate how general trip characteristics and a driver's affective state before starting the drive relate with responsiveness to in-vehicle interventions.
  - We further build machine-learning models to use data available before a drive has started to predict whether a driver is going to respond to or cancel the intervention prompted during the drive. We found that the model predicts whether a participant will start an intervention with an F1 score of 0.81, which is 24.6% greater than a biased-random baseline.
- We evaluate how *in-the-moment* driving situations relate with responsiveness to interventions.
- We identify situations that negate the positive impact of the interventions. We argue that well-being interventions could be triggered when such situations occur to maximize the interventions' effectiveness.
- Finally, we evaluate whether engaging with interventions while driving affects driving behavior.

2 RELATED WORK

In this section, we start by broadly discussing the *timing* of smartphone-based digital health interventions. Next, we discuss the recent status of in-vehicle interventions and the interruptibility of the driver. Finally, we summarize the current state of research and identify opportunities for enabling effective delivery of in-vehicle interventions.

2.1 Just-in-Time Adaptive Interventions

In recent years there has been a significant interest in using ubiquitous technologies to provide interventions related to health and well-being. One increasingly popular framework for delivering such health-related interventions is *Just-in-Time Adaptive Intervention (JITAI)*. JITAI aims to deliver the right amount and type of support, at the right time, while adapting to the users' internal and external contextual change [48, 49]. JITAI relies strongly on the concepts of *states-of-vulnerability*, i.e., moments with increased risk of negative health outcomes when a participant needs an intervention, and *states-of-receptivity*, i.e., moments where a person is able to receive, process, and use the intervention provided. In the UbiComp and HCI communities, there has been a growing research in sensing vulnerable conditions (like stress [25, 45, 46] and depression [44, 67]), and understanding receptivity to interventions [11, 39, 47, 59].

To understand receptivity to interventions, prior work has focused on identifying moments to deliver mHealth interventions on smartphones. Recent work by Künzler et al. noted that participants were less likely to start an intervention with a chat-bot while driving than during other activities, like walking [39]. Intuitively, smartphone-based interventions might not be suitable for use while driving. Given the amount of time an average person spends driving, however, recent research has adapted interventions to be delivered in the context of the car, e.g., audio-based interventions [8], odor-based interventions [17], or ambient-light-based interventions [24]. While in-vehicle interventions provide a unique opportunity to improve a person's well-being, it is important to deliver these interventions at times unlikely to distract the driver from their primary task of driving. Hence, a better understanding of situations when drivers are likely to be available to engage with an intervention could help to design interventions that reduce driver burden and maximize intervention effectiveness.

2.2 In-Vehicle Interventions

Early research on in-vehicle interventions suggested re-purposing existing comfort systems in the vehicle to react to and influence driver states. Hernandez et al. worked to to improve drivers' well-being by changing interior ambient lighting and air conditioning, or by adapting navigation routes (e.g., to avoid traffic jams) [24]. Recently, a major stream of research has focused on bringing guided breathing into the car as an in-vehicle well-being and stress-management intervention. Over several studies, Paredes et al. and Balters et al. developed and validated interventions based on guided breathing exercises for drivers [4, 53, 54]. Guided breathing has shown to help people to cope better with stress [61]. In their first study, Paredes et al. evaluated multiple in-vehicle interventions with 12 participants in a simulated driving environment, before deciding on the use of guided breathing interventions [53]. Next, they validated several breathing exercise patterns combined with a haptic vibrotactile seat cover and compared it to voice-guided breathing exercises, based on results from 24 participants driving on a simulator [54]. Finally, they demonstrated their intervention's effectiveness and compatibility in a study of driving behavior and physiological and psychological stress parameters with 24 participants driving a real car on a closed track [4].

In other work, Dmitrijs et al. explored the effectiveness of subtle, odor-based well-being interventions in the car [17]. In four laboratory studies, Dmitrijs et al. identified different scents for emotion regulation, tested them on drivers in a simulator and found out that pleasant scents can relax drivers, ergo making them safer drivers [17]. Braun et al. conducted a controlled, real-world driving study in which they personalized voice assistants to the driver's personality and driving situation, and found that extroverted people preferred a friendly tone in the assistant's voice – which led to more trust and was perceived as pleasant [8]. In related work, Harris et al. found that the communication style of a voice-based assistant was a relevant factor in its effectiveness to aid drivers in their driving experience [22]. Williams et al. developed the Affective Intelligent Driving Agent (AIDA), a social robot to assist drivers decrease their cognitive load and promote road safety [69]. Gusikhin et al. similarly developed the EmotiveDriver Advisor System (EDAS), focused on personalization and adaptive behavior [20]. Researchers have also explored the potential use of music for emotion regulation [38].

Beyond academia, driver well-being has also garnered interest in the automobile industry. In recent years, two large manufacturers revealed in-vehicle intervention concepts to improve the well-being of drivers. Audi announced the "Audi Fit Driver" concept, which leverages data from smartwatches to deliver interventions [3]. Mercedes already went a step further with offering an optional package called "ENERGIZING" in their cars. The feature uses driver-related data from a smartwatch and vehicle-related data from the car to deliver interventions by adjusting the air conditioning, offering to start seat-massage programs, or adjusting colors of ambient lights [13]. Unfortunately, these manufacturers have not published results about the effectiveness of their interventions.

2.3 In-Vehicle Interruptibility

An intervention delivered at the right moment increases the driver's ability to engage and interact with the intervention, and minimizes the need to burden participants with unwarranted prompts at non-opportune moments. Recent work has attempted to determine moments when drivers might be interruptible for in-vehicle tasks. In a controlled, real-world driving study Kim et al. found that drivers were more likely to engage in peripheral tasks like using the radio, while driving along flat and straight roads, thereby showing signs of lower workload [35]. Semmens et al. confirmed the previous finding related to straight roads, and further found that moments where the vehicle was still (not in motion) were favorable for interactions with drivers [60]. In their controlled, real-world driving study, the authors gauged interruptibility by randomly asking the drivers if that moment was a good time to talk, and the participants responded with a 'yes' or 'no.' The authors found that the 'yes' rate coincided with less brake change and without major steering activity [35]. Further, the authors found that deviating from the navigation system's intended route increased the chances of participants saying 'no.'

In other work, Kim et al. evaluated interruptibility by evaluating the drivers' ability to handle cognitively-demanding auditory-verbal tasks, specifically the n-back counting task, with varying levels of difficulty [33, 34]. In a series of simulated and controlled, real-world driving conditions, the authors found that they could detect interruptible (and safe) moments with an F1-score of 0.74 [33]. Further, they found that drivers' interruptibility relied on a lack of vehicle state changes, i.e., driving in straight roads with steady speed or standing still. The authors also report that drivers frequently overestimated their ability to dual-task, further highlighting the need for accurate models that detect interruptibility [33]. In a later study, Kim et al. found that a more-difficult secondary task led to lower interruptibility while driving, which decreased further when faced with increasingly complex maneuvers (stop, straight, turn, lane-change). The authors concluded that the demand for the secondary tasks and maneuver types negatively impacted the drivers' ability to dual-task [34]; something that could be considered for health-based intervention design to ensure that the intervention tasks are not challenging or demanding.

2.4 Summary of the Current State

To summarize, research on in-vehicle interruptibility is progressing at a rapid pace and shows promising results. This prior work, however, has several limitations. First, none of the work delivered actual interventions; the studies either prompted questions to the drivers [60], asked them to do some cognitive task [33, 34], or used a proxy to determine interruptibility [35]. Second, the data in these prior works was either collected in a simulated environment or in a controlled, real-world driving scenario, in which the participants were asked to drive on a predetermined route for a certain duration. Such controlled driving conditions do not account for the many natural variables in a person's daily drive. In unconstrained scenarios, drivers experience varying situations, such as routine commutes (home → work and work → home), non-routine trips (e.g., weekend outing or a holiday road-trip), challenging or unfamiliar driving conditions (e.g., bad weather or driving at night), or driving with passengers. Hence, it is important to understand when drivers interact with interventions in unconstrained driving conditions. Furthermore, none of the prior studies have conducted longitudinal experiments over the period of many weeks, which is important to sample a variety of trip and driving conditions as well as variations in the driver's affect.

Our work aims to address the aforementioned shortcomings by evaluating when participants interact with actual interventions in unconstrained, free-living conditions over an extended period.

3 METHOD

In this section we discuss the interventions used in the study followed by the study design. We then discuss the data collected during the study, and our analysis plans.

3.1 Intervention Design

We collaborated with psychologists and developed two types of interventions based on proven psychotherapeutic approaches [5, 19, 30, 41] to improve the well-being of participants in the context of driving a car: a *mindfulness* intervention and a *music* intervention. In our prior work we found that both in-vehicle interventions had a positive impact on the affective state of the drivers [36]. Although detailed explanation and evaluation of the interventions are available in our previous work, we briefly outline the intervention components in this section.

3.1.1 General Design Principles. Since the participants were going to drive on public roads, we implemented – for both interventions – four shared design considerations to ensure safety at all time. First, interventions were delivered at the beginning of the trip to ensure that drivers were not surprised or distracted by receiving an alert while driving. We followed this approach because there is little knowledge about the "right" time for interventions in unconstrained real-life driving conditions. Second, each intervention started with a short audio-based warning
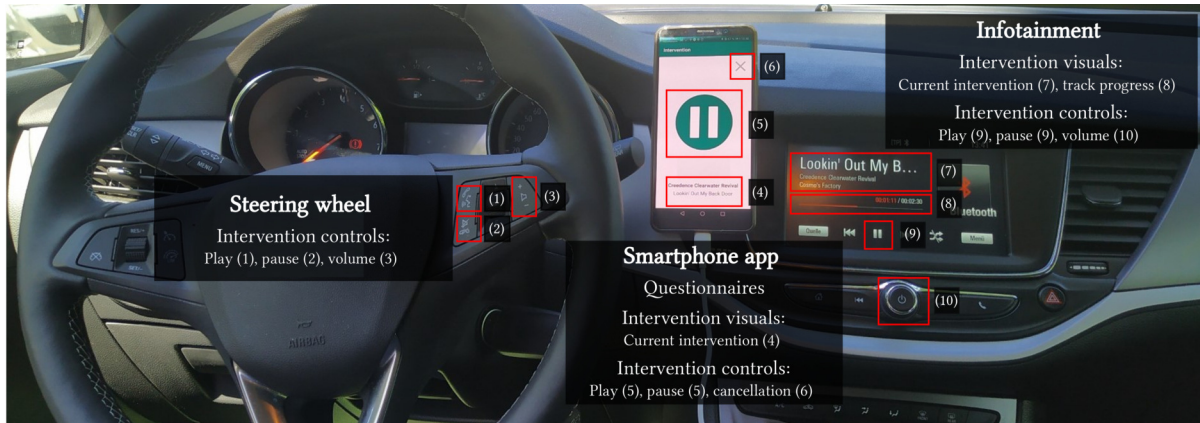
Fig. 1. Car interior with the study setup. Left: Steering wheel buttons to control the intervention playback. Middle: Smartphone with app to answer the questionnaires and to control the interventions. The snapshot shows the app currently playing a music intervention. Right: Car infotainment showing the running intervention.

message. The message reminded the driver to concentrate on the street and to only conduct the intervention if their current driving task would allow it. Third, participants controlled the flow of the intervention. The participants decided when to start, pause or cancel the intervention by simple control elements. And fourth, all interventions were completely audio-based to avoid any visual distraction while driving.

3.1.2 Mindfulness Intervention. Mindfulness exercises have demonstrated consistent effects on improving the overall well-being of people [19, 66]. We thus decided that the first intervention should be a mindfulness exercise. We designed an exercise based on the concept of passive mindfulness, in which the listeners are guided to perform activities more consciously and to increase their perception of the present (e.g. climbing stairs, cooking or showering) [23, 31].

The mindfulness intervention's content and structure followed an instruction-based approach in which a narrator guided the drivers through two major alternating content blocks to conduct the exercise. *Environment*-centric blocks aided drivers to focus on the driving task, the car, and the surrounding traffic. *Self*-centric blocks supported drivers to reflect on their perception of their feelings and emotions. We show an example for both elements in Table 1. Each block consisted of three to six parts separated by longer pauses and were slowly narrated to allow a better concentration on the mindfulness exercise while driving. The intervention started by first playing an environment-centric block, followed by a self-centric block. After a repetition of both the blocks, a last environment-centric block ended the exercise. In total, the mindfulness intervention lasted about 14 minutes. Throughout the study, we used this single exercise, as existing research suggests to repeat a mindfulness exercise several times to unfold the effects [30].

3.1.3 Music Intervention. For the second intervention, we had the participants create an individual playlist with a minimum of ten songs before starting the study. Each song on the playlist had to be explicitly annotated with a positive experience of their life. Our approach follows the evidence that certain songs can induce positive emotions by recalling connected positive autobiographical memories [29]. In the realm of psychotherapy, music is a proven concept [41], and our more specific approach shows its impact in psychotherapeutic practice through anecdotal evidence [5].

Table 1. Examples for both mindfulness content blocks (translated).

Block	Text passage
Environment-centric	"Now focus your attention again on your surroundings, the traffic and the routines in your car. What do you see before you? What is the scene in front of your windshield?" […]
Self-centric	"Now expand your attention to perceive any body feelings you experience. Take a moment to pause and simply allow yourself to observe all the sensations that are present in your body without judging them good or bad and without trying to change them in any way. Just perceive with openness and curiosity what is there." […]

Table 2. Sample excerpt of a playlist for the music intervention by one participant (translated).

Song	Artist	Connected memories
Anywhere	Passenger	"Our marriage vow (song quote: 'I will go anywhere with you')"
You Can't Look Back	Taking Back Sunday	"Song helped me through a difficult time (strokes of my dad)"
Full Steam Space Machine	Royal Republic	"Pushes me a lot when I am stressed"

On average, each participant had 15.6 songs on their playlist. Table 2 shows a excerpt of one participant's playlist. At the beginning of each music intervention, the intervention reminded the drivers by a short voice-based instruction to enjoy the music and recall their memories. This instruction made no explicit links to specific memories. After this short instruction, the playback of the driver's playlist started. The intervention lasted a minimum of 9 minutes or approximately three songs. The intervention terminated once the total playing time exceeded 9 minutes, and the current song ended.

3.2 Study Setup

Prior works have explored and evaluated car-based interventions, and how drivers react to such interventions in simulated or controlled driving conditions (like a pre-determined route/circuit for a short duration). The goal of our work is to investigate when people react to in-car interventions during their regular day-to-day driving on unconstrained public roads.

To this end, we conducted an IRB-approved study with 10 drivers and equipped them with one car each for two months to deliver the interventions during their daily drives. The cars used in the study were prepared for data collection in two ways: a) we installed equipment to record the cars' Controller Area Network (CAN) bus data and b) we retrofitted a smartphone, next to the car infotainment screen, to prompt self-reports, deliver interventions, and collect driver feedback. We illustrate the in-car setup in Figure 1. In this section, we describe the background of the participants, the car's configuration, how we measured interventions' outcome, and how the interventions were delivered.

3.2.1 Participants. To seek participants for our study, we published a call for participation in the social network of a large German company. A total of 54 people responded to the call for participation by completing a survey

Table 3. Summary of study cohort and their driving habits.

n	Age (years)	Gender	Residence	Driving per week (km)	Driving days per week	Distance to work (km)	Driving experience (years)
10	M = 37.2	Female = 4	Urban = 3	M = 327.0	M = 6.8	M = 22.6	M = 20.2
	SD = 8.4	Male = 6	Rural = 7	SD = 103.4	SD = 0.4	SD = 7.0	SD = 8.1
	min = 26			min = 100	min = 6	min = 14	min = 9
	max = 55			max = 500	max = 7	max = 35	max = 38

about their personal circumstances and driving habits. We followed a purposive sampling strategy by weighting demographics, life and family situations, and driving habits to sample a diverse set of daily commuters. We had a final selection of 10 participants (4 female), who could be considered fairly typical commuters. All participants had many years of driving experience and could be expected to interact responsibility while driving. All 10 participants cited the daily commute as the reason to own a car, and they also used their car for daily errands and leisure activities over the week. During the entire two months of study we provided them with a car, including insurance and unlimited mileage. We did not reimburse for fuel costs, and did not provide any monetary incentive to participate in the study, drive the car, or respond to interventions. Table 3 shows a summary of the participants' demographics and driving habits.

3.2.2 Collecting Car Data. Each of the study cars had a dedicated CAN bus recording system that we installed before handing the cars over to the participants. As a central unit, we used an industrial-grade computing system (Compulab IOT-GATE-iMX7) with Linux (Yocto) as the operating system [12, 70]. An external 1-terabyte solid-state disk (SSD) drive was connected to the system so that all car sensor-based data could be recorded over the entire two-month study period. The system passively listened to all messages on two CAN Low and High wires by using a PEAK PCAN-USB Pro FD and two CAN crocodiles. A custom-written software continuously read the raw signal on both busses and deciphered all messages. The observed CAN-busses contained powertrain (e.g., engine and steering wheel), vehicle dynamics (e.g., speed, acceleration, and GPS), and comfort (e.g., air-condition and seat belt usage) messages.[1] Although we recorded all signals in high frequency (> 1 MHz), our subsequent analyses followed best practices with all data re-sampled to 10 Hz [21]. The system booted as soon as the ignition was on and started the recording after 30 seconds to 1 minute. To prevent data loss and reliably stop all data recording, the system was kept alive by a connected Bicker battery system (control unit: UPSI-1208D, battery: BP-LFP-1025D) after the ignition was turned off. After two minutes on battery power, all systems stopped recording and initiated a planned shutdown.

3.2.3 Affective State as a Proxy of Well-being. Empirical evidence indicates that higher levels of well-being lead to higher resilience, and thus contribute to the prevention of mental illness [14, 58]. However, "well-being" is a broad concept as it is interconnected with physical, mental, and social states. In this study we follow the approach of other works and concentrate on *affective* well-being, which is a specific aspect of "well-being" and closely related to mental health [2, 14, 15]. Affective well-being describes individual experiences in everyday life through the frequency and intensity of various emotions such as joy, anger, or affection [14, 68].

Usually, affective well-being is measured by people responding to surveys that assess their affective state at different points in time. In empirical research, the circumplex model is often used [57]. It allows the measurement of the affective state along two dimensions: arousal and valence. Arousal indicates the degree of activation, i.e.,

---

[1]A full list of the collected CAN signals are available in the supplementary material.
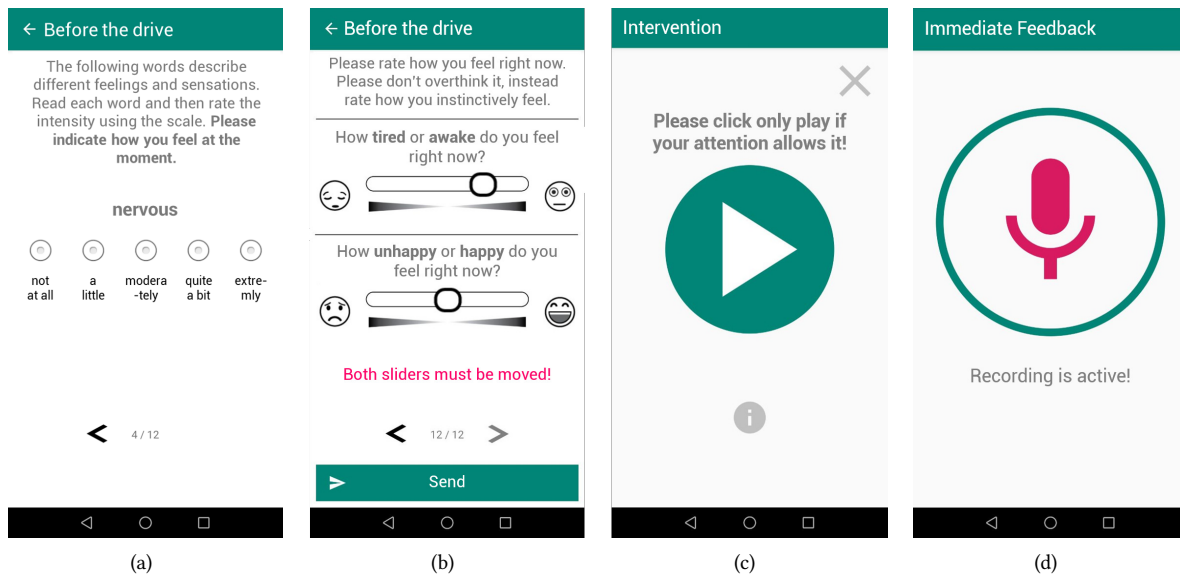
Fig. 2. In-vehicle study smartphone app (screens translated). (a) and (b) show the pre- and post-driving questionnaire screens for PANAS-SF and Affective Slider; (c) shows the intervention controls; and, (d) shows the recording screen for the immediate impact voice questionnaire after an intervention.

how tired or energized someone is, and valence measures the degree of happiness. The composition of different arousal and valence levels can describe more granular emotions such as excitement (high arousal and high valence) or being angry (high arousal and low valence). The circumplex model has two advantages: first, the measurement is rather effortless for participants; and second, arousal and valence measurements are used in many contexts, e.g., for emotions [7], stress [37, 51], and even depression [56].

3.2.4 The In-Vehicle Protocol. In this sub-section, we describe the in-vehicle protocol followed by the participants in each drive. It includes pre- and post-drive affective state self-reports, the actual intervention, and a voice-based feedback to gauge the effectiveness of the intervention.

Pre- and Post-Driving Affective State. For each drive, we asked the drivers about their pre-driving (or baseline) affective state and a post-driving state. The participants had to fill out a self-report using the study smartphone (retrofitted in the car) before and after each driving session. The self-report questionnaire included PANAS-SF (Positive and Negative Affect Schedule–Short Form) [63, 68], followed by the Affective Slider [6] to assess their current affective state. The PANAS-SF is a more concise version of the PANAS questionnaire, in which the participants are asked about their current feeling of five positive affects (attentive, active, alert, determined, and inspired) and five negative affects (upset, afraid, nervous, hostile, and ashamed). The combination of those emotions reflect a more granular estimation of arousal and valence levels. The participants' ratings for each affect could range from 0 ("not at all") to 4 ("extremely") on a Likert scale. In contrast, the Affective Slider, as a modern variant of the Self-Assessment Manikins [7], provides a simpler approach to measure arousal and valence, and is especially suited for modern user interfaces and devices like smartphones. Participants could rate their arousal and valence levels on two continuous scales from 0 (very low) to 100 (very high) with the Affective Slider. Figures 2(a) and 2(b) show screenshots of both questionnaires.
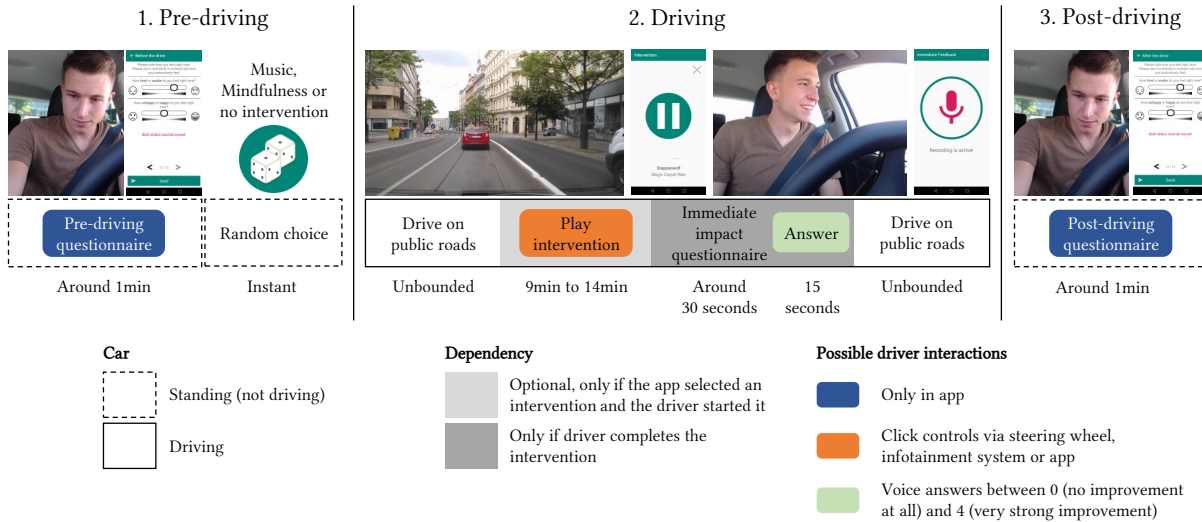
Fig. 3. Study procedure for each trip (screens translated).

The Intervention Delivery. Immediately after the pre-driving self-report, the smartphone would randomly choose among three intervention cases: music intervention, mindfulness intervention, or no intervention. If an intervention option is selected, the smartphone would present a *play* button (as illustrated in Figure 2(c)) to the participant. Participants could start the intervention by clicking on the play button at any time during the drive. Only when the participants click on the play button do they learn the type of intervention (music or mindfulness). For safety purposes, the participants could always pause or cancel a running intervention during the drive.

Immediate Impact on Affect. To evaluate the effectiveness of the interventions, the smartphone would directly ask a simple voice-based questionnaire after participants completed the intervention. The questionnaire consisted of two questions about arousal and valence improvements, based on the Client Oriented Scale of Improvement (COSI) [16]. Explicitly, the app asked the following two questions (translated): "After the intervention, do you feel more aroused (respectively happier, on the second question) than before?" After asking the questions, the smartphone let drivers answer verbally, and recorded the responses. The participants could rate the intervention impact on both dimensions via a Likert scale from 0 ("no improvement at all") to 4 ("very strong improvement"). The smartphone screen during this phase is depicted in Figure 2(d).

Overall, the in-vehicle protocol can be separated into three segments: pre-driving, driving, and post-driving. In Figure 3, we illustrate the three segments with the tasks involved in each.

3.3 Collected Data

We recorded a total of 1,009 trips during the study. Of the 1,009 trips, the smartphone app delivered based on the previous explained random intervention delivery strategy no intervention for 325 trips and either a mindfulness or music intervention for the remaining 684 trips. Given the in-the-wild nature of our study, we experienced some CAN recording outages (e.g., due to problems with the hard drive), resulting in a total of 893 trips with complete data (293 with no interventions, and 600 with interventions). Of the 600 instances where the smartphone prompted an intervention, the participants started the intervention in 402 instances (67% response rate; 214 mindfulness interventions and 188 music interventions). Of the started interventions, the participants canceled 78 mindfulness interventions (37.5%) and 19 music interventions (10.9%); and 26 interventions (9 mindfulness and
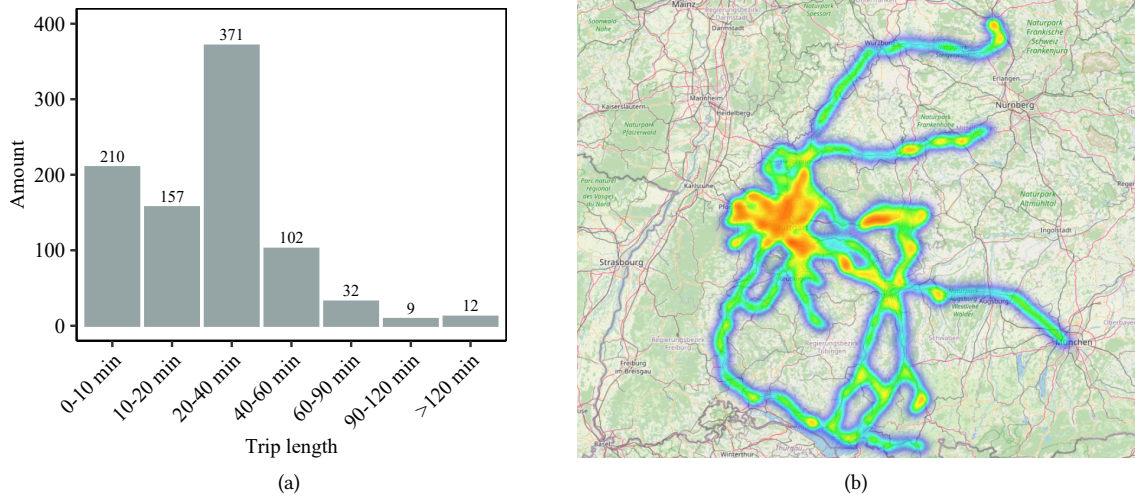
Fig. 4. Trip characteristics. (a) shows the number of trips by trip length intervals, aggregated for all participants; and, (b) shows a GPS heatmap visualizing the trips of all participants in our study.

17 music) were inadvertently closed due to app crashes. Finally, the participants completed 285 interventions (130 mindfulness and 155 music).[2] The complete breakdown is summarized in Table 4.

Further, we provide insights about the number of trips and interventions completed by each participant in Table 5. On average, each driver completed 100.9 trips during the two-month study period. Our app presented each driver an average of 68.4 interventions, which is comparably split between mindfulness and music.

We conducted the study between September and November 2019, during which the participants drove for more than 340 hours and 17,000 kilometers. The participants drove uninstructed on public roads, mostly around the Greater Stuttgart area, and also made several trips across southern Germany, as shown in Figure 4(b). An average trip lasted 22.4 minutes (SD=19.5 minutes) and 18.6 kilometers (SD=25.2 kilometers). On average, participants started the interventions within 1.2 minutes (SD=4.9 minutes) after they started driving. The average drive time remaining after an intervention started was 16.1 minutes (SD=17.5 minutes). We illustrate variation in trip lengths in Figure 4(a).

3.4 Analysis Plan

The goal of our work is to understand when users interact with in-vehicle interventions in daily driving conditions. Our analyses can be divided into four broad categories:

(1) *Impact of general trip characteristics and pre-trip affective state*: We evaluated how certain general trip related information, i.e., information available before a participant starts driving, and a driver's internal affective state relate with drivers' responsiveness to the upcoming in-vehicle interventions. Some examples of trip characteristics include the time of the day, weather, trip distance, and traffic flow.

---

[2]We note that none of the drivers *completed* an intervention before they started driving. They either started driving after starting an intervention or started an intervention at some point they found suitable during their drive.

(2) *Impact of situational driving characteristics*: We evaluated how various *in-the-moment* driving variables, e.g., speed of the car, brake usage, number of stops, sudden brake events, and steering movement, relate with drivers' responsiveness to interventions.

(3) *Identifying "negative" driving situations*: We analyzed and identified driving situations which negate the positive impact of the interventions. These situations could be used as *markers* for delivering in-vehicle affective interventions to maximize the intervention's effectiveness.

(4) *Impact of in-vehicle interventions on driving behavior*: We evaluated whether engaging with interventions while driving affected the participants' driving behavior.

For analyses (1) and (2), we operationalize *responsiveness* to interventions with two variables:

- *Response rate*: The ratio of interventions started by the drivers over all presented interventions.
- *Cancellation rate*: The ratio of interventions explicitly cancelled by the drivers over all the started interventions.

Since we had repeated measurements from participants, our analyses used mixed-effects models to inspect the effect of the various independent variables, with the *participant_id* as the random effect. Given the exploratory nature of our analyses, we calculated several independent variables to capture the complexity of driving behavior, thus increasing the chances of multicollinearity. To address this issue, in all our analyses, we iteratively removed variables with high Variance Inflaction Factor (VIF), i.e., VIF > 10, from our model, as recommended by James et al. [27].

Table 4. Summary of responsiveness to interventions. The percentage of total number of trips is presented in brackets.

	Total Count	Not Started	Started	Cancelled	Completed
Trips	1,009 (100%)				
Intervention / no intervention	684 (68%) / 325 (32%)				
Trips with CAN data	893 (89%)				
Intervention / no intervention	600 (59%) / 293 (29%)				
Trips with CAN data and interventions	600 (59%)	198 (20%)	402 (40%)	97 (10%)	285 (28%)
of those mindfulness / music	-	-	214 (21%) / 188 (19%)	78 (8%) / 19 (2%)	130 (13%) / 155 (15%)

Table 5. Overview of trips and interventions across the participants.

	Mean	CI-95%	Lower limit	Upper limit
Completed trips	100.9	±20.4	80.5	121.3
Delivered interventions	68.4	±14.4	54.0	82.8
Mindfulness	36.1	±8.7	27.4	44.8
Music	32.3	±6.0	26.3	38.3

4 EVALUATION

In this section, we analyze and present our findings on when users interact with in-vehicle interventions, based on the four categories mentioned in Section 3.4.

4.1 Impact of General Trip Characteristics and Pre-trip Affective State

In this section, we evaluate how various pre-trip factors that could be known at the beginning of a trip relate to the drivers' responsiveness to in-vehicle interventions. These factors can be broken into two types: *extrinsic factors*, which include "external" trip information such as time, weather or route, and *intrinsic factors* which describe the drivers' "internal" affective state before the start of the trip. We present a list of the factors in Table 6. We argue that a driver is consciously or subconsciously aware of various factors used in our analysis. When commuting between home and work, people can typically estimate the expected traffic flow, the route they are taking, time spent on highways, and their estimated commute time since this is part of their daily routine. The same also applies to other routine trips (e.g., going for groceries), where several trip-related factors can be known apriori. When driving to completely new (or less-frequented) destinations, it is reasonable to assume they use the navigation system in their car or on their phone, thus getting the relevant trip information.

First, we detail how we calculated the various variables for this analysis.

- *Time related variables*: These were calculated trivially by using the time at which a participant started driving.
- *Weather related variables*: We used a publicly available API to determine the weather at the time and location at which a participant started driving [52]. The weather was categorized as sunny, foggy, or cloudy. In addition to the general weather, we used the CAN data to check if wipers were active during the trip to determine if it was raining.
- *Length of trip*: We used the CAN data to calculate the *total time* taken for the trip and the total *distance* of the trip. We also included a boolean variable to capture if the trip length was longer than the intervention duration.
- *Trip-condition related variables*: Using the CAN data, we calculated the number of times the car stopped during a trip, or if there were other occupants in the car. In addition, we calculated the *traffic flow* for a trip as the ratio of the actual driving speed to the allowed speed limits of the driven roads. We determined the speed limit by querying a publicly available TomTom map service based on the GPS locations of the trips [64]. To account for the repetition of interventions, we included the variable to count the number of times interventions were prompted (for response rate) and started (for cancellation rate) to the participant until that time.
- *Route related variables*: We used the GPS locations from the CAN data to calculate the percentage of time a participant spent on different road types. We again used the TomTom map API to determine road types: minor roads (e.g., parking lots, or one-lane streets), major roads (e.g., major urban roads, or larger countryside roads) and highways (or expressways) [64]. Further, we used the start and end locations for each trip and applied the DBSCAN algorithm to generate location clusters [18]. We generated clusters with a radius of 100 meters and a minimum of 10 samples per cluster. The rationale for 10 samples was to identify frequently visited locations, e.g., grocery store or a favorite restaurant. We marked the most frequently occurring cluster at the starting location of the first drive of each day and the ending location of each day's last drive as *home*. We marked the most frequently occurring cluster at the ending location of drive till 10 a.m. (with no trips for the next two hours, to account for situations where participants might stop by a store or a bakery on their way to work) and starting location between 2:30 p.m. and 5:30 p.m. on weekdays as *work*. We marked all other clusters as *others*.
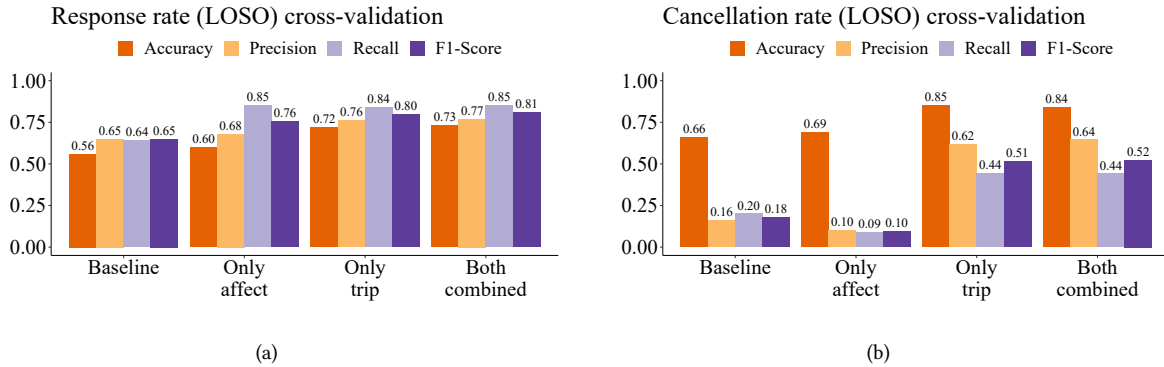
Fig. 5. Prediction results of intervention interactions. In (a), the model predicts whether drivers will start an intervention (response rate); in (b), the model predicts whether drivers will cancel a started intervention (cancellation rate).

To account for routine trips, we merged the location information into larger categories. We included commute variables (home → work, and work → home). We also included a variable to account for *regular* trips, i.e., trips between the identified location clusters that were not "commutes." Finally, trips where the start or end location did not belong to any of the clusters were marked as "others."

• *Affective State*: We used the participant's response to the Affective Slider (AS), and the PANAS-SF questionnaire at the start of the drive to model the participant's internal state to observe if that influences the participant's responsiveness to the intervention.

We report the results of the mixed-effect models for the three outcome variables (response rate, mindfulness cancellation rate, and music cancellation rate) in Table 6. For clarity, we simply highlight the positive or negative associations of the independent variables with the outcome variables. We include the complete model output in supplementary materials.

From Table 6, we observe that participants were more likely to cancel mindfulness interventions in foggy conditions ($p = 0.019$). Next, trip duration had a significant positive association on response rate and a negative association with music intervention cancellation, i.e., participants were more likely to respond in longer trips ($p = 0.041$), and less likely to cancel music interventions during longer trips ($p = 0.028$). Next, we found that if the trip duration was greater than the intervention duration, the participants were more likely to respond to the intervention ($p < 0.001$). This is an interesting finding that suggests that the participants had possibly acclimatized themselves to the intervention duration and hence did not start the intervention if they thought that the trip length duration was shorter than the intervention. Nevertheless, during longer trips the participants were less likely to cancel the mindfulness ($p < 0.001$) or the music interventions ($p < 0.001$).

Further, we observed that traffic flow had a significant positive association with response rate ($p = 0.028$), i.e., fewer traffic congestions led to higher response rate. Next, we found that the presence of passengers in the car significantly reduced the likelihood of responding to the intervention ($p < 0.001$), and significantly increased the likelihood of canceling the mindfulness intervention ($p = 0.02$). Interestingly, the presence of passengers did not have any association with the cancellation of the music intervention. Further, the number of prior interventions had a negative association with the response rate ($p = 0.029$), i.e., the response rate to interventions reduced as the study progressed.

Next, we found that participants were less likely to cancel the music interventions if the drive included a higher percentage of major roads ($p = 0.003$) or highways ($p = 0.01$) when compared to minor roads. Further,

Table 6. Effects of general trip characteristics and affective state on in-vehicle intervention responsiveness. A detailed table with all the variables and the respective coefficients is available in the supplementary document.

			Response rate	Cancellation rate	
			Both	Mindfulness	Music
TIME	Weekend	vs. workday			
	Morning / 5am - noon	vs. rest of day			
	Evening / noon - 8pm	vs. rest of day			
	Daylight	vs. dark			
WEATHER	Fog	vs. cloudy		+*	
	Sunny	vs. cloudy			
	Rain	vs. none			+^
LENGTH	Time	in min.	+*		−*
	Distance	in km			
	Trip length > intervention duration	yes/no	+***	−***	−**
CONDITION	Traffic flow	ratio	+*		
	Stops	count per min.		+^	
	Other occupants	vs. none	−***	+*	
	Number of interventions	count	−*		−^
ROUTE	% major road time	0 to 1			−**
	% highway time	0 to 1			−*
	Regular trip	vs. other trip			
	Commute home → work	vs. other trip	+*		
	Commute work → home	vs. other trip	+*		
AS	Arousal	0 (low) to			
	Valence	100 (high)			
PANAS-SF +	Active				
	Alert	0 (not at all)			
	Attentive	to			
	Determined	4 (extremely)			
	Inspired				
PANAS-SF -	Upset			−*	
	Afraid	0 (not at all)			
	Nervous	to	+^	+*	
	Hostile	4 (extremely)			
	Ashamed			+^	
	n (trips)		402 started vs. 198 not started	78 cancelled vs. 139 completed	19 cancelled vs. 169 completed
	cond. R^2		0.300	0.585	-
	marg. R^2		0.254	0.446	0.215

+ = significant increase, − = significant decrease; ^: $p < 0.1$, *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$

participants were more likely to respond to the interventions during their daily commute, i.e., home → work (p = 0.018) and work → home (p = 0.025). This result suggests that daily commute is a good time to deliver interventions, since participants are more likely to start the intervention. Finally, we found that if the participants were feeling upset, they were less likely to cancel the mindfulness intervention (p = 0.02), but more likely cancel the intervention if they were nervous (p = 0.049).

4.1.1 Predicting Responsiveness Using General Trip Characteristics and Pre-trip Affective State. Since the statistical analysis revealed several significant associations between the general trip characteristics and the driver's pre-trip affective state towards the responsiveness to interventions, we decided to test the feasibility of building machine-learning models for this task. We used Random Forests to build models to predict the participants' response and cancellation.[3] To understand how different factor groups influence model prediction, we built three types of models, (a) using just the affective states of the participant, (b) using just the extrinsic trip characteristics, and (c) using a combination of affective states and extrinsic trip characteristics. The reasoning is that the approximation of trip characteristics such as route or traffic can be achieved by navigational systems or analysis of past travel patterns [26]; detecting the affective state, however, is a more challenging task and is an area of active research [55]. Our models are only based on the features of our prior statistical analysis; we added no additional information.[4]

To evaluate model performance, we conducted leave-one-subject-out (LOSO) cross-validation. To compare the model performance with a baseline, we implemented a biased random classifier, randomly predicting the outcome based on the distribution of labels in the training set (e.g., if the intervention was "started" in 67% of the cases in the training data, then the baseline classifier would randomly predict "started" with 67% probability). We present the results in Figure 5.

We observed – for both model types (response and cancellation) – that the performance by using only the trip characteristics was similar to the performance achieved by a combination of trip characteristics and affective state. When predicting whether a participant would respond to an intervention prompt, the combined model had an F1-score of 0.81, which was 24.6% higher than the baseline F1-score of 0.65. Similarly, when predicting whether a participant would cancel a started intervention, the combined model had an F1-score of 0.52, which was 188.9% higher than the baseline F1-score of 0.18. We present other metrics for evaluation, like accuracy, precision, and recall in Figure 5. To present a balanced outlook, our models were optimized for high F1-score. Depending on the intervention design, the model could be optimized to high recall (risk of higher false positives) or high precision (risk of finding fewer trips to deliver interventions).

We believe that all trip-related factors (distance, duration, traffic flow, weather, occupants, etc.) can be effectively estimated even before a person starts driving, by leveraging a combination of driving history, smartphones (or navigation system), and other in-car sensors. Our results therefore suggest that future intervention systems could predict a person's responsiveness to in-vehicle interventions even before the person starts driving.

4.1.2 Inspecting Model Explainability. To further understand the reasoning behind the machine-learning models' predictions, we investigate the SHAPley Additive exPlanations (SHAP) values for the combined (trip characteristics and affective states) models [42, 43]. Figure 6 summarizes the SHAP values. The plot explains how a prediction is influenced by the features in our dataset. It contains two major types of information: (1) the y-axis lists the features from top to bottom in descending order of importance, and (2) the x-axis shows how each feature value influences model prediction. Each 'dot' represents a data-point from our dataset, and their position on the x-axis informs the degree of positive or negative impact the feature values have on the final model outcome. The color represents the range of the feature values, with 'blue' dots representing the lowest value of the feature and the

---

[3]Random Forests are simple, yet robust classifiers as they perform reasonably well without over-fitting. Further, Künzler et al. demonstrated the effectiveness of using Random Forests in detecting receptivity to Just-in-Time Adaptive Interventions (JITAI) [39].

[4]For cancellations, we combined cancellation of mindfulness and music interventions into one model. We included an additional variable to represent the type of intervention.
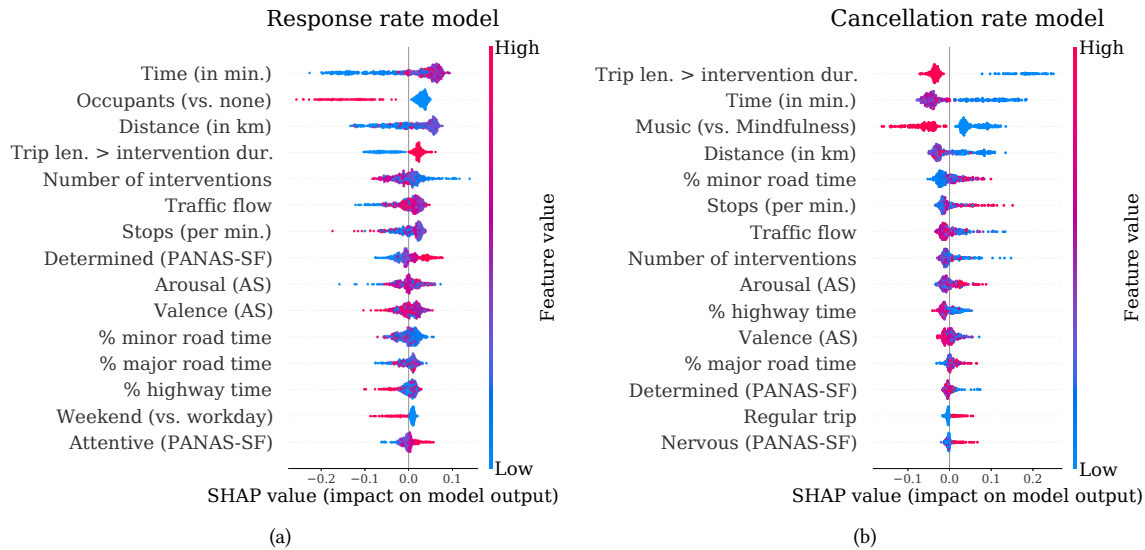
Fig. 6. SHAP values of the 15 most important features for the combined models. (a) shows the SHAP values for the response prediction model and (b) shows the SHAP values for the cancellation prediction model. The y-axis lists the features from top to bottom in descending order of importance, and the x-axis shows how each feature value influences model prediction.

'red' dots representing the highest value of the features. Data samples with equal SHAP values are vertically stacked for better readability. For example, in Figure 6(a), high 'Occupants' value (i.e., presence of occupants) has a negative impact on response rate, suggesting a lower likelihood of response. Using these characteristics of the SHAP plot, we can investigate what data was most relevant for a prediction within our machine-learning models, providing further insights into the responsiveness towards in-vehicle intervention. We found some variables like time, traffic flow, and presence of other passengers showed significant effects in the statistical analysis as well as the model prediction. It is interesting, however, to observe that variables like *Determined*, *Arousal*, and *Attentive*, which did not show a significant association in the statistical analysis, show high importance and good separability in the machine-learning model to predict whether participants will respond to the intervention prompt. Further, commute variables – which showed significant associations during statistical analysis – did not influence the model output. This analysis shows that – beyond the linear relationships between the independent variables and responsiveness metrics – non-linear relations exist, e.g., the commute variable can potentially be substituted by a combination of time of day, day of the week, and % of time on the different road types. Thus, the machine-learning models suggest that a decision to start an intervention could depend on the interplay of several factors.

4.2 Impact of Situational Driving Characteristics

In this section, we evaluate how various dynamically changing in-the-moment driving characteristics related to responsiveness to interventions. Initially, we planned to include all the 402 started interventions for this analysis. However, we noticed some challenges with how participants interacted with the interventions. We found that most of the time, the participants started the intervention before they started driving. Since the goal of this analysis is to evaluate how "driving" characteristics relate to the participant's choice to start or cancel an intervention, we could not analyze all started interventions. Thus, we only considered interventions that were

started or canceled *after* the car started moving, resulting in a total of 103 started intervention prompts and 59 canceled interventions.

To effectively evaluate the momentary characteristics of when participants started an intervention prompt, we followed a window-based approach. We marked the time when the interventions were started as *started* events. We considered a 15-second window before each started event, to reflect their most recent driving experience. In addition, we also considered a 15-second window after the started event, to account for the "to be expected" driving situations. We used the two different windows to calculate several situational characteristics (like acceleration and brake events) and used those characteristics as the independent variables for our analyses.[5] The intent was to find associations by accounting for factors the participants had recently experienced and the factors the participants were about to experience. To compare, we sampled every minute after driving start and before the participant started the intervention and marked them as *waiting-to-start* events. We then considered 15-second windows around these waiting-to-start events. In total, we had 103 started events and 486 waiting-to-start events.

Similarly, to evaluated when participants decided to cancel an intervention; we calculated driving characteristics in the 15-second window before and after each *canceled* event. To compare, we marked each minute between the intervention start time and cancellation time as an *on-going* event, and calculated 15-second windows before and after the on-going events. For our analysis, we had a total of 59 canceled events and 689 on-going events.

In addition to evaluating the association of momentary driving variables on starting and canceling the interventions, we also decided to evaluate moments the drivers thought were suitable for an intervention. We assume that if the drivers did not find the current situation suitable, they would have waited to start the intervention, and if they already started an intervention and encountered unsuitable moments, they would have canceled the intervention. For each minute, we marked the event as *suitable*, if the intervention was running at that time, or *unsuitable*, if the intervention was not running at that time. We followed the aforementioned method of considering 15-second windows before and after each event to capture the driving characteristics. Further, if an intervention running for at least 1 minute was canceled, we assumed that the upcoming times were unsuitable. We thus marked the 10 minutes after such cancellations as *unsuitable* and applied the same window calculation method as before mentioned. For clarity, we diagram how the window were calculated in each of the three analyses in Figure 7.

For all three analyses in this section, in addition to the situational driving characteristics calculated in the 15-second windows, we also included variables to represent the *current state* of the trip and the car at the time of an event, e.g., % of trip distance passed, type of road (highway, major, minor), speed, if the vehicle was accelerating, or if the vehicle had any lateral acceleration (i.e., going through a curve). We report the results for the analyses in Table 7. For brevity, we only include the variables that showed some significant association in Table 7. The complete output of the mixed-effects model is provided in the supplementary material.

From Table 7, we observe, for *starting an intervention*, that participants were more likely to start at times the vehicle was not accelerating ($p = 0.005$). Next, participants were less likely to start if some turn signal events (suggesting turns or overtaking) had recently occurred ($p = 0.039$). We also found that participants were more likely to start the intervention after sudden brake events ($p = 0.004$; potentially resulting in a drop in speed, or maybe even a stop, thereby making the participants more likely to interact).[6] Further, we observed that participants were less likely to start an intervention if there were upcoming lateral acceleration events ($p = 0.011$), i.e., driving through a curve or bend. Further, participants were more likely to start an intervention after higher acceleration ($p < 0.001$) or higher deceleration periods ($p = 0.006$) or they expected to encounter higher acceleration ($p < 0.001$) or deceleration periods ($p = 0.002$) in the next 15 seconds.

---

[5]We provide a detailed list of the various independent variables considered for analyses in this section in the supplementary material.
[6]For certain events like pressing the gas pedal, or braking, we included two levels of measurements. The first was the regular event, i.e., pressing on the gas pedal or pressing the brakes; the other was to measure a sudden increase in the intensity of these signals, i.e., sudden acceleration or sudden braking. We used a peak-detection algorithm on different CAN data elements to determine such "sudden" events.

Table 7. Effects of situational trip characteristics (interactions with interventions only while already driving). A detailed table with all the variables and the respective coefficients is available in the supplementary document.

		Start Intervention	Cancel Intervention	Suitable (vs. unsuitable periods)
		Both	Both	Both
CURRENT SITUATION	% of trip distance		+***	−***
	Highway (vs. major road)			+***
	Accelerate (vs. no change)	−**		
	Lateral activity (vs. none)		−**	
DRIVING BEHAVIOR AND VEHICLE SITUATION (last 15 sec.)	Low speed events			−^
	Low beam ratio			+***
	Turn signal events	−*	−*	−**
	Sudden gas pedal events			−^
	Sudden brake events	+**		
	Avg. speed			+***
	Avg. acceleration	+***	+***	
	Avg. deceleration	+**	+*	
	Avg. steering angle			−**
DRIVING BEHAVIOR AND VEHICLE SITUATION (next 15 sec.)	Steering events		+*	
	Stopped events		+**	
	Low speed events	+^		
	Turn signal events		−^	
	Brake events		−*	
	Lateral acceleration events	−*	+^	
	Sudden brake events	−^		
	Sudden steering events		+***	
	Avg. acceleration	+***	+*	
	Avg. deceleration	+**		
	Avg. steering angle		−***	
	n (windows)	103 started vs. 486 waiting to start	59 cancelled vs. 689 on-going	3,953 suitable vs. 766 unsuitable
	cond. R²	0.418	0.365	0.167
	marg. R²	0.390	0.363	0.075

$+$ = significant increase, $-$ = significant decrease; ^: $p < 0.1$, *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$; / = removed variable due to multicollinearity
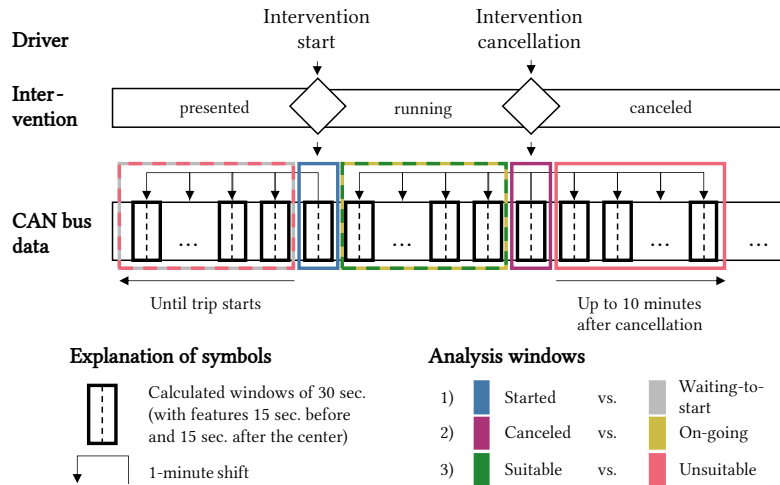
Fig. 7. Window calculation for an example trip where an intervention was started and subsequently canceled.

For *canceling an intervention*, we observed that participants were more likely to cancel interventions later in their trip, and if they were not experiencing any lateral acceleration the time of cancellation. Like starting the intervention, the participants were less likely to cancel if some turn signal events ($p = 0.015$) had recently occurred. Since starting and canceling the intervention require similar actions (i.e., clicking on the smartphone), it is reasonable that the participants are less likely to do such actions if they have recently made a turn or overtook someone. Further, we observed that the participants were more likely to cancel if they expected future steering activity (including sudden steering events, $p < 0.001$) or wanted to stop the car ($p = 0.002$). However, if the expected steering angle was high, e.g., turning at an intersection or parking, they were less likely to click on cancel ($p =< 0.001$; possibly because they were engaged in a demanding driving activity which did not permit any other action).

Finally, we analyzed *suitable vs. unsuitable periods* and found that the interventions were more likely to be running early in the trip ($p = 0.001$). Further, the interventions were more likely to be running on highways than major or minor roads ($p < 0.001$). Finally, we found that interventions were more likely to be running if the participants experienced higher low beam ratio ($p < 0.001$) (i.e., the fraction of time low beam was turned on), fewer turn signal events ($p = 0.003$), higher average speed ($p < 0.001$), and lower steering angles ($p = 0.006$). These results suggest that participants were more likely to be running the interventions if they were not experiencing challenging driving scenarios, i.e., they were driving in a relatively straight path, on highways, with few turns and high average speed. These results are consistent with prior findings that drivers are more open to interruptions in straight roads [33, 34, 60].

4.3 Identifying "Negative" Driving Situations

As mentioned in Section 3.2, after the participants completed an intervention, they provided verbal feedback about the improvement in arousal and valence levels on a scale from 0 ("no improvement at all") to 4 ("very strong improvement"). In our prior work, we used this feedback to evaluate the effectiveness of the intervention [36]. Although we found the intervention led to immediate improvement captured by the driver's voice rating directly after an intervention, we observed no consistent improvement when comparing the pre- and post-driving questionnaires (PANAS-SF and Affective Slider). For example, in the pre-driving questionnaire, a driver rated

their arousal level as 50 (on a scale of 1-100); after the intervention, the driver rated the improvement in arousal as '2' (i.e., moderate improvement); in the post-drive questionnaire, however, the driver rated their arousal level as 40; thus suggesting that even though they found the intervention to be helpful, their arousal levels at the end of the drive decreased. We hypothesize that some events occurred between when the intervention ended and when the drive ended, which negated the interventions' positive impact. We argue that these events could be used as *markers* to deliver in-vehicle affective interventions more targeted in the future to maximize the intervention's effectiveness. In this section, we explore how various driving situations might negate the intervention's effectiveness.

Table 8. Identifying "negative" driving situations. The model for Arousal with intervention impact > 3 had few datapoints, due to which the model did not converge. Hence, we do not report those results. A detailed table with all the variables and the respective coefficients is available in the supplementary document.

Min. immediate intervention impact	Arousal				Valence			
	> 0	> 1	> 2	> 3	> 0	> 1	> 2	> 3
Time (in min.)	+^			-				
Steering events (per min.)				-	—*		/	/
Stopped ratio	—*	—^		-				
Low speed ratio			+^	-			—^	—^
Low speed events (per min.)				-	+**	+*	/	/
Low beam ratio		+*		-				
Turn signal ratio			+^	-				
Wiper active ratio				-	—*	—^		—**
Sudden gas pedal events (per min.)				-				+^
Lateral acceleration events (per min.)		+^		-				
Sudden steering events (per min.)		—^		-	—**	—*	—**	—*
Avg. speed	—*	—*	/	-				
Avg. deceleration	/	/	+*	-	/	/	/	/
Avg. steering angle				-	+**	+*	+**	+*
Avg. steering speed	—*			-			—*	—*
n (trips)	185	143	84	-	187	150	111	63
cond. R²	0.215	-	-	-	-	-	-	0.527
marg. R²	0.174	0.228	0.459	-	0.168	0.173	0.229	0.427

$+$ = significant increase, $-$ = significant decrease; $^$: $p < 0.1$, $*$: $p < 0.05$, $**$: $p < 0.01$, $***$: $p < 0.001$; / = removed variable due to multicollinearity

For this analysis we consider arousal and valence separately, using the delta between the pre- and post-driving affective states as the dependent variable. We evaluated several models for various minimum immediate improvement due to the interventions (i.e., improvement of > 0, > 1, > 2, or > 3). The reasoning behind building multiple models for various improvement levels was that we hypothesized that certain events could "negate" small improvements in the affective state, but may not negate larger improvements. For independent variables, we calculated several driving events from the CAN data. Broadly, we used the same events as Section 4.2, like speed, acceleration, and braking, but changed the time-period represented by these variables. We calculated

these variables over the entire period between when the intervention ended and when the drive ended. Hence, instead of using absolute event counts (as in Section 4.2), we used events per minute, i.e., the fraction unique events in a given period and the total number of minutes in that period. For some variables, we also calculated the event ratio, i.e., the fraction of the total duration of an event in a given time-period and the total length of the time period. For example, in stop-and-go traffic a vehicle stopped ten times in 15-minutes, amounting to a total stopped duration of 5 minutes, resulting in a `stop_event` of 0.66 (10/15), and a `stop_ratio` of 0.33 (5/15). By contrast, in heavy traffic a vehicle could stop once for 15 minutes, resulting in a `stop_event` of 0.06 (1/15) and a `stop_ratio` of 1.0 (15/15). Thus, events per minute and event ratios provide a standardized way to capture driving situations for different periods of time. We present a list of these variables in the supplementary materials.

We present the result of the mixed-effects models in Table 8. Again, for brevity, we only present variables that showed significant associations. The complete output of the models is available in the supplementary material. We found that certain events, like higher stopped ratio (i.e., being stuck in traffic for longer), frequent sudden steering events, higher average speed (possibly driving on the highway), and higher average steering speed, had negative associations with arousal levels, especially when the immediate intervention effect was small (i.e., > 0 or > 1). Similarly, for valence, we observed that frequent steering events, longer wiper active ratio (probably because of rain), frequent sudden steering events, and average steering speed, negative associations with valence levels, even when the immediate intervention effect was high.

We believe these are important findings and future intervention designs could leverage these "negative" variables to be better inform when to deliver more targeted affective well-being interventions, hence maximizing their effectiveness and improving participants' overall well-being.

4.4 Impact of In-Vehicle Interventions on Driving Behavior

The safety of drivers while they interact with in-vehicle interventions is of paramount importance. It is crucial that interacting and engaging with interventions should not cause a major distraction for the drivers, nor should interacting with interventions induce negative driving behaviors among the drivers, e.g., lane deviations, speeding, or sudden braking. In our study design, we randomly prompted the drivers with interventions during 66% of the drives. In the remaining 34% of drives, no interventions were prompted. In this section, we evaluate whether interacting with interventions showed associations with driving behavior: that is, was there a difference in the participants' driving behavior during drives where participants engaged with the intervention compared to drives with no interventions? To this end, we started by comparing the 285 drives where the participants completed an intervention with the 293 drives where no interventions were prompted. We used the same driving events as in Section 4.3 and calculated them over the entire trip duration – we provide a detailed list of all the driving events used for this analysis in the supplementary material. We break down the analysis by intervention type (130 mindfulness interventions and 155 music interventions) and present the results in Table 9. For the mindfulness intervention, we observed the stopped ratio was higher in drives with intervention. For the music intervention, we found drivers had lower high-speed ratio, lower turn signal ratio, lower average steering speed and angle in drives where the interventions were completed. While there are some interesting observations, like high-speed ratio, other significant variables like stopped ratio and turn signal ratio highlight the heterogeneity among the trips. It is possible that the variability in routes being compared lead to the current observations, thus not revealing any significant association with driving behaviors.

Therefore, to account for a more constrained scenario and to reduce the variability in road and trip characteristics, we decided to evaluate associations of intervention and driving only on the commute drives for the participants (home → work and work → home).[7] We present the results in Table 9. We found that participants

---

[7]This analysis only includes events where the participants directly went from home to work or from work to home. We did not consider scenarios where the participants may have stopped somewhere else on their commute.

Table 9. Effects of interventions on driving behavior. A detailed table with all the variables and the respective coefficients is available in the supplementary document.

	All trips		Commute trips	
	Mindfulness	Music	Mindfulness	Music
Stopped ratio	+*			
Stopped events (per min.)				+*
High speed ratio		−*		−*
Turn signal ratio	−^	−**		
Brake events (per min.)	+^		+*	/
Sudden brake events (per min.)				−*
Avg. steering angle		−**		
Avg. steering speed		−*		−*
n (trips)	130 with vs. 293 w/o intervention	155 with vs. 293 w/o intervention	49 with vs. 58 w/o intervention	49 with vs. 58 w/o intervention
cond. R^2	-	0.118	-	-
marg. R^2	0.108	0.104	0.188	0.196

+ = significant increase, − = significant decrease; ^: $p < 0.1$, ∗: $p < 0.05$, ∗∗: $p < 0.01$, ∗ ∗ ∗: $p < 0.001$; / = removed variable due to multicollinearity

had more brake events during drives in which they engaged with the mindfulness intervention, and had lower high-speed ratio, fewer sudden brake events, and lower average steering speed during the drives with the music intervention.

It is reassuring to see that there were not many variables that changed significantly between drives with interventions and drives without interventions. In fact, the results suggest that during the music intervention, the participants had a more "relaxed" drive – lower duration of driving over the speed limit, fewer sudden brake events, and lower average steering speed.

5 QUALITATIVE FEEDBACK BY PARTICIPANTS

To understand the participants' subjective experiences with the interventions, we conducted an hour-long semi-structured interview with each participant at the end of the study. We recorded the interviews and transcribed them for our analysis. We conducted the interview before any quantitative analyses, so neither we nor the participants were aware of our findings during their interview. In this section, we outline the qualitative feedback we received about the participants' interaction with interventions.

All participants were consistent in their evaluation of suitable driving situations to interact with the interventions. Participants preferred highways or straight roads over city roads to interact with interventions. Participants mentioned that it was the complexity of the driving task and the surroundings that influenced their decision (*"Rather not in city traffic with many crossings and traffic lights", "on the highway [with a steady flow] I concentrate on them best"*). P-9 summarized both of these aspects: *"As I said, since my commute is always half an hour anyway, 90% of it is monotonous highway travel. This has actually always fit well. So, if I would have driven half an hour through the city center with many traffic lights and pedestrians, I would have probably found it more than annoying."*

Another factor that echoed across all participants was that they knew their daily commute well and were able to plan their interaction with the intervention. Several participants could reflect on her commute in great detail; P-6 stated, *"Half of the distance is great, country road, about 100 [km/h], it runs fast.... Then I come into the closer district around Stuttgart, which means only red lights, stop and go.... In the morning, usually 35-45 minutes. In the afternoon rather 45-60 minutes."*; P-10 described her commute as, *"I have a three-quarter hour drive [and] the traffic jam situations only come closer to the end."* Finally, participants mentioned the influence of occupants as a barrier to interact with the interventions. Specifically, the mindfulness exercise was mostly perceived by occupants as boring because it was clearly adapted for the driver, P-2 describes the feeling of his wife: *"my wife has often said that she doesn't want to listen to it... as it is boring for her."* The music intervention was more appropriate with others and unpopular only with certain occupants. For example, drivers with children even struggled because P-8 knew that his *"daughter just wants to hear something completely different. That's when [he] didn't even start it."*

Further, some participants mentioned that the mindfulness intervention caused too much distraction and a substantial cognitive burden. For the mindfulness intervention, participants stated that they *"could concentrate well on the exercise"* when they drove on straight streets or with little surrounding traffic (e.g. highways). However, as soon as they *"had to concentrate more on the traffic"* (e.g., in city traffic in which more driving interaction was necessary), they struggled with the intervention. Four participants reported that it was not a serious problem, as they only focused on the traffic in those situations and ignored the exercise. Some drivers thought that the increased cognitive load from the mindfulness intervention in heavy-traffic situations was a potential safety risk. In particular, two participants felt *"extremely distracted at times"* and one even called the mindfulness exercise in such occasions as *"dangerous"*. This observation, however, seems to be driver dependent, e.g., P-6 felt that he was driving more calmly due to the mindfulness intervention, especially compared to his usual *"aggressive"* driving: *"I have sometimes questioned my own behavioral patterns or ways of thinking during the intervention or especially afterwards... just be a bit more relaxed, roll along...."*

In contrast, the participants participants did not have any concerns with the music intervention while driving, due to the lower levels of engagement. P-10 described her experience with these lower levels of engagement with the music intervention as *"it's just that sometimes you don't have to listen."* However, none of the drivers reported any beneficial effects on their driving behavior.

6 DISCUSSION

In this work, we took the first step towards understanding how drivers interact with in-vehicle well-being interventions in unconstrained driving environments. In this section, we discuss the implications of our findings along with the limitations of our work.

6.1 Implications

We found that several general trip characteristics, i.e., information available before a participant starts driving, relate with the drivers' responsiveness to the upcoming in-vehicle interventions. Specifically, we found that traffic flow, trip duration, other occupants, and type of trip (commute) had significant associations with the interventions' response rate. Quantitative evidence showing that participants were more likely to start interventions while commuting is an important result for in-vehicle interventions, given the amount of time people spend traveling to-and-from work. During the post-study interviews, participants confirmed the advantages of interacting with an intervention while commuting as they already *knew* the route, often did it alone, and enjoyed the change in pace during a routine drive.

Further, we found variables like weather, time spent on various road types, and participants' affective states had significant associations with the interventions' cancellation rate. Next, we built machine-learning models to predict participants' responsiveness to interventions and found that these models performed substantially better

than a biased-random classifier. These results suggest that it is possible to predict whether a participant will respond to an in-vehicle intervention even before they start driving. Our findings could inform how researchers design their in-vehicle interventions in the future, by potentially choosing to deliver interventions only during the drives where the participants are more likely to start, and less likely to cancel the intervention.

We also found that several in-situ driving characteristics showed significant associations with: (a) whether the drivers would start the intervention; (b) whether the drivers would cancel the intervention; and (c) whether a period was suitable for an intervention. Overall we found that participants favored straight roads (particularly highways) where they expected few "sudden" events (like braking or acceleration). The qualitative feedback supports this finding as participants mentioned that such conditions were favorable for interventions. These results are consistent with prior findings that drivers are more open to interruptions in straight roads [33, 34, 60]. In the future, researchers could integrate with navigation systems of vehicles to determine periods of such opportune moments to schedule interventions, thus providing the drivers with the best chance to interact and engage with the intervention.

Next, we identified several driving situations that showed significant statistical effects in negating the positive impact of an intervention. Specifically, events like higher stopped ratio, higher average speed, and higher average steering speed had negative associations with arousal; and events like frequent steering events, rainy weather, sudden steering movements, and higher average steering speed had negative associations with valence. We argue that future intervention designs could identify these "negative" events and decide to deliver targeted interventions to maximize affective well-being. In fact, in the realm of Just-in-Time Adaptive Interventions, these "negative" events could be considered as *states-of-vulnerability* that could guide the right type and amount of support [48, 49].

Further, we found that the two in-vehicle interventions employed in this study did not have a negative effect on driving behavior. In fact, we found that when the drivers were listening to the music intervention, they had fewer high-speed periods, fewer sudden brake events, and lower average steering speed, all of which could be considered as markers for a "safer" driving behavior. Such an analysis had not been possible in prior research, in which the participants drove a fixed route for a short time while the research staff watched. We believe that our work provides the first evidence that engaging with audio-based in-vehicle interventions (specifically the music interventions) do not pose a safety risk to the participants. This is a significant result and should encourage researchers to consider deploying their in-vehicle interventions to unconstrained driving conditions for thorough validation. In our qualitative feedback, two participants found the added cognitive load of mindfulness interventions to be dangerous in heavy-traffic situations. Other participants did not mention any serious problems with the mindfulness intervention. However, most participants agreed that they were better able to manage the higher cognitive load of the mindfulness intervention in straight roads and highways, with one participant mentioning that he was driving more calmly due to the mindfulness interventions. The person-specific experiences could be the reason we did not observe major positive or negative associations of mindfulness intervention and driving behavior in our quantitative analysis (unlike the music intervention). Nevertheless, our analysis highlights the importance of our work in delivering interventions at opportune moments – not just to ensure intervention effectiveness, but also to ensure driver safety.

Finally, in all our analyses, we observed statistical differences in participant interaction with the mindfulness and the music interventions. Overall, the participants had a higher cancellation rate for the mindfulness intervention than the music intervention, especially when passengers were present. Based on our quantitative evaluations and qualitative feedback, we highlight two possible reasons. First, *burden*: the mindfulness intervention was tailored explicitly for driving, and required active participation of the driver. Participants noted that this resulted in a higher cognitive load than the music interventions, and hence chose not to interact with mindfulness interventions in heavy-traffic scenarios. Further, passengers did not want to listen to the mindfulness instructions and found them to be boring, hence resulting in higher cancellation rates in presence of passengers. We believe

the second reason could be the lack of personalization: the mindfulness intervention included a standard set of mindfulness activities, and was the same for all participants. The music intervention, however, was created based on the songs chosen by the participants, thus adding a degree of personalization. Future work on intervention design, specifically JITAI, could focus on low-burden and personalized in-vehicle interventions to maximize driver engagement.

6.2 Design Guidelines for In-Vehicle Intervention Delivery

In this section, we discuss how an in-vehicle intervention delivery *system* might work in the future. Based on our results, we envision two complementary methods for in-vehicle intervention delivery:

- *Delivering intervention for a trip*: Determine whether to deliver an intervention in a particular trip, based on the trip-related factors like distance, duration, estimated traffic flow, weather, and occupants. As we mention in Section 4.1, we believe all of these factors can be estimated before a person starts driving, and can be leveraged to make a prediction based on our models in Section 4.1.1.
- *Delivering intervention at a particular moment in the trip*: Determine when to deliver the intervention in a particular trip, based on the in-the-moment driving characteristics, e.g., if the person is driving on the highway with constant speed, then it would be a good time to prompt an intervention. Conversely, if a person is driving through stop-and-go city traffic, then based on our results it would not be a good time for an intervention. For such tasks, models can be trained that continuously check the "likelihood" of engaging in an intervention as the driving conditions change.

Both methods have their benefits, and depending on the type of sensors and data-sources available, future researchers might choose to inspect only one method or a combination of both. We believe it is important to use both these methods in conjunction for maximum effectiveness. The general trip characteristics can be used as an initial filter. Only if the likelihood that a participant will respond to the intervention is above a specific threshold, the in-the-moment model should be activated. Then the in-the-moment model can be used to determine exactly when to deliver an intervention based on the current driving circumstances and the future expectation of traffic and road conditions available from estimated route information determined as part of the trip-related factors.

Similarly, interventions can also be delivered if certain "negative" or "vulnerable" conditions are determined. In such a situation, an intervention *system* could try to estimate such "negative" conditions and plan interventions accordingly. For example, we observe in Table 8 that wiper activity (i.e., rain) had negative associations with valence. In such a circumstance, a smart intervention delivery tool can leverage the trip-related factors to determine whether it is expected to rain on a given route, and could hold-off an intervention until it starts to rain. It is also possible that there are no opportune moments available to deliver an intervention after it started raining (e.g., traffic conditions worsened). In such situations, the system could potentially balance the likelihood of canceling an intervention with the estimated negative impact on valence, and might choose to deliver the intervention anyway. Or, it could delay the intervention to a more opportune moment later in the drive. If no opportune moment is found before the drive is over, then in a "connected-health" environment, the car can *hand-over* to the smartphone asking it to deliver an intervention after the drive is complete.

During the drivers' interactions with the system, an active learning process could further personalize intervention delivery to the users' needs [1]. In this context, our results would serve as a knowledgeable starting point that is adapted over time to each driver. In such a system, we could easily include new interventions and, more importantly, new intervention types that the system can learn from the drivers' interactions.

It is important to note that our results are the first steps towards a sophisticated in-vehicle intervention delivery system. Several studies and deployments are necessary before such a complex system should be realized for real-world traffic situations.

6.3 Limitations

Although we take a significant step forward in understanding long-term driver interaction with in-vehicle interventions by moving out of constrained, fixed-route driving scenarios to an unconstrained, real-world environment, we recognize that we only had 10 participants with long-term driving experience in our study. However, it is important to note that this was an exploratory study, and given the promising results, we plan to conduct studies with a larger number of participants. Furthermore, we only evaluated interaction with two specific types of audio-based interventions with varying cognitive loads. Interaction with other forms of interventions, such as haptic- or odor-based interventions [4, 17] could be different from audio-based interventions and is something we hope to explore in future work. Finally, in terms of analyses, we only considered CAN data. It is possible that additional insights can be generated by including data from street-facing and driver-facing cameras, as indicated by Semmens et al. [60]. We plan to evaluate this in future work.

7 CONCLUSION

In this paper, we evaluated when drivers interact with in-vehicle interventions in unconstrained free-living conditions. We conducted a two-month longitudinal study with 10 participants, in which each participant was provided with a study car for their daily driving needs. We delivered two in-vehicle interventions – each aimed at improving affective well-being – and simultaneously recorded the participants' driving behavior. In our analysis, we found that several pre-trip characteristics (like trip length, traffic flow, and vehicle occupancy) and the pre-trip affective state of the participants had significant associations with whether the participants started an intervention or canceled a started intervention. Next, we found that several in-the-moment driving characteristics (like current road type, past average speed, and future brake behavior) showed significant associations with drivers' responsiveness to the intervention. Further, we identified several driving behaviors that "negated" the effectiveness of interventions and highlight the potential of using such "negative" driving characteristics to better inform intervention delivery. Finally, we compared trips with and without intervention and found that both interventions employed in our study did not have a negative effect on driving behavior. Based on our analyses, we made some solid recommendations on how to deliver interventions to maximize responsiveness and effectiveness and minimize the burden on the drivers.

REFERENCES

[1] Rebecca Adaimi and Edison Thomaz. 2019. Leveraging Active Learning and Conditional Mutual Information to Minimize Data Annotation in Human Activity Recognition. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3, Article 70 (2019), 23 pages. https://doi.org/10.1145/3351228

[2] American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5).* American Psychiatric Association, Washington, DC.

[3] Audi AG. 2016. Audi Fit Driver. https://www.audi-mediacenter.com/en/audi-at-the-ces-2016-5294/audi-fit-driver-5300 Accessed: 05/03/2020.

[4] Stephanie Balters, Matthew L. Mauriello, So Yeon Park, James A. Landay, and Pablo E. Paredes. 2020. Calm Commute: Guided Slow Breathing for Daily Stress Management in Drivers. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 1, Article 28 (2020), 19 pages.

[5] Klaus Bernhardt. 2019. *Depression und Burnout loswerden: Wie seelische Tiefs wirklich entstehen, und was Sie dagegen tun können.* Ariston, Munich, Germany.

[6] Alberto Betella and Paul F. M. J. Verschure. 2016. The Affective Slider: A Digital Self-Assessment Scale for the Measurement of Human Emotions. *PLoS ONE* 11, 2, Article e0148037 (2016), 11 pages.

[7] Margaret M. Bradley and Peter J. Lang. 1994. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry* 25, 1 (1994), 49–59.

[8] Michael Braun, Anja Mainz, Ronee Chadowitz, Bastian Pfleging, and Florian Alt. 2019. At Your Service: Designing Voice Assistant Personalities to Improve Automotive User Interfaces. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. Paper 40.

[9] Michelle Chan and Anthony Singhal. 2015. Emotion matters: Implications for distracted driving. *Safety Science* 72 (2015), 302–309.

[10] Kiron Chatterjee, Samuel Chng, Ben Clark, Adrian Davis, Jonas De Vos, Dick Ettema, Susan Handy, Adam Martin, and Louise Reardon. 2020. Commuting and wellbeing: a critical overview of the literature with implications for policy and future research. *Transport Reviews* 40, 1 (2020), 5–34.

[11] Woohyeok Choi, Sangkeun Park, Duyeon Kim, Youn-kyung Lim, and Uichin Lee. 2019. Multi-Stage Receptivity Model for Mobile Just-In-Time Health Intervention. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 2, Article 39 (2019), 26 pages.

[12] CompuLab. 2020. IOT-GATE-iMX7 - Industrial IoT Gateway. https://www.compulab.com/products/iot-gateways/iot-gate-imx7-nxp-i-mx-7-internet-of-things-gateway/ Accessed: 08/05/2020.

[13] Daimler AG. 2018. ENERGIZING comfort control: Wellness while driving. https://media.daimler.com/marsMediaSite/ko/en/41880672 Accessed: 05/03/2020.

[14] Ed Diener, Shigehiro Oishi, and Louis Tay. 2018. Advances in subjective well-being research. *Nature Human Behaviour* 2, 4 (2018), 253–260.

[15] Ed Diener, Eunkook M. Suh, Richard E. Lucas, and Heidi L. Smith. 1999. Subjective well-being: Three decades of progress. *Psychological bulletin* 125, 2 (1999), 276–302.

[16] Harvey Dillon, Alison James, and Jenny Ginis. 1997. Client Oriented Scale of Improvement (COSI) and its Relationship to Several Other Measures of Benefit and Satisfaction Provided by Hearing Aids. *Journal-American Academy of Audiology* 8 (1997), 27–43.

[17] Dmitrijs Dmitrenko, Emanuela Maggioni, Giada Brianza, Brittany E. Holthausen, Bruce N. Walker, and Marianna Obrist. 2020. CARoma Therapy: Pleasant Scents Promote Safer Driving, Better Mood, and Improved Well-Being in Angry Drivers. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. Paper 49.

[18] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. 1996. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining* (Portland, Oregon) *(KDD'96)*. AAAI Press, 226–231.

[19] Jenny Gu, Clara Strauss, Rod Bond, and Kate Cavanagh. 2015. How do mindfulness-based cognitive therapy and mindfulness-based stress reduction improve mental health and wellbeing? A systematic review and meta-analysis of mediation studies. *Clinical Psychology Review* 37 (2015), 1–12.

[20] Oleg Gusikhin, Erica Klampfl, Dimitar Filev, and Yifan Chen. 2011. Emotive Driver Advisor System (EDAS) *(Informatics in Control, Automation and Robotics)*. 21–36.

[21] David Hallac, Suvrat Bhooshan, Michael Chen, Kacem Abida, and Jure Leskovec. 2018. Drive2vec: Multiscale state-space embedding of vehicular sensor data. In *21st International Conference on Intelligent Transportation Systems (ITSC)*. 3233–3238.

[22] Helen Harris and Clifford Nass. 2011. Emotion regulation for frustrating driving contexts. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. 749–752.

[23] Andrew Hede. 2010. The dynamics of mindfulness in managing emotions and stress. *Journal of Management Development* 29, 1 (2010), 94–110.

[24] Javier Hernandez, Daniel McDuff, Xavier Benavides, Judith Amores, Pattie Maes, and Rosalind Picard. 2014. AutoEmotive: Bringing Empathy to the Driving Experience to Manage Stress. In *Proceedings of the 2014 Companion Publication on Designing Interactive Systems*. 53–56.

[25] Karen Hovsepian, Mustafa Al'Absi, Emre Ertin, Thomas Kamarck, Motohiro Nakajima, and Santosh Kumar. 2015. cStress: Towards a Gold Standard for Continuous Stress Assessment in the Mobile Environment. *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)* (2015), 493–504. https://doi.org/10.1145/2750858.2807526

[26] Ryo Imai, Kota Tsubouchi, Tatsuya Konishi, and Masamichi Shimosaka. 2018. Early Destination Prediction with Spatio-temporal User Behavior Patterns. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4, Article 142 (2018), 19 pages.

[27] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An Introduction to Statistical Learning*. Vol. 112. Springer.

[28] Myounghoon Jeon. 2016. Don't Cry While You're Driving: Sad Driving Is as Bad as Angry Driving. *International Journal of Human–Computer Interaction* 32, 10 (2016), 777–790.

[29] Lutz Jäncke. 2008. Music, memory and emotion. *Journal of Biology* 7, 6, Article 21 (2008), 5 pages.

[30] Jon Kabat-Zinn. 1982. An Outpatient Program in Behavioral Medicine for Chronic Pain Patients Based on the Practice of Mindfulness Meditation: Theoretical Considerations and Preliminary Results. *General Hospital Psychiatry* 4, 1 (1982), 33–47.

[31] Jon Kabat-Zinn. 1994. *Wherever You Go, There You Are: Mindfulness Meditation for Everyday Life.* Hyperion, New York, NY, USA.

[32] Daniel Kahneman, Alan B. Krueger, David A. Schkade, Norbert Schwarz, and Arthur A. Stone. 2004. A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method. *Science* 306, 5702 (2004), 1776–1780.

[33] Auk Kim, Woohyeok Choi, Jungmi Park, Kyeyoon Kim, and Uichin Lee. 2018. Interrupting Drivers for Interactions: Predicting Opportune Moments for In-vehicle Proactive Auditory-verbal Tasks. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 4, Article 175 (2018), 28 pages.

[34] Auk Kim, Jung-Mi Park, and Uichin Lee. 2020. Interruptibility for In-vehicle Multitasking: Influence of Voice Task Demands and Adaptive Behaviors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 1, Article 14 (2020), 22 pages.

[35] SeungJun Kim, Jaemin Chun, and Anind K. Dey. 2015. Sensors Know When to Interrupt You in the Car: Detecting Driver Interruptibility Through Monitoring of Peripheral Interactions. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. 487–496.

[36] Kevin Koch, Verena Tiefenbeck, Shu Liu, Thomas Berger, Elgar Fleisch, and Felix Wortmann. 2021. Taking Mental Health & Well-Being to the Streets: An Exploratory Evaluation of In-Vehicle Interventions in the Wild. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. https://doi.org/10.1145/3411764.3446865

[37] Tobias Kowatsch, Fabian Wahle, and Andreas Filler. 2017. stressOUT: Design, Implementation and Evaluation of a Mouse-based Stress Management Service. In *Designing the Digital Transformation: DESRIST 2017 Research in Progress Proceedings*. 37–45.

[38] Arun Sai Krishnan, Xiping Hu, Jun-qi Deng, Li Zhou, Edith C.-H. Ngai, Xitong Li, Victor C.M. Leung, and Yu-kwong Kwok. 2015. Towards In Time Music Mood-Mapping for Drivers: A Novel Approach. In *Proceedings of the 5th ACM Symposium on Development and Analysis of Intelligent Vehicular Networks and Applications*. 59–66.

[39] Florian Künzler, Varun Mishra, Jan-Niklas Kramer, David Kotz, Elgar Fleisch, and Tobias Kowatsch. 2019. Exploring the State-of-Receptivity for MHealth Interventions. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 4, Article 140 (2019), 27 pages. https://doi.org/10.1145/3369805

[40] Alexander Legrain, Naveen Eluru, and Ahmed M. El-Geneidy. 2015. Am stressed, must travel: The relationship between mode choice and commuting stress. *Transportation Research Part F: Traffic Psychology and Behaviour* 34 (2015), 141–151.

[41] Daniel Leubner and Thilo Hinterberger. 2017. Reviewing the Effectiveness of Music Interventions in Treating Depression. *Frontiers in Psychology* 8, Article 1109 (2017), 21 pages.

[42] Scott M. Lundberg and Su-In Lee. 2017. A unified approach to interpreting model predictions. In *Advances in neural information processing systems*. 4765–4774.

[43] Scott M. Lundberg, Bala Nair, Monica S. Vavilala, Mayumi Horibe, Michael J. Eisses, Trevor Adams, David E. Liston, Daniel King-Wai Low, Shu-Fang Newman, Jerry Kim, and Su-In Lee. 2018. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nature biomedical engineering* 2, 10 (2018), 749–760.

[44] Abhinav Mehrotra and Mirco Musolesi. 2018. Using Autoencoders to Automatically Extract Mobility Features for Predicting Depressive States. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3, Article 127 (2018), 20 pages. https://doi.org/10.1145/3264937

[45] Varun Mishra, Gunnar Pope, Sarah Lord, Stephanie Lewia, Byron Lowens, Kelly Caine, Sougata Sen, Ryan Halter, and David Kotz. 2020. Continuous Detection of Physiological Stress with Commodity Hardware. *ACM Transactions on Computing for Healthcare (HEALTH)* 1, 2, Article 8 (April 2020), 30 pages. https://doi.org/10.1145/3361562

[46] Varun Mishra, Sougata Sen, Grace Chen, Tian Hao, Jeffrey Rogers, Ching-Hua Chen, and David Kotz. 2020. Evaluating the Reproducibility of Physiological Stress Detection Models. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. (IMWUT)* 4, 4, Article 147 (Dec. 2020), 29 pages. https://doi.org/10.1145/3432220

[47] Leanne G Morrison, Charlie Hargood, Veljko Pejovic, Adam WA Geraghty, Scott Lloyd, Natalie Goodman, Danius T Michaelides, Anna Weston, Mirco Musolesi, Mark J Weal, et al. 2017. The Effect of Timing and Frequency of Push Notifications on Usage of a Smartphone-Based Stress Management Intervention: An Exploratory Trial. *PloS one* 12, 1, Article e0169162 (2017), 15 pages.

[48] Inbal Nahum-Shani, Eric B. Hekler, and Donna Spruijt-Metz. 2015. Building health behavior models to guide the development of just-in-time adaptive interventions: A pragmatic framework. *Health psychology* 34 (2015), 1209–1219.

[49] Inbal Nahum-Shani, Shawna N. Smith, Bonnie J. Spring, Linda M. Collins, Katie Witkiewitz, Ambuj Tewari, and Susan A. Murphy. 2017. Just-in-Time Adaptive Interventions (JITAIs) in Mobile Health: Key Components and Design Principles for Ongoing Health Behavior Support. *Annals of Behavioral Medicine* 52, 6 (2017), 446–462.

[50] National Highway Traffic Safety Agency. 2018. Distacted Driving. https://www.nhtsa.gov/risky-driving/distracted-driving Accessed: 08/08/2020.

[51] Albertine J. Oldehinkel, Johan Ormel, Nienke M. Bosch, Esther M. C. Bouma, Arie M. Van Roon, Judith G. M. Rosmalen, and Harriëtte Riese. 2011. Stressed out? Associations between perceived and physiological stress responses in adolescents: The TRAILS study. *Psychophysiology* 48, 4 (2011), 441–452.

[52] World Weather Online. 2020. Weather API. https://www.worldweatheronline.com/developer/api/ Accessed: 08/10/2020.

[53] Pablo E. Paredes, Nur Al-Huda Hamdan, Dav Clark, Carrie Cai, Wendy Ju, and James A. Landay. 2017. Evaluating In-Car Movements in the Design of Mindful Commute Interventions: Exploratory Study. *Journal of Medical Internet Research* 19, 12, Article e372 (2017), 19 pages.

[54] Pablo E. Paredes, Yijun Zhou, Nur Al-Huda Hamdan, Stephanie Balters, Elizabeth Murnane, Wendy Ju, and James A. Landay. 2018. Just Breathe: In-Car Interventions for Guided Slow Breathing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1, Article 28 (2018), 23 pages.

[55] Soujanya Poria, Erik Cambria, Rajiv Bajpai, and Amir Hussain. 2017. A review of affective computing: From unimodal analysis to multimodal fusion. *Information Fusion* 37 (2017), 98–125.

[56] Fabien Ringeval, Björn Schuller, Michel Valstar, Jonathan Gratch, Roddy Cowie, Stefan Scherer, Sharon Mozgai, Nicholas Cummins, Maximilian Schmitt, and Maja Pantic. 2017. AVEC 2017: Real-life Depression, and Affect Recognition Workshop and Challenge. In *Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge*. 3133953, 3–9.

[57] James A. Russell. 1980. A Circumplex Model of Affect. *Journal of Personality and Social Psychology* 39, 6 (1980), 1161–1178.

[58] Carol D. Ryff. 2014. Psychological Well-Being Revisited: Advances in the Science and Practice of Eudaimonia. *Psychotherapy and Psychosomatics* 83, 1 (2014), 10–28.

[59] Hillol Sarker, Moushumi Sharmin, Amin Ahsan Ali, Md. Mahbubur Rahman, Rummana Bari, Syed Monowar Hossain, and Santosh Kumar. 2014. Assessing the availability of users to engage in just-in-time intervention in the natural environment. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)*. ACM, 909–920.

[60] Rob Semmens, Nikolas Martelaro, Pushyami Kaveti, Simon Stent, and Wendy Ju. 2019. Is Now A Good Time? An Empirical Study of Vehicle-Driver Communication Timing. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. Paper 637.

[61] Chen-Hsuan Shih, Naofumi Tomita, Yanick X. Lukic, Álvaro Hernández Reguera, Elgar Fleisch, and Tobias Kowatsch. 2019. Breeze: Smartphone-based Acoustic Real-time Detection of Breathing Phases for a Gamified Biofeedback Breathing Training. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 4, Article 152 (2019), 30 pages.

[62] Alois Stutzer and Bruno S Frey. 2008. Stress that doesn't pay: The commuting paradox. *Scandinavian Journal of Economics* 110, 2 (2008), 339–366.

[63] Edmund R. Thompson. 2007. Development and Validation of an Internationally Reliable Short-Form of the Positive and Negative Affect Schedule (PANAS). *Journal of Cross-Cultural Psychology* 38, 2 (2007), 227–242.

[64] TomTom. 2020. TomTom for Developers. https://developer.tomtom.com/ Accessed: 08/10/2020.

[65] U.S. Department of Transportation. 2019. How Much Time Do Americans Spend Behind the Wheel? https://www.volpe.dot.gov/news/how-much-time-do-americans-spend-behind-wheel Accessed: 8/9/2020.

[66] Nicholas T. Van Dam, Marieke K. van Vugt, David R. Vago, Laura Schmalzl, Clifford D. Saron, Andrew Olendzki, Ted Meissner, Sara W. Lazar, Catherine E. Kerr, and Jolie Gorchov. 2018. Mind the hype: A critical evaluation and prescriptive agenda for research on mindfulness and meditation. *Perspectives on Psychological Science* 13, 1 (2018), 36–61.

[67] Rui Wang, Weichen Wang, Alex DaSilva, Jeremy F. Huckins, William M. Kelley, Todd F. Heatherton, and Andrew T. Campbell. 2018. Tracking Depression Dynamics in College Students Using Mobile Phone and Wearable Sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1, Article 43 (2018), 26 pages. https://doi.org/10.1145/3191775

[68] David Watson, Lee A. Clark, and Auke Tellegen. 1988. Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology* 54, 6 (1988), 1063–1070.

[69] Kenton Williams, José Acevedo Flores, and Joshua Peters. 2014. Affective Robot Influence on Driver Adherence to Safety, Cognitive Load Reduction and Sociability. In *Proceedings of the International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. 1–8.

[70] Yocto Project. 2020. The Yocto Project. https://www.yoctoproject.org/ Accessed: 08/05/2020.