Aalborg University

Department of Computer Science

In-Vehicle Activity Recognition of User Activities Using Smartwatches

Thomas A. C. Hald
`thald12@student.aau.dk`

David H. Junker
`djunke12@student.aau.dk`

Mads Mårtensson
`mmarte12@student.aau.dk`

June 8, 2017

AALBORG UNIVERSITY
STUDENT REPORT

Department of Computer Science
Aalborg University
Software
Selma Lagerløfs Vej 300
9220 Aalborg Øst
`www.cs.aau.dk`

Title:

In-Vehicle Activity Recognition of User Activities Using Smartwatches

Project:

Master Thesis

Project period:

01/02-2017 - 08-06/2017

Project group:

IS103F17

Participants:

Thomas A. C. Hald
David H. Junker
Mads Mårtensson

Supervisor:

Mikael B. Skov

Completed: 08/06-2017

Abstract:

In this thesis, we investigate the use of motion sensors in a commercially available smartwatch as a sensing unit in an in-vehicle role identification system (IRIS). Our contribution is two-fold. Firstly, a system to provide mobile applications with in-vehicle contextual information about a user's role in a car, e.g. driver or passenger, and evaluation of what influences it, e.g. routes and roads. Secondly, an application, Hands-On, that leverages such contextual information to determine whether a driver has their hands in a recommended hand position and exploration of a small group of users' initial reactions to such an application.
With regards to mobile devices, an ongoing challenge is designing for context as mobile contexts are highly dynamic and complex. Previously, mobile applications have been made orientation aware and aware of basic daily activities, such as walking and running. The use of the motion sensors in these devices have shown potential for making the mobile device context aware on a deeper activity level through research in the area known as Human Activity Recognition. Within this area, the recognition in-vehicle activities has been investigated.

,

Contents

Introduction

The proliferation of mobile devices in modern society have reached a point where most people today own several at once, e.g. smartphone or tablet. These type of devices are portable to such a degree that they allow for interaction in many various contexts, e.g. on a train, during a run, or in a gym. To this day, designing for context is an ongoing challenge as mobile contexts are highly dynamic and complex [1]. By this we mean that interaction designed to work well for a static environment, e.g. sitting at a desk, might not necessarily do so for a dynamic environment, e.g. walking in a park. Furthermore, in some contexts interaction with the mobile device at hand may not warrant peoples primary focus, e.g. navigating through traffic. The sensors that mobile devices embed have allowed for the development of highly context-aware systems to such a degree that we have been able to distinguish between daily activities and situations, e.g. smoking [2] and eating [3], to design interaction optimised toward the current activity undertaken. A recent addition to the family of mobile devices is the Smartwatch.

Smartwatches

Smartwatches are wrist-worn mobile devices that allow for fast access and consumption of digital information at a glance. To substantiate this, in a field study of smartwatch use in daily situations from 2016 Pizza et al. state the following;

“ In a straightforward way the watch is more *convenient*, even if it has more limited functionality than other, larger devices... The watch is 'always there' and can be consulted by simply moving your wrist to bring the watch into your visual frame.” ([4])

Furthermore, Pizza et al. describe that the adoption of smartwatches, albeit in its infancy, pose some interesting questions for user research and illustrate the need for further understanding of smartwatches as a device, i.e. their strengths, weaknesses, and unique potential over other mobile devices. Previous research have explored the smartwatch as a device and as a result have identified two main interaction problems related to their small form factor and screen size, fat-fingers and visual occlusion [5]. The former arises from the fact that most peoples' fingers are too “fat” to press icons on the small screen or buttons on the side of the device. The latter arises from the fact that the limited visual real estate tend to cause applications in focus to hide or cover other applications on the screen. Additionally, Xia et al. [6] found that an index finger causes 60% visual occlusion of a standard dimensions smartwatch screen. In an effort to understand, reduce, and potentially solve these problems, the interaction space of smartwatches have been explored and investigated. As a result, several interaction techniques have been developed ranging from adopting existing techniques from other mobile devices [7, 8] to more extreme approaches, such as turning the user's skin into an interaction surface [9, 10] or the use of gestures which leverage the embedded sensors in the device [11, 12, 13].

Besides its shown use for interacting with smartwatches, gesture recognition through smartwatches have opened new doors for recognising activities through performed hand gestures. Especially, due to the fact that this type of device is always on-body due to its wrist-worn characteristic and always present nature as Pizza et al. [4] remarked. Incidentally, Fang and Chang [14] found that the wrist is perceived by the elderly as the most unobtrusive on-body location. Using on-body sensors, e.g. smartwatch sensors, and gesture recognition to infer user activities is not a new area of research, it is known as Human Activity Recognition (HAR)[15].

Human Activity Recognition

Activity recognition with a smartphone has seen a variety of work over recent years. The main incentive for researching and using activity recognition in software systems is that they provide systems with the possibility of becoming context aware, through knowing what activities the user is currently performing. Enabling context can provide software with the knowledge of when to prompt users and when not to do so if users are occupied with tasks that require a lot of attention.

The healthcare sector can benefit from activity recognition by monitoring patients activities, how many times and whether or not they perform a certain required task like taking their medicine. Lau et al. [16] creates an activity recognition system which identifies the activity of walking and compares walking patterns with others using embedded smartphone sensors. Such a system can possibly be used to identify problems with a user's legs. One key contribution of the paper is;

" The accelerometer in a smartphone can be used as a suitable sensor device for activity recognition, particularly as a non-obtrusive device for potential patient monitoring services." ([16])

Such a finding illustrates the benefit and possibility of using an accelerometer for activity recognition. Because of the recent advancement with smartwatches and the advantages it brings, new possibilities can be realised within activity recognition. Researchers have utilised the wrist-worn sensors, e.g. from smartwatches, in multiple ways for recognising activities like smoking or eating gestures, as well as recognising more overall activities.

In [2], Parate et al. develops a system capable of recognising the activity of smoking. The paper demonstrates that it is possible to accurately recognise smoking gestures based on smartwatch sensor data. Another paper working with activity recognition with smartwatches is [3]. Specifically the paper focuses on the eating activity, here Thomaz et al. recognises the gesture of picking something off a plate and putting it in the mouth.

In [17], Lee et al. uses activity recognition of various tasks, like watching TV, eating, cleaning and so on, to determine the indoor location of a user in their home. Like other activity recognition systems, the authors use the accelerometer to detect and recognise activities. Additionally, the system makes use of the microphone to analyse ambient noise, like the noise coming from brewing coffee or watching TV. Additionally, researchers have studied detection of in-vehicle activities and events.

In-Vehicle Activity Recognition

With regards to in-vehicle activity recognition, several aspects have been investigated with the purpose of detecting unsafe driving behaviour, driver inattention, car crashes, and drunk driving.

Akin et al. [18], developed a car-independent system that utilised sensor data from a smartphone to identify safe and unsafe driving behaviour in drivers. Dougherty et al. [19], develop a system that from a user's smartphone can detect car collisions to reduce first responder times. To accomplish this, the motion sensors, networking infrastructure, and microphone of the smartphone were leveraged. Hong et al. [20] develop a sensing platform to recognise aggressive driving behaviour using a smartphone instead of heavyweight and expensive systems. Additionally, Dai et al. [21], developed a system which facilitate detection of drunk driving through sensors in a mobile phone placed in the middle of the car.

Lastly, in-vehicle activity recognition research have also investigated the potential for distinguishing between driver and passengers based on the activity they perform [22, 23, 24]. A system that could recognise the in-vehicle role of a user, e.g. driver, allows for applications to be developed targeted at the role of interest. In the papers, the use of smartphones [22, 24, 23] and wrist-worn motion sensors [23] have been explored as sensing units. However, the use of a commercially available smartwatch for recognising in-vehicle roles have not been explored

Research Questions

This master thesis investigates aspects of human activity recognition in direct relation to the concept of being in a vehicle, specifically what users are doing in said vehicle. Using mobile sensors, we intend to contribute to the field by improving the in-vehicle contextual awareness of mobile devices. To this end, we present the following two research questions.

1. How can we detect in-vehicle roles using smartwatch technology?
2. How can a system leverage the contextual information of knowing the in-vehicle role of users, and what are their reaction to one such system?

The first question encapsulates the investigation and understanding on how a smartwatch can be of use to identify the in-vehicle role of a user, i.e. driver or passenger. Additionally, we will also investigate what influences the recognition of in-vehicle roles, such as different types of routes and road segments (i.e. roundabout, straights, right/left turns).

The second question encapsulates the focus of developing a system for users who are driving in a vehicle. The question pursues opportunities which have been made feasible by the first research paper, namely, inferring the in-vehicle role. The focus will be on analysing an in-vehicle safety perspective and, through embedded sensors in a smartwatch, seeking an initial system for coping with it. Furthermore, we explore initial reactions from test subjects exposed to such a system.

Research Contribution

Research Paper 1

IRIS: Employing Machine Learning and Smartwatch Gesture Recognition for In-vehicle Role Detection

Previous research regarding the identification of a user's in-vehicle role have shown that it is possible through the use of either smartphones or wrist-worn motion sensors. While the identification accuracy of previous attempts is arguable high, they suffer from a number of assumptions about how users are placing their mobile devices. Furthermore, in the evaluation of previous in-vehicle role identification systems none have yet to investigate the impact that conditions such as individuals, routes, and road segments, have on their system accuracies. Lastly, there have yet to be a role identification system developed that solely uses a commercially available smartwatch as a sensing unit.

On the basis of these three observations about the previous research, the purpose of this paper is to research whether assumptions regarding in-vehicle infrastructure or a fixed smartphone position can be eliminated by using a commercially available smartwatch. We developed an In-vehicle Role Identification System (IRIS) that employs machine learning on collected accelerometer data, from a smartwatch, to identify whether the user is a driver or passenger. To train our machine learning algorithm, we collected accelerometer data from 97 real-life driving sessions with 10 different individuals in 7 different vehicles. In an experiment, we evaluated the accuracy and detection time of IRIS and the influence that various conditions, such as routes, road segments, tasks, and individuals, have on them. 24 participants (12 drivers, and 12 passengers) were recruited through Aalborg University, personal connections, and our supervisors network. In some cases, snowball sampling occurred. None of our experiment participants were involved in training data collection for the machine learning algorithm. Additionally, 9 different vehicles were used. In the experiment, 3 routes were driven. On each route, drivers drove and passengers performed tasks given to them verbally and sporadically by an experiment leader.

Based on our experiment results, we achieve an overall accuracy of 87.0% (93% for driver and 80.9% for passenger) with an average detection time of 44.5 seconds. We show that the accuracy can be further improved to 93.3% at the cost of an increase in detection time to 57.5 seconds on average. We only found a statistical significant impact on identification accuracy between individuals. This concludes that individual behavior is an important factor to take into account when designing and evaluating in-vehicle activity recognition systems. Interestingly, our findings indicate that straight road segments only have slightly lower accuracies than road segments with turn moments (roundabouts, right/left turns). The previous research rely on turn moments which can result in longer detection times on routes with few turns. With regards to task performed by passengers in the experiment, we found that ones that were trained on achieved a higher classification accuracy than untrained tasks. However, we show that IRIS is still capable of distinguishing between passengers and drivers when faced with the untrained tasks. This result and the fact that passengers achieved an accuracy 12.1% points lower than drivers, suggest that the passenger activity is complex and hard to recognise compared drivers, because of the freedom of passenger behavior.

Research Paper 2

Hands-On: Raising Awareness of Driver Hand Position whilst Driving

With the ability to identify whether a user of a smartwatch is the driver or a passenger of a vehicle, developers can begin creating applications directed at these roles. The development of such systems could among others, help drivers maintain vehicle control through notifications when drowsiness or bad habits are recognised. Based on the possibilities mentioned, we formulate the following research question;

Previous research have investigated driver hand position on a steering wheel and found that despite recommended hand positions for optimal steering and maneuverability, drivers tend to adopt suboptimal hand positions instead. We argue that being able to notify drivers of their hand positions has merit for raising hand position awareness and assist them in maintaining a proper hand position.

To answer the research question of how users react to such applications, we develop an in-vehicle system, Hands-On, capable of notifying/warning a driver, through vibrations and sound, when the user positions their left hand in a less desirable position on the steering wheel. Hands-On utilises a commercially available smartwatch and the watches incorporated accelerometer. In an effort to explore the use of Hands-On we conduct an explorative user study with 6 participants.

Through the user study, we found that users reacted naturally to Hands-On feedback by repositioning their hand position on the steering wheel. This indicates that a system capable of warning users of bad hand positioning can indirectly increase drivers control of the vehicle. Furthermore, although participants said that they would not willingly use a system, such as Hands-On, some still argued that they would not disable it if it was integrated into the smartwatch. Additionally, participants saw merit in Hands-On as a system to help create better hand position habits in new drivers. On the basis of this, we provide three design guidelines for designing hand position systems:

- *Wait a bit and be accurate*, i.e. Allow drivers the opportunity to correct themselves before attempting to raise hand position awareness.
- *Do not come on too strong*, i.e. when considering notification strategies pick one that increases in severity (weak to strong) and frequency (slow to fast) within a reasonable limit to incrementally raise hand position awareness.
- *Avoid sound*, i.e. When notifying a user, do so through vibrations and avoid the distraction of sounds.

Conclusion

In this master thesis, we set out to investigate the use of smartwatch sensors in relation to in-vehicle activity recognition to increase the in-vehicle contextual awareness of future mobile device applications. Additionally, we set out to explore how this heightened in-vehicle contextual awareness can be leveraged to assist users whilst traveling in a vehicle. This led to the formulation of two research question, and subsequent two research papers that answer their respective question. In this section, we conclude on our overall contribution and whether or not our research have been sufficient to answer both research questions.

1. **How can we detect in-vehicle roles using smartwatch technology?**

We show that it is possible to detect in-vehicle roles, driver and passenger, through the use of an accelerometer in a commercially available smartwatch and machine learning through our developed in-vehicle role identification system, IRIS. In an experiment, we found the accuracy of IRIS to be 87.0% with a detection time of 44.5 seconds. Furthermore, the use of threshold results in a higher classification accuracy of 93.3% at the expense of a longer detection time of 57.5 seconds on average. We found that in-vehicle roles can be identified regardless of the driven route and road segment. Additionally, individuals have a significant impact on the accuracy of the proposed system, and are an important aspect to consider when designing and developing future in-vehicle activity recognition systems. Lastly, IRIS have been proven usable for passenger tasks unaccounted for in the design of the system, showing the clear strength of utilising machine learning for recognising human activities through sensing units.

Based on the findings of the first research paper, and the resulting smartwatch in-vehicle role identification system, we formulated the second research question:

2. **How can a system leverage the contextual information of knowing the in-vehicle role of users, and what are their reaction to one such system?**

We utilised IRIS to create a new system, Hands-On, capable of notifying drivers when they adopt a less reasonable hand position on the steering wheel. To evaluate the applicability of such a system we conducted an explorative user study with 6 participants. Through the study we show that all participants adjusted their hand position, to a recommended position, after receiving a notification from Hands-On. Thus we show that such a system can leverage the contextual information of in-vehicle roles.

In conclusion, through our in-vehicle activity recognition system we were able to provide in-vehicle contextual information in the form of users' role. With the development of a driver hand position awareness notification system, Hands-On, we have shown that such information can be leveraged to develop an assistive application targeted at drivers. Lastly, we argue that in-vehicle activity recognition systems has the potential to open new doors in the design of in-vehicle safety and attention systems.

Bibliography

[1] Jesper Kjeldskov. "Mobile computing". In: *The Encyclopedia of Human-Computer Interaction, 2nd Ed. Aarhus, Dänemark: The Interaction Design Foundation. Verfügbar unter https://www. interaction-design. org/encyclopedia/mobile_ computing. html, Stand* 10 (2013), p. 2016.

[2] Abhinav Parate et al. "RisQ: Recognizing Smoking Gestures with Inertial Sensors on a Wristband". In: *Proceedings of the 12th Annual International Conference on Mobile Systems, Applications, and Services*. MobiSys '14. Bretton Woods, New Hampshire, USA: ACM, 2014, pp. 149–161. ISBN: 978-1-4503-2793-0. DOI: 10.1145/2594368.2594379. URL: http://doi.acm.org/10.1145/2594368.2594379.

[3] Edison Thomaz, Irfan Essa, and Gregory D. Abowd. "A Practical Approach for Recognizing Eating Moments with Wrist-mounted Inertial Sensing". In: *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. UbiComp '15. Osaka, Japan: ACM, 2015, pp. 1029–1040. ISBN: 978-1-4503-3574-4. DOI: 10.1145/2750858.2807545. URL: http://doi.acm.org/10.1145/2750858.2807545.

[4] Stefania Pizza et al. "Smartwatch in vivo". In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM. 2016, pp. 5456–5469.

[5] Kuo-Ying Huang. "Challenges in Human-Computer Interaction Design for Mobile Devices". In: *Proceedings of the World Congress on Engineering and Computer Science*. Vol. 1. 2009.

[6] Haijun Xia, Tovi Grossman, and George Fitzmaurice. "NanoStylus: Enhancing Input on Ultra-Small Displays with a Finger-Mounted Stylus". In: *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. UIST '15. Daegu, Kyungpook, Republic of Korea: ACM, 2015, pp. 447–456. ISBN: 978-1-4503-3779-3. DOI: 10.1145/2807442.2807500. URL: http://doi.acm.org/10.1145/2807442.2807500.

[7] Mitchell Gordon, Tom Ouyang, and Shumin Zhai. "WatchWriter: tap and gesture typing on a smartwatch miniature keyboard with statistical decoding". In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM. 2016, pp. 3817–3821.

[8] Aske Mottelson et al. "Invisiboard: Maximizing Display and Input Space with a Full Screen Text Entry Method for Smartwatches". In: *Proceedings of the 18th*

International Conference on Human-Computer Interaction with Mobile Devices and Services. MobileHCI '16. Florence, Italy: ACM, 2016, pp. 53–59. ISBN: 978-1-4503-4408-1. DOI: 10.1145/2935334.2935360. URL: http://doi.acm.org/10.1145/2935334.2935360.

[9] Gierad Laput et al. "Skin Buttons: Cheap, Small, Low-powered and Clickable Fixed-icon Laser Projectors". In: *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*. UIST '14. Honolulu, Hawaii, USA: ACM, 2014, pp. 389–394. ISBN: 978-1-4503-3069-5. DOI: 10.1145/2642918.2647356. URL: http://doi.acm.org/10.1145/2642918.2647356.

[10] Yang Zhang et al. "SkinTrack: Using the Body as an Electrical Waveguide for Continuous Finger Tracking on the Skin". In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM. 2016, pp. 1491–1503.

[11] Shaikh Shawon Arefin Shimon et al. "Exploring Non-touchscreen Gestures for Smartwatches". In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM. 2016, pp. 3822–3833.

[12] Anusha Withana et al. "zSense: Enabling shallow depth gesture recognition for greater input expressivity on smart wearables". In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM. 2015, pp. 3661–3670.

[13] Simon T Perrault et al. "Watchit: simple gestures and eyes-free interaction for wristwatches and bracelets". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM. 2013, pp. 1451–1460.

[14] Yu-Min Fang and Chien-Cheng Chang. "Users' psychological perception and perceived readability of wearable devices for elderly people". In: *Behaviour & Information Technology* 35.3 (2016), pp. 225–232.

[15] Andreas Bulling, Ulf Blanke, and Bernt Schiele. "A tutorial on human activity recognition using body-worn inertial sensors". In: *ACM Computing Surveys (CSUR)* 46.3 (2014), p. 33.

[16] Sian Lun Lau et al. "Supporting patient monitoring using activity recognition with a smartphone". In: *Wireless communication systems (ISWCS), 2010 7th international symposium on*. IEEE. 2010, pp. 810–814.

[17] Seungwoo Lee et al. "Non-obstructive Room-level Locating System in Home Environments Using Activity Fingerprints from Smartwatch". In: *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. UbiComp '15. Osaka, Japan: ACM, 2015, pp. 939–950. ISBN: 978-1-4503-3574-4. DOI: 10.1145/2750858.2804272. URL: http://doi.acm.org/10.1145/2750858.2804272.

[18] Kari Torkkola, Noel Massey, and Chip Wood. "Driver inattention detection through intelligent analysis of readily available sensors". In: *Intelligent Transportation Systems, 2004. Proceedings. The 7th International IEEE Conference on*. IEEE. 2004, pp. 326–331.

[19] Jules White et al. "Wreckwatch: Automatic traffic accident detection and notification with smartphones". In: *Mobile Networks and Applications* 16.3 (2011), p. 285.

[20] Jin-Hyuk Hong, Ben Margines, and Anind K. Dey. "A Smartphone-based Sensing Platform to Model Aggressive Driving Behaviors". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '14. Toronto, Ontario, Canada: ACM, 2014, pp. 4047–4056. ISBN: 978-1-4503-2473-1. DOI: 10.1145/2556288.2557321. URL: http://doi.acm.org/10.1145/2556288.2557321.

[21] J. Dai et al. "Mobile phone based drunk driving detection". In: *2010 4th International Conference on Pervasive Computing Technologies for Healthcare*. 2010, pp. 1–8. DOI: 10.4108/ICST.PERVASIVEHEALTH2010.8901.

[22] Yan Wang et al. "Sensing Vehicle Dynamics for Determining Driver Phone Use". In: *Proceeding of the 11th Annual International Conference on Mobile Systems, Applications, and Services*. MobiSys '13. Taipei, Taiwan: ACM, 2013, pp. 41–54. ISBN: 978-1-4503-1672-9. DOI: 10.1145/2462456.2464447. URL: http://doi.acm.org/10.1145/2462456.2464447.

[23] Luyang Liu et al. "Toward Detection of Unsafe Driving with Wearables". In: *Proceedings of the 2015 Workshop on Wearable Systems and Applications*. WearSys '15. Florence, Italy: ACM, 2015, pp. 27–32. ISBN: 978-1-4503-3500-3. DOI: 10.1145/2753509.2753518. URL: http://doi.acm.org/10.1145/2753509.2753518.

[24] Hon Lung Chu et al. "In-vehicle driver detection using mobile phone sensors". In: *ACM MobiSys*. 2011.

Appendices

Appendix A

Research Paper 1

IRIS: Employing Machine Learning and Smartwatch Gesture Recognition for In-vehicle Role Detection

Thomas A. C. Hald, David H. Junker, Mads Mårtensson
Aalborg University,
Department of Computer Science
Selma Lagerløfs Vej 300, DK-9220
Aalborg, Denmark
thald12@student.aau.dk, djunke12@student.aau.dk, mmarte12@student.aau.dk

ABSTRACT
This paper explores the potential for commercially available smartwatches to identify the in-vehicle role of users, driver or passenger, without relying on in-vehicle sensors or infrastructure. In particular, we study how an accelerometer sensor, in a smartwatch, can be used to collect data about steering wheel usage and other hand movements to distinguish between drivers and passengers through machine learning. To this end, we develop and evaluate an in-vehicle role identification system (IRIS) through a user experiment. Systems, such as IRIS, open new possibilities for tracking the behaviour of drivers and passengers. The behaviour of drivers can be analysed to identify unsafe driving patterns for use in in-vehicle safety applications. Additionally, these systems allow for interaction to be tailored to either role, e.g. locking of a driver's phone whilst a car is in motion. Using threshold-based classification, a field experiment shows that IRIS can identify the in-vehicle role of users with 93.3% accuracy. Surprisingly, we also find that straight road segments achieve similar accuracies compared to turns.

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous; See http://acm.org/about/class/1998/ for the full list of ACM classifiers. This section is required.

Author Keywords
Authors' choice; of terms; separated; by semicolons; include commas, within terms only; required.

INTRODUCTION
Inference of users' activities through embedded sensors in mobile devices provides important contextual information into their daily life, habits, and behaviour. Previously, research have explored the use of such information from a health [1, 2] or in-vehicle safety perspective [3, 4]. To illustrate, the ability to distinguish between drivers and passengers holds value for developing safety applications.

Recognising the in-vehicle roles through sensors is related to Human Activity Recognition (HAR), that is, recognising activities through on-body sensors. This has been shown valuable in numerous cases. Among others, smoking, eating, and sleeping [5, 6, 7], where systems have been developed and explored to aid users in their everyday life. Real-life activities, such as eating, requires recognition of specific hand gestures to confidently identify said activity. When an activity encompass gestures performed by the user's hand, wrist-worn motion sensors, especially from smartwatches, is crucial to be able to provide a realistic picture of what activity a user is conducting. The inference of said activities from sensors in smartphones and smartwatches is typically facilitated through the use of machine learning to identify patterns in data which is unique for an activity.

Previous research have explored the possibility of recognising in-vehicle roles, driver and passenger, using sensors in smartphones and wrist worn wearables [8, 9, 3]. Despite these preliminary studies, several aspects from an HCI perspective have yet to be investigated for such systems in a user experiment. Firstly, the use of a commercially available smartwatch as a sensing unit. Secondly, the impact of real-life driving conditions such as turns. Thirdly, the impact of individual users.

We show that it is possible to recognise in-vehicle roles, using solely a low-budget commercial smartwatch through the development of a system, IRIS. In a field experiment, we evaluate the achieved accuracy of IRIS under various driving conditions, such as routes, road segments. We found that role identification using smartwatches can be done for any part of a drive even in absence of turns as previous research have utilised [9, 3]. Additionally, we found that individual users have a significant impact on the accuracy of role identification. Finally, we compare our system with previous work and show strengths and weaknesses of the proposed in-vehicle role identification system, IRIS.

RELATED WORK

In this section, smartwatches in relation to small screens and embedded sensors will be described. Additionally, recognising physical activities performed by users, and the opportunities it can have will be outlined followed by a summary of approaches towards identifying in-vehicle roles using mobile device sensors.

Smartwatches

Smartwatches have been proven beneficial in a variety of ways, in part due to their convenient placement on the wrist and wide array of sensors. In recent years, research concerning interaction with smartwatches and their small screens have been in focus. The small screens are the cause of two inherent interaction problems, namely the fat-finger and the occlusion problem [10]. Researchers have tried tackling these problems in three ways, by extending the hardware of smartwatches to allow new interaction possibilities [11, 12], by creating unique interaction for the small touchscreen [13, 14, 15], or by exploring the possibility of gestures [16, 17, 18].

Despite smartwatches being problematic due to their screen sizes, they have been used to some extent because of their sensors. Gesture recognition systems have been developed as a solution to the fat-finger and occlusion problem. Kim et al. developed a system capable of letting users define their own gestures across multiple mobile devices, including smartwatches [19]. The system builds on accelerometer data from various mobile devices. In a usability study of their system, 20 participants were able to complete a total of 59 out of 60 tasks, and participants generally reported that they understood the system and its concept of acceleration based gesturing. Guo et al. [11] developed a system that is able to recognise which direction a user tilts a smartwatch towards. The motivation behind tilting is based on the possibility of having hands-free interaction with smartwatches. Through a user study Guo et al. [11] found that users were able to use device tilting as a hands-free interaction technique to interact with a smartwatch.

Besides exploring the possibility of using smartwatch sensors to recognise gestures for smartwatch interaction, sensors have been utilised to unobtrusively capture user dynamics in order to recognise the activities that users' are engaged in.

Human Activity Recognition

Human activity recognition (HAR) is a research area related to HCI which explores the automatic detection of physical activities using machine learning [20]. Previously, researchers have explored the recognition of gestures to infer activities using video in constrained environments [21, 22]. Recently, a shift has occurred within the area from video to motion sensors in an effort to investigate daily activities or real life situations in an unobtrusive and unconstrained manner [20].

To facilitate early detection and timely treatment of smoking addiction, Parate et al. [5] argue that there is a need for a smoking detector with high sensitivity, specificity, and is easy to wear on a day-to-day basis. To develop such a detector, wrist-worn motion sensors and machine learning was used to detect and classify smoking gestures in-real world settings. Based on the impracticality of previous food intake monitoring systems requiring multiple on-body sensors or specialised devices, Thomaz et al. [6] simply leverage an accelerometer found in a commercially available smartwatch and machine learning to develop a practical automated everyday food intake monitoring system. Bao and Intille [23] found that two acceleration sensors worn on the thigh and wrist were adequate to recognise a slew of daily activities, e.g. walking, folding laundry, and watching tv.

By recognising user activities, application could be developed to assist users or provide them with important contextual information about the activity they are undertaking, e.g. an application for smokers that track their smoke count and average puff duration. Recently, another area in HAR that have been explored and investigated is in-vehicle driving recognition due to the safety implications associated with the activity.

Driver and Passenger Identification

An in-vehicle role in relation to HAR can be seen as recognising the conducted activity uniquely associated to each role, driver or passenger. Mobile sensors have previously been used to identify if a person is the driver of a vehicle [8, 9, 3].

Chu [8] identified in-vehicle roles using only a smartphone. They analysed different positions common for a smartphone while being in a vehicle, and through machine learning an overall accuracy of 84.67% was achieved based on five different test subjects. They further discussed the assumption and issues with other phone positions, not taken into account and possible solutions. They reported a detection time of approximately 3 minutes and 15 seconds. Their system has a well-defined life-cycle which is initiated when an entry swing is detected and is terminated when no vehicular motion can be detected.

In another approach, Wang et al. [9], employ sensors from a smartphone to recognize in-vehicle roles during turns. A fixed reference point in the middle of the car was required to differentiate between motion sensor data from drivers and passengers. After one turn, they achieved an accuracy of 80% with a detection time between 38 seconds and 48 seconds.

Using a smartphone and a wearable motion sensor, Liu et al. [3] identified in-vehicle roles. Adopting the approach from Wang et al. [9], they also focused on detecting in-vehicle roles during turns. Additionally, Liu et al. explore the idea of having passengers perform tasks, specifically using the phone or eating. The system was evaluated for 280 turns, 239 of these were in a controlled area with the passenger being stationary (no arm movements), and 41 turns in a real world setting with the passenger being tasked with playing on their phone and eating snacks. Using machine learning a classification accuracy of 96.1% with a detection time of 21.13 seconds is achieved.

In this section, we have shown three previous approaches to in-vehicle role identification with varying levels of success. The experiments and evaluations carried out have not investigated the impact that different participants, vehicles, or routes have on in-vehicle role identification.

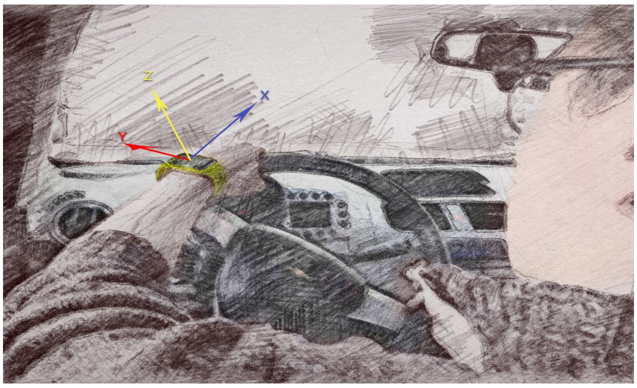


Figure 1: Concepts of arm movements from a driver

IN-VEHICLE ROLE IDENTIFICATION SYSTEM (IRIS)

We designed and implemented an In-vehicle Role Identification System (IRIS) which is a system capable of classifying a user's in-vehicle role, based on accelerometer data from a smartwatch. It is designed for users which wear their smartwatch on their left hand, and drive cars manufactured for left-hand drive.

We chose to model a user's in-vehicle role as either, driver or passenger. In Figure 1, the user steers the car and the smartwatch on his hand follows the curvature of the steering wheel. In this scenario, the watch captures the acceleration which represents the steering gesture. Similarly, a passenger could tune the radio and this gesture would also be captured in acceleration from the watch. Based on the acceleration, the performed gestures can be distinguished from each other, and leveraged to infer their respective in-vehicle role in IRIS.

Overview

We have designed IRIS to be a system that does not require any direct user interaction to initiate or terminate. Figure 2 depicts the complete life cycle of IRIS and its internal components. The life-cycle is initiated when IRIS detects that users are in a moving vehicle and terminates when the system detects steps.

When a user has been detected to be in a moving vehicle, 1, the sensor logger component in the smartwatch collects and transfers accelerometer data continuously to the smartphone. After a window of time (15 seconds), the collected data is fed

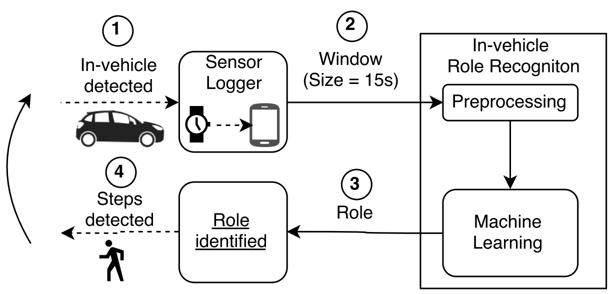


Figure 2: The life-cycle and the components required for IRIS to identify the in-vehicle role.

to a preprocessing subcomponent, 2, which is a part of the In-vehicle Role Recognition component. In the preprocessing subcomponent, we transform the received acceleration data into an interpretable format for our machine learning algorithm. After preprocessing, the data is transferred to a trained machine learning subcomponent for role identification. In this component, IRIS interprets the collected and preprocessed acceleration data using Hidden Markov Models (HMMs), a probabilistic and temporal model [24, p. 578-583]. This type of machine learning model was chosen because it utilises the temporal aspects in sensor data opposed to more traditionally applied methods such as Support Vector Machines [25]. After the in-vehicle role recognition component has classified the user's role 3, IRIS enters a passive state. When steps are detected 4, IRIS leaves the passive state and terminates its lifecycle. IRIS automatically starts again when Google's API detects a user to be in-vehicle 1.

Implementation

We implemented IRIS through a number of steps; in-vehicle and step detection, logging and transferring of sensor data, and machine learning classification. IRIS is mainly developed in Java 6.0 for Android 4.4.

Firstly, we use Google's activity recognition API [26] to facilitate in-vehicle and step detection, depicted as 1 and 4 in Figure 2.

Secondly, the sensor logger component illustrated in Figure 2 facilitates the activation of sensors in the smartwatch. The sensor data from the smartwatch is transferred to the smartphone using bluetooth and Android's message passing API. Initially, this was facilitated using a Java library developed as a project by IBM [27]. However, the library was unfinished and did not support many sensors. We fetched the source code and extended the library to support additional motion sensors which potentially could provide value to the goal of identifying the in-vehicle roles. Additionally, in the library, we discovered that the timestamping method used for individual sensor measurements had a propensity to drift. Consequently, we changed the timestamping method to one without drifting.

To train and evaluate IRIS offline, sensor data from the sensor logger component was transferred to a remote server. The remote server was developed in Python 2.7 and consisted of a developed rest-API endpoint which made it possible for IRIS to transfer sensor data. Furthermore, we employed a client-server consistency pattern which ensured that data was successfully transferred.

Finally, we implemented the machine learning subcomponent in Figure 2 using a java library, JAHMM [28]. The library was chosen because it made it possible to facilitate seamless integration for mobile devices running Android.

Machine Learning and Role Recognition

The use of machine learning in relation to role recognition is clarified in the following section. Figure 3 depicts the internal steps for identifying the in-vehicle role described formally in this section. The figure is a visual elaboration of the *In-vehicle*

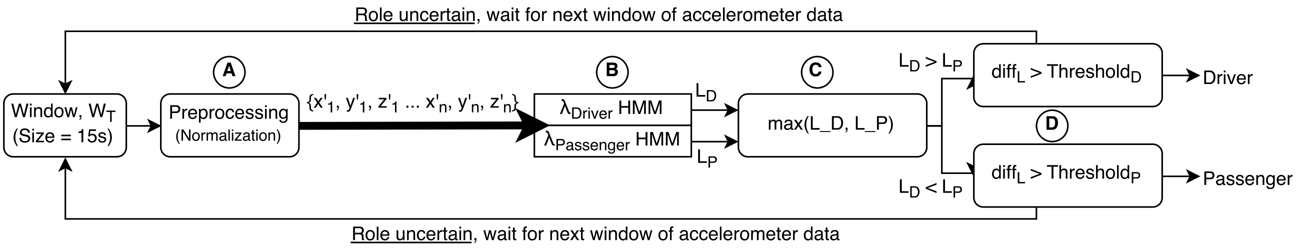


Figure 3: The internal flow of IRIS for classifying the in-vehicle role.

Role Recognition component in Figure 2. We will formally go through the figure in the following section.

Given a time window of accelerometer data defined as, $W_t = \{x_1, y_1, z_1, ..., x_n, y_n, z_n\}$, we preprocess the the data by applying L2-normalization [29, p. 36] to each of the three dimensions of the accelerometer data, formally defined in (1),

$$x' = \frac{x}{\sqrt{x^2 + y^2 + z^2}} \tag{1}$$

This is depicted A in Figure 3. $\{x, y, z\}$ is a three dimensional accelerometer vector, in this case x is the dimension which is normalised. The numerator is then changed for each dimension. L2-Normalization scales the dimensions of a vector to their unit norm, ie, the sum of the dimensions would equal 1. Scaling each accelerometer vector in a window provides a more generalised representation of the window, which can aid the accuracy of machine learning.

We employ a HMM model for each role, driver and passenger, denoted as λ_{Driver} and $\lambda_{Passenger}$, as seen at B. Each HMM outputs a value (Likelihood) describing how likely it is that a window W_t is generated by the specific HMM, formally defined as;

$$\mathcal{L} = P(W_t|\lambda) \tag{2}$$

In IRIS, given a window W_t, the driver HMM and passenger HMM will output a likelihood L_D and L_P, respectively describing how likely it is that the individual model generated W_t. This is formally defined in (3).

$$\begin{aligned} \mathcal{L}_D &= P(W_t|\lambda_{Driver}), \\ \mathcal{L}_P &= P(W_t|\lambda_{Passenger}) \end{aligned} \tag{3}$$

A role is selected by believing in the model with the highest likelihood for a given window, defined in (4) and vice versa, if the passenger likelihood is the highest then we believe the user is a passenger. This process is depicted in Figure 3, at C.

$$Highest_{Role} = max(\mathcal{L}_D, \mathcal{L}_P) \tag{4}$$

In situations where the difference between the likelihoods is low, the uncertainty of the predicted role is high. Therefore, we introduce threshold-based classification in IRIS. The difference between the likelihoods from the driver and passenger model needs to exceed a specified threshold before taking a decision. This is formally defined in (5), and shown at D.

$$Role = Highest_{Role} \wedge diff_{\mathcal{L}} > Threshold_{Role} \tag{5}$$

In other words, the formal definition describes, that given the driver HMM yields the highest likelihood, we verify that the difference between the likelihoods is above a certain driver threshold given in 5, and vice versa, if the passenger model yields the highest likelihood. The difference, $diff_{\mathcal{L}}$, is defined in (6) which describes the difference between the highest likelihood and the lowest likelihood.

$$diff_{\mathcal{L}} = max(\mathcal{L}_D, \mathcal{L}_P) - min(\mathcal{L}_D, \mathcal{L}_P) \tag{6}$$

The likelihood for a HMM does not describe a relationship between the driver and passenger models (e.g. 95% for driver, 5% for passenger), it solely describes its isolated belief of whether it has generated a similar window before. The product of probabilities results in likelihoods which are substantially small and usually not interpretable by a computer. Therefore, logarithm of probabilities and value scaling is performed in Jahmm to make the likelihood value comparable and interpretable by the machine [30][p. 27-30].

Collected Data for Training from Real Driving

We trained the two HMMs with either a driver or passenger dataset. Each dataset was crafted through a self-developed Android application which gathers and stores accelerometer data from a smartwatch. Data were collected in 97 driving sessions, where each individual session consists of accelerometer data for precisely one driver or passenger. Furthermore, GPS data from the phone was collected to be able to approximate the driven distance and number of occurred turns. A Sony Smartwatch 3 and a Google Nexus 5x smartphone was used to collect data for the training set.

The collected data were divided into multiple windows of 15 seconds of acceleration data using a sliding time window technique. Sliding can be defined as moving a time window a bit at a time, which in our system, is moved by one second. This makes it possible to generate a much larger set of training examples, and to catch different parts of the same individual event, e.g. a left turn. However, the optimal length for such a window intuitively varies based on the task being classified.

Shoaib et al. [31] argues that for complex activities it is favorable to use window lengths between 15-30 seconds. In that particular paper, a simple activity is defined as a simple repetitive activity such as biking, whereas a complex activity is described as activities which are not repetitive in nature and can differ substantially. We argue, that a passenger role is a vaguely defined activity and that it is complex due to the few restrictions to allowed hand movements.

Data collection for training was conducted in seven different cars by ten different individuals as either drivers or passengers in different parts of Denmark, though primarily in Aalborg. We instructed drivers to drive as they normally would, whereas passengers were instructed to periodically do one of four tasks. Primarily, passengers were tasked with sitting with their hands in their lap. Additionally, passengers were instructed to; interact with the radio, drink from a bottle, and write messages on their smartphone. We selected these tasks to model passenger behaviour in the dataset. The GPS data from each driving session has been used to automatically derive turns based on the latitude and longitude bearing over a sequence of GPS coordinates. The collected data for drivers amounted to a distance of 310 kilometers, for an driving time of 596 minutes with an approximate of 602 turns. Similarly, the data for passengers amounted to 210 kilometers, for a driving time of 337 minutes with an approximate of 385 turns. Consequently, this resulted in 23749 and 14789 training examples (windows) for driver and passenger respectively.

To test IRIS and the assumptions we initially made regarding the gestures performed by drivers and passengers, as well as to test the classifier, we developed an application capable of delivering classifications live during a drive.

EXPERIMENT

The purpose of the experiment was to examine the ability of IRIS to recognise the role of users in cars. We conducted a user experiment with participants as either drivers or passengers. Drivers were instructed to drive their own vehicle, and passengers were instructed to complete tasks given by an experiment leader. In particular, we studied the classification accuracy of IRIS under different route, task, and road segment conditions to study their impact.

Participants

24 people (8 female) participated in the experiment, 12 drivers and 12 passengers. None of the participants were previously used in the training data collection. Participants' age ranged between 19 to 70 years, (Mean = 26.7). Out of all participants, 21 were right-handed and 12 reported prior experience with smartwatches. In regards to drivers, all of them had a driver's license and driving experience ranging from 2 to 52 years, (Mean = 11.8). Furthermore, on average drivers drove at least four days a week with trips between 15 to 30 minutes. 22 out of 24 participants reported that they would prefer to have the smartwatch on their left wrist. Participants were recruited among personal connections, through the university, and our supervisor's network. Furthermore, in some cases participants were recruited through snowball sampling.

Routes

Three distinct routes were selected and given an appropriate name based on the following criterias; speed, distance, and amount of turns.

Rural (High speed, Few turns)
The route is defined as rural due to the majority of the road being highway with a speed limit of 80 km/h. The distance is 4.6 km with 7 turns (3 roundabouts, 1 right, 3 lefts). On average there is 1.52 turns per km. Additionally, the route has five segments of straight road with an average length of 784 meters. The first route was selected due to its high speed, long distance, and low number of turns. It starts from a parking lot at Aalborg University, and ends at a parking lot at a local supermarket.

Suburban (Low speed, Many turns)
The route is defined as suburban because of its residential surroundings. The distance of the route is 1.7 km with 12 turns, (2 roundabouts, 4 rights, 6 lefts), on average, 7.05 turns per km. The second route was selected due to its low speed, short distance, and high number of turns. The route starts and ends from the end point in route 1, the parking lot at the local supermarket. On the route, the speed limit is 50 km/h. We assume that the actual speed will be significantly lower as most of the route is in a residential area, and also due to the number and frequency of turns.

Semi Rural (Low speed, Few turns)
The route is defined as semi rural based on its mix of residential surroundings and highway. The distance of the route is 1.9 km with 5 turns (1 roundabout, 2 right turns, 2 left turns), on average, 2.63 turns per km. rural based on its mix of residential surroundings and highway. The third route was selected due to its medium speed, short distance, and low number of turns. The route starts from a parking lot at a local supermarket and ends at another parking lot at Aalborg University. The speed limit is, for the most part, 50 km/h.

Tasks

For the passengers, we incorporated the same tasks as the ones performed in the training data collection, i.e. texting, drinking, and radio tuning. Additionally, we included two new tasks as well to challenge the classification accuracy of IRIS on untrained tasks. The new tasks were; talking on their phone and eating chips from a bag. Lastly, possible self-imposed movements by passengers were not explicitly annotated during the experiment and are therefore reported under the category; None Task.

Procedure

In the experiment, each participant went through three distinct stages; briefing, a driving session, and debriefing.

We briefed our participants with a short explanation of the experiment and their role. We also provided a flyer which contained information about the experiment and contact information for each of the researchers. A consent form was handed out for participants to sign. By signing the form, participants agreed to having listened, read, and understood the information provided to them about the experiment. Participants with the role of a passenger were handed a bag of chips and a bottle of water to be used in the driving session. Lastly, participants were equipped with a Sony smartwatch 3 on their left hand while a researcher held the paired Nexus 5x smartphone.

In the driving session, drivers had to drive whilst following directions and passengers performed tasks which were given verbally and sporadically by an experiment leader, and all tasks were conducted for by all passenger. When a task was performed, an observer annotated the duration of the task with timestamps on the Nexus 5x smartphone. When arriving at the destination of a route, each participant regardless of role were told to exit the car and walk around until the system detected that they were traveling on foot. To avoid legal issues with insurance claims in the case of accidents, drivers were asked to drive their own car and passengers were driven around in a Chevrolet Spark. All cars used in the experiment were equipped with manual gears. The size of the cars ranged from micro to medium. In total 9 different car models were used during the experiment, since some drivers drove in the same vehicle.

In the debriefing, participants answered a questionnaire about their use of watches, driving experience, and smartphone use as a driver.

Data Collection and Analysis

The required data for evaluating IRIS was collected during the experiment. To facilitate this, the Android application used for collecting training data was reused and extended to allow an observer to explicitly note the start and end time of passenger tasks. Additionally, the application saved timestamps for each new step in the execution flow of IRIS, i.e. started driving, in-vehicle detected, role detected, stopped driving, and walking detected.

As with the training dataset, each session was divided into 15 second windows of accelerometer data for the analysis. We consider all windows individually, i.e. a window does not affect the next. Afterwards, each window was annotated with further information. Firstly, the role as recognised by IRIS and likelihood from each HMM. Secondly, the name of a task was annotated if at least two seconds of said task had occurred within the window. Thirdly, whether the road at that time was a straight, a roundabout, or a turn using the same criteria as with tasks. Task annotation was accomplished through each start and end timestamp created during passenger sessions. For road annotation, one researcher manually created road segments for each driving session using the smartphone GPS coordinates which were verified afterwards by another researcher. Each segment was given one of four types (straight, roundabout, or right/left turn), as well as a start and end timestamp based on the first and last GPS coordinate timestamp in the segment, respectively. To clarify, a passenger window is annotated with a left turn and the drinking task, if the window occurred in a left turn while the passenger drank from a bottle.

We carry out Chi-square tests of independence between the classification accuracy and the following variables; routes, road segments, tasks, and participants.

Threshold Values for the Experiment

In the description of IRIS, we mentioned that the difference between likelihoods is compared against a threshold for both driver and passenger. As mentioned, each window is annotated with a likelihood for driver and passenger.

We divided the windows into four equal groups (quartiles) based on the difference between the likelihoods from driver and passenger for each window. Quartiles are values that divide a dataset into quarters, each representing 25% of the data. Quartiles are denoted lower, middle, and upper quartile. The lower quartile divides the 25% of the windows with the lowest differences from the windows with the highest 75%, whereas middle quartile divide the differences in half. We chose to use the lower and middle quartile value as thresholds. In short, using the lower quartile as threshold means that we exclude 25% of all windows with the lowest difference in likelihood between driver and passenger from the dataset.

RESULTS

We evaluated IRIS with respect to its accuracy and detection time by using no threshold, the lower quartile as a threshold, and the middle quartile as a threshold. The accuracy of IRIS was determined by the total amount of windows that has been classified correctly. This includes windows containing all types of road segments and trained tasks, if any occurred, i.e. drinking, texting and interacting with the radio. Our dataset consists of 1093 windows (596 driver and 497 passenger) for no threshold, 883 (472 driver and 421 passenger) for the lower quartile, and 585 (306 driver and 279 passenger) for the middle quartile.

Overall Accuracy and Detection Time

With respect to the overall accuracy and detection time of IRIS, we found that no threshold results in a lower accuracy but faster detection time. Furthermore, using middle quartile as a threshold results in a higher accuracy but slower detection time.The accuracies of IRIS can be seen in Table 1.

Threshold	Accuracy (N = 24)	Avg. time (N = 24)	Driver (N = 12)	Passenger (N = 12)
None	**87.0%**	15s	93%	80.9%
Lower	**91.7%**	18.6s	94.7%	88.6%
Middle	**93.3%**	28.0s	97.4%	89.2%

Table 1: **The table shows the accuracies for IRIS using no threshold, lower quartile and middle quartile as threshold. For drivers calculations are made from 12 participants, and likewise for passengers we consider 12 participants separate from driver participants.**

Our findings show that when using no threshold, IRIS achieves an accuracy of 87.0%, i.e. 13.0% of the windows was misclassified. Looking at drivers and passengers separately, using no threshold, IRIS achieved an accuracy of 93.0% for drivers, and an accuracy of 80.9% for passengers. Using thresholding with the lower quartile, the overall classification accuracy of IRIS is 91.7% (94.7% for drivers and 88.6% for passengers). When applying thresholding with the middle quartile, IRIS achieves an overall accuracy of 93.3% (97.4% for drivers and 89.2%

for passenger). In short, excluding windows where the HMMs yields likelihoods close to each other showed an increase in accuracy.

With regards to detection time, we found that applying thresholds resulted in an increased detection time. Furthermore, Google's in-vehicle detection time is a bottleneck. The detection time was 15.0 seconds for no threshold, on average 18.6 seconds for the lower threshold, and on average 28.0 seconds for the middle threshold. The total detection time for IRIS is further influenced by the time it took Google's in-vehicle API to detect that participants were in a vehicle. The average detection time was 29.5 (SD = 20.17) seconds which translates to a loss of two classification windows in IRIS. As an example, the total detection time for IRIS was on average 44.5 seconds for the no threshold configuration. Additionally, it took Google's API 26.9 (SD = 6.57) seconds to detect when a person was on foot which included participants getting out of the car first. For the first three participants, step detection did not function, and therefore their detection time was not included.

To elaborate on our overall accuracies, we show results for driver and passenger data separately.

Road Influence and Driver Behaviour

We report on drivers and their accuracies on three different variables, Routes, Road segments, and participants. Recall that among the 12 drivers we have collected 596, windows of 15 second acceleration data.

Routes

Between routes, the classification accuracy of IRIS only deviates slightly as seen in Table 2. The accuracies across routes are all above 90%. We expected that IRIS would achieve a higher classification accuracy for routes with a high number of turns over ones with few. Interestingly, this was not the case. The Suburban route only achieved the second highest classification accuracy.

	Rural (N = 262)	**Suburban** (N = 180)	**Semi-rural** (N=154)
Drivers (N = 12)	92.4%	93.3%	93.5%

Table 2: **The table shows the accuracy for drivers of IRIS in relation to Routes: Rural, Suburban, and Semi-rural. The amount of windows representing the specific Route is also shown.**

A Chi-square test of independence was conducted to examine the relation between the routes and prediction of windows for each route. The statistical relation between the variables were shown not to be significant, X^2 (2, 596) = 0.249, p = 0.883. Interestingly, IRIS performs almost equally on the chosen routes. Threshold-based classification for both the lower and middle quartile showed an increase in accuracy for all routes, but Chi-square tests of independence still showed no statistical significance.

Road Segments

Surprisingly, we found that the accuracies for straight road segments are only slightly lower than the road segments with occurring turn moments with no thresholding. Additionally, thresholding resulted in an increase in classification accuracy and with straight road segments overtaking right/left road segments.

We evaluated windows containing the four different road segments (straights, roundabouts, right turns, and left turns) for drivers. Accuracies for road segments, for no thresholding, with windows excluded under the lower quartile, and with windows excluded under the middle quartile, can be seen in Table 3. Note that 29 windows contained more than one turn or roundabout, and are not included.

Threshold	**Straights**	**Round-about**	**Right**	**Left**	**Avg.**
None (N = 567)	91.6%	91.9%	97.3%	95.5%	94.1%
Lower (N = 452)	93.4%	94.9%	98.2%	96.3%	95.2%
Middle (N = 298)	97.1%	100%	97.1%	95.7%	97.5%

Table 3: **The table shows the accuracies for road segments, calculated with no threshold, thresholding at the lower quartile, and thresholding at the middle quartile.**

Both straights and roundabouts are subject to a large increase in accuracy. Straights go from 91.6% with no threshold to 97.1% when using a middle quartile threshold. Roundabouts go from a 91.9% accuracy with no threshold to 100% using the middle quartile as a threshold, suggesting that classifying drivers in roundabouts to be the most beneficial. Interestingly neither right or left turns change much in prediction accuracy, only increasing with almost one percent point going from no threshold to the lower quartile threshold, whereas going from no threshold to a middle quartile threshold yields almost equivalent accuracies.

A Chi-square test of independence for road segments and predictions made on each individual road segment, showed that there is no significant relation between the two, X^2 (3, 567) = 4.358, p = 0.225. Surprisingly, the result suggests that classifications made for drivers on straight segments are as useful as classifications made during turns. Likewise, Chi-square tests of independence for each of the thresholds showed no statistical significant relation.

Driver Behavior

We found that accuracies across test subjects for drivers deviated substantially from each other regardless of the used threshold.

For no threshold, a Chi-square test of independence showed a statistically significant relation between driver participants and the prediction of each window for the individual drivers, X^2 (11, 596) = 60.122, p < 0.001. Likewise, Chi-square tests of independence for both thresholds showed a statistically significant relation between the two variables as well. The result suggests that one or more participant characteristics impact classification accuracy. On further inspection of individual participants' classification accuracy, we found that the accuracies of driver participants vary with 25% points, ranging from 75% for participant number 6 to a 100% for participant number 17, 20, 21, 23 and 24.

Task Influence and Passenger Behaviour

In the same manner as for drivers, we report on passengers on two conditions, tasks, and participants. Recall that among the 12 passengers, we have collected 497 windows of 15 second acceleration data. Out of the 497 windows, 144 had been annotated as encapsulating a trained tasks, i.e. drinking, texting, and radio. Additionally, we collected 125 windows containing untrained tasks, i.e. eating from a bag of chips, and talking on the phone. Lastly, two windows out of the 497 were annotated with multiple tasks and were excluded.

Trained vs. Untrained Tasks

Unsurprisingly, we found that untrained tasks achieved a lower accuracy than their counterpart.

We separated windows encapsulating tasks. As mentioned, we define a window encapsulating a task when at least two seconds of a task occurs within the given window. Table 4 shows accuracies for the performed tasks, separated in trained and untrained tasks. The untrained tasks, calling and eating, were not included in the training dataset and are the tasks with the lowest accuracy. We see a difference of 11.2% points in accuracy going from 81.8% accuracy when no task is performed down to 70.6% accuracy for the task of when passengers received a call. Different observations on tasks between participants were made. Our definition of "No task" is that the participant was not asked to do anything specific in these periods. This means that participants were able to do what they desired. However, for the most part we observed that they would place their hands in their lap in different positions.

For trained passenger tasks with no threshold, a Chi-square test of independence on classifications on each individual task window and tasks showed no significance between the two variables, X^2 (3, 495) = 0.621, p = 0.892, i.e. no particular trained task differs significantly from the others with regards to classification accuracy.

Additionally, we ran a Chi-square test of independence on classifications between trained and untrained tasks. The test showed a statistical significance between the two variables X^2 (2, 622) = 6.255, p = 0.012. This show that the classifications differ statistically between untrained and trained tasks. Consequently, this suggest that incorporating tasks in the training of passengers improves upon the overall system accuracy as tasks model passenger behaviour.

	Task	Accuracy
Trained (N=144)	None (N=351)	81.8%
	Drinking (N=49)	77.6%
	Texting (N=58)	79.3%
	Radio (N=37)	81.1%
Untrained (N=125)	Eating (N=76)	71.1%
	Calling (N=51)	70.6%

Table 4: **The table shows the accuracy for passengers in relation to trained (None, Drinking, Texting, Radio) and untrained (Eating, Calling) tasks. The amount of windows representing the specific tasks is also shown.**

During the experiment, we observed that passengers performed tasks in various ways. Taking a look at how passengers performed a single task, the complexity and diversity of execution becomes apparent. As an example, the task of drinking from a bottle can be divided into three steps, grabbing the bottle, unscrewing the cap, and drinking from it. Looking at two different participants and the way they completed the task, gave us an insight into the diversity of the task completion. Participant 8 grabbed the water bottle with his left hand, swapped the bottle to his right and unscrewed the cap with his left, and drank with his right. In contrast, participant 9 took the the water bottle with his right hand, swapped it over to his left and unscrewed the cap with his right after which he drank with his left. Both participants were right handed.

Passenger Behavior

Similarly to drivers, we found that accuracies across passengers deviated substantially from each other regardless of the configuration of IRIS.

A chi-square test of independence between the passenger participants and predictions made on windows for each individual participant showed a statistical significant relation, X^2 (11, 497) = 107.309, p < 0.001. The result suggests that one or more participant characteristics impact classification accuracy. Further inspection of individual participants' classification accuracy, we found that passenger accuracies varies with 50.8% points, going from 46.9% for participant number 2, to 97.7% for participant number 15. This range is more than double the range of accuracies between driver participants, which was 25% points.

Diving into the classification of windows with excluded cases under the lower quartile, we see that a total of 59 windows are misclassified for passengers. We found that participant 1, 2, and 10 combined, were the least accurate passengers holding 57 out of 95 misclassified windows between them. This means that three participants were responsible for 60% of all the misclassifications.

DISCUSSION

We developed a system, IRIS, to facilitate the identification of users' in-vehicle role using a low-budget commercially available smartwatch. From a field experiment with 24 participants (12 drivers, 12 passengers), we found that IRIS achieved an overall accuracy of 87.0%, recognising drivers with 93% accuracy and passengers with 80.9%. Our results indicate that the passenger role is more complex than the driver role from an activity recognition perspective. Subsequently, we found that applying threshold-classification generally increases the accuracy of the roles, but at the cost of a higher detection time. Surprisingly, we found that straight road segments achieve similar accuracies to segments with turn moments, i.e. roundabouts, rights, and lefts.

Accuracy, Detection Time, and Experiment Design

By comparing the classification accuracy of IRIS (87.0%) and average detection time (44.5 seconds) to previously developed systems capable of identifying in-vehicle roles, we argue that our system achieves comparable results.

Chu et al. [8], achieved a classification accuracy of 84.67% and a best case detection time of 3 minutes and 15 seconds using a smartphone. The accuracy was obtained through cross-validation of test data collected from five participants. Wang et al. [9] achieved a classification accuracy of 80% on the first turn with a detection time between 38 to 48 seconds using a smartphone. Their system is evaluated based on data collected on two distinct routes with two different vehicles. Liu et al. [3], achieved a classification accuracy of 96.7% and an average detection time of 21.13 seconds using a smartphone and wrist-worn wearable.

Bulling et al. [20] described that a common challenge within activity recognition is achieving high intraclass variability, that is, taking into account that different individuals perform gestures differently. As our overall accuracies for each role indicate, we have taken intraclass variability into account and succeeded to do so for drivers. But, the passenger still shows room for improvement. The previous research have not taken intraclass variability into account as the purpose in these studies has been to show proof of concepts for the use of different sensing units and approaches in role identification systems. To specify, Wang et al. do not use different test subjects, and Chu et al. and Liu et al. both use cross validation with few or no test participants for evaluation of their accuracies. As a result, the reported accuracies may have been increased because the training and test data is based on the same individuals which results in low intraclass variability [20].

A Fixed In-Vehicle Reference Point

Compared to previously developed in-vehicle role identification systems, IRIS does not require a fixed in-vehicle reference point, such as the use of additional hardware, in-vehicle infrastructure, or a fixed smartphone position. IRIS solely makes use of an off-the-shelf smartwatch for identifying hand gestures associated with in-vehicle roles, i.e. the in-vehicle placement of the paired smartphone is irrelevant.

Both Wang et al. [9], and Liu et al. [3] require a fixed reference point in the vehicle to represent car acceleration. Liu et al. argue that car acceleration produces too much noise in wrist-worn wearable acceleration data to accurately classify the in-vehicle role, and thus remove it. In Wang et al. [9], the reference point either requires additional hardware or the use of in-vehicle infrastructure not present in all vehicles. Liu et al. explore a system not reliant on in-vehicle infrastructure, and thus the smartphone is used as the reference point. Consequently, this required the device to be fixed. Chu et al. [8], identified in-vehicle driving role using only a smartphone. Inherently, the system assumes that the device is fixed to the user, e.g. in a user's pocket.

Modelling Passenger Behaviour through Tasks

Previously, Liu et al. [3] explored the idea of having passengers conducting tasks, specifically using the phone and eating. In IRIS, we decided to model passenger behaviour based on a list of tasks that they could perform.

We chose to impose a variety of tasks on passengers to test the system on realistic actions performed by passengers. We chose 3 tasks which we knew were in the training data, adjusting the radio, drinking from a bottle, and writing on a mobile phone, acquiring 81.1%, 77.6%, and 79.3%, in accuracy respectively. Additionally we created 2 tasks which was not in the training data, eating from a bag of chips, and answering a call on the mobile phone, gaining 71.1%, and 70.6% in accuracy respectively. Unsurprisingly IRIS had a harder time recognising the untrained tasks opposed to the trained tasks. However, it is still impressive that IRIS is able to identify untrained tasks with over 70% accuracy. We argue that based on untrained tasks being harder to classify, the passenger role and their potential tasks should be thoroughly investigated to understand the range of tasks that can be used to adequately encapsulate the behaviour of a passenger.

The Impact of Individual Behavior

We have shown that there is a statistically significant difference between test subjects both for drivers and passengers. The largest difference in accuracy between two drivers is 25% points and 50.8% points for passengers. This result strengthens the argument that driving a car is less complex activity [31], opposed to the complexity of actions available to a passenger. Furthermore, the results also strongly suggest that one person's behavior is not necessarily representative of another's, meaning that to represent individuality in the classification models, one would have to look at a diverse training set. Additionally, the results corroborate that intraclass variability is a challenge which is present in human activity recognition systems [20].

As reported earlier, we found that participant 1, 2, and 10, were the cause of 60% of all misclassifications for passengers. During our experiment, we observed that participant number 1 and 2 would gesticulate while speaking with us, and participant 2 had a tendency to turn around looking at the backseat when speaking. Participant number 1 used an especially long amount of time when performing the drinking task, resulting in him sitting with his left hand raised for an extended period. For participant number 10, no notable deviations were observed. Note that these observations are what we see as special occurrences and can be part in why these participants were misclassified to an extended degree.

Interestingly, participants who had the largest amount of misclassifications did not do anything they were not allowed to, and we did not observe anything that we would qualify as abnormal behavior. Even though our system had trouble classifying these particular participants, it is not unrealistic that it could have succeeded with more diversity in the training data.

Machine Learning within HCI

Dove et al. [32], found that surveyed and interviewed UX designers had no previous education or limited experience with machine learning. Consequently, from a UX design perspective little is understood about the potentials, limitations, and challenges to working with the technology. Furthermore, Dove et al. argue that for many UX designers machine learning takes on an otherworldly or magical character. In a call to action, Dove et al. encourage designers and researchers to be more open about the challenges they face when working with machine learning, in particular with regards to data gathering and labeling. Responding to the call, we discuss our insight into working with machine learning within HCI.

For supervised machine learning, it is important to distinctly consider two phases, training and evaluation of the system to be developed. Acquiring a diverse training set is necessary for recognising a specific pattern such as hand gestures through the use of sensing. Collecting training data for an in-vehicle environment is a time consuming task because it requires multiple people to actually drive around equipped with mobile sensing units. Simply, collecting sensor data for such an environment in a laboratory will yield different results, due to the environmental impact.

The collection of data to train our HMMs required us to recruit participants with different car models to ensure that the resulting dataset had as many different examples of driving as possible. Secondly, mobile devices currently have insufficient computational power to rapidly train a model. We had to setup a server to store sensor data and train our HMMs. Even with a server that had vastly more resources than the smartphone (16 GB of memory and 3.6 Ghz CPU), some training sessions of the HMMs took up to 72 hours to complete. Lastly, in the training phase of IRIS several possibilities was explored to achieve a higher classification accuracy. First, we attempted to use other sensors from the smartwatch in combination with acceleration data. Secondly, we experimented with different model parameters, and lastly we experimented with a completely different machine learning algorithm. Each change in model parameters or machine learning algorithm required new models to be trained.

Design Ideas

In-vehicle activity recognition systems, such as IRIS, enable designers of in-vehicle systems to design applications targeted at drivers. As an example, systems that target bad driver habits which can be a safety concern, e.g. texting and one handed steering.

Mobile Phone Blocking

Our first design idea is an application which blocks notifications from drivers. Mobile phones have been found to be a source of distraction that negatively affect attention on the primary driving task [33, 34, 35, 36]. Using IRIS, we propose the development of an in-vehicle role-aware safety application to lock mobile devices for drivers but keep them unlocked for passengers. Additionally, such an application, could potentially explore the concept of dynamically changing the lock state of a driver's device depending on contextual information, e.g. lock or unlock based on a speed threshold.

Steering Wheel Handling

Our second design idea is a notification application which can help drivers maintain optimal hand positions on the steering wheel. Studies on drivers' hand position have found that drivers regardless of age, gender, and experience consciously, or unconsciously, tend to adopt a hand position which is sub-optimal for steering and maintaining control over their vehicle in emergency situations [37, 38, 39]. We propose the development of a hand position awareness application that can help drivers become aware of their hand position in-real time.

Driver Fatigue Detection

Our third design idea, is an application which can detect if drivers is getting drowsy during driving by monitoring heart rate and whether the driver has their hand on the steering wheel.

LIMITATIONS

Because our experiment design included imposing tasks on passengers, we argue that this might have influenced some passenger participants to conduct themselves differently than they would have as passengers in a real life driving situation.

With regards to our findings about the influence that routes, road segments, tasks, and individuals have on the accuracy of IRIS, and the fact that it is designed for users with their smartwatch on the left hand in left hand drive cars is a limitation. We are unable to conclude if our aforementioned findings are applicable for countries with right-hand drive or users wearing their smartwatch on the right hand.

CONCLUSION

In this paper, we presented an activity recognition system, IRIS, for in-vehicle role identification of users. Our work corroborates previous insights on in-vehicle role identification, i.e. wearable device motion sensors and machine learning can be used to develop systems that can accurately classify in-vehicle roles. In particular, our work shows that role identification systems based on data collection from commercially available smartwatches can lead to high accuracies for distinguishing between drivers and passengers. We show through a field experiment with applied threshold-based classification, that IRIS achieves an overall accuracy of 93.3%. We found that individual behaviour influence classification accuracy, especially for passengers. Surprisingly, IRIS as a system and its accuracies for different types of road segments show that the whole part of a route, straights included, has merit for in-vehicle activity recognition systems.

Through our work, we have identified two topics for possible future research. Firstly, applications and systems that can leverage in-vehicle role identification have to be developed and studied to better understand their potential in an in-vehicle context. Secondly, a further understanding of the impact of individual user behavior and vehicle types in recognising in-vehicles roles.

References

[1] Agnes Grünerbl et al. "Monitoring and Enhancing Nurse Emergency Training with Wearable Devices". In: *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*. UbiComp/ISWC'15 Adjunct. Osaka, Japan: ACM, 2015, pp. 1261–1267. ISBN: 978-1-4503-3575-1. DOI: 10.1145/2800835.2807941. URL: http://doi.acm.org/10.1145/2800835.2807941.

[2] Juha Parkka et al. "Activity classification using realistic data from wearable sensors". In: *IEEE Transactions*

on information technology in biomedicine 10.1 (2006), pp. 119–128.

[3] Luyang Liu et al. "Toward detection of unsafe driving with wearables". In: *Proceedings of the 2015 workshop on Wearable Systems and Applications*. ACM. 2015, pp. 27–32.

[4] Jin-Hyuk Hong, Ben Margines, and Anind K. Dey. "A Smartphone-based Sensing Platform to Model Aggressive Driving Behaviors". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '14. Toronto, Ontario, Canada: ACM, 2014, pp. 4047–4056. ISBN: 978-1-4503-2473-1. DOI: 10.1145/2556288.2557321. URL: http://doi.acm.org/10.1145/2556288.2557321.

[5] Abhinav Parate et al. "Risq: Recognizing smoking gestures with inertial sensors on a wristband". In: *Proceedings of the 12th annual international conference on Mobile systems, applications, and services*. ACM. 2014, pp. 149–161.

[6] Edison Thomaz, Irfan Essa, and Gregory D Abowd. "A practical approach for recognizing eating moments with wrist-mounted inertial sensing". In: *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM. 2015, pp. 1029–1040.

[7] Zhenyu Chen et al. "Unobtrusive Sleep Monitoring Using Smartphones". In: *Proceedings of the 7th International Conference on Pervasive Computing Technologies for Healthcare*. PervasiveHealth '13. Venice, Italy: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2013, pp. 145–152. ISBN: 978-1-936968-80-0. DOI: 10.4108/icst.pervasivehealth.2013.252148. URL: http://dx.doi.org/10.4108/icst.pervasivehealth.2013.252148.

[8] Hon Lung Chu et al. "In-vehicle driver detection using mobile phone sensors". In: *ACM MobiSys*. 2011.

[9] Yan Wang et al. "Sensing vehicle dynamics for determining driver phone use". In: *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*. ACM. 2013, pp. 41–54.

[10] Kuo-Ying Huang. "Challenges in Human-Computer Interaction Design for Mobile Devices". In: *Proceedings of the World Congress on Engineering and Computer Science*. Vol. 1. 2009.

[11] Anhong Guo and Tim Paek. "Exploring Tilt for No-touch, Wrist-only Interactions on Smartwatches". In: *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*. MobileHCI '16. Florence, Italy: ACM, 2016, pp. 17–28. ISBN: 978-1-4503-4408-1. DOI: 10.1145/2935334.2935345. URL: http://doi.acm.org/10.1145/2935334.2935345.

[12] Gierad Laput et al. "Skin Buttons: Cheap, Small, Low-powered and Clickable Fixed-icon Laser Projectors". In: *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*. UIST '14. Honolulu, Hawaii, USA: ACM, 2014, pp. 389–394. ISBN: 978-1-4503-3069-5. DOI: 10.1145/2642918.2647356. URL: http://doi.acm.org/10.1145/2642918.2647356.

[13] Mitchell Gordon, Tom Ouyang, and Shumin Zhai. "WatchWriter: tap and gesture typing on a smartwatch miniature keyboard with statistical decoding". In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM. 2016, pp. 3817–3821.

[14] Stephen Oney et al. "ZoomBoard: a diminutive qwerty soft keyboard using iterative zooming for ultra-small devices". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM. 2013, pp. 2799–2802.

[15] Aske Mottelson et al. "Invisiboard: Maximizing Display and Input Space with a Full Screen Text Entry Method for Smartwatches". In: *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*. MobileHCI '16. Florence, Italy: ACM, 2016, pp. 53–59. ISBN: 978-1-4503-4408-1. DOI: 10.1145/2935334.2935360. URL: http://doi.acm.org/10.1145/2935334.2935360.

[16] Shaikh Shawon Arefin Shimon et al. "Exploring Non-touchscreen Gestures for Smartwatches". In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM. 2016, pp. 3822–3833.

[17] Anusha Withana et al. "zSense: Enabling shallow depth gesture recognition for greater input expressivity on smart wearables". In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM. 2015, pp. 3661–3670.

[18] Simon T Perrault et al. "Watchit: simple gestures and eyes-free interaction for wristwatches and bracelets". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM. 2013, pp. 1451–1460.

[19] Ju-Whan Kim, Han-Jong Kim, and Tek-Jin Nam. "M. Gesture: An Acceleration-Based Gesture Authoring System on Multiple Handheld and Wearable Devices". In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM. 2016, pp. 2307–2318.

[20] Andreas Bulling, Ulf Blanke, and Bernt Schiele. "A tutorial on human activity recognition using body-worn inertial sensors". In: *ACM Computing Surveys (CSUR)* 46.3 (2014), p. 33.

[21] Jake K Aggarwal and Michael S Ryoo. "Human activity analysis: A review". In: *ACM Computing Surveys (CSUR)* 43.3 (2011), p. 16.

[22] Ronald Poppe. "A survey on vision-based human action recognition". In: *Image and vision computing* 28.6 (2010), pp. 976–990.

[23] Ling Bao and Stephen S Intille. "Activity recognition from user-annotated acceleration data". In: *International Conference on Pervasive Computing*. Springer. 2004, pp. 1–17.

[24] Peter Norvig Stuart Russel. *Artificial Intelligence. A Modern Approach*. Prentice Hall, 2013.

[25] Angelo Maria Mannini Andrea; Sabatini. "Machine Learning Methods for Classifying Human Physical Activity from On-Body Accelerometers". In: *Sensors 10, no. 2: 1154-1175* (2010).

[26] *DetectedActivity*. URL: https://developers.google.com/android/reference/com/google/android/gms/location/DetectedActivity.

[27] *IBM Wearables SDK for Android*. URL: https://github.com/ibm-wearables-sdk-for-mobile/ibm-wearables-android-sdk.

[28] Jean-Marc Francois. *JAHMM: An implementation of hidden Markov models in Java*. 2010.

[29] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. http://www.deeplearningbook.org. MIT Press, 2016.

[30] Mikael Nilsson. *First Order Hidden Markov Model Theory and Implementation Issues*. Tech. rep. Blekinge Institute of Technology, 2005.

[31] Muhammad Shoaib et al. "Complex human activity recognition using smartphone and wrist-worn motion sensors". In: *Sensors* 16.4 (2016), p. 426.

[32] Graham Dove et al. "UX Design Innovation: Challenges for Working with Machine Learning As a Design Material". In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. CHI '17. Denver, Colorado, USA: ACM, 2017, pp. 278–288. ISBN: 978-1-4503-4655-9. DOI: 10.1145/3025453.3025739. URL: http://doi.acm.org/10.1145/3025453.3025739.

[33] Dan Basacik, Nick Reed, Ryan Robbins, et al. *Smartphone use while driving: a simulator study*. IHS, 2011.

[34] Aubrey Samost et al. "Comparing the Relative Impact of Smartwatch and Smartphone Use While Driving on Workload, Attention, and Driving Performance". In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol. 59. 1. SAGE Publications Sage CA: Los Angeles, CA. 2015, pp. 1602–1606.

[35] Jeff K Caird et al. "A meta-analysis of the effects of texting on driving". In: *Accident Analysis & Prevention* 71 (2014), pp. 311–318.

[36] World Health Organization et al. "Mobile phone use: a growing problem of driver distraction". In: (2011).

[37] M Fourie, D Walton, and JA Thomas. "Naturalistic observation of drivers' hands, speed and headway". In: *Transportation research part F: traffic psychology and behaviour* 14.5 (2011), pp. 413–421.

[38] Bertil Jonsson. "Hand position on steering wheel during driving". In: *Traffic injury prevention* 12.2 (2011), pp. 187–190.

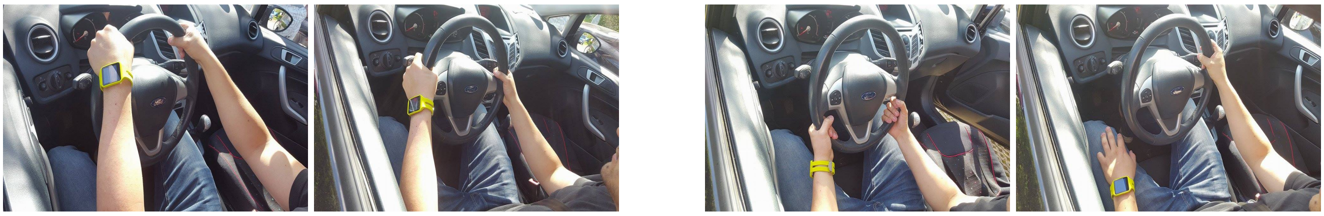
[39] Dick De Waard, Thigri GMPR Van den Bold, and Ben Lewis-Evans. "Driver hand position on the steering wheel while merging into motorway traffic". In: *Transportation research part F: traffic psychology and behaviour* 13.2 (2010), pp. 129–140.

Appendix B

Research Paper 2

Hands-On: Raising Awareness of Driver Hand Position whilst Driving

Thomas A.C. Hald, David H. Junker, Mads Mårtensson
Aalborg University,
Department of Computer Science
Selma Lagerløfs Vej 300, DK-9220
Aalborg, Denmark



ABSTRACT

In several countries, there is a recommended position for a driver to keep their hands on a steering wheel. Primarily, the recommended positions are meant to ensure optimal steering for evasive maneuvers in emergency situations, e.g. swerving to avoid a sudden foreign object on the road. However, the literature have shown that drivers tend to mostly drive with one hand off the steering wheel due to other factors than just arm fatigue. Clearly, this present a problem as drivers either consciously or unconsciously go against recommendations meant to potentially save their lives. We develop and evaluate a system, Hands-On, that in real-time can determine if the left hand position of driver is in a recommended position, and notify them in cases when their hand is not. Mainly, we found that our participants were open to the use of devices pre-installed with Hands-On, citing it as a cool feature rather than a nuisance in this case.

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous; See `http://acm.org/about/class/1998/` for the full list of ACM classifiers. This section is required.

Author Keywords

Authors' choice; of terms; separated; by semicolons; include commas, within terms only; required.

INTRODUCTION

With regards to in-vehicle safety, research has shown that the position of a driver's hands on/off the steering wheel is important for vehicle maneuverability in emergency situations [1]. Conversely, the position of hands on a steering wheel can be described by the means of a clock face - 12 O'clock signifying the top of the wheel. Several countries recommend drivers to have their hands on the wheel at all times in either a 10-2 or 9-3 position for optimal steering and safety [1]. However, research have shown that it is more common for drivers to operate the steering wheel with one hand despite safety implications [2, 3, 4, 5]. By not following hand position recommendations, drivers unintentionally endanger themselves, their passengers, and other trafficants.

To Illustrate, a driver, Bob, traveling down a rural highway and no other trafficants in sight adopts a relaxed hand position. His left hand is resting in the window ceiling and his right hand is in a 4 O'clock position on the steering wheel. Suddenly, a thick branch falls down onto the road ahead and Bob has to perform an emergency avoidance maneuver. But, due to his hand position Bob reacts too late and with his limited control over the car crashes into the branch. Bob's car crash could have been avoided if he had his hands in a recommended position.

To notify drivers of the position of their left hand, we develop an audio and vibrotactile application, Hands-On. Using a smartwatch, the system determines whether or not a driver's left hand is in a recommended position, i.e. at 9 or 10 O'clock. Other Placements of the left hand will after a short delay result in a slight vibration from the smartwatch, and a short high pitched sound from a paired smartphone. Through a user study of Hands-On, we investigate driver hand placement in cars, and whether the system has a positive impact on this.

RELATED WORK

Previously, research have investigated driver hand position for different types of drivers, vehicles, and driving scenarios. We provide a short summary of the published work and findings on driver hand position.

Research has linked driver hand position to workload demand of the primary task of driving a vehicle [2, 4, 3]. In 2005, Walton and Thomas, [3], found some evidence for the use of driver hand position to measure perceived risk of the road context as they were shown to vary depending on speed and complexity of the driving in a naturalistic observation study. In 2007, Walton and Thomas [4] found in a study of SUV and car drivers that both groups recognised that placing both hands on the top part of the steering wheel to be safest in a tense situation. In a relaxed situation, drivers were more likely to drive with only one hand on the top of the wheel. Additionally, they found that drivers did not associate improper hand position negatively as with other observable driver behaviour, e.g. seatbelt use or speed. Hand position has also been shown to be impacted by various driver characteristics, age [6], gender [7], expertise [2], and sitting position [8].

More recently, Schiro et. al, [9], found in a low-speed maneuver driver simulation study that most subjects either crossed one or both hands on a steering wheel due to their hand position. This is a safety issue for modern cars with airbags as such behaviour could cause severe injuries to the upper extremities in the event of airbag deployment [10].

In regards to methods used to study driver hand position, no consensus has been reached on the topic. Both driving simulation [2, 11] and naturalistic driver [6, 1, 3] methods have been adopted. Within the methods, hand position data collection approaches have ranged from independent observers observing vehicle from vantage points looking over roads [3] to detection of hand position on a steering wheel [11]. Approaches such as independent observers are simple but only capture the upper part of the steering wheel; whereas hand position detection are more extreme but technically demanding.

Through the literature, an understanding of driver hand position has been built. Commonly, the research has found portions of participants in their studies to drive with their hands in other positions than the recommended ones [2, 7, 1]. To our knowledge, no published research has of yet developed and explored a system which raise awareness of driver hand position whilst driving using smartwatches.

HANDS-ON

Hands-On is an audio and vibrotactile smartwatch application for Android. It can determine if a driver has their left hand placed in a recommended steering wheel position in real-time using a commercially available smartwatch and -phone. The system is designed to consider the upper half of a steering wheel to be a correct place to put one's left hand and the lower half as an incorrect one, as seen in Figure 1. As further illustration of this, a series of different hand positions that a driver can adopt whilst on the road can be seen on the titlepage. The two pictures above the left column are recommended hand position whereas the two above the right column are examples of incorrect left hand placement in accordance with Hands-On and research [1, 3]. To clarify, a user that is i driving with their left hand in a recommended position, can after a while adopt a more relaxed one with their left hand below the recommendation. Consequently, Hands-On begins to remind the user to adjust their hand position through audio and vibrotactile notifications.

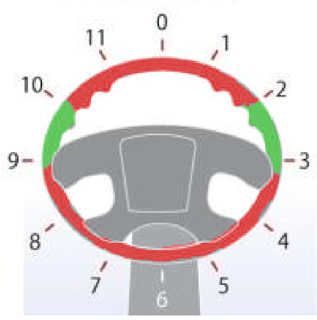


Figure 1: Visualisation of a steering wheel with drawn correct (green) and incorrect (red) zones for hand position.

The system consists of two crucial steps; identification of whether the user is a driver and hand position identification. To accomplish these two steps, Hands-On leverages an In-vehicle Role Identification System (IRIS) [12] developed for Android. IRIS collects acceleration data from a smartwatch and then uses a classifier on a paired smartphone to determine the role of a user in a car. Hand position identification is conducted if IRIS determines a user to be a driver. For hand position identification, we leverage the fact that IRIS is a system based on data where drivers drove with their left hand placed in a recommended position. In short, a driver will be determined to be a passenger by IRIS if they drive with their left hand in any position other than the recommended ones, e.g. 9-3 and 10-2. Through a pilot study, we found that two second windows can be used to determine left hand position effectively.

Consequently, after a user has been identified as a driver by IRIS, Hands-On waits for three consecutive two second windows where IRIS determines the user to be a passenger. Hands-On exploits this information to infer that the driver has their hand placed incorrectly. In such cases, Hands-On prompts the user with an audio and vibrotactile notification to make them aware of their left hand placement.

EVALUATION - APPROACH AND FINDINGS

To explore Hands-On and hand position awareness, a field experiment was conducted where qualitative data was collected through semi-structured interviews and observations about participants' driving habits and reactions to the system. In this section, we describe our approach and findings.

Approach

6 participants were recruited for the study, all were male. The age of the participants ranged from 24 to 26, with driving experience ranging between 6 and 8 years. All participants had a valid driver's license. Three participants reported having a job requiring them to drive regularly or having had one.

Participants reported driving an average of 2.2 times a week, with an average driving period of 35 minutes. The study had three parts - a briefing, a driving session, and a debriefing. In an effort to curb possible unease occurred by wearing unfamiliar technology, participants was selected on the basis of their previous driving experience whilst wearing a smartwatch.

In the briefing, the participants were informed about the study and asked to sign a consent form declaring their willingness to participate given the supplied information. Firstly, The participants were told that they would be going out on a short drive ($\sim$ 15 min.) on a route by their own choosing in, and around, the city of Aalborg, Denmark. Secondly, participants were told to drive as they normally would but also to obey the law. Lastly, participants were given a short introduction to Hands-On as a system and especially its audio/vibrotactile features. During the drive, drivers wore a Sony Smartwatch 3 equipped with a version of Hands-On whilst the paired LG Nexus 5x smartphone was placed in the middle of the vehicle. The experiment leader engaged in small talk with the driver as a passenger would, and an experiment observer took notes. In the debriefing, the experiment leader sat down with participants and conducted a semi-structured interview and the experiment observer took notes. All interviews were audio recorded.

Findings

Findings reported in this section is based on observations we made during the six driving sessions and the interviews we conducted after the sessions. We will report on three aspects of Hands-On; reception and thoughts, impact, and future work towards hand position systems.

We found that participants received between 4 and 15 notifications during their driving session. The trip length for each participant were over the initial 15 minutes, but no longer than 25 minutes.

Participants Initial Reactions to Hands-On

We observed that all participants had one of three approaches when using the system. Participants approached the system with either, curiosity, indifference, or respect. Curious participants set out to identify what the system determined to be incorrect hand position. They did this by playing with the placement of their left hand, ranging from placing it at the bottom parts of the steering wheel to not having their left hand on the wheel at all. 3 out of 6 participants were indifferent, and drove as they normally would, making vibrations reported by Hands-On interesting in relation to realistic everyday driving. All of the indifferent participants did at some point adopt a hand position that was determined to be incorrect by Hands-On, thus triggering a notification.

To outsmart Hands-On, curious participants tried to find a position where the hand position was close to a recommended one, but still slightly off. Attempted left hand positions by participants ranged from bottom to the very top of the steering wheel as well as completely off it. Exploration did not last for the whole driving session and participants were generally surprised when the watch vibrated because of an unconscious readjustment of their hand position.

Respectful participants were noticeably affected by the briefing. This was evident through observations as they sat more rigidly with their hands in a recommended position while also being hesitant to move their hands from this position. For these participants, the experiment leader asked them to explain as to why Hands-On had not yet triggered. Most of these participants answered that it was due to them driving correctly which afterwards resulted in them being more inquisitive and curious about what constituted incorrect driving. As a result, these participants also began to shift their hand position and trigger a Hands-On notifications.

Hands-On Notifications Affect Driver Behavior

We observed that all participants reaction to vibrations during the driving sessions were to reposition their left hand to what they assumed was a recommended hand position, e.g. the 10-2 position. An example of this behavior was that participant A placed his elbow in the left side window and gripped the steering wheel in a lacks grip, Hands-On send a notification and the participant immediately transitioned to another hand position. Even though this was the general behavior, Participant E consciously chose to ignore the notifications for a longer period at the end of their driving session. To illustrate, he was adamant in keeping his left hand at an unrecommended position despite more than 10 notifications reminding him to correct his hand position.

We asked participants if Hands-On affected their driving habits during the driving session. Participant C stated that the first vibration he received surprised him, making him change his hand position. Additionally 3 out of 6 participants, reported that they readjusted their hands following a vibration. Participant A stated that he moved his hands after a vibration because it was "sufficiently annoying". Participant B stated that he thought that, even though the system could be used as a reminder of holding one's hands at a recommended position, he would probably forget about it after five minutes and then the system would have to renotify him. Additionally, 2 out of 6 participants stated that sound from the phone was distracting and it diverted their attention from the primary task of driving.

Although rarely, Hands-On on occasion misclassified participants' hand position. In one instance, the system misclassified the hand position of participant C to not be in a recommended position due to a long turn in a roundabout where the third exit was taken. Interestingly, we observed that participant A with his left hand at a, 12 'o'clock position did not trigger any notifications from the system.

Changes and Uses for Future Hand Position Systems

Hands-On was met with mixed reviews, and participants provided constructive feedback towards changes and the use of Hands-On and similar systems. In relation to the use of audio and vibrations, 3 out of 6 participants stated that vibrations would be sufficient. Additionally 2 out of 6 participants commented on the frequency of consecutive notifications, which is at most once every two seconds. These participants stated that a single notification every minute or every two minutes at most would be sufficient, getting them too frequently was seen as a source of irritation.

When asked whether the participants would voluntarily install and use systems like Hands-On, all answered no. However, 2 out of 6 participants stated that if the system was embedded as an integral part of mobile devices then they would perceive it as an interesting feature and therefore not actively disable it. Despite general unwillingness to use the system for themselves, participants still found scenarios, cases, or groups of drivers that would potentially benefit or willingly use Hands-On. 4 out of 6 participants said that Hands-On could be useful for users interested in safety. Additionally, 3 out of 6 participants, saw Hands-On as helpful for promoting good driving habits in new drivers. Furthermore, participant F saw a potential in Hands-On as a system that intervenes when drivers remove their hands off the steering wheel due to drowsiness.

CONCLUSION

The reception of Hands-On, a driver behaviour awareness system, was well-received. However, as our findings show there are room for improvements. Firstly, notifications due to misclassifications were shown to be confusing. Secondly, the frequency and severity of notifications was a source of distraction and annoyance. Thirdly, the use of audio was found to be an annoyance, especially for participants that regularly listen to radio. Based on our findings, we have derived the following design guidelines for designers of such systems;

- *Wait a bit and be accurate*, i.e. Allow drivers the opportunity to correct themselves before attempting to raise hand position awareness.
- *Do not come on too strong*, i.e. when considering notification strategies pick one that increases in severity (weak to strong) and frequency (slow to fast) within a reasonable limit to incrementally raise hand position awareness.
- *Avoid sound*, i.e. When notifying a user, do so through vibrations and avoid the distraction of sounds.

In this paper, we have shown that the creation of systems which classify hand positions is feasible, and we show that it can provide insightful information to users during a drive. We observed that through Hands-On notifications participants at some point chose to readjust their hand position from a relaxed one to a recommended one. We observed that through Hands-On notifications participants at some point chose to readjust their hand position from a relaxed one to a recommended one. Through our interviews, we found that Hands-On as a hand position awareness system can be used for various purposes, such as creating good driving habits in new drivers, to alarm drivers about onset drowsiness, and to monitor truck drivers. Hand position awareness systems and similar driver behavioural systems have to be further studied to understand their implications in relation to raising attention through distraction, especially during longer than 15 minutes drives.

References

[1] James Hartley. "Look: No hands! Driving on the motorway". In: *Transportation research part F: traffic psychology and behaviour* 42 (2016), pp. 558–561.

[2] Dick De Waard, Thigri GMPR Van den Bold, and Ben Lewis-Evans. "Driver hand position on the steering wheel while merging into motorway traffic". In: *Transportation research part F: traffic psychology and behaviour* 13.2 (2010), pp. 129–140.

[3] D Walton and Joan A Thomas. "Naturalistic observations of driver hand positions". In: *Transportation Research Part F: Traffic Psychology and Behaviour* 8.3 (2005), pp. 229–238.

[4] Joan A Thomas and D Walton. "Measuring perceived risk: Self-reported and actual hand positions of SUV and car drivers". In: *Transportation research part F: traffic psychology and behaviour* 10.3 (2007), pp. 201–207.

[5] Jane Stutts et al. "Driver's exposure to distractions in their natural driving environment". In: *Accident Analysis & Prevention* 37.6 (2005), pp. 1093–1101.

[6] M Fourie, D Walton, and JA Thomas. "Naturalistic observation of drivers' hands, speed and headway". In: *Transportation research part F: traffic psychology and behaviour* 14.5 (2011), pp. 413–421.

[7] Bertil Jonsson. "Hand position on steering wheel during driving". In: *Traffic injury prevention* 12.2 (2011), pp. 187–190.

[8] Bertil Jonsson et al. "Seat adjustment–capacity and repeatability among occupants in a modern car". In: *Ergonomics* 51.2 (2008), pp. 232–241.

[9] J Schiro et al. "Steering wheel hand position in low-speed maneuvers". In: *Transportation Research Part F: Traffic Psychology and Behaviour* 21 (2013), pp. 133–145.

[10] Claude Serra et al. "Upper limb traumatic lesions related to airbag deployment: a case report and review of literature". In: *Journal of Trauma and Acute Care Surgery* 65.3 (2008), pp. 704–707.

[11] GB Piccinini et al. "Drivers' hand positions on the steering wheel while using Adaptive Cruise Control (ACC) and driving without the system". In: *Proceedings of the human factors and ergonomics society Europe* (2013), pp. 207–216.

[12] David Junker, Thomas Hald, and Mads Mårtensson. "IRIS: Employing Machine Learning and Smartwatch Gesture Recognition for Driver Detection". diploma thesis. University of Aalborg, 2017.

Appendix C

Resume

Denne specialeafhandling forsker i udnyttelsen af sensorer i smartwatches relateret til en kontekst foregående i et køretøj. Formålet er, at forhøje den kontekstuelle bevidsthed af fremtidige applikationer som relaterer sig til en bilkontekst. Bidraget i denne afhandling er todelt. Først viser vi, at det er muligt at udvikle et system, der kun ved brug af sensorer i et smartwatch kan identificere den kontekstuelle rolle for en bruger i et køretøj. Dernæst viser vi, at det førnævnte resultat kan udnyttes til at udvikle applikationer, der er kontekstuelt bevidste i køretøjer.

Vi redegører for eksisterende forskning indenfor forhøjning af den kontekstuelle bevidsthed af mobile applikationer brugt i køretøjer, specifikt i relation til brugen af sensorer i mobile enheder til at genkende hvilken rolle brugere har i et køretøj. Vi identificerede tre aspekter som endnu ikke var kortlagt i relation til følgende; brugen af kommercielle smartwatches, indflydelsen af realistiske betingelser under kørsel og indflydelse af individuelle brugere. Gennem udvikling af et system, *In-vehicle Role Identification System* (IRIS), som udnytter maskinlæring på indsamlet accelerometer data fra et smartwatch til at genkende hvorvidt en bruger er en fører eller en passager i den bil de befinder sig i. For at indsamle træningsdata til vores maskinlæringsalgoritme udførte vi en træningsdata indsamlings fase. I denne fase indsamlede vi accelerometer data fra 97 kørselssessioner med 10 forskellige mennesker i syv forskellige biler. I et felteksperiment evaluerede vi nøjagtigheden og genkendelsestiden af IRIS i relation til indflydelsen af forskellige betingelser, herunder, forskellige ruter, vej segmenter, passager opgaver og individuel opførsel. 24 testdeltagerer (12 førerer, 12 passagerer) blev rekrutteret gennem Aalborg Universitet, personlige forbindelser og vores vejleders netværk. Under udførelsen af eksperimentet, kørte testdeltagere 3 forskellige ruter hvor passagerer udførte opgaver givet af en eksperimentleder.

Vi opnår en overordnet nøjagtighed på 87,0% med en gennemsnitlig genkendelsestid på 45,5 sekunder. Ydermere viser vi, at nøjagtigheden kan forhøjes til 93,3% dog på bekostning af forhøjelse af genkendelsestiden (57.5 sekunder). Vi evaluerede IRIS i relation til de førnævnte betingelser og fandt ud af, at kun individers opførsel havde en statistisk signifikant indflydelse på genkendelsensnøjagtigheden. Dette konkluderer at det er vigtigt at tage hensyn til individers opførsel når denne type systemer skal designes, udvikles og evalueres.

Muligheden for at kunne genkende forskel på føreren og en passager muliggør udvikling af systemer som er kontekstuelt beviste. Vi påviser dette ved udvikling af Hands-On, som facilitere realtids bevidsthed om førerens venstre hånd ved brug af et kommercielt smartwatch. Dette muliggør at brugere kan notificeres, gennem vibration og lyd, når deres hånd position på rattet afviger fra en anbefalet position. Yderligere udførte vi et brugerstudie, hvor vi undersøgte seks brugeres reaktion til et system som Hands-On. Vi observerede at brugere reagerede naturligt til de notifikationer Hands-On leverede, hvilket resulterede i at de flyttede deres hænder tilbage i en anbefalet position.

Udviklingen og evaluering af IRIS og Hands-On påviser, at det er muligt, udelukkende ved brug af sensorer i et smartwatch i kombination med maskinlæring, at forhøje den kontekstuelle bevidsthed i en kørende bil.