Delft University of Technology

Your Car Knows Best

van Gent, P.

DOI

Publication date
2021

Document Version
Final published version

Citation (APA)

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Your Car Knows Best

Paul van Gent

Delft University of Technology

Your Car Knows Best

Dissertation

for the purpose of obtaining the degree of doctor

at Delft University of Technology

by the authority of the Rector Magnificus Prof.dr.ir. T.H.J.J. van der Hagen,

chair of the Board for Doctorates,

to be defended publicly on

Friday 9th, July 2021 at **15:00** o'clock

by

Paul VAN GENT

Master of Science, Leiden University, the Netherlands

born in Noordwijkerhout, the Netherlands

This dissertation has been approved by the promotors.

Composition of the doctoral committee:
Rector Magnificus,   chairperson
Prof.dr.ir. B. van Arem   Delft University of Technology, promotor
Prof.dr.ir. C.N. van Nes   Delft University of Technology, promotor
Dr.ir. H. Farah   Delft University of Technology, copromotor

Independent members:
Dr. A.M. Brouwer   TNO
Prof.dr. L. Boyle   University of Washington, United States
Prof.dr.ir. R. Mugge   Delft University of Technology
Prof.dr. M.P. Hagenzieker Delft University of Technology

Printed in the Netherlands

Dedicated to

Hannah Loth

Zonder jou was deze weg niet mogelijk geweest

Intelligence is the ability to adapt to change
Stephen Hawking

Somewhere, something incredible is waiting to be known
Carl Sagan

Every one of us is, in the cosmic perspective, precious. If a human disagrees with you, let him live. In a hundred billion galaxies, you will not find another.
Carl Sagan

Acknowledgements

I joined Delft University of Technology in September 2015 in the 'Taking the Fast Lane' project. Doing a Ph.D has been a wonderful journey that both shaped me and was a deeply humbling experience. A journey of this magnitude, however, is never undertaken alone.

I would like to express my deepest gratitude to my advisors during the journey: Haneen Farah, Nicole van Nes and Bart van Arem. Through personal and professional guidance they helped me develop and helped build this thesis into what it has become.

Haneen, you've been always supportive, full of energy and have a sharp mind. The ways you have helped me develop my scientific side are hard to summarize in such a brief paragraph. Know I'm forever grateful for your endless support and guidance.

Nicole, I still remember the day in the summer of 2015 at SWOV, when you told me: "Paul, At TU Delft they're looking for a technical Psychologist, and I think you might be who they're looking for". In bridging the worlds of psychology and technology I've always been a bit of 'the odd duck', and you helped me find a place where I could truly develop both sets of skills. You have my deepest thanks for this, as well as your guidance and advice along the way.

Bart, you have a talent of 'steering without steering'. With a light touch you helped guide the direction my Ph.D. took without intruding upon the academic freedom to pursue what was necessary to pursue in order to complete this work. At the right moments you held up a mirror or had advice ready that helped me develop and continue.

I am grateful to Linda Boyle, Anne-Marie Brouwer, Ruth Mugge, and Marjan Hagenzieker for being members of my doctoral committee, for taking the time to read through the draft manuscript, and for your comments that without doubt helped improve the manuscript and research.

I would like to also thank all my colleagues and friends at the Transport & Planning department of Delft University of Technology, and my former colleagues at the Dutch Road Safety Research Institute (SWOV) for all advice, fun and friendship we had along the way. In particular my (former) office mates Hamid, Hari, Lin, Mo, Na, Raeed, Shengye, and Silvia. There were always fun and interesting moments and working with you was a pleasure. Edwin and Peter, the morning coffee was always warm and the breaks 'gezellig', thank you for not just this but also the crucial work you do in keeping our ict and equipment inventory in working order.

My thanks extend to all those involved in the project who provided an inspirational environment. In particular Victor Knoop, Hari Nagular Subraveti, Shengyue Yao, Christiaan Tiberius, Peter de Bakker, and Riender Happee.

Along the way I had the opportunity to set up and teach in a deep learning workshop for academic staff and Ph.D students for two years. This was a great experience that I learned a lot from. My thanks to Jan van Gemert for contributing his expertise to the course and teaching the theory, as well as to Osman Kayhan for his help in fine-tuning the exercises and helping teach the practical part of the course. Conchita van der Stelt offered valuable help in the practical side of organizing the course.

A special thanks goes to Bert van Wee for his valuable comments on the literature review included in this thesis.

I would like to acknowledge and thank also my close friends Bas, Inge, Jorry, Nathalie, Robin, and Tessa. We've been through so much over the last 18-odd years, and your support, ideas, and endless 'moah moah's have kept me sane in sometimes trying times.

Coen, Merel, from being hungover in the early morning classes, we all graduated and went on to become parents. The wheels of time grind on with you two being the constant. Thank you for your friendship, support, advice and tips along the way, even before this particular adventure started.

My appreciation and thanks extends to everyone I omit in these pages, since I still need to include a thesis in this these pages as well, who nonetheless have been a valued part of my life.

Lastly that brings me to you, Hannah. My life partner and support through all this. It's no longer just the two of us. You and our beautiful little rascal Thijmen have been my encouragement and kept me going through the difficult parts. Your names deserve to be on this work as much as mine.

<div align="center">Thank you all.</div>

Research Funding and Context

Content

List of Figures

List of Tables

Chapter 1.

Introduction

Abstract

Congestion is a major issue in traffic systems around the globe, with high economical, societal, and health-related costs incurred. Mitigating congestion is a difficult issue. Counterintuitively, building more roads might not lead to a congestion reduction. This is because, as more road area becomes available and congestion reduces, more people might choose to take up their car as a primary mode of transportation, again increasing congestion.

This chapter describes another solution to the problem: reducing congestion through a change in human behaviour. First the costs and previously attempted congestion reduction strategies will be discussed, followed by an overview of how congestion might arise through moving bottlenecks. Finally, lane-specific control as a possible solution to the congestion problem is discussed.

1.1   Introduction

Around the globe, traffic jams are a source of stress and irritation (Hennessy & Wiesenthal, 1999) to drivers. 2019 saw a 17% increase of congestion on Dutch roads[1] compared to 2018, when it rose 20% compared to the year before. Congestion reduced by 63% during the 2020 COVID-19 pandemic[2]. Although it remains an open question how the post-pandemic situation will develop, it is at least conceivable that congestion levels will again increase. It cost the Dutch transport sector roughly €1.5 bln in 2019 alone (Economische Wegwijzer 2020).

Aside from being expensive, congestion increases accident rates, which in turn raise the costs, both human and financial, of congestion further. There exists a U-shaped relation between accident rates and the traffic volume-to-capacity ratio (Zhou & Sisiopiku, 1997), where both low and high traffic volume to capacity ratios lead to increased accident rates. Low volume traffic offers a monotonous driving environment, which potentially contributes to fatigue or distraction, or allows for more opportunities to speed causing run-off-road accidents, thereby increasing accident rates. Furthermore, increases in speed variability lead to an increase in accidents (Quddus, 2013), and especially at the tail end of congestion there are large differences in speed.

Lastly, besides cost and increased accident rates, congestion asserts a cost on human health as well. Lower speeds increase vehicle emissions. For example, Requia et al. (2018) link increased particular matter emissions due to congested traffic to 206 yearly deaths in the Toronto and Hamilton area in Canada alone. Worldwide, particulate matter pollution has been linked to roughly 4.2 million yearly premature deaths (Forouzanfar et al., 2016). Congestion related increases in emissions of CO_2 (Barth & Boriboonsomsin, 2008) needlessly exacerbate already critical climate issues related to carbon dioxide concentrations warming the planet, as well as NOx emissions from even the latest euro-6 engines (Ko et al., 2019). Clearly, ways to reduce congestion need to be found not just to alleviate driver irritation and economic damage, but to protect the health of ourselves as well as that of our planet.

This dissertation is not about the need for more asphalt, nor about the need to reduce mobility, rather it explores how to persuade drivers to use the already existing roads more effectively through lane-specific advice, thereby alleviating congestion on (Dutch) multi-lane motorways. The present chapter describes the context of the work developed in this thesis. The next sections will outline examples of efforts taken and planned to reduce congestion (1.1), how congestion may arise (1.2), how inefficient lane use creates problems (1.3), and how lane-specific control might reduce the problem (1.4).

1.2   Efforts to Solve Congestion

Congestion has long been a part of the worldwide traffic system despite considerable research and efforts to mitigate it, which indicates the complexity of the challenge. In the Netherlands and abroad subtantial work has gone into expanding infrastructure, yet unfortunately congestion keeps growing year on year.

---

[1] https://www.anwb.nl/verkeer/nieuws/nederland/2019/december/knelpunten-2019

[2] https://www.anwb.nl/verkeer/nieuws/nederland/2020/december/thuiswerken-leidt-tot-63-procent-minder-files-in-2020

Is this surprising? It turns out that reducing congestion is not simply a matter of expanding the road network or adding more lanes to existing roads. In a study published in 2011, Duranton, & Turner (2011) analysed two decades of traffic data (1983 - 2003). They showed that the number of vehicle kilometres travelled increases proportionally with the number of road kilometres available. In other words: building more road kilometres does not lead to less congestion. This is explained by a latent travel demand. When roads get congested frequently, a modal shift might happen for parts of the population: those who can, will choose to commute by bike, public transport, travel at another time of day, or choose to change their working location. The same is true for the transportation industry: congestion-related delays are expensive, and as congestion frequency increases, other modes of transport become more cost-effective and attractive.

When the available kilometres of road are expanded, the latent demand will manifest as more people switch travel modes, and congestion will not be reduced substantially. To repeat: building more roads invites more traffic onto them, not necessarily leading to a reduction in congestion rates. This has been dubbed the 'fundamental law of road congestion'. Although the law was formulated based on US data, it has been found to hold in European contexts as well with an elasticity of between 0.7 and 1.0 (Garcia-Lopez, Pasidis, & Viladecans-Marsal, 2017). This does not mean governments should stop investing in roads. A well connected and maintained road network is crucial, but the focus should not be only on building more roads, but more on other means of reducing congestion and through this reducing the negative impacts of congestion.

One such initiative to reduce congestion was the 'Spitsmijden' project ('avoiding rush hour travel' in English) (Meurs et al., 2015). It rewards drivers financially for travelling outside of rush hours. This relies on extrinsic motivation to change behaviour: rather than wanting to change the behaviour themselves (intrinsic motivation), compensation is offered to those who change their behaviour. The 'spitsmijden' project reported relapses after rewards stopped, but the relapses were only partial. These relapses were to be expected, as for example it is known from research that offering extrinsic rewards as motivation will only lead to short-term behavioural change, and can actually undermine intrinsic motivation (Deci, Koestner, & Ryan, 1999). The project reported how low-effort behavioural changes like changing route, travelling at a different time of day, or working from home, showed a high rate of relapsing to old behaviours once the rewards stopped. Interestingly, behaviours that required substantial effort from participants, such as changing mode to bike, e-bike, or public transit, showed much lower relapse rates. Encouraging participants to put in effort themselves is a clever way of exploiting the sunk cost effect (Arkes & Blumer, 1985), an effect whereby there is a greater tendency to continue with a behaviour if an investment of time, effort, or money has already been made. The 'spitsmijden' project is part of the larger 'optimising use' initiative ('beter benutten' in Dutch), which concluded in 2018 and showed overall positive results on encouraging modal shifts, improving travel time and reducing emissions (Programma Beter Benutten Vervolg | Eindrapportage, 2018). It turns out reducing congestion through other means than expanding infrastructure can be effective, too.

Further progress can likely be made by replacing the current fixed-rate vehicle tax in the Netherlands with a kilometre-based tax. This may help reduce the number of car kilometres by as much as 25% (Ubbels, Rietveld, & Peeters, 2002) as well as reduce emissions from cars by up to 70%, by stimulating the move to cleaner cars and different modes of transit. After initial support from the government (Besseling, Geurs, Hilbers, Lebouille, & Thissen, 2008), the political will for the plan unfortunately evaporated completely in 2016.

Before discussing how to reduce congestion through other means, the next section first explains how a particular type of congestion can arise, and how the occurrence of these types of congestion can be reduced.

1.3   'Ghosts in the Machine', or Something Else?

Congestion usually occurs at a bottleneck, and 'spookfiles' ('phantom traffic jams') are a special type of bottleneck: when traffic density increases, headway decreases as vehicles start travelling closer together. At that point any small disturbance, such as a mild braking action by a single car, may trigger the car behind it to brake stronger (Calvert, Van Den Broek, & Van Noort, 2011; Van Den Broek, Netten, Hoedemaeker, & Ploeg, 2010), followed by the next car braking stronger still. This causes shockwaves in the traffic stream, causing a fluctuation of alternating braking and accelerating traffic. These shockwaves propagate upstream over the road at a speed of approximately 18 km/h (Lu & Skabardonis, 2007), often snowballing along the way to the point of causing congestion. Small disturbances might be absorbed by the traffic system without leading to congestion (Schakel, Arem, & Netten, 2010), but the risk of shockwaves causing a breakdown in traffic flow increases as traffic density increases (Sugiyamal et al., 2008). It is estimated that about 20% of Dutch traffic jams are 'spookfiles' (Suijs, Wismans, Krol, & Van Berkum, 2015).

Human behaviour contributes to phantom traffic jams in several ways. Aside from braking actions, lane changes can create disturbances in traffic flow, as well as reduce capacity because a vehicle briefly occupies two lanes (Coifman, Mishalani, Wang, & Krishnamurthy, 2006). Often, drivers choose to change lanes under the assumption that another lane is travelling faster, an assumption that may be more illusion than reality (Redelmeier & Tibshirani, 1999). These lane changes can induce braking actions by other vehicles in the adjacent lane (Ahn & Cassidy, 2007), initiating a shockwave that can, again, cause a breakdown in traffic flow resulting in congestion (Banks, 2002) that seemingly appears out of nothing.

1.4   Inefficient Lane Usage Driving Congestion

As traffic density on a road segment increases, the distribution of traffic over available lanes changes with it. For example, Knoop et al. found that the lane distribution of vehicles significantly changes with speed (Knoop, Duret, Buisson, & Van Arem, 2010). Under normal 100km/h conditions, the outside lane (right-most lane) is underutilised because many drivers choose to drive on the other (faster) lanes. This creates an apparent paradox: there is sufficient road capacity available, but nonetheless congestion occurs frequently.

The problem is that, although the road still has spare capacity available, the left lane is already at capacity and may become unstable. Shockwaves can now form that eventually cripple flow on all lanes through minor braking actions or lane changes as discussed in the previous section, especially when a lane is at or near capacity. Once congestion sets in and a small traffic queue forms the problem worsens immediately: the total capacity of the road drops between 6% (Hall & Agyemang-Duah, 1991) and 25% (Yuan, Knoop, & Hoogendoorn, 2014). The location of capacity drop now acts as a moving bottleneck and the congested area rapidly grows. This occurs because the queue discharge rate is lower than the capacity of the bottleneck area. In other words: there are more vehicles entering the congested section than there are exiting, leading to a growing traffic jam. Several reasons are a cause of this, and include increased time headways (Treiber & Helbing, 2003), increased lane change behaviour under congested conditions (Laval & Daganzo, 2006), and drivers not accelerating out of the queue efficiently

(Schakel & Van Arem, 2014). As the queue grows upstream past other potential bottlenecks such as off-ramps, traffic that would not normally pass the moving bottleneck becomes included in the congestion as well, further compounding the growth of the congested section. The key to preventing the occurrence of these phantom traffic jams while the road is not yet at capacity, then, is to somehow make sure all available lanes are utilized properly.

1.5   Reducing Congestion Through Lane-Specific Control

Lane specific control may offer a solution to the described shockwave congestion (Yao, Knoop, & van Arem, 2017). Lane specific control means directing single vehicles to specific lanes, which allows for fine-grained control of the traffic state. By proactively distributing traffic over the available road space, the road area is better utilised and the ability of dense traffic to absorb arising shockwaves can be improved, leading to a reduction in congestion. Several challenges need to be solved to allow for lane specific control.

First, to control vehicles on a lane-level requires real-time information of the specific on which each vehicle is currently traveling. This is an issue because standard GPS has an accuracy error which is worse than the lane width, making lane-level positioning impossible. High accuracy GPS systems are available but have significant disadvantages. Differential GPS relies for example on a (non-existent) dense network of beacons to allow for precise positioning, and dual-frequency GPS receivers are prohibitively expensive. Recently, advances in single-frequency precise point positioning (SF-PPP) algorithms have been proposed that mitigate many of the issues, allow for both a quick time to get a location fix, and use low-cost hardware (Knoop, De Bakker, Tiberius, & Van Arem, 2017). It allows for fast lane-level positioning even in situations where there may be no clear view of the full sky (de Bakker & Tiberius, 2017).

Second, traffic states need to be predicted in advance based on a lane-level. Traffic data in the Netherlands comes from loop detectors. These provide information on traffic counts at specific points along the road. Going from this data to a prediction of traffic state is another research area in the project (Subraveti, Knoop, & Van Arem, 2018, 2020; Yao et al., 2017), based on methods developed earlier (Schakel & Van Arem, 2014).

Finally, any lane-specific direction or advice will have to be followed by a driver, who has their own goals and probably wants to get to a destination as fast as possible. The lane-specific advice given will not necessarily be in the benefit of this driver. An optimal usage of available lanes will mean some drivers will be asked to move to a slower lane than they might prefer to drive on. This will be in the benefit of the traffic flow on the road segment as a whole, not necessarily in the benefit of the individual, and so getting drivers to follow these requests could prove to be difficult. This is the main research goal of this thesis as will be outlined in the next chapter.

References

Ahn, S., & Cassidy, M. J. (2007). Freeway Traffic Oscillations and Vehicle Lane-Change Maneuvers. Transportation and Traffic Theory 2007. Papers Selected for Presentation at ISTTT17, 691–710.

Arkes, H. R., & Blumer, C. (1985). The psychology of sunk cost. Organizational Behavior and Human Decision Processes, 35(1), 124–140. https://doi.org/10.1016/0749-5978(85)90049-4

Banks, J. H. (2002). Review of empirical research on congested freeway flow. Transportation Research Record, (1802), 225–232. https://doi.org/10.3141/1802-25

Barth, M., & Boriboonsomsin, K. (2008). Real-world carbon dioxide impacts of traffic congestion. Transportation Research Record, (2058), 163–171. https://doi.org/10.3141/2058-20

Besseling, P., Geurs, K., Hilbers, H., Lebouille, R., & Thissen, M. (2008). Effecten van omzetting van de aanschafbelasting op personenauto's in een kilometerprijs. Retrieved from https://www.cpb.nl/sites/default/files/omnidownload/CPB-PBL-Boek-24-Kansrijk-woonbeleid.pdf

Calvert, S. C., Van Den Broek, T. H. A., & Van Noort, M. (2011). Modelling cooperative driving in congestion shockwaves on a freeway network. IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, 614–619. https://doi.org/10.1109/ITSC.2011.6082837

Coifman, B., Mishalani, R., Wang, C., & Krishnamurthy, S. (2006). Impact of lane-change maneuvers on congested freeway segment delays pilot study. Transportation Research Record, (1965), 152–159. https://doi.org/10.3141/1965-16

de Bakker, P. F., & Tiberius, C. C. J. M. (2017). Real-time multi-GNSS single-frequency precise point positioning. GPS Solutions, 21(4), 1791–1803. https://doi.org/10.1007/s10291-017-0653-2

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. Psychological Bulletin, 125(6), 627–668; discussion 692-700. https://doi.org/10.1037/0033-2909.125.6.627

Duranton, G., & Turner, M. A. (2011). The fundamental law of road congestion: Evidence from US cities. American Economic Review, 101(6), 2616–2652. https://doi.org/10.1257/aer.101.6.2616

Economische Wegwijzer 2020. (2020), https://www.tln.nl/app/uploads/2020/11/TLN_EcoWegwijzer_2020_A4_2P_DEF_RGB_HR.pdf.

Forouzanfar, M. H., Afshin, A., Alexander, L. T., Biryukov, S., Brauer, M., Cercy, K., … Zhu, J. (2016). Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. The Lancet, 388(10053), 1659–1724. https://doi.org/10.1016/S0140-6736(16)31679-8

Garcia-Lopez, M.-A., Pasidis, I., & Viladecans-Marsal, E. (2017). The 'fundamental law of highway congestion' and air pollution in Europe's cities.

Hall, F. L., & Agyemang-Duah, K. (1991). Freeway Capacity Drop and the Definition of Capacity. Transportation Research Record, 1320, 91–98.

Hennessy, D. A., & Wiesenthal, D. L. (1999). Traffic congestion, driver stress, and driver aggression. Aggressive Behavior, 25(6), 409–423. https://doi.org/10.1002/(SICI)1098-2337(1999)25:6<409::AID-AB2>3.0.CO;2-0

Knoop, V. L., De Bakker, P. F., Tiberius, C. C. J. M., & Van Arem, B. (2017). Lane Determination with GPS Precise Point Positioning. IEEE Transactions on Intelligent Transportation Systems, 18(9), 2503–2513. https://doi.org/10.1109/TITS.2016.2632751

Knoop, V. L., Duret, A., Buisson, C., & Van Arem, B. (2010). Lane distribution of traffic near merging zones influence of variable speed limits. IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, 485–490. https://doi.org/10.1109/ITSC.2010.5625034

Ko, J., Myung, C. L., & Park, S. (2019). Impacts of ambient temperature, DPF regeneration, and traffic congestion on NOx emissions from a Euro 6-compliant diesel vehicle equipped with an LNT under real-world driving conditions. *Atmospheric Environment*, *200*(April 2018), 1–14. https://doi.org/10.1016/j.atmosenv.2018.11.029

Laval, J. a., & Daganzo, C. F. (2006). Lane-changing in traffic streams. Transportation Research Part B: Methodological, 40(3), 251–264. https://doi.org/10.1016/j.trb.2005.04.003

Lu, X., & Skabardonis, A. (2007). Freeway Traffic Shockwave Analysis: Exploring NGSIM Trajectory Data. Transportation Research Board 86th Annual Meeting, (January 2007), 19.

Meurs, H., Stelling, C., & Haaijer, R. (2015). Belonen voor spitsmijden: effecten van mobiliteitsprojecten. Tijdschrift Vervoerswetenschap, 51(4), 63–86.

Programma Beter Benutten Vervolg | Eindrapportage. (2018). Retrieved from https://beterbenutten.nl/assets/upload/files/BeterBenutten_Vervolg_rapport_2018 def.pdf

Quddus, M. (2013). Exploring the Relationship Between Average Speed, Speed Variation, and Accident Rates Using Spatial Statistical Models and GIS. Journal of Transportation Safety & Security, 5(1), 27–45. https://doi.org/10.1080/19439962.2012.705232

Redelmeier, D. A., & Tibshirani, R. J. (1999). Why cars in the next lane seem to go faster. Nature, 401(6748), 35. https://doi.org/10.1038/43360

Requia, W. J., Higgins, C. D., Adams, M. D., Mohamed, M., & Koutrakis, P. (2018). The health impacts of weekday traffic: A health risk assessment of PM2.5 emissions during congested periods. Environment International, 111(December 2017), 164–176. https://doi.org/10.1016/j.envint.2017.11.025

Schakel, W. J., Arem, B. Van, & Netten, B. D. (2010). Effects of Cooperative Adaptive Cruise Control on Traffic Flow Stability. Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference On, (Idm), 759–764. https://doi.org/10.1109/ITSC.2010.5625133

Schakel, W. J., & Van Arem, B. (2014). Improving traffic flow efficiency by in-car advice on lane, speed, and Headway. IEEE Transactions on Intelligent Transportation Systems, 15(4), 1597–1606. https://doi.org/10.1109/TITS.2014.2303577

Subraveti, H. H. S. N., Knoop, V. L., & Van Arem, B. (2018). Rule based control for merges: Assessment and case study. IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, 2018-Novem, 3006–3013. https://doi.org/10.1109/ITSC.2018.8569221

Subraveti, H. H. S. N., Knoop, V. L., & Van Arem, B. (2020). Improving Traffic Flow Efficiency at Motorway Lane Drops by Influencing Lateral Flows. Accepted for 99th Annual Meeting of the Transportation Research Board.

Sugiyamal, Y., Fukui, M., Kikuchi, M., Hasebe, K., Nakayama, A., Nishinari, K., … Yukawa, S. (2008). Traffic jams without bottlenecks-experimental evidence for the physical mechanism of the formation of a jam. New Journal of Physics, 10. https://doi.org/10.1088/1367-2630/10/3/033001

Suijs, L. C. W., Wismans, L. J. J., Krol, L., & Van Berkum, E. C. (2015). Phantom jam avoidance through in-car speed advice. Transportation Research Procedia, 8, 227–236. https://doi.org/10.1016/j.trpro.2015.06.057

Treiber, M., & Helbing, D. (2003). Memory effects in microscopic traffic models and wide scattering in flow-density data. Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics, 68(4), 1–8. https://doi.org/10.1103/PhysRevE.68.046119

Ubbels, B., Rietveld, P., & Peeters, P. (2002). Environmental effects of a kilometre charge in road transport: An investigation for the Netherlands. In Transportation Research Part D: Transport and Environment (Vol. 7). https://doi.org/10.1016/S1361-9209(01)00023-2

Van Den Broek, T. H. A., Netten, B. D., Hoedemaeker, M., & Ploeg, J. (2010). The experimental setup of a large field operational test for cooperative driving vehicles at the A270. IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, 198–203. https://doi.org/10.1109/ITSC.2010.5625050

Yao, S., Knoop, V. L., & van Arem, B. (2017). Optimizing traffic flow efficiency by controlling lane changes : collective , group and user optima. Transportation Research Record: Journal of the Transportation Research Board.

Yuan, K., Knoop, V. L., & Hoogendoorn, S. P. (2014). Capacity Drop : a Relation Between the Speed In Congestion And The Queue Dicharge Rate. Transportation Research Record: Journal of the Transportation Research Board, 2491, 72–80.

Zhou, M., & Sisiopiku, V. P. (1997). Relationship between volume-to-capacity ratios and accident rates. Transportation Research Record, (1581), 47–52. https://doi.org/10.3141/1581-06

Chapter 2.

Research Overview

Abstract

Lane-specific control is proposed as a solution to congestion, as described in the previous chapter. In order for lanes-specific control to work, the cooperation from the human driver in control of the vehicle is required. To elicit this cooperation, we can try using persuasive approached to persuade the driver to follow directions from the lane-specific control system. This chapter introduces the problem statement of the thesis, which centers around getting a driver to follow an advice that is in the benefit of all drivers, but not necessarily themselves. For the road segment as a whole to experience less congestion, some individual drivers will need to make small sacrifices such as staying in a slower lane. This thesis is about how to communicate with the driver in such a way that compliance to such a lane-specific advice is maximized.

The contributions of this thesis -both scientific and practical- are outlined in this chapter, followed by a reading ouline for the rest of the dissertation.

2.1 Problem Statement and State of the Art

As discussed in the previous chapter congestion can arise even though enough capacity remains on a road segment. In denser traffic the right lane remains underutilised as more drivers choose to drive on the left-most lane, creating a situation where although road capacity remains on the other lanes, the left-most lane might already be near or at capacity and become unstable. Minor events like braking actions or lane changes can create a breakdown in the traffic flow which then spreads across lanes, leading to the onset of congestion.

A potential solution is lane-specific control: by redistributing traffic more efficiently across the available lanes, the available road area can be better utilised and congestion could be avoided. Precise and affordable GPS solutions (SF-PPP) are available that enable lane-specific control on a technical level through lane-specific advices presented to drivers.

The challenge that needs to be solved is getting drivers to follow directions or an advice that may not be in their immediate benefit. Although advices ultimately aim to reduce congestion and thus are in the collective benefit of all drivers on a road segment, an advice may require a single driver to move temporarily to a slower lane. This might be an undesirable action from their point of view.

To convince drivers of the benefits of following lane specific advice, this thesis will develop a persuasive lane-specific advice system with the aim of reducing congestion. What is needed is a way to maximise the persuasive effectiveness of the system, while ensuring the safety of road users. The latter is of particular importance, since advices will need to be given in nearly congested traffic conditions, when driver workload is likely to be high. This section will describe the problem statement and go into the state of the art of both persuasive technology as well as requirements to predict driver workload.

2.1.1 Problem Statement: Winning the Congestion Game

An unbalanced distribution, an upcoming on-ramp or lane drop, or an incident upstream may require a redistribution of traffic to ensure continued flow and avoid congestion. Because the lane-specific advice system focuses on optimising the traffic flow of a road segment, the generated advices will be in the collective benefit of drivers on the specific stretch of the road in terms of minimising the total travel time. This means the advices might not be in the benefit of individual drivers receiving the advice (e.g. 'stay behind this slow truck for now'), creating a potential problem of drivers not wanting to follow these messages (Risto & Martens, 2012).

The problem is compounded by where congestion forms relative to the driver that causes it: if a driver brakes or changes lane and causes a shockwave that eventually leads to congestion, the traffic queue forms behind them. No delay is experienced by the driver and they will not be affected personally, and so most drivers are not even aware of the negative consequences of their behaviour. In a lane specific advice scenario, any given driver will only benefit from the behaviour of others ahead of them, and in turn their behaviour only affects traffic behind them. This is a key characteristic of the problem, and can be seen as a variant of the prisoner's dilemma (Axelrod, 1980).
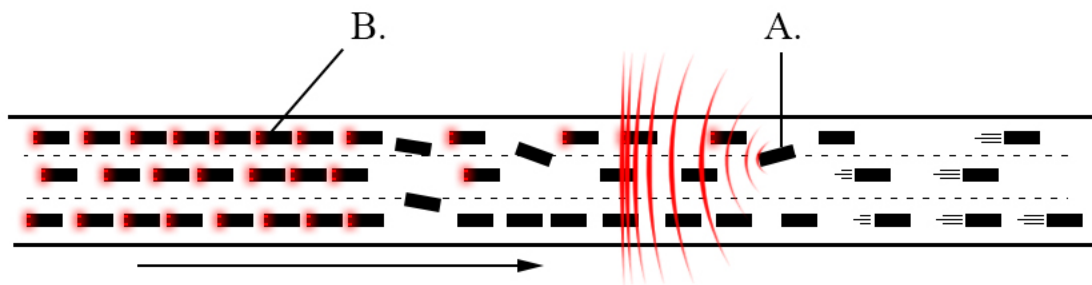
Figure 2.1 - Visualisation of the dilemma that a driver faces when following an advice: any other driver ahead (for example driver A.) disregarding an advice can cause congestion, effectively making the behaviour of drivers behind (for example driver B.) irrelevant.

The prisoner's dilemma is an example of a cooperative game between two players. The setting is that both have participated in a crime and are being questioned in separate rooms. There are two moves are possible: cooperate with the other player by not telling the interrogator anything, or defecting by confessing both of your involvement in the crime to the interrogator, effectively betraying the other player in exchange for a reduction in punishment. The important part is the possible cooperation without knowledge of the other player's actions: both players will gain something if both choose to cooperate. If both players defect nothing will be gained. However, if one of the players defects while the other does not, then the defecting player will gain more (a reduction in sentence) than if both players had cooperated. This means it is in the benefit of the individual player to defect if they want to maximise their gains, but there is the risk that if others do the same, nothing will be gained.

Putting the lane-specific control scenario in terms of the prisoner's dilemma: any driver can choose to cooperate (follow an advice) or defect (not follow an advice). Cooperating is in the collective benefit of all drivers on a road segment: if most drivers follow their advices, congestion can be avoided. It may also require a small sacrifice as some drivers are required to move to or stay in a slower lane. Defecting means choosing to take individual gains, such as staying on a faster lane, at the cost of the collective goal of avoiding congestion. As congestion can form behind a defecting driver, for any driver to choose to cooperate, this implies a certain level of trust that those drivers ahead will also follow their advice. After all, why would any driver invest effort in following an advice and move to a slower lane, if it may be for nothing because another driver ahead defects and causes congestion anyway? Figure 2.1 displays a scenario where driver A defects, causing a shockwave that leads to congestion upstream and resulting in driver B getting stuck in traffic.

This type of blind choice is key to the prisoner's dilemma. Humans can employ various strategies to cope with the uncertainties in the dilemma. Research into human cooperative strategies in these situations has generally found 'tit for tat' reciprocity ('I will help you, if you also help me') to be the predominant strategy (Trivers, 1971). In this strategy potential freeloaders (i.e. those never following an advice and always defecting) will be punished: as people learn about the freeloader, less and less will choose to cooperate. Interestingly, humans are willing to enact this form of punishment even at cost to themselves (Milinski & Rockenbach, 2008; Sigmund, 2007). In traffic the other person is anonymous, anyone can be a freeloader or

a co-operator. Care needs to be taken that congestion is not perceived as the failure to follow advices by other drivers, otherwise each time congestion is encountered, such a willingness to enact punishment may lead to less and less drivers following advices in a negative feedback loop (i.e. 'nobody ever seems to follow their advices so why should I?'). Somehow, enough drivers need to be persuaded to follow advices so that congestion levels are meaningfully affected, and drivers can observe that other drivers follow the given advices, and that it indeed reduces congestion.

2.1.2   Persuading the Driver to Cooperate

Persuading drivers to follow advices requires influencing their behaviour. This can be done using persuasive technology (B.J. Fogg, 2003; Oinas-kukkonen, 2010). Persuasive methods can broadly be divided into: Gamification, Behavioural Economics, and Captology.

Gamification is a relatively new method. Video games create an environment that motivates players to perform specific behaviours in order to reach a goal. Gamification is about lifting those game design elements and applying them outside of video games, in the hopes of creating persuasive situations outside of games that successfully influence behaviour. Examples of often used design elements are leader boards, achievements and challenges (Hamari, Koivisto, & Pakkanen, 2014; Hamari, Koivisto, & Sarsa, 2014).

Behavioural economics seeks to 'understand behaviour by incorporating insights from behavioural sciences into economics' (Avineri et al., 2010). It turns out that humans are not rational when it comes to behaviour and decision making. Rather than rationality, we use a range of heuristics and display biases that act as shortcuts (Kahneman, 2003). Although this allows complex behaviour without processing all the details of each encountered situation, it is not universally the best approach and has been shown to lead to reasoning errors in many cases (Ayton & Fischer, 2004; Bornstein & Emler, 2000; Gino, Moore, & Bazerman, 2011; Kahneman, 2013; Samuelson & Zeckhauser, 1988). For example, framing something in terms of a loss instead of a gain works because the emotions attached to a loss typically weigh stronger in decision making compared to a gain (Avineri, 2011). The choice set offered also has an influence (Lee, Kiesler, & Forlizzi, 2011): pairing a choice with a less attractive alternative will increase the perceived value of the primary choice. Some evidence suggests these reasoning methods may be hardwired into the brain (Martino, Kumaran, Seymour, & Dolan, 2006).

Captology (acronym: computers as persuasive technology) was introduced by Fogg (B.J. Fogg, 1998), and centres around using technology to change behaviour (B.J. Fogg, 2003; Bj J Fogg, 2009). The Fogg Behavioural Model (FBM) (B.J. Fogg, 2009) is prominent in the field of persuasion. It postulates that in order for a persuasive intervention to be successful, three factors need to converge: the person needs to be able to perform the behaviour ('ability'), be motivated to perform the behaviour ('motivation'), and finally a trigger should be present to elicit the behaviour. Targeting simple behaviours has a higher likelihood of success (Bj J Fogg, 2009). A thorough overview of persuasive methods and the way these are applied in this dissertation is given in chapter 3, which deals with my conceptual model and the theoretical foundations of the research.

2.1.3   Being Fair About It: Persuasive Ethics

Persuasive technology refers to technology that is designed to influence or change behaviour. More importantly, for persuasive technology to be considered persuasive it should be free of

coercion, deception, and manipulation (B.J. Fogg, 2003; Oinas-kukkonen, 2010; Smids, 2012). Smids (Smids, 2012) argues that the degree to which a persuasive technology leaves room for voluntary behaviour is an important consideration when deciding whether it is persuasive, coercive, or manipulative. An example he gives is that of the 'fasten your seatbelt' warning systems present in most modern cars. While the driver is still free not to fasten his seatbelt, the loud, persistent and highly irritating noise invariably leads to the driver fastening their seatbelt to stop the noise, regardless of behavioural intent. This is closer to coercion rather than persuasion: the threat of the persistent noise essentially forces the driver to perform a certain behaviour *or else*. This makes the example of the seat belt warning coercive rather than persuasive, as the alternative of not complying to the system's goal is so undesirable that there is really no free choice. In the end whether this is undesirable and something to be avoided is application-specific. There is no question that seatbelts save lives, and by forcing people to wear them every time, save even more lives.

Berdichevsky & Neuenschwander (1999) describe an 'ethics of persuasive technology'. In the work they define principles of ethical persuasive design. The principles call for transparency in the persuasive attempt, strict privacy regulations for user data, and ethical considerations. For example, would the persuasion also be considered ethical if performed by a person rather than a machine? In that context, the fasten your seatbelt warning clearly becomes problematic: having a co-driver screaming repeatedly until the driver fastens their seatbelt would be considered rather unacceptable.

The persuasive lane-specific advice system that is being studied aims to be persuasive in nature, not coercive or manipulative. This means that the goal of the advice should always be clear to drivers: reducing congestion. Any metrics communicated relating to the effects of driver behaviour, such as travel time saved or lost, need to be truthful. At all times the driver should be free to follow or not follow advice, and have control to switch the system off, should they so desire.

2.1.4   When to Bother the Driver? Driver Workload Prediction

The proposed lane-specific advices will be given under specific circumstances. Advices will not be necessary when little traffic is on the road, as sufficient room will be available on all lanes and traffic flow will not be affected. Once traffic becomes dense the lane distribution of vehicles changes and the risk for shockwaves and congestion increases. It is under these dense traffic conditions that an advice needs to be provided to the driver. Dense traffic conditions also increase the workload of the driver (de Waard, 1996). This offers a particular challenge, since there is the potential of raising driver workload and thereby creating unsafe situations. It needs to be determined when it is possible to communicate with the driver, and when it is better not to.

The human capacity for information processing is limited and driving imposes certain demands that partly fill this capacity. This demand is quantified as driver workload (de Waard, 1996), which can be defined as "the level of attentional resources required to meet both objective and subjective performance criteria" (Stanton, Hedge, Brookhuis, Salas, & Hendrick, 2004). The workload arises from an interplay between the demands placed on the driver by the driving task, the complexity of the driving environment in, and the driver's capacity to meet those demands (Fuller, 2005). Workload has been measured in both driving simulators and in real driving situations, usually using physiological measures and self-reported measurements. Although possible, it is not straightforward (Brookhuis & de Waard, 2010). Research to date

has produced mixed results, with different studies pointing to different variables as important for measuring driver workload (Matthews, Reinerman-Jones, Barber, & Abich, 2014; Mehler, Reimer, & Coughlin, 2012; Mehler, Reimer, & Wang, 2011). Among measures collected, heart rate is consistently measures and shown to be related to workload (Mehler et al., 2012; Reimer & Mehler, 2011; Wiberg, Nilsson, Lindén, Svanberg, & Poom, 2015). Recently, supervised learning approaches (i.e. machine learning) have provided indications that workload prediction is possible (Ferreira et al., 2014; Haapalainen, Kim, Forlizzi, & Dey, 2010; Liang, Reyes, & Lee, 2007; Rusnock, Borghetti, & McQuaid, 2015).

To determine when it is safe to communicate with a driver, their workload needs to be predicted. Chapter 4 describes the development of a workload-predictor that can predict workload on a non-binary scale. Because accurate heart rate (variability) analysis software was lacking in the open source domain for PPG recordings, chapters 5 and 6 detail the development and validation of a heart rate analysis toolkit performed within the context of the research.

2.2   Research Objectives and Research Questions

The effectiveness of the lane-specific advice system to prevent or reduce congestion depends on the number of drivers following its advices. The objective of this research is to identify ways of persuading drivers to follow given directions or advices, effectively and safely. This led to the following main research question:

How can we persuade a driver to follow a lane-specific advice without enforcing behaviour?

Fundamental to persuading drivers to follow an advice is the transfer of information. We need to inform drivers of the reason for the advice, the goal of the advice, and possibly the results of their behaviour. This led us to subdivide the main research question into three sub-questions related to communication between the persuasive lane-specific advice system and the driver:

Sub-question 1:
How to communicate with the driver? Fundamental requirements for a persuasive system to be effective and safe.

Sub-question 2:
When to communicate with the driver? Timing messages to low workload periods is safer and more likely to persuade.

Sub-question 3:
What to communicate with the driver? Design of a persuasive HMI system.

These three sub-questions will be answered in sequence in this dissertation. The next section will give an overview of the dissertation outline used to answer these questions.

2.3   Contributions

2.3.1   Scientific Contributions

A conceptual model to describe the use of persuasicve technology in driving contexts. The conceptual model consists of three interacting layers that describe effects of persuasive technology on drivers' decisions and behaviour in driving contexts, based on existing literature. The model describes driver persuasion from a system level, information transfer level, and driver level, and can be used to guide persuasive in-car system design and research efforts. The model is applied to this thesis' problem of giving drivers lane-specific advice to illustrate how it can be used.

Driver workload prediction using off-the-shelf and non-intrusive sensing. This thesis presents a generic machine learning based approach to predict driver workload in real-time. The literature on workload prediction is divided at best with mixed results. By exploring data-driven approaches together with different workload-inducing circumstances, it is shown that workload prediction is possible for individual and group-based models, but that for predicting the workload of to the model unknown drivers, only extremes in workload could be predicted well.

Development and validation of an open-source, noise-resistant heart rate analysis toolkit. The development of HeartPy, a toolkit aimed at analysing noisy photoplethysmogram (PPG) and electrocardiogram (ECG) data, is presented in this thesis. PPG data can be obtained nonintrusively at the wrist, earlobe, finger, or even to some extent contactless through video cameras, meaning the data can be collected unintrusively in scientific studies. This, combined with the increasing availability of low-cost sensors, enables research groups to conduct studies including heart rate data at very low cost. However, low-cost sensors often introduce extra noise in the signal, which complicates analysis. HeartPy was developed to handle the (sometimes noisy) PPG data collected in both the lab and real-world scenarios, and is available for use open source by researchers.

Development and evaluation of a persuasive in-car system. Finally, this thesis develops and presents a persuasive in-car advice system. The conceptual model informs the focus of further research into driver workload, as well as how to implement persuasive messages and information transfer to the driver. Through an end-user driven process the system characteristics such as location of the message, preferred modality of the message, and preferred way of presenting the information to drivers are confirmed and refined. This is applied to a system design and experimentally evaluated in a driving simulator. By contrasting a gamified version, a socially cooperative version, and a control version, possible ways of effectively applying persuasion are identified.

2.3.2   Practical Contributions

A conceptual model to describe the use of persuasicve technology in driving contexts. For car manufacturers and in-car system designers, the developed conceptual framework provides insight into how design choices can affect driver behaviour and safety. Applying this framework early in the design process of a (persuasive) in-car advice system can lead to better system effectiveness and safety.

Driver workload prediction using off-the-shelf and non-intrusive sensing
The presented research on driver workload prediction shows mainly that individualized and group-based models work well for workload prediction in driving settings, but that generalizing to unknown drivers was only successful for extremes in workload. This provides a good direction for practical application of data-driven workload prediction, depending on the application and what range needs to be predicted.

Development and validation of an open-source, noise-resistant heart rate analysis toolkit.
HeartPy is an algorithm developed to handle noisy, real-world PPG and ECG data collected by both medical-grade and low cost off-the-shelf sensors. HeartPy's focus on accuracy makes and open source availability makes it suitable for use in both rapid prototyping as well as real-world projects requiring accurate heart rate analysis on the fly.

Development and evaluation of a persuasive in-car system
For policy makers, the study into driver persuasion using gamification or cooperation show ways of nudging drivers to change their behaviour for the betterment of everyone on the road system, even if the individual behaviours are not directly beneficial to the drivers themselves. The study showed that both using gamified and cooperative approaches lead to significantly higher rates of message compliance than simply asking a driver to do something. These approaches can potentially help to reduce not just congestion, but can help nudge drivers away from dangerous behaviours such as speeding or red light negation as well.

2.4   Dissertation Outline

The dissertation will discuss how to approach driver persuasion. It is divided into three main areas of contributions: **how** to communicate with the driver, **when** to communicate with the driver, and **what to communicate** to the driver.

Chapter 3 discusses the theoretical foundations of the research. It details my theoretical framework for safe driver persuasion. The framework seeks to embed the persuasive lane-specific advice system into literature on safety, (driver) behaviour, and persuasion. This section is about **how** to communicate with the driver, and will be based on the work:

van Gent, P., Farah, H., van Nes, N., & van Arem, B. (2019). A conceptual model for persuasive in-vehicle technology to influence tactical level driver behaviour. Transportation Research Part F: Traffic Psychology and Behaviour, 60, 202–216. https://doi.org/10.1016/j.trf.2018.10.004

Chapters 4-6 discuss **when** to communicate to the driver. A workload estimator is developed that could be used to determine when the driver workload is at safe levels. An ideal moment of low workload could then be chosen to communicate with the driver safely, without risk of overloading them. Chapter 4 describes the development of this online workload estimator.

Chapter 5 describes the development of an open-sourced heart rate analysis toolbox capable of analysing noiy PPG data from low-cost sensors. Chapter 6 details the analysis and validation of the developed toolbox. These chapters are based on the following works:

van Gent, P., Farah, H., van Nes, N., & van Arem, B. (2019). Analysing Noisy Driver Physiology Real-Time Using Off-the-Shelf Sensors: Heart Rate Analysis Software from the Taking the Fast Lane Project. Journal of Open Research Software, 7(1). https://doi.org/10.5334/jors.241

van Gent, P., Farah, H., van Nes, N., & van Arem, B. (2019). HeartPy: A novel heart rate algorithm for the analysis of noisy signals. Transportation Research Part F: Traffic Psychology and Behaviour, 66, 368–378. https://doi.org/10.1016/j.trf.2019.09.015

van Gent, P., Melman, T., Farah, H., van Nes, N., & van Arem, B. (2018). Multi-Level Driver Workload Prediction Using Machine Learning and Off-The-Shelf Sensors. Transportation Research Record: Journal of the Transportation Research Board. https://doi.org/10.1177/0361198118790372

Chapter 7 discusses **what to communicate** to the driver. Based on two questionnaire studies, a persuasive Human-Machine Interface (HMI), an avatar to encourage drivers, and a web-portal where drivers can view and monitor their performance are developed. Persuasive messages are then designed and evaluate the HMI in a driving simulator study. This section will be based on:

Van Gent, P., Farah, H., van Nes, N., & van Arem, B. (2020). The Persuasive Automobile: Design and Evaluation of a Persuasive Lane-Specific Advice Human Machine Interface. Manuscript submitted for review.

Finally, in chapter 8 the main findings are summarized (8.1), followed by a discussion regarding the reasons for several methodological choices made in this thesis and their consequences (8.2), the main findings are then put in context in both science (8.3) and practice (8.4). The dissertation ends with recommendations for future research (8.5).

2.5   Conclusion

In conclusion, the main problem underlying this thesis is that congestion can arise on roads despite that the capacity of the road has not yet been reached. The emergence of this shockwave congestion is driven through inefficient lane usage. Lane-specific advices, where drivers are encouraged to change to or stay in a lane, can be employed to reduce the occurrence of shockwave congestion. The goal of this thesis is to contribute to the development of methods that can help reduce congestion through lane-specific advices given to individual drivers.

References

Avineri, E. (2011). Applying behavioural economics in the design of travel information systems. 43rd Universities Transport Study Group Conference, 1–12. https://doi.org/10.1017/CBO9781107415324.004

Avineri, E., Chatterjee, K., Darnton, A., Goodwin, P., Lyons, G., Musselwhite, C., … Wiltshire, P. (2010). Individual Behaviour Change : Evidence in transport and public health.

Axelrod, R. (1980). Effective Choice in the Prisoner's Dilemma. The Journal of Conflict Resolution, 24(1), 3–25.

Ayton, P., & Fischer, I. (2004). The hot hand fallacy and the gambler's fallacy: Two faces of subjective randomness? Memory and Cognition, 32(8), 1369–1378. https://doi.org/10.3758/BF03206327

Berdichevsky, D., & Neuenschwander, E. (1999). Toward an ethics of persuasive technology. Communications of the ACM, 42(5), 51–58. https://doi.org/10.1145/301353.301410

Bornstein, B. H., & Emler, A. C. (2000). Rationality in medical decision making: a review of the literature on doctors' decision-making biases. 97–107. https://doi.org/10.1371/journal.pone.0056962

Brookhuis, K. A., & de Waard, D. (2010). Monitoring drivers' mental workload in driving simulators using physiological measures. Accident Analysis and Prevention, 42(3), 898–903. https://doi.org/10.1016/j.aap.2009.06.001

de Waard, D. (1996). The Measurement of Drivers ' Mental Workload. https://doi.org/10.1016/j.apergo.2003.11.009

Ferreira, E., Ferreira, D., Kim, S., Siirtola, P., Röning, J., Forlizzi, J. F., & Dey, A. K. (2014). Assessing real-time cognitive load based on psycho- physiological measures for younger and older adults. Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB).

Fogg, B.J. (1998). Persuasive computers: Perspectives and research directions. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, (April), 225–232. https://doi.org/10.1145/274644.274677

Fogg, B.J. (2003). Persuasive Technology: Using computers to change what we think and do. Elsevier Science & Technology.

Fogg, B.J. (2009). A behavior model for persuasive design. Proceedings of the 4th International Conference on Persuasive Technology - Persuasive '09, 1. https://doi.org/10.1145/1541948.1541999

Fogg, Bj J. (2009). Creating Persuasive Technologies : An Eight-Step Design Process. Technology, 91, 1–6. https://doi.org/10.1145/1541948.1542005

Fuller, R. (2005). Towards a general theory of driver behaviour. Accident Analysis and Prevention, 37(3), 461–472. https://doi.org/10.1016/j.aap.2004.11.003

Gino, F., Moore, D. A., & Bazerman, M. H. (2011). No Harm, No Foul: The Outcome Bias in Ethical Judgments. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.1099464

Haapalainen, E., Kim, S., Forlizzi, J. F., & Dey, A. K. (2010). Psycho-Physiological Measures for Assessing Cognitive Load. Proceedings of the 12th ACM International Conference on Ubiquitous Computing, 301–310. https://doi.org/10.1145/1864349.1864395

Hamari, J., Koivisto, J., & Pakkanen, T. (2014). Do Persuasive Technologies Persuade?-A Review of Empirical Studies. Persuasive Technology, 8462, 118–136. https://doi.org/10.1007/978-3-319-07127-5_11

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work? - A literature review of empirical studies on gamification. Proceedings of the Annual Hawaii International Conference on System Sciences, 3025–3034. https://doi.org/10.1109/HICSS.2014.377

Kahneman, D. (2003). Maps of Bounded Rationality : Economicst Psychology for Behavioral. The American Economic Review, 93(5), 1449–1475. https://doi.org/10.1257/000282803322655392

Kahneman, D. (2013). Thinking Fast and Slow. In Journal of Chemical Information and Modeling. https://doi.org/10.1017/CBO9781107415324.004

Lee, M. K., Kiesler, S., & Forlizzi, J. (2011). Mining Behavioral Economics to Design Persuasive Technology for Healthy Choices. Conference on Human Factors in Computing Systems, 325–334. https://doi.org/10.1007/11755494

Liang, Y., Reyes, M. L., & Lee, J. D. (2007). Real-time detection of driver cognitive distraction using support vector machines. IEEE Transactions on Intelligent Transportation Systems, 8(2), 340–350. https://doi.org/10.1109/TITS.2007.895298

Martino, B. De, Kumaran, D., Seymour, B., & Dolan, R. J. (2006). Decision-Making in the Human Brain. (August), 684–688.

Matthews, G., Reinerman-Jones, L. E., Barber, D. J., & Abich, J. (2014). The Psychometrics of Mental Workload: Multiple Measures Are Sensitive but Divergent. Human Factors: The Journal of the Human Factors and Ergonomics Society. https://doi.org/10.1177/0018720814539505

Mehler, B., Reimer, B., & Coughlin, J. F. (2012). Sensitivity of Physiological Measures for Detecting Systematic Variations in Cognitive Demand From a Working Memory Task: An On-Road Study Across Three Age Groups. Human Factors: The Journal of the Human Factors and Ergonomics Society, 54(3), 396–412. https://doi.org/10.1177/0018720812442086

Mehler, B., Reimer, B., & Wang, Y. (2011). Comparison of heart rate and heart rate variability indices in distinguishing single task driving and driving under secondary cognitive workload. Proceedings of the Sixth International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design, 590–597.

Milinski, M., & Rockenbach, B. (2008). Human behaviour: Punisher pays. Nature, 452(7185), 297–298. https://doi.org/10.1038/452297a

Oinas-kukkonen, H. (2010). Behavior Change Support Systems: A Research Model and Agenda. International Conference on Persuasive Technology, 4–14. https://doi.org/10.1007/978-3-642-13226-1

Reimer, B., & Mehler, B. (2011). The impact of cognitive workload on physiological arousal in young adult drivers: a field study and simulation validation. Ergonomics, 54(10), 932–942. https://doi.org/10.1080/00140139.2011.604431

Risto, M., & Martens, M. H. (2012). Improving traffic flow on motorways through individual driver advice: A social dilemma? TRAIL-Beta Congress 2012, (October).

Rusnock, C., Borghetti, B., & McQuaid, I. (2015). Objective-Analytical Measures of Workload – the Third Pillar of Workload Triangulation? Lecture Notes in Computer Science, 9183, 124–135. https://doi.org/10.1007/978-3-319-20816-9

Samuelson, W., & Zeckhauser, R. (1988). Status quo bias in decision making. Journal of Risk and Uncertainty, 1(1), 7–59. https://doi.org/10.1007/BF00055564

Sigmund, K. (2007). Punish or perish? Retaliation and collaboration among humans. Trends in Ecology and Evolution, 22(11), 593–600. https://doi.org/10.1016/j.tree.2007.06.012

Smids, J. (2012). The Voluntariness of Persuasive Technology. In M. Bang & E. L. Ragnemalm (Eds.), Persuasive Technology. Design for Health and Safety (pp. 123–132). Berlin, Heidelberg: Springer Berlin Heidelberg.

Stanton, N., Hedge, A., Brookhuis, K., Salas, E., & Hendrick, H. (2004). Handbook of human factors and ergonomics methods. https://doi.org/10.1201/9780203489925

Trivers, R. L. (1971). The Evolution of Reciprocal Altruism. The Quarterly Review of Biology, 46(1), 35–57. https://doi.org/10.1086/406755

Wiberg, H., Nilsson, E., Lindén, P., Svanberg, B., & Poom, L. (2015). Physiological responses related to moderate mental load during car driving in field conditions. Biological Psychology, 108, 115–125. https://doi.org/10.1016/j.biopsycho.2015.03.017

Chapter 3.

A Conceptual Model for Persuasive In-Vehicle Technology to Influence Tactical Level Driver Behaviour

Abstract

Persuasive in-vehicle systems aim to intuitively influence the attitudes and/or behaviour of a driver without forcing them. The challenge of using these systems in a driving setting is to maximise the persuasive effect without infringing upon the driver's safety.

This chapter proposes a conceptual model for driver persuasion targeting the tactical driving level (i.e. the driver manoeuvring level, such has lane-changing and car-following behaviour). The main focus of the conceptual model is to describe how to safely persuade a driver to change their behaviour, and how persuasive systems may affect driver behaviour.

This chapter explores available driver behaviour models along with persuasive models and aims to integrate these into a framework for safe driver persuasion. The developed model is applied to a case study of a lane-specific advice system that aims to reduce travel time delay and traffic congestion, by advising some drivers to change lanes in order to achieve a better distribution of traffic over the motorway lanes.

3.1   Introduction

3.1.1   The Problem and Scope

The way drivers interact with their vehicles is changing (Damiani, Deregibus, & Andreone, 2009; Ulrich et al., 2013). Modern vehicles are more and more equipped with advanced driver assistance systems (ADAS) that can assist the driver, as well as in-vehicle information systems (IVIS) that provide the driver with traffic information or driving advice. Increases in IVIS/ADAS in-vehicle systems mean that the driving environment becomes more information rich, and more systems compete for the driver's attention.

One field of development within IVIS is that of persuasive systems. Persuasive systems employ techniques or incentives to change drivers' voluntary attitudes or behaviours (Fogg, 2010). The implementation of such persuasive systems in the driving environment can for example help reduce speeding and improve driver engagement during monotonous driving (Steinberger, Proppe, Schroeter, & Alt, 2016). Persuasive systems have also been used to encourage drivers to adopt a more eco-friendly driving style (Ecker, Holzer, Broy, & Butz, 2011), or a safer driving style (Shi, Lee, Kurczak, & Lee, 2012).

While persuasive systems can positively influence driver behaviour and increase safety, they might also introduce new risks (van Nes & Duivenvoorden, 2017). For example, the use of these systems can lead to indirect behavioural adaptations (unwanted and unplanned side-effects) (Martens & Jenssen, 2012), such as the anti-lock braking system (ABS) which when implemented led drivers to maintain shorter headways (Sagberg, Fosser, & Sætermo, 1997). Additionally, increasing the number of in-vehicle systems can negatively influence traffic safety by overloading or distracting the driver at inappropriate times (Reyes & Lee, 2004; Mark S. Young, Brookhuis, Wickens, & Hancock, 2015).

To our knowledge, a conceptual model tying driver persuasion to safety and behavioural outcomes has not been developed yet. In this study, we aim to fill this research gap by developing a conceptual model that describes the effects of in-vehicle persuasive systems on driver behaviour, with the goal of effectively and safely persuading the driver. We will focus specifically on IVIS systems aiming at persuading drivers to change their behaviour at the tactical level. Examples of such systems include lane-specific advice to improve traffic flow (Malte Risto & Martens, 2013; Schakel & Van Arem, 2014), and systems that encourage eco-driving with the goal of reducing pollution (Ecker et al., 2011).

3.1.2   Context of the Developed Framework

The framework was developed in the context of a lane-specific advice system. The goal of this system is to reduce travel time delay and congestion by encouraging a better distribution of the vehicles over the available motorway lanes. This means advising drivers on which lane to take, depending on external factors. For instance, an unbalanced distribution, an upcoming on-ramp or lane drop, or an incident upstream may require a redistribution of traffic to ensure continued flow and avoid congestion. Because the system focuses on optimising traffic flow of a road segment, the generated advices will be in the collective benefit of drivers on the specific stretch of the road in terms of minimising the total travel time. This means the advices might not be in the benefit of individual drivers receiving the advice (e.g. 'stay behind this slow truck for now'), creating a potential problem of drivers not complying with these messages (Malte Risto & Martens, 2012). We incorporated persuasive strategies into the framework to engage drivers

with the system and to also stimulate adherence to lane-specific advices, especially when they are not in the driver's own benefit. The goal of the applied persuasive techniques, is to make the advices attractive enough and convince drivers to follow them. Various ways of accomplishing this are discussed in section 3.

We apply the model to the design of our lane-specific advice system, as described in section 6 of this paper. The developed model can be applied in a broader sense, for example to cooperative driving systems that require drivers to behave in a certain way (Lütteken, Zimmermann, & Bengler, 2016; M Risto, 2014), eco-driving systems (Ecker et al., 2011; Magana & Organero, 2011), or systems stimulating safer driving styles (Rodríguez et al., 2014; Steinberger et al., 2016).

The proposed model is essentially a system-centric model, where a traffic system decides upon for example an ideal traffic distribution, or on which set of driving styles are 'safe', and subsequently stimulates the driver to conform to these types of behaviours. This is in line with persuasive technology, which aims to stimulate certain attitudinal or behavioural patterns over others (Fogg, 1998).

3.1.3   Why Target Driver Behaviour at the Tactical Level?

Driver behaviour is often divided into three levels: the strategic, tactical and control level (Evans & Michon, 1985). The strategic level considers high-level choices related to driver's route choice behaviour, which is generally constant over longer periods of time. At the tactical level, drivers decide upon and perform manoeuvres (e.g. change lane, take exit, overtake car) considering the observable and anticipated part of the road network to reach their strategic goals. At the control level, the driver performs actions to operate the vehicle (e.g. change gears, press accelerator pedal, turn on blinker).

Our conceptual model will focus on safely persuading driver behaviour at the tactical level. From a persuasive perspective, targeting short-term behavioural responses (e.g. adjusting speed, changing lane) increases the effectiveness of the persuasion (see for example Fogg, 2009a; 2009b, Oinas-Kukkonen, 2013, section 3.2, 4.2). From a safety perspective, it is important to manage the demands placed on the driver. According to the Task-Capability Interface model (TCI) by Fuller (Fuller, 2005), driving demands that exceed driver capability might lead to risky situations such as loss of control or a collision. Persuasive effectiveness and safety need to be balanced: targeting tactical level behaviours such as changing lane, especially in demanding traffic conditions, does carry risk and requires careful implementation of in-car interfaces to not become distracting or affect driving adversely. Managing task demands is crucial and a key element in ensuring driver safety when applying persuasive approaches, or when communicating information to the driver.

In order to keep task demands low, a persuasive system should focus on short term behavioural responses that are low effort. These low-effort behaviours can be identified through the behaviour taxonomy of Rasmussen (Rasmussen, 1983). The taxonomy divides driver behaviour into three levels: skill-based, rule-based, and knowledge-based. Skill-based behaviour is highly automatic and can be performed without much attentional demands. Tasks at the control level fall into this category, and for experienced drivers likely some highly automated behaviours at the tactical level as well in non-complex traffic conditions (e.g. lane changing, overtaking, merging). In rule-based behaviour, a response or a set of responses is selected based on earlier learned rules. Knowledge-based behaviour is applied in mostly unknown situations when novel

behavioural responses are needed. Required attentional demands and effort increase from skill-based to rule-based to knowledge-based behaviour. Since behaviour at the tactical level (mostly) consists of skill-based and rule-based behaviours, changing these types of behaviours carries the least risk of imposing high demands on the driver (Birrel, Young, Staton, & Jennings, 2017). Aside from the targeted behaviours, the context and complexity of the driving environment may influence the difficulty of the tactical level manoeuvres as well. An example of a low effort behavioural response is requesting a driver to reduce speed in response to downstream traffic disturbance (skill-based, control level). On the other hand, asking a driver to take a different route along a busy unknown road is likely to place higher demands on the driver, since the execution of a task at the strategic level (knowledge-based behaviour) also involves the tactical (rule-based), and operational level (skill-based) (Alexander & Lunenfeld, 1986).

We hypothesise that a trade-off exists between persuasive effectiveness and the described task complexity. As task complexity increases, persuasive effectiveness should decrease in theory, based on the work of Rasmussen (1983) and Fogg (2009b, section 3.1), and on research showing how a lower perceived ability to perform a target behaviour can lower the intention to perform the behaviour, as well as the likelihood of that behaviour (Elliott, Thomson, Robertson, Stephenson, & Wicks, 2013). The decision which behaviours to select depends on the driver workload, as under- or overloading the driver can create dangerous situations (M S Young, Brookhuis, Wickens, & Hancock, 2014). In combination with for example a driver monitoring system (Aghaei et al., 2016; van Gent, Melman, Farah, van Nes, & van Arem, 2018b), it becomes possible to monitor a driver's state and make inferences about which advices a driver likely can or cannot safely handle. If at any point during the generation of the advice or the execution of the behaviour driver workload exceeds safe levels, the system might decide not to display the advice, retract it, or recommend termination of the execution of the advice.

We first conduct a critical overview of available behavioural models and select the model most applicable to driver behaviour. We then describe driver behaviour at the tactical level and present the general requirements for an in-vehicle persuasive system. Following this, in section 4, we investigate the different persuasive approaches used in the literature and discuss how these approaches fit into the driving environment. Finally, in section 5 we describe the proposed conceptual model and its relation to the current literature. As an example, we apply the conceptual model to the design of a persuasive lane-specific advice system currently in development.

3.2   Describing Behaviour at the Tactical Level

In order to develop our persuasive conceptual model, a behavioural model capable of describing the effects of persuasion on driver behaviour at the tactical level is needed. We have searched the literature for behavioural models that have been used in connection with behavioural change. The search engines used were Google Scholar, Scopus and Web of Science, with the keywords: "behaviour* model AND behaviour* change OR persuasi*". We limited the results to papers of 2005 and newer. Backward snowballing was performed to find the original papers proposing the models. This led to the Social Learning Theory (SLT) (Bandura, 1971), Self-Determination Theory (SDT) (Deci & Ryan, 1985), the Trans-Theoretical Model (Norcross, Krebs, & Prochaska, 2011), and the Theory of Planned Behaviour (TPB) (Ajzen, 1991). For each model, we reviewed their applicability to the driving task, ability to explain the relatively short-term changes in behavioural patterns resulting from persuasion at the tactical level, longer

term attitudes towards the use of the system, as well as the ability to accommodate the effects of persuasive efforts.

3.2.1   Overview of Behavioural Models

The *Social Learning Theory (SLT)*, also known as Social Cognitive Theory, suggests that human behaviour emerges from a constant interaction between environmental, behavioural and cognitive influences (Bandura, 1971; Fluegge, 2016). It incorporates elements of operant conditioning to explain how behaviours are learned through social interactions with others (Watkins, 2016). SLT has been applied to a wide range of fields, including how unwanted behaviours may arise (criminal, drug misuse, smoking, traffic violations) and ways to induce a positive change (Hoeben & Weerman, 2016; Lochbuehler, Schuck, Otten, Ringlever, & Hiemstra, 2016; Watkins, 2016; Zaso et al., 2016), how public perception is formed and influenced (Fluegge, 2016) and students' tendencies to procrastinate (Gadong & Chavez, 2016). The model is directed at describing how learning experiences are shaped by cognitive and social factors.

The *Self-Determination Theory (SDT)* is often cited for its use of intrinsic and extrinsic motivation to explain behaviour (Deci & Ryan, 1985), but actually postulates three basic psychological needs that drive behaviour: autonomy (being in control of one's decisions and behaviour), competence (feeling able to attain behavioural outcomes) and relatedness (feeling understood and respected by others) (Ridgway, Hickson, & Lind, 2016). This model has mostly been applied to behavioural change towards healthier behaviours in the health domain (Friederichs, Bolman, Oenema, Verboon, & Lechner, 2016; Lekes, Houlfort, Milyavskaya, Hope, & Koestner, 2016; Niven & Markland, 2015; Sebire et al., 2016; Staunton, Gellert, Knittle, & Sniehotta, 2015), to medical training (Hoffman, 2014), and to volunteering behaviours (Wu, Li, & Khoo, 2015). The SDT describes behavioural motivation at the macro level (Niven & Markland, 2015).

The *Trans Theoretical Model (TTM)* describes behaviour as consisting of five stages: pre-contemplation (not thinking about changing behaviour), contemplation (thinking about changing behaviour), preparation (making preparations for changing behaviour), action (changing behaviour) and maintenance (keeping changed behavioural patterns intact) (Norcross et al., 2011). The model originated as a fusion of models from several fields of therapy. Like the SDT, the TTM is a macro model of behaviour, describing high level behavioural processes (see for example Brick, Velicer, Redding, Rossi, & Prochaska, 2016; Kushnir, Godinho, Hodgins, Hendershot, & Cunningham, 2015; Prochaska et al., 1994; Yusufov et al., 2016).

The *Theory of Planned Behaviour (TPB)*, based on the Theory of Reasoned Action (Fishbein & Ajzen, 1975), posits that behaviour is directly predicted by 'behavioural intention' and 'perceived behavioural control' (the perceived volitional control over the behaviour). 'Behavioural intention' is predicted by 'attitude towards behaviour', 'social norms regarding the behaviour' as well as 'perceived behavioural control'. The model is displayed in Figure 3.1. In the traffic domain, the TPB has been used to predict traffic violations (Castanier, Deroche, & Woodman, 2013), speeding behaviour (Elliott, Armitage, & Baughan, 2005), evaluating engagement in distracting secondary tasks (Chen, Donmez, Hoekstra-atwood, & Marulanda, 2016), and aggressive driving (Efrat & Shoham, 2013). It has also been used successfully in experiments with the goal of behavioural change (Chorlton & Conner, 2012). It describes how situational constraints and long-term attitudes can influence behaviour.

3.2.2   Representing Persuasive Effects on Tactical Driver Behaviour

We have selected the TPB as a behavioural basis for the conceptual model. This is because this theory can explain both short-term behaviour at the tactical level in the driving setting, as well as the long-term social and attitudinal factors acting on behavioural patterns, which might be relevant when explaining variables like continued system usage. The other reviewed models were either geared more towards changing long-term behavioural patterns (SLT, SDT), describing behaviour at a macro level (SDT, TTM), or describing (changing) behaviour in clinical settings (SDT, TTM). The TPB also plays a central role in models of technology acceptance and trust, such as the Technology Acceptance Model (TAM) (F. D. . Davis et al., 1989; F. D. Davis, 1986) and the UTAUT (Venkatesh, Morris, Davis, & Davis, 2003; Vlassenroot, Brookhuis, Marchau, & Witlox, 2010), which adds usefulness in the context of persuasive systems that need to be trusted and accepted before they can have an effect. In this study, we will utilise the TPB (Figure 3.1) as a behavioural basis for the conceptual model.
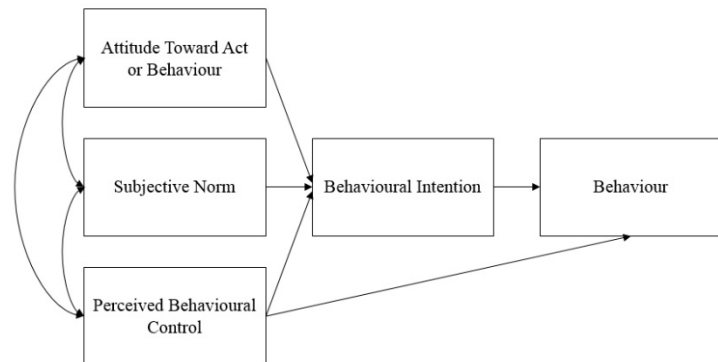
Figure 3.1: The Theory of Planned Behaviour.

In more detail, the TPB posits that behaviour is directly predicted by two factors: 'Behavioural Intention' (BI) and 'Perceived Behavioural Control' (PBC). PBC reflects the degree to which the individual perceives to have volitional control over its own behaviour. In other words, whether the individual believes they are able to successfully perform the target behaviour. PBC directly influences behaviour as well as the intention to perform a behaviour. In some studies, PBC has been split into self-efficacy (perceived ability to perform target behaviour) and perceived controllability (perceptions about whether the person has control over the behaviour or outcomes), with only the self-efficacy component being related to changes in the intent to perform a behaviour and the actual behaviour (Elliott et al., 2013). This indicates that PBC is more closely related to 'ability' from the Fogg Behaviour Model (FBM, see 3.1), rather than to a locus-of-control type of evaluation. BI is predicted by 'Attitude Towards Behaviour', 'Subjective Norms' regarding the behaviour and PBC. The attitude towards the behaviour represents how the behaviour is appraised not only in terms of the act, but also in relation to the possible outcomes of displaying the behaviour, such as potential rewards, or the averting of negative consequences. 'Subjective norms' refers to how displaying the behaviour is evaluated by the social network around the individual, and how displaying the behaviour might affect social relationships.

3.3   Influencing Behaviour at the Tactical Level

We searched the literature for persuasive methods that were used or have the potential to be used in the traffic domain. The search engines used were Google Scholar, Scopus and Web of Science, with the keywords: "driver persuasion AND system OR ivis OR adas", "persuasi* AND traffic OR in-car", "persuasive systems OR persuasive technology", "persuasive methods". We limited the results to experimental papers of 2010 and newer. For methodological papers proposing persuasive methods, no time frame was used. Forward and backward snowballing was performed. This resulted in the persuasive categories of Gamification, Behavioural Economics and Captology. These different methods often overlap to some degree

in the persuasive elements used. In this section, we discuss these persuasive methods and motivate our choice for the models we adopt for developing the conceptual model.

3.3.1   Persuasive Methods

The persuasive methods we reviewed can broadly be divided into Gamification, Behavioural Economics and Captology, although these fields show some overlap in the persuasive elements used or approaches taken.

Gamification is a term that has emerged relatively recently. Video games create an environment in which the player is highly motivated to perform certain behaviours to achieve game-related goals (finishing a level, getting a high score). Gamification takes the elements that elicit this motivational behaviour and applies them to other situations (Deterding, Dixon, Khaled, & Nacke, 2011). The most often and successfully applied game design elements are leader boards, achievements and challenges (Hamari, Koivisto, & Sarsa, 2014). Gamification may work through raising the driver's implicit motivation, by inducing group-effects such as in-group/out-group bias – simply assigning people to a group, induces positive feelings to other group members (Baron & Dunham, 2015) and a motivation to help achieve group goals (Musicant, Lotan, & Grimberg, 2015) –, as well as through a 'fear of missing out' effect (Przybylski, Murayama, DeHaan, & Gladwell, 2013). For example, Musicant et al. (2015) found that, when offering financial incentives and inducing a common group goal of collecting as many safe driving miles as possible, motivation to use a driving safety app on a smartphone was high over a period of more than a hundred days, as indicated by app usage and the active recruitment of friends as users. App usage dropped significantly once the group goal was achieved, indicating that any persuasive system should be cautious with formulating group goals and financial incentives. A quite extensive review of previous studies found that generally the effects of gamification are positive, although this is moderated by the context in which gamification is used as well as the users that are targeted (Hamari, Koivisto, & Sarsa, 2014). Gamification effectiveness might also be reduced over time due to a novelty-like-effect (Farzan et al., 2008a), although motivation can remain high when using group-based goals as long as these goals remain active (Musicant et al., 2015). Examples of gamification applied to the transportation domain include EcoChallenge (Ecker et al., 2011): a reward and competition-based system to persuade drivers to engage in a more eco-friendly behaviour, I-GEAR (McCall & Koenig, 2012): a system to change driver behaviour by providing small financial and non-financial rewards, and 'Driving Miss Daisy' (Shi et al., 2012): a gamified solution to help drivers improve their driving skills by providing a virtual passenger that occasionally comments on driving styles.

Behavioural economics has been defined as the 'body of work seeking to understand behaviour by incorporating insights from behavioural sciences into economics' (Avineri et al., 2010). Rather than being rational thinkers, people use a range of heuristics and display biases that often work well, but can lead to reasoning errors in certain situations (Kahneman, 2003). An overview can be found for instance in the work of Kahneman (Kahneman, 2013) or Cialdini (Cialdini, 2006). Persuasive elements from Behavioural Economics applied to the transportation domain can be found in for example the design of travel information systems (Avineri, 2011), approaches to promoting safe driving behaviours (Millar & Millar, 2000), and methods analysing travel behaviour (Metcalfe & Dolan, 2012).

Captology (acronym: computers as persuasive technology) was introduced by Fogg (1998). It is a field of study which uses computers to influence behaviour in various ways (Fogg, 2010).

The Fogg Behavioural Model (FBM) (Fogg, 2009a) is prominent in the field of persuasion. It postulates that in order for a persuasive intervention to be successful, three factors need to converge: the person needs to be able to perform the behaviour ('ability'), be motivated to perform the behaviour ('motivation'), and finally a trigger should be present to elicit the behaviour. Targeting simple behaviours has a higher likelihood of success (Fogg, 2009b). In the context of driver persuasion: making sure 'ability' is high means requesting short, simple to perform behaviours such as a speed change, an overtaking manoeuvre, a lane change, or a merging manoeuvre, as well as timing persuasive attempts to moments when driver workload is not high and when traffic conditions allow for the requested behaviour (e.g. don't request a lane change when the neighbouring lane is crowded). 'Motivation' can be raised by using persuasive techniques (see also 3.2). The FBM has been applied to the traffic setting, for instance it has been applied in a persuasive intervention that successfully reduced texting behaviour while driving (Miranda et al., 2013).

3.3.2   Integrating Persuasive Methods

The Persuasive Systems Design model (PSD) (Oinas-Kukkonen & Harjumaa, 2008) presents a systematic framework for designing and evaluating persuasive systems. It brings concepts from Gamification, Behavioural Economics and Captology together. The PSD states that a system can be made persuasive by providing the user with support in distinct categories: primary task support, dialogue support, system credibility support and social support.

Primary task support shows many of the principles put forth by the FBM and Behavioural Economics. The focus is on supporting the user by making the behavioural tasks more manageable, personal and transparent. Making the tasks more manageable by reducing complex behaviour to a series of steps and then leading the user through them is especially important when considering in-vehicle systems. Apart from increasing the system's persuasive power, this approach reduces task demands placed on the driver, which in turn increases system safety (Fuller, 2005; Wickens, 2002). An example of primary task support can be a lane change system that guides the driver through the steps of finding a gap, matching speed and merging. There is a growing similarity between primary task support and ADAS, such as lane-change assist systems (Habenicht, Winner, Bone, Sasse, & Korzenietz, 2011), as ADAS become more capable. In primary task support, one way of increasing persuasiveness is reducing complex behaviour to a series of steps and guiding the user through them, which is similar to what for example lane-change assistance systems do (Habenicht et al., 2011).

Dialogue support is aimed at keeping users moving towards their goals. This support level contains elements from Gamification, Behavioural Economics and the FBM. Offering praise and rewards can increase motivation, which is an important factor for persuasion in the FBM (Fogg, 2009a). If applicable, providing reminders for target behaviour or suggesting certain behavioural responses may be a way to increase behavioural effects by facilitating the creation of habits. Habits are a main factor in making persuasive effects last over time (Lally & Gardner, 2013). Further important factors in dialogue support are similarity and liking (Fogg, 2010), which can increase trust and intentions to comply to system requests.

System credibility support is mainly important from the perspective of trust and acceptance. It is about showing the driver that the system makes correct decisions and recommendations. Trust and acceptance are major factors in whether a persuasive system's suggestions or advices will be considered by the driver (Malte Risto & Martens, 2013; Vlassenroot et al., 2010). Factors at this support level relate to the accuracy of the information presented, its transparency, and how

users will evaluate it. This in turn is important for forming and maintaining trust in the system (Lee & Moray, 1992; Martens & Jenssen, 2012). The need for trust in a persuasive system is underscored by the work of Risto (Malte Risto & Martens, 2013), who reported that, in their study, drivers constantly tried to verify the accuracy of system requests before following them, and refused to follow messages they interpreted as incorrect. Apart from validity of the advices, acceptance can also be influenced by what modality is used (Donmez, Boyle, Lee, & Mcgehee, 2006).

Social support aims at persuading users by increasing motivation using social factors. This level has parallels with Gamification. It includes factors to incentivise behavioural change by allowing performance comparison with other users, facilitating cooperation and/or competition, creating transparency in behaviour-result relationships of other users and even applying forms of normative social pressure (see for example Lütteken, Zimmermann, & Bengler, 2016). Social factors vary in importance and effects on different age groups (McEachan, Conner, Taylor, & Lawton, 2011), which is important for instance when targeting specific demographic groups.

To summarise, Gamification has been shown to be effective in motivating people to change their behaviour. However, some studies report that its effectiveness might reduce over time (Farzan et al., 2008a, 2008b). Behavioural Economics as a field has many applicable concepts that can persuade drivers effectively, and the FBM presents a view of how driver motivation and ability need to converge in the presence of a trigger for persuasive influence to be effective. The PSD model unifies these persuasive methods using the described four support groupings. These provide persuasive elements that can be used depending on the type of system and the context in which it is intended to be applied. For example, in a cooperative system, which is social by nature, the 'social support level' provides ways to add persuasive elements to the social aspects present in the system (see Lütteken et al., 2016). More generally: system credibility can assist persuasion in most systems by increasing trust in the validity of the messages over time, which has been shown to be a large factor in whether a driver responds to the advice or not (Abe & Richardson, 2006; Malte Risto & Martens, 2013), or even a factor in determining system usage over time (Martens & Jenssen, 2012).

3.4   Considerations for Safe Driver Persuasion

The driving task is complex, requires constant attention from the driver (de Waard, 1996) and presents frequent distractions. Stutts and Gish (2003) report that drivers engaged in distracting activities for 16.10% of the time the car was moving (31.42% if in-car conversations were included). Poorly designed or implemented persuasive in-vehicle systems may increase this percentage by providing more distractions to a driver (Hibberd, Jamson, & Carsten, 2010), potentially increasing driver workload (Horberry, Anderson, Regan, Triggs, & Brown, 2006), inducing behavioural adaptation (Martens & Jenssen, 2012), or otherwise creating unsafe situations. Safety, therefore, is an important characteristic of a persuasive in-vehicle system. An effective but unsafe system is not likely to be used long term, either through consumer choice or through changing legislation. In this section, we discuss how improving safety can also increase persuasive effectiveness in the short and long run.

3.4.1   Safety, Driver Demand, and Unsafe Situations

A persuasive system needs to communicate with the driver. At the very least this means transmitting information to the driver, and in more complex cases it may require interaction. One way of limiting negative effects of this communication on driving performance, based on

the TCI (Fuller, 2005), is by ensuring that the demands placed on the driver do not create dangerous high workload situations. Although this is a broad statement, it can be assessed using for example environmental variables that may affect the driver, such as the proximity of other vehicles, traffic conditions and weather conditions, and driver variables such as driving demand and driver workload as well.

Driver workload results from the interplay between the demands placed on the driver by the driving task, the complexity of the environment, and the driver's capacity to meet those demands (de Waard, 1996). It is an important factor in terms of safety, since under- or overload can influence a driver's performance and create hazardous situations (Mark S. Young et al., 2015). In section 1.2 we have discussed how targeting the tactical level for persuasive attempts will likely limit the impact on driver demand (compared to targeting the strategic level), and by extension, on driver workload. Despite this, a poorly designed persuasive system targeting tactical-level behaviours can still result in high driving demand and/or workload. The Multiple Resource Theory (MRT) by Wickens (Wickens, 2002) can help understand why, even when a persuasive in-vehicle system targets simple-to-change behavioural tasks, high driver demand or workload may still result.

In the MRT, interference from a secondary task is most likely when it accesses the same resources as the primary task. Since driving is mainly a visual task, transmitting information to the driver through a visual channel may cause interference. For instance, diverting the eyes from the road for extended time has serious consequences for driving performance and lane-keeping ability (Peng, Boyle, & Hallmark, 2013). Heads-Up-Displays do not require the driver to take his eyes off the road and can be a better alternative (Liu & Wen, 2004), but do not mitigate all negative effects, and can introduce some new potential problems related to sharing visual resources (Wickens, 2002) and to characteristics of the human visual system, such as involuntary accommodation responses from the eye that cause the driver to temporarily lose optical focus of the road scene, even though both the HUD and the road scene are in the same field of view (Edgar, 2007). Competing resource types are not the only factor in the MRT that can lead to reduced task performance: if the demands of one or both tasks are higher than what the driver can handle, two tasks that use very different resources are still likely to cause dual-task interference and degrade driving performance. In terms of a persuasive in-car system, minimising the effect on workload therefore means choosing the correct modality to transmit information to the driver, keeping the cognitive demands of the interaction low to prevent interference with the main driving task, and timing the messages to periods when the driver can accommodate them. If the cognitive demands of the main task (driving) are already high, per the MRT a simple secondary task may create dual-task interference even when using a different modality from the main task, degrading the performance on the main task and thereby potentially compromising driver safety. Adaptive interfaces (Birrel et al., 2017; Park & Kim, 2015) try to counter this by changing either the complexity of messages presented, the modality used to convey the message to the driver, or by suppressing messages in conditions where safety or workload may be dangerously affected, safety can be improved.

Unsafe situations can still arise from persuasive in-vehicle systems even when changes induced in driver demand and workload are minimal. A system that distracts the driver at the wrong moment may create a potentially dangerous situation (K. Young & Regan, 2007), highlighting the importance of timing the communication with the driver. Unsafe situations may also arise from the way drivers accommodate the functions of in-vehicle devices into their driving habits, giving rise to behavioural adaptation effects (Martens & Jenssen, 2012; Smiley, 2000). As an example of an unintended behavioural effect, in response to having Anti-Lock Braking (ABS)

and Airbag systems installed, headways decreased and seatbelt usage reduced (Sagberg et al., 1997). Overreliance on a system is another potential problem, and research has shown that the degree of reliance by human operators doesn't always match the system capabilities (Parasuraman & Riley, 1997). For example, with a lane-change advice system: if a driver places too much trust in the lane change advice system, a lane change may be initiated when the system gives an advice, without the driver checking whether it is actually safe to change lane.

3.4.2 Persuasive Attempts and Acceptance

A persuasive in-vehicle system needs to be able to consistently persuade the driver. According to the Fogg Behaviour Model (FBM) (Fogg, 2009a), persuasive interventions timed to periods when both motivation and ability are high, have a higher chance of resulting in changed behavioural outcomes. In terms of an in-vehicle system, an advice that is given to a driver when there is a high motivation to follow it, will have a higher probability to be complied to. Similarly, an advice given at a time when the driver ability is high, i.e. when the driver perceives they can follow the advice, will be more likely to result in the target behaviour. This again underscores the importance of targeting behaviours that require less effort to change, such as tactical level driver behaviour: not only it is safer, persuasive effectiveness is also likely to increase when doing so (Fogg, 2009a). In the traffic context, the FBM's 'ability' to follow a persuasive advice can be impacted by multiple factors and conditions, such as weather conditions, traffic conditions, secondary tasks or driver states (de Waard, Kruizinga, & Brookhuis, 2008). One such driver state is driver workload, which needs to be considered for the effectiveness of persuasion as well as for safety. When driver workload is high, presenting an advice and/or requesting an action from the driver may increase the difficulty of the driving task further, in turn reducing the likelihood that the driver complies to the persuasive request because the requested behaviour is seen as difficult or impossible given the circumstances. In other words, high workload is likely counterproductive when trying to persuade the driver.

In addition to persuading a driver effectively, a persuasive system needs to be, and keep on being, used. To a large degree, this usage will depend on the acceptance of a system (Vlassenroot et al., 2010). Without taking steps to ensure acceptance, there is the risk that a persuasive in-vehicle system falls into disuse or works counterproductively (Martens & Jenssen, 2012). This is especially damaging if the system relies on a user base to function, as for example with cooperative (lane change) systems (Lütteken et al., 2016). To describe the acceptance of new technology several models have been developed, such as the Unified Theory of Acceptance and Use of Technology (UTAUT) (Venkatesh et al., 2003) and the Technology Acceptance Model (TAM) (F. D. Davis, 1986).

3.4.3 Technical Feasibility and In-Car Persuasion

As discussed, persuading the driver safely assumes that an in-car system has awareness of the driver ability and the driving context, so that messages can be transmitted at the right time (i.e. the driver has available capacity). Various technological building blocks exist that facilitate this, as briefly outlined in the following paragraphs.

Driver workload is mentioned as an important factor both for safety and persuasive effectiveness. Using various approaches, on-line driver workload predictors have been proposed and tested based on physiological characteristics and driver performance measures (Kim, Chun, & Dey, 2015; Solovey, Zec, Garcia Perez, Reimer, & Mehler, 2014; van Gent, Melman, et al., 2018b). These predictors often take at least heart rate into account. We recently

developed a toolkit that allows for online analysis of (noisy) heart rate data collected in in-car settings (van Gent, Farah, van Nes, & van Arem, 2018). This allows the reliable collection of this type of input for the workload predictors.

Capturing driver ability can be done through surrogate measures, for example by combining the workload prediction with traffic conditions. Camera-based systems exist that can detect and label other road users accurately (Ashraf, Wu, Iandola, Moskewicz, & Keutzer, 2016). This opens up the possibility to automatically consider aspects like traffic density and position of nearby vehicles into account. In this case, advising a lane change when driver workload is predicted to not be high and when a sufficiently large gap is detected on an adjacent lane, provides a way to select safe situations where 'driver ability' is estimated to be high as well.

The motivation to follow an advice is difficult to capture. The role of the persuasive system is to raise the motivation of the driver to follow an advice, by using one or several of the discussed persuasive techniques. It is conceivable that the system monitors the results of different persuasive attempts made and optimises the methods used to each driver individually based on performance statistics, but more research would be required to determine the optimal performance statistics.

3.4.4   Persuasion in Time

Any persuasive attempt will need time to be successful: the persuasive message needs to be generated, transmitted to the driver, interpreted by the driver, and finally followed if the driver decides to. Whether messages are time critical or not depends on the implementation. For example, a persuasive eco-driving application as described in (Ecker et al., 2011; McCall & Koenig, 2012) is not time critical. However, in the context of the lane-specific advice system described in this paper, a correct advice depends on current traffic conditions, and therefore is time critical. Traffic conditions are dynamic, meaning that if persuasion takes too long in this case, the advice might be obsolete. Advices incongruent with the surroundings are not only a problem for the functioning of the system, but also harm drivers' trust in the system (Malte Risto & Martens, 2013). In our research we aim to generate advices with a time validity of approximately two minutes. Within these two minutes, the lane change system will determine, based on the current and predicted traffic distribution, the optimal distribution to work towards. The main challenge from a traffic modelling point of view is to predict the risk of unstable traffic congestions far enough ahead (i.e. 2 minutes) to allow the driver enough time to follow any advices.

3.5   The Conceptual Model for Driver Persuasion at the Tactical Level

In this section, we present the proposed conceptual model for driver persuasion at the tactical level using in-vehicle systems. The model is meant to help guide the development of persuasive in-car systems by integrating persuasive methods as well as behavioural models. It has three levels: The System Level, the Information Transfer Level and the Driver Level. The System Level is where the persuasive strategy is formed and safety checks are performed. It incorporates the defined safety criteria (4.1, 4.2) and the four support levels from the Persuasive Systems Design model discussed earlier (3.2). The Information Transfer Level is where communication with the driver takes place. It incorporates elements from Wickens' MRT and Fuller's TCI Model. The Driver Level describes the behavioural effects of the persuasive attempt. It incorporates the TPB (2.2), along with considerations regarding effects on driver workload, indirect behavioural effects and driver safety (4.1, 4.2). Design of factors at the

system and information transfer level should take human factors described in the driver level into account, and reflect the desired outcomes of the system-driver interaction (safety, persuasive effectiveness). The following sub-sections detail these levels and how they are built up from the existing models and theories in the literature.

In the conceptual model three types of relationships are indicated. Solid lines indicated relationships that have been empirically validated and are known from meta-analyses. We have added the reported correlation coefficients and R2 statistics of these relationships to the model. The two types of dashed lines indicate relationships that are established in the literature, and hypothesised relationships. The basis for the hypotheses relationships is discussed in the corresponding sections (5.1-5.3).
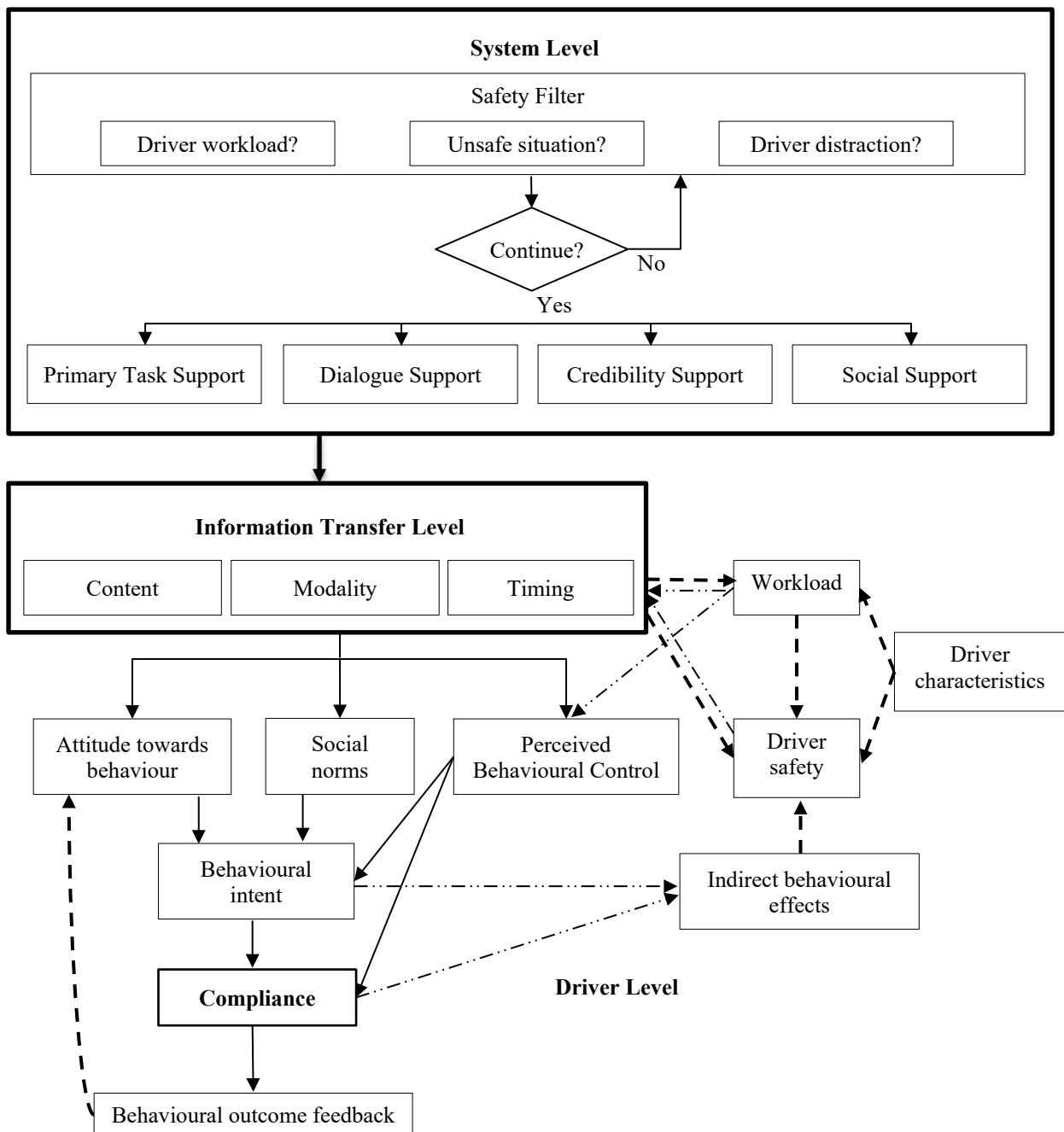
Figure 3.2: Proposed conceptual model for influencing tactical driver behaviour. Solid arrows indicate relationships known from the literature, dashed arrows indicate hypothesised relationships. Statistical properties are given where available from meta analyses in the format (correlation coefficient, R²).

3.5.1   Planning Driver Persuasion: The System Level

The System Level represents the back-end of the persuasive in-vehicle system. It is built up from the PSD model (3.2) and the considerations of driver safety and the persuasiveness (4).

Safety is central to the persuasive system design and operation. This is explicitly reflected in the model, where the first evaluation made is whether it is safe to initiate an information transfer to the driver. An existing type of driver monitoring system could perform this role effectively (Aghaei et al., 2016; van Gent et al., in press). Ideally, the persuasive system should (either directly or indirectly) take driver workload into account, should not create unsafe traffic situations, and should aim not to distract the driver at the wrong time. For example, a lane-change system designed to assist the driver in dense traffic, needs to take into account not only the surrounding traffic but also the driver state, when deciding on whether to continue or abort a lane-change manoeuvre (Habenicht et al., 2011). In situations where safety criteria are not met, they must be re-evaluated until they are met, represented in the model by the conditional loop. Ways to automatically evaluate these safety criteria exist, such as in systems that monitor on-coming traffic (Curry et al., 2010), label nearby road users (Ashraf et al., 2016), detect weather conditions (Green, 2004), and systems that attempt to estimate driver state (Ferreira et al., 2014; Liang, Reyes, & Lee, 2007; van Gent, Melman, et al., 2018b).

Once it is determined that interacting with the driver does not pose a safety risk, tactical driver advice may be given to persuade the driver. The PSD described in this paper combines persuasive techniques into four support levels. These four levels of support are included as possible routes to persuasion (see also Oinas-kukkonen & Harjumaa, 2009; Oinas-Kukkonen & Harjumaa, 2008, 3.2). A recent meta-analysis study by Hamari et al. (2014) report that from 95 empirical studies looking at persuasive techniques applied to diverse fields, the majority report positive (52 studies) or partially positive (36 studies) results. Many of the included papers utilise the PSD framework. This indicates the viability of using persuasive methods to achieve behavioural change effectively.

3.5.2   Interacting with the Driver: The Information Transfer Level

The information transfer level comprises the communication between the persuasive system and the driver. Usually this communication takes place through a type of interface (visual, auditory, tactile or multimodal). The information transfer level and its effects on behaviour (driver level, 5.3) are built up from the TPB, MRT, TCI and FBM discussed in the previous sections. The information transfer itself is operationalised as having 'content' ('what' is in the message?), 'modality' ('how' is the message transmitted to the driver?) and 'timing' ('when' is the message transmitted?) as factors. The modality used to convey the message could be dependent on the type of information being transmitted (Donmez, Boyle, Lee, & City, 2006), and can influence the acceptance of the advices as well (Donmez, Boyle, Lee, & Mcgehee, 2006). In the conceptual model, the information transfer influences driver workload, driver safety and the behavioural determinants of the TPB (attitude, social norms and perceived behavioural control). This impact on the behavioural determinants is the goal of the conceptual model: in order to affect behavioural change, the system needs to influence these motivations (Elliott et al., 2013). Here we discuss these effects in terms of the impact on safety and the impact on persuasive potential.

From a safety perspective, the model shows an effect of the information transfer on 'workload' and 'perceived behavioural control' based on the TPB and MRT. According to the MRT, dual-task interference is likely when two concurrent tasks use the same modality, or when the

cognitive load from one or both tasks is high. Dual-task interference reduces performance on the main (driving) task and increases demands placed on the driver, which in turn can raise workload. As demands and workload rise, we hypothesised that the perceived behavioural control of the driver lowers. As we discussed in section 2.2 PBC reflects self-efficacy (judgement of being able to perform the target behaviour), not perceived level of control over the behaviour (Elliott et al., 2013). This means that the higher the (perceived) driver workload, the lower the driver's appraisal of being able to comply with persuasive messages will be. This appraisal of ability is important in the persuasive context: if a driver lacks the confidence to follow up on a persuasive advice, the persuasive attempt will likely not succeed. In addition, this could lead to a degradation of driver performance, or even undesirable situations such as a loss of control or a collision (TCI, 4.1, 5.3). A direct link to driver safety is also included, which includes for example situations where the information transfer leads to eyes-off-road situations (Dozza, 2013; Peng et al., 2013) or to distraction at a critical moment.

From the persuasion perspective, the FBM (Fogg, 2009a) specifies that motivation and ability need to be high at the moment of a behavioural trigger, in order for persuasion to have a high chance of being successful. The goal of the persuasive techniques used ('content') is to raise motivation to perform a behaviour, for instance by using social support to increase motivation to comply to a message. Making sure 'ability' is high, essentially means timing the information transfer to situations where the driver's PBC is high (Elliott et al., 2013, see also section 5.3). In a driving setting, the PBC term implicitly includes an environmental component (e.g. give a lane change request only when there is sufficient room on the adjacent lane), and a driver component (a high workload will result in lower PBC). Both components are important for persuasion and safety. For example, if a lane-specific advice system requests a lane-change when a driver does not feel capable of performing the requested manoeuvre, it is unlikely the persuasion will have an effect. Alternatively, if an already overloaded driver complies with the requested behaviour, unsafe or outright dangerous situations can result.

3.5.3   Human Factors: The Driver Level

The driver level provides a basis to describe expected behavioural effects of the persuasion. In this section, we describe how the TPB fits into the model, how workload relates to both safety and persuasion, its dependence on driver characteristics and factors on the information transfer level, possible behavioural effects and the importance of outcome feedback.

As argued in the previous section, both motivation and ability need to be high in order for persuasive systems to actually persuade (Fogg, 2009a). In the conceptual model, motivation is captured by the TPB terms 'attitude towards behaviour' and 'social norms'. The attitude and social norms influence driver behaviour through the 'behavioural intent' (BI) (Ajzen, 1991; Armitage & Conner, 2001; McEachan et al., 2011). The ability to follow persuasive advices is captured through 'perceived behavioural control' (PBC) and its interaction with workload. PBC affects the intent to perform a behaviour as well as the behaviour directly (Armitage & Conner, 2001; McEachan et al., 2011), and additionally we hypothesize that it acts as a modulator of workload on behaviour. As discussed previously, this hypothesis is based on earlier work showing that PBC relates to the perceived ability a person has to perform a given behaviour, rather than a locus of control-like evaluation of whether the behaviour lies within the control of the individual (Elliott et al., 2013). This means that with a high PBC the driver feels competent and able to perform a requested behaviour, whereas a low PBC will negatively influence the likelihood of a persuasion from resulting in the desired behaviour (Armitage & Conner, 2001; McEachan et al., 2011).

Apart from the information transfer (5.2), driver workload is also affected by 'driver characteristics'. This component is a broad term meant to capture the heterogeneity of the drivers and how this relates especially to driver workload and driver safety. For example, driver ability is not static and varies between and within individuals over time (Mark S. Young et al., 2015), which may cause workload experienced by two different drivers in a comparable situation to be very different. 'Driver characteristics' also includes differences in inherent driver safety. For example, some age groups display more risky behaviour (Carter, Bingham, Zakrajsek, Shope, & Sayer, 2014), there may be sex differences or geographical differences in driver behaviour and capability (Twisk & Stacey, 2007; Vlakveld, 2011), or individual differences in driver aggression (Hennessy & Wiesenthal, 2001). These characteristics may result in some classes of drivers being exposed to higher risk while driving, especially in combination with in-car systems.

'Indirect behavioural effects' (Martens & Jenssen, 2012) were discussed in 4.2. These refer to changes in driver behaviour or intentions to perform behaviours that are not intended by the designers of the (persuasive) system. An often-cited example of indirect behavioural effects is that of the anti-lock braking system (ABS), which helps reduce stopping distances of the cars in which it is installed. Positive effects were offset by behavioural effects: adaptation was reported from drivers choosing to driver faster on wet surfaces (Smiley, 2000) or with shorter headway and varying seatbelt usage (Sagberg et al., 1997). When developing and implementing a persuasive in-car system it is imperative to include these possible indirect behavioural effects in experiments to evaluate it.

The last undiscussed term in the model is feedback about behavioural outcomes. This feedback, including information on the behaviour-result relationships in other drivers, is expected to influence the driver's attitude towards future behaviours in a feedback loop (see also Lütteken et al., 2016). For instance, if a driver observes that complying to an in-vehicle system has resulted in shorter travel times on previous occasions or with other drivers, this might bias the driver to comply more with the system's advices in the future. This ties into the "system credibility support" level of the PSD (Oinas-Kukkonen & Harjumaa, 2008). It is also in line with an earlier study into compliance to tactical driving advice (Malte Risto & Martens, 2013), where drivers were observed attempting to evaluate the validity of tactical advice in the context of what they observed on the road and the history of the system's accuracy.

3.6   Application to a Lane-Specific Advice System

In this last section, we present a case study based on a lane-specific advice system, in which we apply the developed model to the system and discuss how this helps structure system design for safety and persuasion.

The goal of the system is to reduce travel time delay and congestion by encouraging a better distribution of the vehicles over the available motorway lanes. This means advising drivers on what lane to take, depending on external factors. For instance, an unbalanced distribution, an upcoming on-ramp or lane drop, or an incident upstream may require a redistribution of traffic to ensure continued flow and avoid congestion. The system's advices will be in the collective benefit of drivers on a specific stretch of road (minimised total travel time), but will sometimes not be in the benefit of individual drivers receiving the advice (e.g. stay behind this slow truck for now), creating a potential problem (Malte Risto & Martens, 2012). The challenge is to persuade drivers to follow the advices that are in the benefit of the collective rather than the individual. We aim to apply the persuasive techniques to engage drivers with the system and to

also stimulate adherence to lane-specific advices, especially when they are not in the individual's benefit. By applying the various persuasive techniques described in the paper, driver motivation and the attractiveness of the advices are hypothesised to increase. We will verify this experimentally. The designed system will consist of an in-vehicle part and a back-end that predicts traffic states and approximates the optimal lane use situation.

The developed conceptual model described in this paper helped to direct our research in several ways. At the 'System Level', a safety filter is required. Early in the design phase, this redirected the process from focusing mostly on the effectiveness of the persuasive design, to an approach that considered potential effects on safety and on the driver as well. As a result, we are developing an affordable driver monitoring system to estimate driver state (Gent, Farah, Nes, & Arem, 2017; van Gent, Melman, Farah, van Nes, & van Arem, 2018a). In combination with environmental sensing systems built into the vehicle, this provides a safety filter that will suppress messages to drivers that are estimated not to respond (safely) to the persuasion. The result of this message filtering, we argue, is two-fold (see 4.2, 5.2, 5.3): apart from increasing the safety of the system, it works to increase persuasive effectiveness and facilitate long-term usage of the system as well.

Persuasive strategies are outlined in the four support levels from the PSD model (Oinas-Kukkonen & Harjumaa, 2008, see also 4.2). These support levels offer persuasive strategy elements from which a selection can be made. We selected strategies mainly from primary task support and dialogue support, with some elements from the other two support levels. The system will support the driver by breaking down a requested lane-change into smaller steps, and guiding the driver through them (primary task support: 'reduction' and 'tunnelling'). This will increase persuasive power and make the task less demanding, benefitting both safety and persuasion (Fuller, 2005; Wickens, 2002, see also 4.1). Second, the system will provide the user with transparent information regarding obtained benefits in terms of travel time saved in relation to the performed behaviour through either an app or a web-portal (primary task support: 'self-monitoring'). Providing a means of 'self-monitoring' of on-going benefits increases immediate persuasive effects, but also works to increase 'trustworthiness' and 'verifiability' of the system (credibility support). As discussed in 3.1 the effectiveness of persuasive methods might decrease over time. In one study, especially the presence of clear (group) goals was found to keep system usage high (Musicant et al., 2015). In the case of our lane-specific advice system, the group goal is to reduce congestion on the road that the user is driving on, which is a relevant goal along the whole drive. Whether the use of group-based incentives can be implemented will be evaluated at a later stage of the system design.

At the information transfer level, an advice is communicated to the driver, the effects of which are described at the driver level (5.3). As described in 5.2, in the model the information transfer between system and driver is operationalised as having content, modality and timing. The model shows how these factors mediate safety and persuasive effectiveness through workload and perceived behavioural control (see also 4.2, 5.2). This means that, in further development of our lane-specific advice system, our research will focus on how driver workload and perceived behavioural control are influenced by content, modality and timing decisions with our lane-specific advice system. Additionally, it simplifies the scope of our research: in order to estimate the effects on the behavioural outcome, we only need to investigate how the three information transfer factors influence the 'attitude towards behaviour', the perceived 'social norms' and the PBC. How these three factors in turn influence BI and Behaviour is known from several exhaustive meta analyses (Armitage & Conner, 2001; McEachan et al., 2011; Notani, 1998). To assist in estimating how our persuasive system influences these factors, it is useful

to point out that guidelines have been formulated on how to operationalise these constructs (Ajzen, 2010; French & Hankins, 2003).

In this section, we have applied the model to the design of our persuasive lane-specific advice system, and have discussed how this helped shift the focus of our research away from one emphasizing persuasion, to one that includes the driver's behaviour and traffic safety as well. We have shown how this shift will benefit not just traffic safety but the persuasive effectiveness of the system as well.

3.7   Conclusion

In this paper, we have proposed a conceptual model to help guide the design of persuasive in-vehicle systems with the aim of influencing driver behaviour at the tactical level. The model was designed with safety and persuasion as core elements, and explains how a persuasive in-vehicle system is expected to affect driver behaviour, workload, and safety. The model contains four 'support levels' from the PSD from Oinas-Kukkonen (2009), that can be used as guidelines for implementing specific persuasive elements in persuasive in-vehicle systems. Similarities exist between ADAS and for example the primary task support level from the PSD, and similarities will likely increase as ADAS become more complex. This provides an interesting possibility for the integration of persuasive driver methods using existing systems.

The proposed model is split into three levels explaining the different elements of the information chain: the system level where the persuasive strategy is formed after a safety check, the information transfer level where communication with the driver takes place, and the driver level where the act of presenting advice impacts driver behaviour, workload and safety in several ways. The focus while designing the model was on safely attaining effective driver persuasion. As a behavioural basis, the Theory of Planned Behaviour was selected. The persuasive elements come from the PSD model. We have discussed how the PSD is built from elements in Gamification, Behavioural Economics and Captology. We have also included elements from Wickens' MRT Model and Fuller's TCI that help explain why the timing and modality of the information transfer are key factors in both safety and persuasive effectiveness. Finally, we have applied the model to a persuasive system which aims to reduce travel time delay and congestion by encouraging a better distribution of the cars over the available motorway lanes, to illustrate how the application of the model guided our research efforts and helped shape a safe and effective design.

Future work will focus on evaluating the best set of persuasive techniques for driver persuasion, as well as the most promising delivery method ('content', 'modality' and 'timing') to ensure persuasive effectiveness as well as safety and low distraction caused by the advice.

Other opportunities for research still exist within the model apart from our planned future work. For example, the building blocks of the described 'safety filter' exist as discussed in the paper, but a unified application that takes the driver into account as well is still lacking.

Several relationships are unique to each specific implementation of an in-car persuasive system and can only be evaluated in that specific context. For example, the indirect effects on behaviour will be different for different systems, and as such will need to be determined every time a new persuasive system is developed and tested.

Persuading drivers is a complex task, especially since the driving environment requires extra considerations in terms of safety, and because the demands the environment places on drivers are highly dynamic. In the near future persuasion might become easier to accomplish once vehicle automation becomes more prevalent: as drivers get used to sharing the driving task with their vehicle, it is likely they develop a stronger sense of trust, which may favour complying with generated advices. The presented work and model in this paper aim to assist those working on driver persuasion by providing a theoretical framework within which persuasive systems can be developed.

References

Abe, G., & Richardson, J. (2006). Alarm timing, trust and driver expectation for forward collision warning systems. Applied Ergonomics, 37(5), 577–586. https://doi.org/10.1016/j.apergo.2005.11.001

Aghaei, A. S., Donmez, B., Liu, C. C., He, D., Liu, G., Plataniotis, K. N., … Sojoudi, Z. (2016). Smart Driver Monitoring: When Signal Processing Meets Human Factors: In the driver's seat. IEEE Signal Processing Magazine, 33(6), 35–48. https://doi.org/10.1109/MSP.2016.2602379

Ajzen, I. (1991). The theory of planned behavior. Orgnizational Behavior and Human Decision Processes, 50, 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Ajzen, I. (2010). Constructing a theory of planned behavior questionnaire. Biofeedback and Selfregulation, 17, 1–7. https://doi.org/10.1016/0749-5978(91)90020-T

Alexander, G., & Lunenfeld, H. (1986). Driver Expectancy in Highway Design and Traffic Operations. NHTSA Technical Report. https://doi.org/FHWA-TO-86-1

Armitage, C. J., & Conner, M. (2001). Efficacy of the Theory of Planned Behaviour : A meta-analytic review. British Journal of Social Psychology, (40), 471–499.

Ashraf, K., Wu, B., Iandola, F. N., Moskewicz, M. W., & Keutzer, K. (2016). Shallow Networks for High-Accuracy Road Object-Detection. https://doi.org/10.5220/0006214900330040

Avineri, E. (2011). Applying behavioural economics in the design of travel information systems. 43rd Universities Transport Study Group Conference, 1–12. https://doi.org/10.1017/CBO9781107415324.004

Avineri, E., Chatterjee, K., Darnton, A., Goodwin, P., Lyons, G., Musselwhite, C., … Wiltshire, P. (2010). Individual Behaviour Change : Evidence in transport and public health.

Bandura, A. (1971). Social learning theory. Social Learning Theory. https://doi.org/10.1111/j.1460-2466.1978.tb01621.x

Baron, A. S., & Dunham, Y. (2015). Representing 'Us' and 'Them': Building Blocks of Intergroup Cognition. Journal of Cognition and Development, 0(0), 1–22. https://doi.org/10.1080/15248372.2014.1000459

Birrel, S., Young, M., Staton, N., & Jennings, P. (2017). Using Adaptive Interfaces to Encourage Smart Driving and Their Effect on Driver Workload, 484, 764. https://doi.org/10.1007/978-3-319-41682-3

Brick, L. A. D., Velicer, W. F., Redding, C. A., Rossi, J. S., & Prochaska, J. O. (2016). Extending Theory-Based Quantitative Predictions to New Health Behaviors. International Journal of Behavioral Medicine, 23(2), 123–134. https://doi.org/10.1007/s12529-015-9506-y

Carter, P. M., Bingham, C. R., Zakrajsek, J. S., Shope, J. T., & Sayer, T. B. (2014). Social norms and risk perception: Predictors of distracted driving behavior among novice adolescent drivers. Journal of Adolescent Health, 54(5), S32–S41. https://doi.org/10.1016/j.jadohealth.2014.01.008

Castanier, C., Deroche, T., & Woodman, T. (2013). Theory of planned behaviour and road violations: The moderating influence of perceived behavioural control. Transportation Research Part F: Traffic Psychology and Behaviour, 18, 148–158. https://doi.org/10.1016/j.trf.2012.12.014

Chen, H. W., Donmez, B., Hoekstra-atwood, L., & Marulanda, S. (2016). Self-reported engagement in driver distraction : An application of the Theory of Planned Behaviour. Transportation Research Part F: Psychology and Behaviour, 38, 151–163. https://doi.org/10.1016/j.trf.2016.02.003

Chorlton, K., & Conner, M. (2012). Can enforced behaviour change attitudes: Exploring the influence of Intelligent Speed Adaptation. Accident Analysis & Prevention, 48, 49–56. https://doi.org/10.1016/j.aap.2010.06.007

Cialdini, R. B. (2006). Influence: The Psychology of Persuasion. London: Harper Collins Publishers Ltd. https://doi.org/10.1017/CBO9781107415324.004

Curry, R., Blommer, M., Tijerina, L., Greenberg, J., Kochhar, D., Simonds, C., & Watson, D. (2010). A mental model perspective of a driver workload system. Proceedings of the Human Factors and Ergonomics Society, 3, 2096–2100. https://doi.org/10.1518/107118110X12829370264682

Damiani, S., Deregibus, E., & Andreone, L. (2009). Driver-vehicle interfaces and interaction: Where are they going? European Transport Research Review, 1(2), 87–96. https://doi.org/10.1007/s12544-009-0009-2

Davis, F. D. (1986). A technology acceptance model for empirically testing new end-user information systems: Theory and results. https://doi.org/oclc/56932490

Davis, F. D. ., Bagozzi, R. P. ., Warshaw, P. R., Davis, F. D. ., Bagozzi, R. P. ., Bagozzi, R. P. ., … Warshaw, P. R. (1989). User Acceptance Of Computer Technology: A Comparison Of Two Theoretical Models. Management Science, 35(8), 982. https://doi.org/10.2307/2632151

de Waard, D. (1996). The Measurement of Drivers ' Mental Workload. Alphen aan den Rijn: Drukkerij Haasbeek. https://doi.org/10.1016/j.apergo.2003.11.009

de Waard, D., Kruizinga, A., & Brookhuis, K. A. (2008). The consequences of an increase in heavy goods vehicles for passenger car drivers' mental workload and behaviour: A simulator study. Accident Analysis and Prevention, 40(2), 818–828. https://doi.org/10.1016/j.aap.2007.09.029

Deci, E. L., & Ryan, R. M. (1985). Intrinsic Motivation and Self-determination in Human Behavior. New York: Plenum.

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From Game Design Elements to Gamefulness: Defining "Gamification." Proceedings of the 15th International Academic MindTrek Conference on Envisioning Future Media Environments - MindTrek '11, 9–11. https://doi.org/10.1145/2181037.2181040

Donmez, B., Boyle, L. N., Lee, J. D., & City, I. (2006). The Impact of Distraction Mitigation Strategies on Driving Performance, 48(4), 785–804.

Donmez, B., Boyle, L. N., Lee, J. D., & Mcgehee, D. V. (2006). Drivers ' attitudes toward imperfect distraction mitigation strategies, 9, 387–398. https://doi.org/10.1016/j.trf.2006.02.001

Dozza, M. (2013). What factors influence drivers' response time for evasive maneuvers in real traffic? Accident Analysis and Prevention, 58, 299–308. https://doi.org/10.1016/j.aap.2012.06.003

Ecker, R., Holzer, P., Broy, V., & Butz, A. (2011). Ecochallenge: a race for efficiency. Proceedings of the 13th International Conference on Multimodal Interfaces - ICMI '11, 13, 91–94. https://doi.org/10.1145/2037373.2037389

Edgar, G. K. (2007). Accommodation, cognition, and virtual image displays: A review of the literature. Displays, 28(2), 45–59. https://doi.org/10.1016/j.displa.2007.04.009

Efrat, K., & Shoham, A. (2013). The theory of planned behavior, materialism, and aggressive driving. Accident Analysis and Prevention, 59, 459–465. https://doi.org/10.1016/j.aap.2013.06.023

Elliott, M. a., Armitage, C. J., & Baughan, C. J. (2005). Exploring the beliefs underpinning drivers' intentions to comply with speed limits. Transportation Research Part F: Traffic Psychology and Behaviour, 8(6), 459–479. https://doi.org/10.1016/j.trf.2005.08.002

Elliott, M., Thomson, J., Robertson, K., Stephenson, C., & Wicks, J. (2013). Evidence that changes in social cognitions predict changes in self-reported driver behavior: Causal analyses of two-wave panel data. Accident Analysis and Prevention, 50, 905–916. https://doi.org/10.1016/j.aap.2012.07.017

Evans, L., & Michon, J. A. (1985). A critical view of driver behavior models: what do we know, what should we do? In L. Evans & J. A. Michon (Eds.), Human behavior and traffic safety (pp. 485–520). New York: Plenum Press. https://doi.org/10.1007/978-1-4613-2173-6

Farzan, R., Dimicco, J. M., Millen, D. R., Brownholtz, B., Geyer, W., & Dugan, C. (2008a). Results from Deploying a Participation Incentive Mechanism within the Enterpris. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'08), 563–572. https://doi.org/10.1145/1357054.1357145

Farzan, R., Dimicco, J. M., Millen, D. R., Brownholtz, B., Geyer, W., & Dugan, C. (2008b). When the experiment is over : Deploying an incentive system to all the users. Proceedings of the AISB 2008 Symposium on Persuasive Technology, 33–38. https://doi.org/10.1145/2145204.2145362

Ferreira, E., Ferreira, D., Kim, S., Siirtola, P., Röning, J., Forlizzi, J. F., & Dey, A. K. (2014). Assessing real-time cognitive load based on psycho- physiological measures for younger and older adults. In Computational Intelligence, Cognitive Algorithms, mind, and Brain (CCMB).

Fishbein, M., & Ajzen, I. (1975). Belief; attitude, intention and behavior: An introduction to theory and research. MA: Addison-Wesley.

Fluegge, K. (2016). Social learning theory and public perception of GMOs: What Blancke et al . (2015) and other plant biotechnologists are missing. Journal of the Science of Food and Agriculture, 96(9), 2939–2940. https://doi.org/10.1002/jsfa.7634

Fogg, B. J. (1998). Persuasive computers: Perspectives and research directions. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, (April), 225–232. https://doi.org/10.1145/274644.274677

Fogg, B. J. (2009a). A behavior model for persuasive design. Proceedings of the 4th International Conference on Persuasive Technology - Persuasive '09, 1. https://doi.org/10.1145/1541948.1541999

Fogg, B. J. (2009b). Creating Persuasive Technologies : An Eight-Step Design Process. Technology, 91, 1–6. https://doi.org/10.1145/1541948.1542005

Fogg, B. J. (2010). Persuasive Technology: Using computers to change what we think and do. Elsevier Science & Technology.

French, D. P., & Hankins, M. (2003). The expectancy-value muddle in the theory of planned behaviour - and some proposed solutions. British Journal of Health Psychology, 8(Pt 1), 37–55. https://doi.org/10.1348/135910703762879192

Friederichs, S. A. H., Bolman, C., Oenema, A., Verboon, P., & Lechner, L. (2016). Exploring the working mechanisms of a web-based physical activity intervention, based on self-determination theory and motivational interviewing. Internet Interventions, 3, 8–17. https://doi.org/10.1016/j.invent.2015.11.003

Fuller, R. (2005). Towards a general theory of driver behaviour. Accident; Analysis and Prevention, 37(3), 461–472. https://doi.org/10.1016/j.aap.2004.11.003

Gadong, E. S. A., & Chavez, M. R. (2016). Procrastination and sense of community: Patterns and relationships in a blended learning setting. Journal of Institutional Research South East Asia, 14(1), 5–16. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-84964916782&partnerID=tZOtx3y1

Gent, P. Van, Farah, H., Nes, N. Van, & Arem, B. Van. (2017). Towards Real-Time, Nonintrusive Estimation of Driver Workload: A Simulator Study. In Road Safety and Simulation 2017 Conference Proceedings.

Green, P. (2004). Driver distraction , telematics design , and workload managers : safety issues and solutions. Proc. Int. Congr. Transp. Electron, 165–180. https://doi.org/10.1167/3.9.157

Habenicht, S., Winner, H., Bone, S., Sasse, F., & Korzenietz, P. (2011). A maneuver-based lane change assistance system. IEEE Intelligent Vehicles Symposium, Proceedings, 375–380. https://doi.org/10.1109/IVS.2011.5940417

Hamari, J., Koivisto, J., & Pakkanen, T. (2014). Do Persuasive Technologies Persuade?-A Review of Empirical Studies. Persuasive Technology, 8462, 118–136. https://doi.org/10.1007/978-3-319-07127-5_11

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work? - A literature review of empirical studies on gamification. In Proceedings of the Annual Hawaii International Conference on System Sciences (pp. 3025–3034). https://doi.org/10.1109/HICSS.2014.377

Hennessy, D. A., & Wiesenthal, D. L. (2001). Gender, driver aggression, and driver violence: An applied evaluation. Sex Roles, 44(11–12), 661–676. https://doi.org/10.1023/A:1012246213617

Hibberd, D. L., Jamson, S. L., & Carsten, O. M. J. (2010). Managing in-vehicle distractions. Proceedings of the 2nd International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '10, 2, 4–11. https://doi.org/10.1145/1969773.1969775

Hoeben, E. M., & Weerman, F. M. (2016). Why Is Involvement in Unstructured Socializing Related To Adolescent Delinquency?*. Criminology, 54(2), 242–281. https://doi.org/10.1111/1745-9125.12105

Hoffman, B. D. (2014). Using Self-Determination Theory to Improve Residency Training: Learning to Make Omelets Without Breaking Eggs. Academic Medicine : Journal of the Association of American Medical Colleges, XXX(X), 1–3. https://doi.org/10.1097/ACM.0000000000000523

Horberry, T., Anderson, J., Regan, M. a., Triggs, T. J., & Brown, J. (2006). Driver distraction: The effects of concurrent in-vehicle tasks, road environment complexity and age on driving performance. Accident Analysis and Prevention, 38(1), 185–191. https://doi.org/10.1016/j.aap.2005.09.007

Kahneman, D. (2003). Maps of Bounded Rationality : Economicst Psychology for Behavioral. The American Economic Review, 93(5), 1449–1475. https://doi.org/10.1257/000282803322655392

Kahneman, D. (2013). Thinking Fast and Slow. Journal of Chemical Information and Modeling. Penguin Books Ltd. https://doi.org/10.1017/CBO9781107415324.004

Kim, S., Chun, J., & Dey, A. K. (2015). Sensors Know When to Interrupt You in the Car. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15, 487–496. https://doi.org/10.1145/2702123.2702409

Kushnir, V., Godinho, A., Hodgins, D. C., Hendershot, C. S., & Cunningham, J. A. (2015). Motivation to Quit or Reduce Gambling: Associations between Self-Determination Theory and the Transtheoretical Model of Change. Journal of Addictive Diseases, 0(ja), 00–00. https://doi.org/10.1080/10550887.2016.1107315

Lally, P., & Gardner, B. (2013). Promoting habit formation. Health Psychology Review, 7, S137–S158. https://doi.org/10.1080/17437199.2011.603640

Lee, J., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. Ergonomics, 35(10), 1243–1270. https://doi.org/10.1080/00140139208967392

Lekes, N., Houlfort, N., Milyavskaya, M., Hope, N. H., & Koestner, R. (2016). The role of intrinsic values for self-growth and community contribution at different life stages: Differentially predicting the vitality of university students and teachers over one year. Personality and Individual Differences, 98, 48–52. https://doi.org/10.1016/j.paid.2016.03.093

Liang, Y., Reyes, M. L., & Lee, J. D. (2007). Real-time detection of driver cognitive distraction using support vector machines. IEEE Transactions on Intelligent Transportation Systems, 8(2), 340–350. https://doi.org/10.1109/TITS.2007.895298

Liu, Y. C., & Wen, M. H. (2004). Comparison of head-up display (HUD) vs. head-down display (HDD): Driving performance of commercial vehicle operators in Taiwan. International Journal of Human Computer Studies, 61(5), 679–697. https://doi.org/10.1016/j.ijhcs.2004.06.002

Lochbuehler, K., Schuck, K., Otten, R., Ringlever, L., & Hiemstra, M. (2016). Parental Smoking and Smoking Cognitions among Youth: A Systematic Review of the Literature. European Addiction Research, 22(4), 215–232. https://doi.org/10.1159/000446022

Lütteken, N., Zimmermann, M., & Bengler, K. J. (2016). Using gamification to motivate human cooperation in a lane-change scenario. IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, 899–906. https://doi.org/10.1109/ITSC.2016.7795662

Magana, V. C., & Organero, M. M. (2011). The Impact of Using Gamification on the Eco-driving Learning. Advances in Intelligent and Soft Computing, 92, 141–148. https://doi.org/10.1007/978-3-642-19937-0

Martens, M. H., & Jenssen, G. D. (2012). Behavioral Adaptation and Acceptance. In A. Eskandarian (Ed.), Handbook of Intelligent Vehicles (pp. 117–138). London: Springer-Verlag London Ltd. https://doi.org/10.1007/978-0-85729-085-4

McCall, R., & Koenig, V. (2012). Gaming concepts and incentives to change driver behaviour. 2012 the 11th Annual Mediterranean Ad Hoc Networking Workshop, Med-Hoc-Net 2012, 146–151. https://doi.org/10.1109/MedHocNet.2012.6257115

McEachan, R. R. C., Conner, M., Taylor, N. J., & Lawton, R. J. (2011). Prospective prediction of health-related behaviours with the Theory of Planned Behaviour: a meta-analysis. Health Psychology Review, 5(2), 97–144. https://doi.org/10.1080/17437199.2010.521684

Metcalfe, R., & Dolan, P. (2012). Behavioural economics and its implications for transport. Journal of Transport Geography, 24, 503–511. https://doi.org/10.1016/j.jtrangeo.2012.01.019

Millar, M. G., & Millar, K. U. (2000). Promoting Safe Driving Behaviors: The Influence of Message Framing and Issue Involvement. Journal of Applied Social Psychology, 30(4), 853–866. https://doi.org/10.1111/j.1559-1816.2000.tb02827.x

Miranda, B., Jere, C., Alharbi, O., Lakshmi, S., Khouja, Y., & Chatterjee, S. (2013). Examining the efficacy of a persuasive technology package in reducing texting and driving behavior. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 7822 LNCS, 137–148. https://doi.org/10.1007/978-3-642-37157-8_17

Musicant, O., Lotan, T., & Grimberg, E. (2015). Potential of Group Incentive Schemes to Encourage Use of Driving Safety Apps. Transportation Research Record: Journal of the Transportation Research Board, 2516(2516), 1–7. https://doi.org/10.3141/2516-01

Niven, A. G., & Markland, D. (2015). Using self-determination theory to understand motivation for walking: Instrument development and model testing using Bayesian structural equation modeling. Psychology of Sport and Exercise, 23, 90–100. https://doi.org/10.1016/j.psychsport.2015.11.004

Norcross, J. C., Krebs, P. M., & Prochaska, J. O. (2011). Stages of change. Journal of Clinical Psychology, 67(2), 143–154. https://doi.org/10.1002/jclp.20758

Notani, A. (1998). Moderators of Perceived Behavioral Control's Predictiveness in the Theory of Planned Behavior: A Meta-Analysis. Journal of Consumer Psychology, 7(3), 247–271. https://doi.org/10.1207/s15327663jcp0703_02

Oinas-Kukkonen, H. (2013). A foundation for the study of behavior change support systems. Personal and Ubiquitous Computing, 17(6), 1223–1235. https://doi.org/10.1007/s00779-012-0591-5

Oinas-kukkonen, H., & Harjumaa, M. (2009). Persuasive Systems Design : Key Issues , Process Model , and System Features Persuasive Systems Design : Key Issues , Process Model , and System Features. Communications of the Association for Information Systems, 24(1).

Oinas-Kukkonen, H., & Harjumaa, M. (2008). A Systematic Framework for Designing and Evaluating Persuasive Systems. Persuasive Technology, Third International Conference, PERSUASIVE 2008, Oulu, Finland, June 4-6, 2008. Proceedings, 5033, 164–176. https://doi.org/10.1007/978-3-540-68504-3

Parasuraman, R., & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. Human Factors: The Journal of the Human Factors and Ergonomics Society, 39(2), 230–253. https://doi.org/10.1518/001872097778543886

Park, H. S., & Kim, K. H. (2015). Adaptive multimodal in-vehicle information system for safe driving. ETRI Journal, 37(3), 626–636. https://doi.org/10.4218/etrij.15.0114.1104

Peng, Y., Boyle, L. N., & Hallmark, S. L. (2013). Driver's lane keeping ability with eyes off road: Insights from a naturalistic study. Accident Analysis and Prevention, 50, 628–634. https://doi.org/10.1016/j.aap.2012.06.013

Prochaska, J. O., Velicer, W. F., Rossi, J. S., Goldstein, M. G., Marcus, B. H., Rakowski, W., … Rossi, S. R. (1994). Stages of Change and Decisional Balance for 12 Problem Behaviors. Health Psychology, 13(1), 39–46. https://doi.org/10.1037/0278-6133.13.1.39

Przybylski, A. K., Murayama, K., DeHaan, C. R., & Gladwell, V. (2013). Motivational, emotional, and behavioral correlates of fear of missing out. Computers in Human Behavior, 29(4), 1841–1848. https://doi.org/10.1016/j.chb.2013.02.014

Rasmussen, J. (1983). Skills, Rules, and Knowledge; Signals, Signs, and Symbols, and Other Distinctions in Human Performance Models. IEEE Transactions on Systems, Man, and Cybernetics, 13(3), 257–266.

Reyes, M. L., & Lee, J. D. (2004). The Influence of IVIS Distractions on Tactical and Control Levels of Driving Performance. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 48(September 2004), 2369–2373. https://doi.org/10.1177/154193120404801935

Ridgway, J., Hickson, L., & Lind, C. (2016). Decision-making and outcomes of hearing help-seekers: A self-determination theory perspective. International Journal of Audiology, 2027(February), 1–10. https://doi.org/10.3109/14992027.2015.1120893

Risto, M. (2014). Cooperative In-Vehicle Advice. University of Twente.

Risto, M., & Martens, M. H. (2012). Improving traffic flow on motorways through individual driver advice: A social dilemma? TRAIL-Beta Congress 2012, (October).

Risto, M., & Martens, M. H. (2013). Factors Influencing Compliance to Tactical Driver Advice: An Assessment Using a Think-Aloud Protocol. Proceedings of the 16th International IEEE Annual Conference on Intelligent Transportation Systems (ITSC 2013), 1923–1928. https://doi.org/10.1109/ITSC.2013.6728510

Rodríguez, M. D., Ibarra, J. E., Roa, J. R., Curlango, C. M., Bedoya, L. F., & Montes, H. D. (2014). Ambient gamification of automobile driving to encourage safety behaviors. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 8867(2002), 37–43. https://doi.org/10.1007/978-3-319-13102-3_8

Sagberg, F., Fosser, S., & Sætermo, I. A. F. (1997). An investigation of behavioural adaptation to airbags and antilock brakes among taxi drivers. Accident Analysis and Prevention, 29(3), 293–302. https://doi.org/10.1016/S0001-4575(96)00083-8

Schakel, W. J., & Van Arem, B. (2014). Improving traffic flow efficiency by in-car advice on lane, speed, and Headway. IEEE Transactions on Intelligent Transportation Systems, 15(4), 1597–1606. https://doi.org/10.1109/TITS.2014.2303577

Sebire, S. J., Kesten, J. M., Edwards, M. J., May, T., Banfield, K., Tomkinson, K., … Jago, R. (2016). Using self-determination theory to promote adolescent girls' physical activity: Exploring the theoretical fidelity of the Bristol Girls Dance Project. Psychology of Sport and Exercise, 24, 100–110. https://doi.org/10.1016/j.psychsport.2016.01.009

Shi, C., Lee, H. J., Kurczak, J., & Lee, A. (2012). Routine Driving Infotainment App: Gamification of Performance Driving. Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI'12), 181–183.

Smiley, A. (2000). Behavioral adaptation, safety, and intelligent transportation systems. Transportation Research Record, (1724), 47–51. https://doi.org/10.3141/1724-07

Solovey, E. T., Zec, M., Garcia Perez, E. A., Reimer, B., & Mehler, B. (2014). Classifying driver workload using physiological and driving performance data. Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems - CHI '14, 4057–4066. https://doi.org/10.1145/2556288.2557068

Staunton, L., Gellert, P., Knittle, K., & Sniehotta, F. F. (2015). Perceived Control and Intrinsic vs. Extrinsic Motivation for Oral Self-Care: A Full Factorial Experimental Test of Theory-Based Persuasive Messages. Annals of Behavioral Medicine, 49(2), 258–268. https://doi.org/10.1007/s12160-014-9655-2

Steinberger, F., Proppe, P., Schroeter, R., & Alt, F. (2016). CoastMaster: An ambient speedometer to gamify safe driving. Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, 83–90. https://doi.org/10.1145/3003715.3005412

Stutts, J., & Gish, K. (2003). Distraction in Everyday Driving. AAA Foundation for Traffic Safety. https://doi.org/10.1037/e364102004-001

Twisk, D., & Stacey, C. (2007). Trends in young driver risk and countermeasures in European countries. Journal of Safety Research, 38, 245–257. https://doi.org/10.1016/j.jsr.2007.03.006

Ulrich, T. A., Spielman, Z., Holmberg, J., Hoover, C., Sanders, N., Gohil, K., & Werner, S. (2013). Playing Charades With Your Car - The Potential of Free-form and Contact-based Gestural Interfaces for Human Vehicle Interaction. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 57(1), 1643–1647. https://doi.org/10.1177/1541931213571365

van Gent, P., Farah, H., van Nes, N., & van Arem, B. (2018). Analysing Noisy Driver Physiology Real-Time Using Off-the-Shelf Sensors: Heart rate analysis software from the Taking the Fast Lane Project. Submitted to Journal of Open Research Software.

van Gent, P., Melman, T., Farah, H., van Nes, N., & van Arem, B. (2018a). Multi-Class Driver Workload Prediction Using Machine Learning and Off-The-Shelf Sensors. Submitted to Transportation Research Board Conference 2018.

van Gent, P., Melman, T., Farah, H., van Nes, N., & van Arem, B. (2018b). Multi-Level Driver Workload Prediction Using Machine Learning and Off-The-Shelf Sensors. Transportation Research Record: Journal of the Transportation Research Board. https://doi.org/10.1177/0361198118790372

van Nes, N., & Duivenvoorden, K. (2017). Safely towards self-driving vehicles. https://doi.org/R-2017-2E

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User Acceptance of Information Technology: Toward a Unified View. MIS Quarterly, 27(3), 425–478.

Vlakveld, W. (2011). Hazard anticipation of young novice drivers. SWOV Institute for Road Safety Research.

Vlassenroot, S., Brookhuis, K., Marchau, V., & Witlox, F. (2010). Towards defining a unified concept for the acceptability of Intelligent Transport Systems (ITS): A conceptual analysis based on the case of Intelligent Speed Adaptation (ISA). Transportation Research Part F: Traffic Psychology and Behaviour, 13(3), 164–178. https://doi.org/10.1016/j.trf.2010.02.001

Watkins, W. C. (2016). A Social Learning Approach to Prescription Drug Misuse Among College Students. Deviant Behavior, 37(6), 601–614. https://doi.org/10.1080/01639625.2015.1060799

Wickens, C. D. (2002). Multiple resources and performance prediction. Theoretical Issues in Ergonomics Science, 3(2), 159–177. https://doi.org/10.1080/14639220210123806

Wu, Y., Li, C., & Khoo, S. (2015). Predicting Future Volunteering Intentions Through a Self-determination Theory Perspective. Voluntas, 1266–1279. https://doi.org/10.1007/s11266-015-9570-6

Young, K., & Regan, M. (2007). Driver distraction : A review of the literature. Distracted Driving, 379–405. https://doi.org/10.1201/9781420007497

Young, M. S., Brookhuis, K. A., Wickens, C. D., & Hancock, P. A. (2014). State of science: mental workload in ergonomics. Ergonomics, 0139(January 2015), 1–17. https://doi.org/10.1080/00140139.2014.956151

Young, M. S., Brookhuis, K. A., Wickens, C. D., & Hancock, P. A. (2015). State of science: mental workload in ergonomics. In Ergonomics (Vol. 58, pp. 1–17). https://doi.org/10.1080/00140139.2014.956151

Yusufov, M., Rossi, J. S., Redding, C. A., Yin, H. Q., Paiva, A. L., Velicer, W. F., … Prochaska, J. O. (2016). Transtheoretical Model Constructs' Longitudinal Prediction of Sun Protection Over 24 Months. International Journal of Behavioral Medicine, 23(1), 71–83. https://doi.org/10.1007/s12529-015-9498-7

Zaso, M. J., Park, A., Jueun, K., Gellis, L. A., Kwon, H., & Maisto, S. A. (2016). The associations among prior drinking consequences, subjective evaluations, and subsequent alcohol outcomes, 21(3), 193–201. https://doi.org/10.1016/j.molmed.2014.11.008.Mitochondria

Chapter 4.

Multi-Level Driver Workload Prediction Using Machine Learning and Off-The-Shelf Sensors

Abstract

This chapter presents a multi-level driver workload prediction model that can work with low-cost off-the-shelf sensor equipment. Driver workload was chosen as an important factor in timing messages to the driver, as described in the system level and information transfer level of the conceptual model in the previous chapter (3). The presented approach relies on measures that can be obtained unobtrusively in the driving environment, thus minimizing effects on the experimental interventions. To develop the prediction models two driving simulator studies were performed, one used regular driving conditions to induce workload, and one induces workload artificially induced with a demanding lane-keeping task. Individual and group-based models were trained and evaluated on both datasets. For the group-based models the generalizing capability was assessed using a leave-one-out cross validation. Results show that multi-level workload prediction on both the individual and group level can work well, achieving

high correct rates and accuracy scores. Generalizing to unknown individuals proved difficult using the realistic driving conditions, however generalization proved possible in the more demanding lane-keeping task. Reasons for this discrepancy along with future research directions are discussed in this chapter.

This chapter is based on an edited version of the following paper:
van Gent, P., Melman, T., Farah, H., van Nes, N., & van Arem, B. (2018). Multi-Level Driver Workload Prediction Using Machine Learning and Off-The-Shelf Sensors. Transportation Research Record: Journal of the Transportation Research Board. https://doi.org/10.1177/0361198118790372

4.1   Introduction

Research into driver workload has been conducted for at least three decades (Aasman, Mulder, & Mulder, 1987; de Waard, 1996). Recently, research efforts have shifted to using powerful Machine Learning (ML) methods, giving promising results (Jarvis, Putze, Heger, & Schultz, 2011; Solovey, Zec, Garcia Perez, Reimer, & Mehler, 2014). ML methods have been used for other driver-related classification problems, such as driver distraction (Liang, Reyes, & Lee, 2007), driver interruptibility (Kim, Chun, & Dey, 2015) or driver identification (Moreira-matias & Farah, 2017). The present study aims to fill the gaps in the existing research on predicting driver workload using ML methods in several ways, as will be explained in the next paragraphs.

First, ML studies into predicting driver workload often focus on a binary classification problem (high workload vs. low workload). A more fine-grained prediction of workload may be desirable to enable adaptive interfaces for in-vehicle advice systems (IVIS), systems that may simplify their content (Birrel, Young, Stanton, & Jennings, 2017), or driver assistance systems that may incrementally increase their level of support based on the level of driver workload. The experiments described in this paper attempt to predict workload on 7- and 10-point workload scales.

Second, studies to date often use intrusive sensors or measure variables (i.e. electroencephalogram, EEG) that are not practical in the driving environment (see for example (Jarvis et al., 2011; Solovey et al., 2014)). Additionally, it is unknown how well results obtained by the high-grade intrusive sensors used in experiments translate to low-cost sensors. This work uses low-cost sensors that can be integrated into the real-world driving environment, and uses measures that can be obtained non-intrusively. This is important, since especially low-cost sensors are likely to be integrated into the driving environment in real-world applications.

Lastly, the models generated in most studies are not generally publicly available for use by the research community. The models developed in this study will be made available for scientific use after publication of results (https://github.com/paulvangentcom).

4.1.1   Research Objectives

The previous section outlined the main research gaps and ways to add to the present literature. This led to the formulation of three criteria for predicting driver workload in the present work: The main goal is to develop a workload algorithm that (A) has usable accuracy when predicting multiple workload levels, while generalising among individuals, (B) uses data that can be measured with available low-cost sensors that can be integrated into the driving environment, and (C) is implementable on embedded hardware (for example in a smart steering wheel).

The first criterion (A), predicting workload at a higher resolution than the binary low/high found in previous literature while generalising among individuals, is addressed in the experimental design and data analysis presented in subsequent sections.

The second criterion (B) entails using sensor inputs from readily available, low-cost sensors that are easy to implement in the driving environment. By using low-cost sensors, which are likely to present more noise in the signal compared to high-end sensors, results will give a better reflection of real-world performance compared to studies using high-end sensors. Apart from having been used successfully in other workload prediction studies, selected variables should be measurable non-intrusively in the driving environment. This led to the selection of heart rate, skin response, blink rate and several performance measures (for an overview of the selection process, see (van Gent, Farah, van Nes, & van Arem, 2017)). This criterion ensures any results are directly applicable to in-car settings at a low cost, and that results obtained are likely to translate well to real-world applications.

Criterion C, ensuring the model is implementable on an embedded system, means it must be efficient both in memory use as well as computational requirements. Two machine learning algorithms were selected that can satisfy this criterion: 'Random Forest' and 'Support Vector Machine' algorithms. Random Forests (Breiman, 2001) are computationally efficient (Sventnik et al., 2003) but can have a large memory footprint. Solutions have been proposed that allow embedded implementations while maintaining performance (Mishina, Murata, Yamauchi, Yamashita, & Fujiyoshi, 2015), making it a suitable algorithm to use. Support Vector Machines (Cortes & Vapnik, 1995) implementations can suffer from computational complexity, as well as high memory footprint for more complex models. Methods have been proposed, however, that achieve remarkable efficiency increases without sacrificing performance (Bajaj, Chiu, & Allebach, 2014; Theocharides & Member, 2016), making SVM's also a suitable candidate algorithm.

Two experiments were conducted to evaluate the feasibility of the previously defined criteria. First, a simulator experiment was performed, where workload was induced using realistic driving situations. Results of this experiment were explored futher using a dataset obtained from another driving simulator experiment that induced workload with a demanding lane-keeping task. Finally, results of both experiments are discussed and future steps are outlined.

4.2   Estimating Workload in a Realistic Driving Scenario Study

To assess the feasibility of predicting driver workload in realistic driving settings, a simulator study was performed. The main goal was to evaluate the prediction of multi-level driver workload in realistic driving conditions.

4.2.1   Methods

4.2.1.1   *Equipment*
The study was performed in a fixed-base, medium-fidelity driving simulator. A dashboard mockup with three 4K-displays (resolution 4096*2160 px) provided roughly 180-degree vision. Actuators consisted of a Fanatec steering wheel and pedals, and a custom blinker control. The simulation ran in Unity3D. The simulated vehicle had an automatic gearbox and a top speed of 165 km/h. Figure 4.1(A) illustrates the set-up.

Physiological data were recorded at 100Hz, using low-cost sensors powered by an Atmel ATMega328p embedded processor board. Heart rate was recorded using a photoplethysmographic (PPG) method (Jae Baek et al., 2009) at the left index finger. Skin response was recorded at the middle and ring finger of the same hand (see figure 4.1(B)). Additionally, blink data were recorded using a GoPro HERO+ camera on the dashboard, running at 1080p@30Hz. Simulator data were logged at 50Hz.
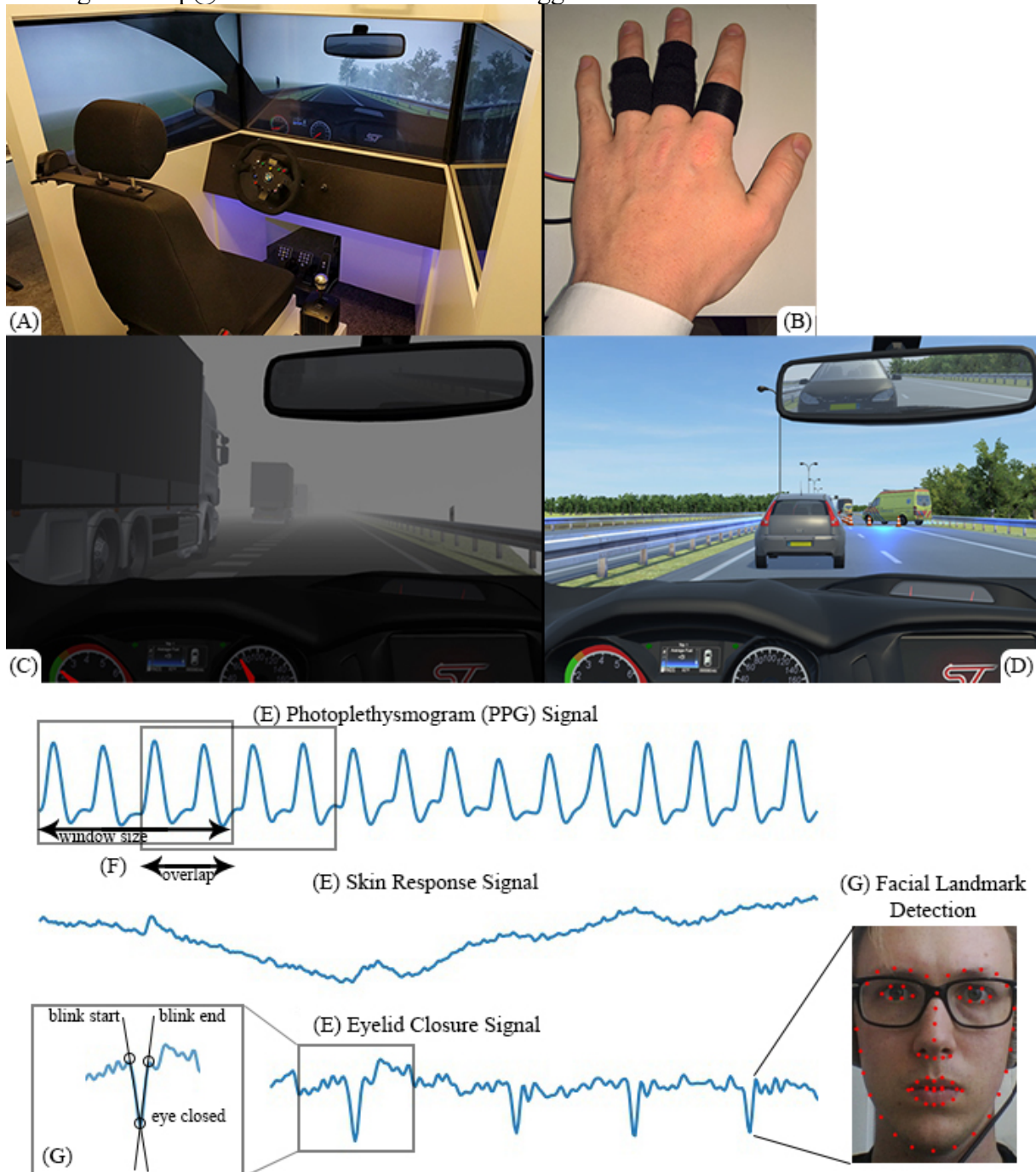
Figure 4.1 - Figure showing the simulator set-up (A), physiological sensors (B), the merging between a platoon of trucks in dense fog (C) and the accident site at the end of the 'high workload' scenario (D). Examples of the raw signal data are shown (E), the concepts of window size and overlap factor (F), an example of the facial landmark detection and the resulting process of analysing the blink rate signal (G).

4.2.1.2   Simulator Scenarios

Two scenarios were created in Unity3D, one scenario with situations likely to induce high workload ('high' workload' scenario) and one with situations that are not likely to induce high workload ('low workload' scenario). Road geometry was based on a part of the Cooperative-ITS (C-ITS) corridor in the Netherlands: the A67, a two-lane highway between Eindhoven and Venlo with speed limit of 130km/h. Three weather conditions were designed for each scenario: clear weather, and two levels of fog with visibility of approx. 150 meters ('light fog') and below 25 meters ('heavy fog'). This gave a total of six scenarios.

To accurately design the road geometry, CAD drawings of the road segments were secured from the open data program of the Dutch government (https://data.overheid.nl). Using Autodesk 3DS Max, the data in the CAD files were converted to 3D models and textured. The surrounding terrain was generated using height map data obtained from the Microsoft Bing Maps API (https://www.bingmapsportal.com/). Canals and wooded areas were extracted automatically from satellite imagery, and adjusted by hand where necessary. The location, shape, and content of traffic signs was inferred from Google Streetview, designed in 3DS Max and manually placed at the corresponding locations in the scenario.

The 'high workload' scenario was 15.9 km in length, and ran between Eindhoven and Someren. Participants would encounter several workload-inducing 'events' spread out across the scenario. After accelerating across an on-ramp, the first event was encountered: particpants had to merge into a dense platoon of trucks (4-5 meters headway, Figure 4.1(C)), a manoeuvre shown to increase workload on the driver (de Waard, Kruizinga, & Brookhuis, 2008). The second event was encountered two kilometres downstream and consisted of a segment of slow moving traffic on the right lane, designed to nudge the participants to drive in the left lane. While passing the slow-moving traffic, an ambulance approached from behind exhibiting auditory and visual signals, travelling at the legally allowed max speed of 170km/h in the Netherlands (max. 40km/h difference with other traffic). This placed the participant in the demanding situation of quickly having to find a gap in the much slower moving lane to the right and perform a merging manoeuvre. The third event was a game of '20 questions' (Kun, Shyrokov, & Heeman, 2013), intended to simulate an engaging (phone) conversation. By asking at most 20 polar (yes/no) questions, participants had to guess which animal, object or person the experimenter had in mind. The final event came near the end of the scenario. The right lane was closed off due to an accident, with slow moving (< 15 km/h) traffic on the left lane (Figure 4.1 (D)). The 20 questions game was played until the accident site was reached. If participants finished early, the game was restarted with a different subject. After this, participants took the next exit and stopped the car.

The 'low workload' scenario consisted of self-paced driving in light traffic for 20.5km. The simulated road was a replica of the A67 road between Someren and Venlo. There were no events. Participants drove until reaching a designated exit, where they stopped the car.

4.2.1.3   Experimental Procedure

Approval for the study was obtained from the ethics committee at Delft University of Technology. Participants drove the six scenarios spread out over three separate days, each day driving one randomly assigned 'high workload' and one 'low workload' scenario. This approach was taken because physiological measures can vary from day to day, as well as to avoid a fatigue effect from occurring when asking participants to drive six 10-15-minute scenarios consecutively.

In the 'high workload' scenario, participants were asked to rate their experienced mental effort and task difficulty on a 7-point scale after each event, leading to six workload data points per run. In the 'low workload' scenario, the questions were asked at fixed positions in the scenario, leading to four workload data points per run. The exact questions were 'How much mental effort did the driving task take in the last few moments, on a scale of 1-7?' and 'How difficult was the driving task in the last few moments, on a scale of 1-7?'. Scale labels ranged from very low/easy, to very high/difficult, and were explained to participants before the experiment started. Note that we did not use a standardised workload scale such as the NASA TLX or RSME, since we wanted to keep interaction time with and demands on the driver to a minimum.

Participants that registered for the experiment received a copy of the informed consent. It was signed and brought to the first session. After being seated in the simulator, a relaxation period of three minutes was given to the participants. This was to allow the physiological measures of each participant to return to its baseline. Sensors were attached, after which the signal quality was checked. A physiological baseline was recorded first. After the baseline, it was briefly explained to the participant that there would follow a drive on a segment of the A67 highway. Participants were instructed to drive at their own pace, but not exceed the speed limit as indicated on road-side signs. If a participant was unfamiliar with '20 questions', a test round was played to familiarise them with the game.

4.2.2   Data Analysis

Participants were asked to rate their mental effort and driving task difficulty on a 7-point scale. Since querying the driver might influence workload, the 'high workload' scenario was constructed in such a way that at least one minute of driving was between each two events, to allow signals to return to baseline. The data recorded between two events were not used in the analysis. In the case of the 'low workload' scenario, one minute of data following each question were excluded from the analysis.

4.2.2.1   *Preprocessing of Physiological Data*

An algorithm was developed to extract the most commonly used features from the measured heart rate signal (van Gent, 2016, 2017), using a sliding window approach (see Figure 4.1F). The output measures are divided into time-domain (Reimer et al., 2013) and frequency-domain measures (Montano et al., 2009). In the time-domain, the measures included are BPM (beats per minute), IBI (inter-beat interval), MAD (median absolute deviation of intervals between heart beats), SDNN (standard deviation of intervals between heart beats), RMSSD (root mean square of successive differences between neighbouring heart beat intervals), SDSD (standard deviation of successive differences between neighbouring heart beat intervals), and the pNN50 and pNN20 (proportion of differences between successive heart beats greater than 50ms and 20ms, resp.) In the frequency domain, included measures are LF (the low frequency band: 0,04-0,15Hz), which is related to short-term blood pressure variation, and HF (the high frequency band: 0,16-0,5Hz), which reflects breathing rate, and the LF/HF ratio, a measure of sympathetic-parasympathetic balance (Billman, 2011; Montano et al., 2009).

Skin response consists of a tonic and phasic component (Lim et al., 1997). Tonic represents the long-term, slow variation in the signal, indicative of general psycho-physiological arousal (Seitz, Daun, Zimmermann, & Lienkamp, 2012). Phasic reflects relatively quick responses to discrete external stimuli, occurring generally between 1-3 seconds after stimulus onset (Seitz et al., 2012). Power in the frequency spectrum of skin reponse between 0.03Hz-0.5Hz has been linked to short term workload changes (Shimomura et al., 2008). The mean, max-min

difference, MAD (median absolute difference), and 0.03-0.5Hz frequency spectrum were extracted from the GSR signal, using the same window approach as for heart rate. Frequency spectra were extracted using a trapezoidal integration of the area under corresponding frequency bands in the power spectrum.

Blink data were detected offline from recorded video data. An algorithm was developed to extract blink number, blink duration and inter-blink-interval. It functioned by detecting 68 'facial landmarks' (Köstinger, Wohlhart, Roth, & Bischof, 2011), then calculating eyelid distance for each frame. Blinks were detected in the resulting signal by finding large slopes, then finding the lowest point of reversal. The process is displayed visually in Figure 4.1 (G).

4.2.2.2   *Driver Performance Data*

Performance measures reflect how the control the driver exerts over the vehicle varies across conditions. We included steering wheel angle, steering wheel reversals, speed, variation in lateral and longitudinal position, and headway and time to collision when available (for more information, see (van Gent et al., 2017)).

4.2.2.3   *Generating Machine Learning Sets*

Machine learning sets were generated from the raw data and labelled based on self-report data, by varying window size and overlap factor. Window size refers to how much data is used for the calculation of features, overlap factor refers to how much data any window W_i shares with the previous window W_{i-1}. Both concepts are visualised in Figure 4.1 (F). Window sizes of 5, 10 and 30 seconds, and overlap factors of 0% and 50% were used, leading to a total of 6 sets.

4.2.2.4   *Model Development and Evaluation*

Two different machine learning algorithms were used: A Random Forest Regressor (RFR), and a Support Vector Machines Regressor (SVR). The RFR creates an ensemble (forest) of regression trees in which each tree is trained on a random subset of the features. They have been used in for example (Miyaji, Danno, Kawanaka, & Oguri, 2008). Support Vector Machines function by mapping the data to a higher dimensional space, and solving an optimization problem to identify a set of hyperplanes that separate the training data into classes. They have been used in for example (Liang et al., 2007; Moreira-matias & Farah, 2017). With the SVR, the Polynomial kernel (SVR(poly)), and the Radial Basis Function kernel (SVR(rbf)) were evaluated. Algorithms that were used are taken from the SciKit-Learn repository (Pedregosa et al., 2012).

The resulting models were evaluated using several metrics. Model error was evaluated using mean absolute error ($AE\mu$) and median absolute error ($AE\mu1/2$), both measures of the accuracy of the predictions. The coefficient of determination ($R2$) was also computed as a goodness-of-fit measure. Performance for class-based predictions was also evaluated, expressed as correct rate.

4.2.3   Results

4.2.3.1   *Participants*

19 participants took part in the experiment. Data from one participant were excluded because of a failure to understand some tasks due to a language barrier. This left 18 participants, of which 12 were males and 6 were females. The average age was 34.56 years (SD 10.09). Of the 18 participants, 12 owned a car and reported using it three to four times a week on average, and travelling between 2500 and 15000km annually. All participants held a valid driver's license.

No simulator sickness severe enough to terminate a driving session was reported. Reported mental effort and perceived difficulty correlated with weather conditions and with scenario type independently and in line with expectations, although no interaction effect was present (van Gent et al., 2017).

4.2.3.2   *Individual Models*
The training and testing sets for the individual models were generated by dividing the dataset of each driver into training and testing sets with an 80%/20% split ratio, respectively. This split ratio was chosen to ensure sufficient training data, since individual datasets were relatively small.

The results indicated that the models functioned well, with the RFR outperforming the SVR. For all individual models with a window size of 5s and overlap of 0%, the AEμ was 0.343, the AEμ1/2 was 0.129, R2 was 0.679, Correct Rate (CR) was 76.30% when predicting discrete classes, and 93.80% when miss-by-one errors were allowed (CR+/-1). This indicated that on average, predictions were off by 0.343, and that half the predictions had an error less than 0.129, from a total scale of 7 classes. See Table 4.1 for an overview of all results. Model performance increased with a larger overlap factor. This was expected, since a larger overlap creates a larger training set to fit the model to, and because a larger overlap factor indicates more shared variance between adjacent samples. Interestingly, an inverse relationship between window size and model performance was observed, contrary to what has been reported previously (Solovey et al., 2014). Miss-by-one errors indicate predictions that are 'almost correct', and still contain enough information about the true workload states. For example, if workload is predicted as '6' while the true value is '7', the information in the prediction is still useful: in either case workload is on the high end.

Table 4.1 Performance Metrics RFR models.

Window Size	5 sec		10 sec		30 sec	
Overlap Factor	0%	50%	0%	50%	0%	50%
Individual Model						
AE_μ	0.343	0.219	0.431	0.280	0.613	0.492
$AE_{\mu1/2}$	0.129	0.565	0.296	0.109	0.490	0.291
R^2	0.679	0.8716	0.590	0.794	0.071	0.306
CR	76.30%	85.21%	67.88%	80.77%	49.68%	60.82%
CR +/- 1	93.80%	97.61%	92.93%	96.13%	85.81%	89.55%
Group Model						
AE_μ	0.605	0.455	0744	0.553	0.898	0.801
$AE_{\mu1/2}$	0.406	0.250	0.565	0.344	0.628	0.652
R^2	0.661	0.774	0.564	0.709	0.372	0.504
CR	57.40%	69.57%	46.12%	62.48%	40.47%	43.82%
CR +/- 1	90.60%	93.81%	87.02%	91.42%	80.60%	84.56%
Generalising Model						
AE_μ	1.522	1.536	1.457	1.519	1.375	1.424
$AE_{\mu1/2}$	1.163	1.201	1.199	1.253	1.174	1.230
R^2	-0.538	-0.623	-0.460	-0.602	-0.299	-0.396
CR	20.07%	20.05%	19.81%	20.21%	20.21%	20.47%
CR +/- 1	55.18%	55.19%	55.46%	54.94%	57.21%	55.89%

4.2.3.3   Group Models

The second step was to estimate the model performance within the entire group. The dataset containing data from all drivers was split into training- and testing sets with a 60%/40% split ratio. Since the size of the group dataset is much larger compared to individual dataset, a more stringent split ratio could be chosen while maintaining a sufficiently large training set.

Results indicated group models performed well. The $AE\mu$ for the model with window size 5s and 0% overlap was 0.605, the $AE\mu1/2$ 0.406, R2 0.661, CR 57.40%, and CR+/-1 90.60%.

4.2.3.4   Generalising Group Models

The last step was to assess how models would perform in a realistic setting, e.g. a setting where workload from an unknown driver is predicted based on data from a pool of other drivers. To achieve this, data were sampled using a k-fold approach, with k = Nparticipants. For every ki, the training set consisted of all data except the held out participant ki. Workload for participant ki was then predicted and model performance evaluated. This method simulated how the trained models would perform when predicting data from previously unseen individuals. This obtained performance measure reflects real-world settings, where it is impractical for models to be trained on all possible drivers and generalising power is thus preferable.

Results showed that models did not perform well when generalizing to unknown drivers. The $AE\mu$ for all individual models with window size 5s and 0% overlap was 1.522, $AE\mu1/2$ was 1.163, R2 was -0.538, CR 20.07%, and CR+/-1 55.18%. The strongly negative coefficient of determination suggests unsatisfactory performance (the mean of the data is a better predictor than the trained model). The relatively low (though above chance level, not satisfactory) absolute error rates given R2 are explained by a class imbalance in the dataset, where two classes (workload level 1 and 2) dominate. To assess whether this was a possible cause for the poor performance of the models, data were resampled using SMOTE (Synthetic Minority Over-Sampling Technique) (Chawla, Bowyer, Hall, & Kegelmeyer, 2002). This had little discernible effect on the model performance, and it was concluded that low performance was not due to the class imbalance in the dataset. It was also observed that R2 increases slightly with increasing window size, in accordance with earlier studies (Solovey et al., 2014) and contrary to the individual and group models in the present study.

4.2.4   Conclusion

The results of this study showed that predicting self-reported workload in a simulated realistic environment was possible at the individual and group level, but proved difficult when generalising to unknown drivers. Several causes can be identified. The simulated scenarios might not have induced sufficient workload to be measurable with performance or physiological measures. Indeed, most participants indicated that driving in the simulator felt very different from actual driving, and was not that difficult at all. Since a self-report measure was used, which is a subjective measure, it is possible that different participants had biased response tendencies. Lastly, it might also be the case that different physiological response patterns to workload exist, in which case the sample size of 18 could have been too small to account for all occurring patterns.

This raises the question whether workload prediction is at all possible on non-binary scales, while generalising across drivers. To further explore this possibility, a dataset from a study with a lane-keeping task was obtained. This study and the results are discussed in the next section.

4.3 Estimating Workload in a Forced-Pace Simulator Study

A dataset was re-used from a previously executed study by Melman et al. (Melman, Abbink, van Paassen, de Boer, & Winter, 2018) to further assess multi-level workload prediction in drivers. The study featured a challenging lane-keeping task, which had the potential to induce higher workload than the previous study. The same physiological and performance measurements were used in as in the previously described simulator study.

4.3.1 Method

4.3.1.1 Equipment
The study was performed in a fixed-base driving simulator at the faculty of Aerospace Engineering, Delft University of Technology. The simulator consisted of a mockup dashboard with three LCD projectors (BenQ W1080ST 1080p) that provided roughly 180-degree vision. The simulated vehicle had an automatic gearbox and a top speed of 210 km/h.

Physiological data were logged using a biosignalsPlux wireless hub at 1000Hz. Heart rate was recorded using three pre-gelled Ag/AgCl electrodes at the heart's v3-node. Skin response was measured using the same pre-gelled electrodes, placed inside the palm and on the wrist of both hands. Simulator data were logged at 100Hz.

4.3.1.2 Scenarios
The scenarios used to induce workload in drivers each consisted of a 25km long, single-lane road. The road was divided into four 6km sections of different lane width (3.6m, 2.8m, 2.4m, 2.0m). Each section had seven curves, five with an inner radius of 750m and two with a 500m radius. Transitions between sections of different width always took place in a 750m radius curve, and were preceded by a road sign indicating a narrowing road. The four sections were identical, with the exception that the curves of segments 2 and 4 four were mirrored with respect to section 1 and 3.

Cones were placed 8m apart on the road markings on both sides of the road. The main task was to hit as few cones as possible. A cone hit was indicated to the participant visually by a red dot on the side of the car where the cone was hit, and by a loud auditory beep. Extra difficulty in lane-keeping was induced by a perturbation added to the vehicle's lateral motion. This perturbation was an unpredictable multi-sine signal with five frequencies between 0.067Hz and 0.25Hz, with a maximum summed amplitude of 1,000N. Without the perturbation, lane keeping (especially on straight segments) was not considered challenging enough. The width of the simulated vehicle was 1.8m.

Three runs were driven with the aim of inducing different levels of workload: a self-paced run and two forced-pace runs of 90km/h and 130km/h. In the self-paced run, participants had full longitudinal control over the car and could drive at their own pace. In the forced-pace conditions, however, the car's speed was automated and kept constant at 90km/h and 130km/h. This would push participants into curves at high speeds, with the goal of raising their workload significantly. The three runs were presented to the participants in randomised order.

4.3.1.3 Procedure
Participants read and signed an informed consent form, informing them of the purpose and procedure of the study. Participants were instructed that the main task was to minimise the total number of cone hits. Furthermore, participants were informed that during the experiment, a

beep would sound every 20 seconds. At the sounding of this beep, participants were asked to verbally answer the question "From 0 to 10, how much effort does the current driving task take you?", with 0 being 'no effort', 5 being 'moderate effort' and 10 being 'a lot of effort'.

Before the experiment started, participants were familiarised with the simulator and the procedure by driving two 3.7km trial runs. The first trial run was self-paced, the second was forced-pace with speed at 110km/h. After the trial run, any question the participant had was answered. The electrodes were attached, and a one-minute baseline was recorded.

4.3.2   Analysis

Participants rated their mental effort on a scale of 0-10, every 20 seconds. This rating was annotated by the experimenter and added to the dataset. What data were logged, data preprocessing, ML set generation, model development and evaluation are identical to what has been described in the previous study.

4.3.3   Results

4.3.3.1   Participants

In total twenty-four participants took part in the experiment (17 male, 7 female). The average age was 24.6 years (SD 2.4). Participants reported driving multiple times a week (11 participants), at least once a month (7 participants) or less than one month (6 participants). All participants held a valid driving license. Reported mental effort was sensitive to the lane width variations, although regarding speed only to 130 km/h forced-pace condition (Melman et al., 2018).

Table 4.2 Performance Metrics RFR Models.

Window Size	5 sec		10 sec		30 sec	
Overlap Factor	0%	50%	0%	50%	0%	50%
Individual Model						
AE_μ	1.046	0.823	1.213	0.853	1.127	0.870
$AE_{\mu1/2}$	0.662	0.511	0.833	0.518	0.959	0.694
R^2	0.635	0.763	0.600	0.675	0.561	0.735
CR	40.74%	50.31%	33.93%	45.83%	20.83%	40.28%
CR +/- 1	77.31%	84.34%	70.83%	81.94%	65.83%	81.48%
Group Model						
AE_μ	0.904	0.730	0.984	0.808	1.084	0.876
$AE_{\mu1/2}$	0.638	0.482	0.722	0.546	0.792	0.663
R^2	0.774	0.830	0.740	0.802	0.718	0.811
CR	41.61%	51.30%	35.12%	46.44%	34.22%	37.87%
CR +/- 1	82.30%	88.18%	80.32%	85.88%	73.21%	82.41%
Generalising Model						
AE_μ	1.878	1.988	1.988	1.989	1.809	1.717
$AE_{\mu1/2}$	1.831	1.844	1.718	1.741	1.680	1.568
R^2	0.118	0.079	0.196	0.177	0.411	0.433
CR	14.09%	13.45%	12.62%	13.44%	15.72%	15.21%
CR +/- 1	41.92%	40.70%	44.15%	42.29%	47.16%	46.32%

4.3.3.2   *Individual Models*

As in the previous study, training and testing sets for the individual models were generated by dividing the dataset into two stratified sets. More data per participant were collected than in the previous experiment, so data were split with the more stringent 60%/40% split ratio.

Results were similar to the previous study, and indicated that the models performed well, with RFR outperforming SVR. An inverse relationship between model performance and overlap factor was observed, as well as increasing performance with increasing overlap factors, both as in the previous experiment. For all individual models with a window size of 5s and overlap of 0%, the AEμ was 1.046, the AEμ1/2 0.662, R2 0.635, CR 40.74%, and CR+/-1 77.31%. The relatively larger absolute errors, compared to individual models in the previous study, might have resulted from the wider workload scale, the different nature of the driving task, or the more frequent reporting of mental workload. More information is displayed in Table 4.2.

4.3.3.3   *Group Models*

To evaluate performance at the group level, data were split with a 60%/40% split ratio. Results indicated group models attained high performance. For the model with window size 5s and 0% overlap, the AEμ was 0.904, the AEμ1/2 0.638, R2 0.774, CR 41.61%, and CR+/-1 82.30%. Table 4.2 displays the full results. Performance increased with larger overlap factors, and again an (weak) inverse relationship between performance and window size was observed.

4.3.3.4   *Generalising Group Models*

Model performance when generalising to unknown individuals was then assessed, which did not perform well in the first simulator experiment. Data sampling methods were identical to the previous study.

Results indicated models performed moderately well. For the best performing model with window size 30s and 50% overlap, the AEμ was 1.717, the AEμ1/2 1.568, R2 0.433, CR 15.21%, CR+/-1 46.32%. Although model absolute error is relatively large, the coefficient of determination indicated a moderate relationship between model and data. Figure 4.2 below displays the predicted and true values for the first four participants. Individual model performance varied, with workload being predicted well for some participants, while for others showed a correct trend but with a constant offset error. These offset errors inflated the absolute error rates and deflated the predictive accuracy despite good model performance. Generally, a decreased performance with increased overlap factor was observed (except for the largest window size of 30s), as well as increased performance with increased window size. The effect is similar to results for the model generalisation step in the previous study, but more pronounced. The effect also corresponds with what has been reported before (Solovey et al., 2014).
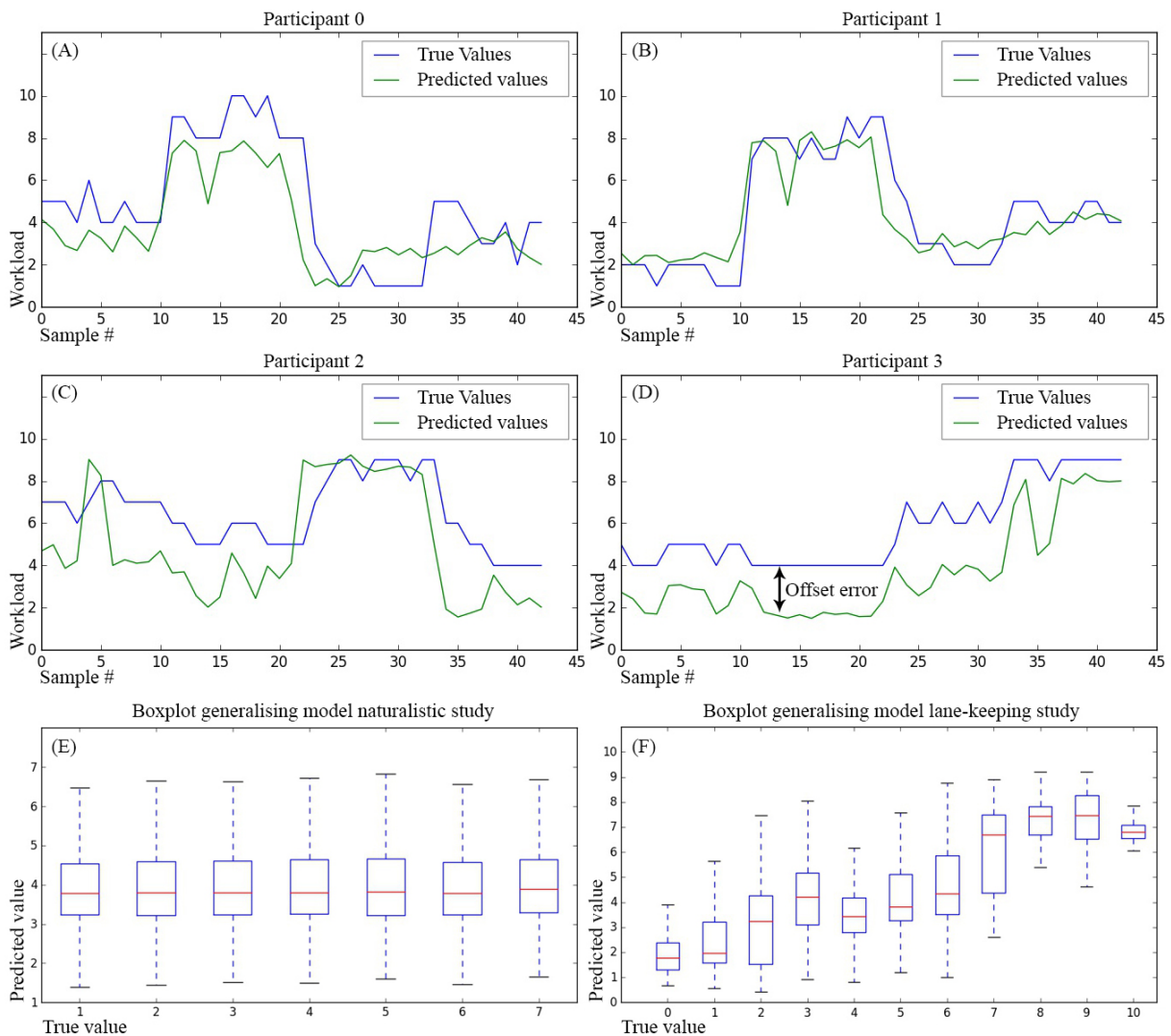
Figure 4.2 – The top four windows (A-D) show plots from the first four participants indicate that the models performed well, with the relatively large absolute errors likely resulting from individual scaling problems in the predictions. These offset errors are indicated in (D): the general trend is predicted well but there is a constant offset error. The last two windows (E-F) show box plots, further exploring the generalising models from both studies.

4.3.4   Conclusion

The results of this study show similarities with the previous study for individual and group-based models. Additionally, this second experiment shows that, when predicting multi-level workload (11 classes), generalising performance was satisfactory, although still with room for improvement.

This study seems to indicate that indeed non-binary workload prediction that generalises to unknown individuals is possible using ML methods. Although models generalising between individuals showed variations in performance based on which individual's workload was being predicted, including constant offset errors in several participants, overall performance was promising.

4.4   Overall Conclusion and Discussion

The present study tried to model driver workload using machine learning techniques that can run on embedded systems, with data collected from low-cost-sensors. Results have shown that individual models and within-group models functioned well in both a realistic driving setting as well as an artificial lane-keeping task setting. When generalising to unknown drivers, only the lane-keeping study produced usable results. As displayed in Figure 4.2 (E-F), in the first study the generalised model learns to predict values around the mean to optimize accuracy, in the second study the model learns to predict based on the reported workload.

Since the data gathered in the study are time-series human physiological and performance data, it likely exhibits strong autocorrelation from one sample to the next. This might be a potential explanator for the higher performance in the individual and group models in both studies. Since with random sampling, shared variance between samples from the training set and the prediction set might bias the classifier towards a higher accuracy. To better assess performance, training cases were included where the models had to generalize to unknown individuals. These give a more accurate indication of performance, since with this approach there is no shared variance between training set (all participants minus participant k) and the testing set (participant k). As such, only the generalizing training case offers a reliable index of performance. This is an important distinction, since it shows that although using machine learning to predict driver workload can lead to promising results, care must be taken when interpreting the results. Without care in selecting the sampling techniques used, model performance might be inflated.

Possible reasons for the discrepancy in generalizing performance between both studies could include that the workload induced in the realistic settings was too low to be reflected in the physiological or performance signals, that workload induced by artificial tasks is more easily measurable than that induced by more realistic tasks, or that different physiological response patterns to workload might exist and that the sample in the first study was either too small or contained too much individual variation.

Possible limitations of the present study are that we employed a self-report measure as ground truth of the experienced mental workload of the drivers. We did not employ standardised workload scales such as NASA TLX, to keep interaction time and demand with the driver to a minimum. However, this may have contributed to lower model performance through participant response tendencies, and leaves some doubt as to what degree the data captures workload. In addition to this, we did not look at compensatory behaviour drivers might employ to manage their workload, such as reducing speed in complex or demanding situations.

Future directions are planned. These include feature space normalisation of the dataset to attempt to reduce the offset errors observed in some individuals, as well as exploring additional feature extraction methods. After this, on-road testing is planned to explore model performance in real-world driving settings. Lastly, development of an embedded variant of the model is planned.

References

Aasman, J., Mulder, G., & Mulder, L. J. M. (1987). Operator effort and the measurement of heart-rate variability. Human Factors, 29(2), 161–170. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-0023330102&partnerID=tZOtx3y1

Bajaj, N., Chiu, G. T. C., & Allebach, J. P. (2014). Reduction of memory footprint and computation time for embedded Support Vector Machine (SVM) by kernel expansion and consolidation. IEEE International Workshop on Machine Learning for Signal Processing, MLSP. https://doi.org/10.1109/MLSP.2014.6958875

Billman, G. E. (2011). Heart rate variability - A historical perspective. Frontiers in Physiology, 2 NOV(November), 1–13. https://doi.org/10.3389/fphys.2011.00086

Birrel, S., Young, M., Stanton, N., & Jennings, P. (2017). Using Adaptive Interfaces to Encourage Smart Driving and Their Effect on Driver Workload. 484, 764. https://doi.org/10.1007/978-3-319-41682-3

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16, 321–357. https://doi.org/10.1613/jair.953

Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. Machine Learning, 20(3), 273–297. https://doi.org/10.1023/A:1022627411411

de Waard, D. (1996). The Measurement of Drivers ' Mental Workload. https://doi.org/10.1016/j.apergo.2003.11.009

de Waard, D., Kruizinga, A., & Brookhuis, K. A. (2008). The consequences of an increase in heavy goods vehicles for passenger car drivers' mental workload and behaviour: A simulator study. Accident Analysis and Prevention, 40(2), 818–828. https://doi.org/10.1016/j.aap.2007.09.029

Jae Baek, H., Bit Lee, H., Soo Kim, J., Min Choi, J., Keun Kim, K., & Suk Park, K. (2009). Nonintrusive Biological Signal Monitoring in a Car to Evaluate a Driver's Stress and Health State. Telemedicine and E-HEALTH, 15(2), 182–189.

Jarvis, J., Putze, F., Heger, D., & Schultz, T. (2011). Multimodal person independent recognition of workload related biosignal patterns. Proceedings of the 13th International Conference on Multimodal Interfaces - ICMI '11, 205. https://doi.org/10.1145/2070481.2070516

Kim, S., Chun, J., & Dey, A. K. (2015). Sensors Know When to Interrupt You in the Car. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15, 487–496. https://doi.org/10.1145/2702123.2702409

Köstinger, M., Wohlhart, P., Roth, P. M., & Bischof, H. (2011). Annotated facial landmarks in the wild: A large-scale, real-world database for facial landmark localization. Proceedings of the IEEE International Conference on Computer Vision, 2144–2151. https://doi.org/10.1109/ICCVW.2011.6130513

Kun, A. L., Shyrokov, A., & Heeman, P. a. (2013). Interactions between human-human multi-threaded dialogues and driving. Personal and Ubiquitous Computing, 17(5), 825–834. https://doi.org/10.1007/s00779-012-0518-1

Liang, Y., Reyes, M. L., & Lee, J. D. (2007). Real-time detection of driver cognitive distraction using support vector machines. IEEE Transactions on Intelligent Transportation Systems, 8(2), 340–350. https://doi.org/10.1109/TITS.2007.895298

Lim, C. L., Rennie, C., Barry, R. J., Bahramali, H., Lazzaro, I., Manor, B., & Gordon, E. (1997). Decomposing skin conductance into tonic and phasic components. International Journal of Psychophysiology, 25(2), 97–109. https://doi.org/10.1016/S0167-8760(96)00713-1

Melman, T., Abbink, D. A., van Paassen, M. M., de Boer, E. R., & Winter, J. C. F. (2018). What determines drivers' speed? A replication of three behavioural adaptation experiments in a single driving simulator study. Ergonomics. https://doi.org/10.1080/00140139.2018.1426790

Mishina, Y., Murata, R., Yamauchi, Y., Yamashita, T., & Fujiyoshi, H. (2015). Boosted Random Forest. IEICE Transactions on Information and Systems, E98(D), 1630–1636.

Miyaji, M., Danno, M., Kawanaka, H., & Oguri, K. (2008). Driver's cognitive distraction detection using adaboost on pattern recognition basis. Proceedings of the 2008 IEEE International Conference on Vehicular Electronics and Safety, ICVES 2008, 51–56. https://doi.org/10.1109/ICVES.2008.4640853

Montano, N., Porta, A., Cogliati, C., Costantino, G., Tobaldini, E., Casali, K. R., & Iellamo, F. (2009). Heart rate variability explored in the frequency domain: A tool to investigate the link between heart and behavior. Neuroscience and Biobehavioral Reviews, 33(2), 71–80. https://doi.org/10.1016/j.neubiorev.2008.07.006

Moreira-matias, L., & Farah, H. (2017). On Developing a Driver Identification Methodology Using In-Vehicle Data Recorders. IEEE Transactions on Intelligent Transportation Systems, submitted. https://doi.org/10.1109/TITS.2016.2639361

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, É. (2012). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825–2830. https://doi.org/10.1007/s13398-014-0173-7.2

Reimer, B., Donmez, B., Lavallière, M., Mehler, B., Coughlin, J. F., & Teasdale, N. (2013). Impact of age and cognitive demand on lane choice and changing under actual highway conditions. Accident Analysis and Prevention, 52, 125–132. https://doi.org/10.1016/j.aap.2012.12.008

Seitz, M., Daun, T. J., Zimmermann, A., & Lienkamp, M. (2012). Measurement of Electrodermal Activity to Evaluate the Impact of Environmental Complexity on Driver Workload. Proceedings of the FISITA 2012 World Automotive Congress, 245–256. https://doi.org/10.1007/978-3-642-33741-3

Shimomura, Y., Yoda, T., Sugiura, K., Horiguchi, A., Iwanaga, K., & Katsuura, T. (2008). Use of frequency domain analysis of skin conductance for evaluation of mental workload. Journal of Physiological Anthropology, 27(4), 173–177. https://doi.org/10.2114/jpa2.27.173

Solovey, E. T., Zec, M., Garcia Perez, E. A., Reimer, B., & Mehler, B. (2014). Classifying driver workload using physiological and driving performance data. Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems - CHI '14, 4057–4066. https://doi.org/10.1145/2556288.2557068

Sventnik, V., Liaw, A., Tong, C., Culberson, J. C., Sheridan, R. P., & Feuston, B. P. (2003). Random forest: a tool for classification and regression in compound classification and QSAR modeling. Journal of Chemical Information and Computer Science, 43(6), 1947–1958. https://doi.org/10.1016/j.rse.2008.02.011

Theocharides, T., & Member, S. (2016). Embedded Hardware-Efficient Real-Time Vector Machines. IEEE Transactions on Neural Networks and Learning Systems, 27(1), 99–112. https://doi.org/10.1109/TNNLS.2015.2428738

van Gent, P. (2016). Analyzing a Discrete Heart Rate Signal Using Python. Retrieved from http://www.paulvangent.com/2016/03/15/analyzing-a-discrete-heart-rate-signal-using-python-part-1/

van Gent, P. (2017). Python Heart Rate Analysis Toolkit. Retrieved from GitHub Repository website: https://github.com/paulvangentcom/heartrate_analysis_python

van Gent, P., Farah, H., van Nes, N., & van Arem, B. (2017). Towards Real-Time, Nonintrusive Estimation of Driver Workload: A Simulator Study. Road Safety and Simulation 2017 Conference Proceedings.

Chapter 5.

Analysing Noisy Driver Physiology Real-Time Using Off-the-Shelf Sensors: Heart Rate Analysis Software

Abstract

For the prediction of driver workload as described in the previous chapter (4), robust heart rate analysis algorithms were required capable of handling noisy PPG data collected 'in the wild'. This chapter describes the development and functioning of HeartPy: a heart rate analysis tookling designed to handle noise photoplethysmogram (PPG) data from low-cost sensors. Most openly available algorithms are designed to handle electrocardiogram (ECG) data, which has different signal properties and morphology, creating a problem when trying to analyse PPG data. These ECG-based algorithms typically do not function well on PPG data, especially not on the more noisy PPG data collected in experimental settings using low-cost devices. To solve this issue, HeartPy was developed to be a noise-resistant algorithm tailored to PPG data. It has been implemented in Python and is available on both GitHub and through Python's package manager (PIP). C-based versions are available for Arduino and other embedded systems as well. This provides researchers with both pc-based and wearable implementations to be used in human factors experiments.

5.1 Introduction

In the field of transportation research one of the main goals is to get to a point where zero traffic fatalities occur (Belin, Tillgren, & Vedung, 2012). The rise of smart in-car systems makes reaching this goal possible. For example, systems exist that automatically take over safety critical tasks of drivers when needed, such as autonomous emergency braking systems. When a driver fails to spot a hazard on the road in front of the vehicle, these systems intervene to avert a collision. However, these are reactions to outside events, whereas another improvement to traffic safety can be made by changing the way drivers and their cars interact. Human error, attentional failures, or driver states that are incongruent with the driving task (fatigue, overload) are a major cause of traffic accidents (Kaplan, Guvensan, Yavuz, & Karalurt, 2015). Sensing when a driver is underloaded, overloaded, distracted or tired can improve safety by enabling dynamic adjustments in the way in-car systems interact with the driver. For example, by timing when navigational or other in-vehicle information systems relay information to the driver, or by adapting the content of their messages to match the current driver state, safety can be improved (van Gent, Farah, van Nes, & van Arem, 2018a).

Human factors research into driver states is an active field. To estimate driver states, physiological measures are often taken together with performance measures (Brookhuis & de Waard, 2010). Heart rate data is collected in many studies, as it is sensitive to changes in workload (Aasman, Mulder, & Mulder, 1987; Bruce Mehler, Reimer, Coughlin, & Dusek, 2010; Stuiver et al., 2012) and general driver state (Danisman, Bilasco, Djeraba, & Ihaddadene, 2010). However, capturing and analysing heart rate in the -often noisy- conditions of either a simulator or an on-road setting can be difficult or costly (Brookhuis & de Waard, 2010). The recent advances in wearable technology and open hardware platforms, such as the Arduino[3] and Raspberry Pi[4], create new possibilities for collecting and analysing physiological data at low cost, given that validated algorithms exist to analyse and process it. In this paper we describe the development of such an algorithm named HeartPy, which we validated as described in (van Gent, Farah, van Nes, & van Arem, 2018b).

5.1.1 Overview of Project Context

Within the 'Taking the Fast Lane' project[5], we are working towards lane-specific advice generation. One possible application is the reduction of congestion by using driver advices to distribute traffic across the available lanes more efficiently. This means advising drivers on where to drive. When interacting with a driver, the timing of messages to the driver is crucial not only for safety but also for the effectiveness of the advices (van Gent, Farah, et al., 2018a). Advising on those moments that workload is low and the driver can accommodate the advice, gives a higher chance of the driver following the advice. However, this means the driver state needs to be known. For this reason, a driver state monitoring system is being developed.

To facilitate the on-line capture and analysis of physiological data, a noise-resistant heart rate collection and analysis toolkit was developed. We created an easy to use, open source analysis toolkit that could handle the collection and analysis of data from available low-cost photoplethysmogram (PPG) sensors, as we could not find an openly available, robust analysis

---

[3] See http://www.arduino.cc
[4] See http://www.raspberrypi.org
[5] See http://tfl.tudelft.nl/

toolkit for this. In this paper we present its development. The toolkit has been used in two simulator studies that modelled driver workload (van Gent, Farah, van Nes, & van Arem, 2017; van Gent, Melman, Farah, van Nes, & van Arem, 2018), as well as in a study looking at the cognitive effects of monitoring automated driving in different conditions (Stapel, Mullakkal-Babu, & Happee, 2017, 2018).

5.1.2   Similar Software

Similar heart rate analysis software exists. These can be divided into commercial and open source variants.

Psychlab and Biopac both offer lab-based solutions including both hardware and software for psychophysiological and medical research. Both offer validated devices and algorithms that have been widely cited. These are, however, not openly available and come at substantial cost.

Kubios offers both a paid software version for HRV analysis, as well as a free version. The free version lacks peak detection functionality and instead requires pre-detected RR-intervals from which to calculate HRV measures.

Physionet (Goldberger et al., 2000) is a large open medical database of Electrocardiogram (ECG) recordings. They also implement WFDB, a software package to retrieve data from their online database and perform waveform analysis. Python bindings are available. WFDB is feature rich, however uses a custom data format and can be technical to implement. It also doesn't handle PPG data well.

HRVAS is a heart rate analysis package for Matlab. It offers many features and is openly available. It, however, still requires Matlab or an older version of its runtime to run, which is not always available. It suffers from the same setback of not handling PPG data well.

The presently discussed heart rate analysis toolkit aims to add to the current body of available software by providing a toolkit for both desktop written in Python, and for (embedded) open hardware platforms written in C. The toolkit focuses on Photoplethysmogram (PPG) recordings but handles ECG data as well.

5.2   Implementation and Architecture

HeartPy has been developed to be sensor-independent, with the use of embedded systems with low computational resources in mind. We have tried to create a fast method of extracting heart beats, that is resistant to types of noise frequently occurring when recording ECG or PPG in field-based studies with low-cost sensors. A Python version is available for PC-based research, as well as limited implementations for several popular Arduino and ARM-based boards that assist in data collection, pre-processing and offer methods of real-time analysis.

5.2.1   Measuring the Heart Rate Signal

Two often used ways of measuring the heart rate are the electrocardiogram (ECG) and the Photoplethysmogram (PPG). The ECG measures the electrical activations that lead to the contraction of the heart muscle, using electrodes attached to the body, usually at the chest. The PPG uses a small optical sensor in conjunction with a light source to measure the discoloration of the skin as blood perfuses through it after each heartbeat.
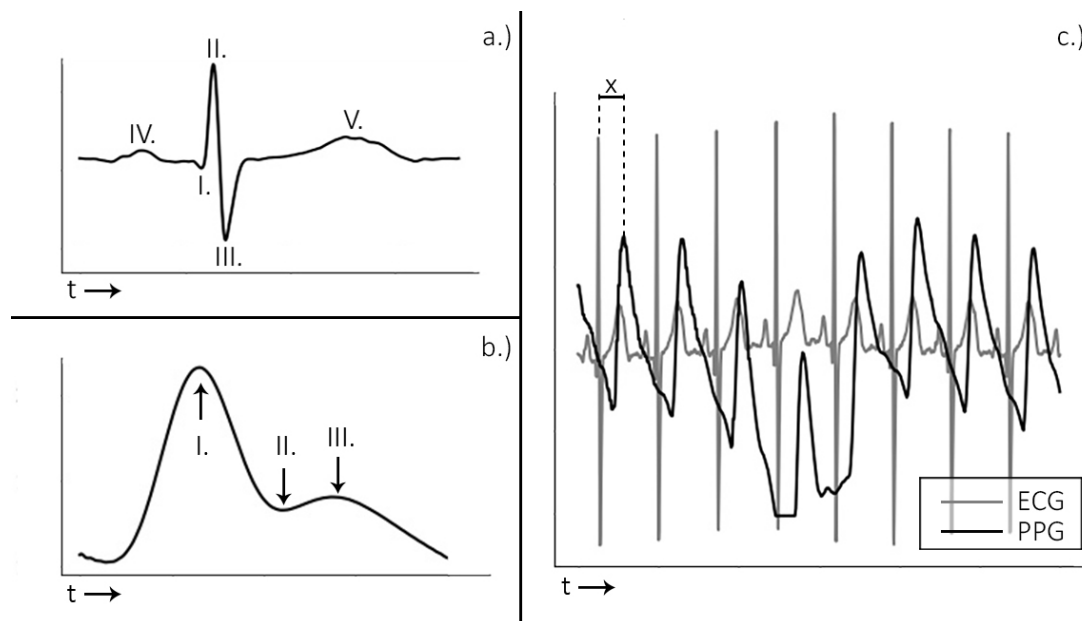
Figure 5.1 – a. and b. display the ECG and PPG waveform morphology, respectively. The ECG is divided into distinct waves (a, I-V), of which the R-wave (a, II) is used for heart beat extraction. With the PPG wave, the systolic peak (b, I) is used. The plot in c. shows the relationship between ECG and PPG signals.

Most notable in the ECG is the QRS-complex (Figure 5.1a, I-III), which represents the electrical activation that leads to the ventricles contracting, expelling blood from the heart muscle. The R-peak is the point of largest amplitude in the signal. When extracting heart beats, these peaks are marked in the ECG. Advantages of the ECG are that it provides a good signal/noise ratio, and the R-peak that is of interest generally has a large amplitude compared to the surrounding data points (Figure 5.1c). The main disadvantage is that the measurement of the ECG is invasive in terms of human factors studies[6]. It requires the attachment of wired electrodes to the chest of the participant, which can interfere with experimental tasks such as driving. This can be undesirable because it can influence participant behaviour, or create potentially dangerous situations for example when driving.

---

[6] Note that the definition of 'invasive' in human factors studies refers to intrusion into the person's privacy, personal space or thoughts. It differs from the medical definition, where 'invasive' indicates that a foreign object intrudes into the body.
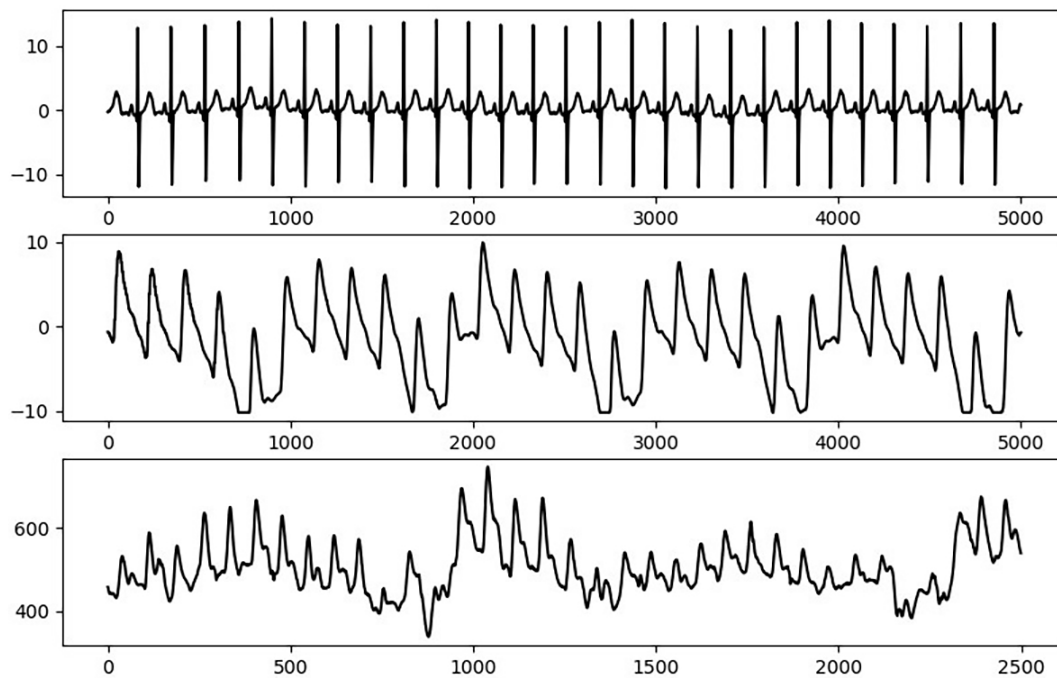
Figure 5.2 – The ECG signal (a.) shows a strong QRS complex together with little amplitude variation. The PPG signal measured simultaneously while the patient is at rest in a hospital bed (b.) shows some amplitude variation but relatively stable morphology. When measuring PPG in a driving simulator using low-cost sensors (c.), strong amplitude and waveform morphology variation is visible.

The PPG measures the discoloration of the skin as blood perfuses through the capillaries and arteries after each heartbeat. The signal consists of the systolic peak (Figure 5.1-b, I), dicrotic notch (II), and the diastolic peak (III). When extracting heart beats, the systolic peaks (I) are used. PPG sensors offer a less invasive way of measuring heart rate data, which is one of their main advantages. Usually the sensors are placed at the fingertip, earlobe, or on the wrist using a bracelet. Contactless camera-based systems have recently been demonstrated (Bousefsaf, Maaoui, & Pruski, 2014; Lewandowska, Ruminsky, Kocejko, & Nowak, 2011; Sun, Hu, Azorin-Peris, Kalawsky, & Greenwald, 2012). These offer non-intrusive ways of acquiring the PPG signal. PPG signals have the disadvantages of showing more noise, large amplitude variations, and the morphology of the peaks displays broader variation (Figure 5.2b, c). This complicates analysis of the signal, especially when using software designed for ECG, which the available open source tools generally are. The toolkit described in this paper aims to provide an efficient means of analysing noisy PPG signals.

5.2.2   Heart Rate and Heart Rate Variability Measures

Analysis of the heart signal is split into heart rate (HR) and heart rate variability (HRV) measures. The heart rate is a simple measure of the heart period, expressed in the beats per minute and the inter-beat interval. Heart rate variability measures describe how the heart rate signal varies over time, and can be divided into time-domain measures and frequency-domain measures (B Mehler, Reimer, & Wang, 2011; Montano et al., 2009).

When extracting heart beats from a signal, a marker is chosen that can reliably be detected at the same position on all heartbeat complexes in the signal. In the ECG the R-peak is often taken (Figure 5.1a-II), in the PPG signal the maximum of the Systolic wave is usually marked (Figure 5.1b-I). Common measures expressing the HR found in the literature are the beats per minute (BPM) and the mean inter-beat interval (IBI). HRV is expressed in the median absolute deviation of intervals between heart beats (MAD), the standard deviation of intervals between heart beats (SDNN), the root mean square of successive differences between neighbouring heart beat intervals (RMSSD), the standard deviation of successive differences between neighbouring heart beat intervals (SDSD), and the proportion of differences between successive heart beats greater than 50ms and 20ms (pNN50, pNN20, resp.). HRV can also be expressed in the frequency domain, where two frequency bands are usually included: low frequency (LF, 0,04-0,15Hz), which is related to short-term blood pressure variation (Bernardi et al., 1994), and high frequency (HF, 0,16-0,5Hz), which is a reflection of breathing rate (Montano et al., 2009).

5.2.3   Analysis Overview

This section describes the architecture of the algorithm and gives an overview of how the heart rate signal is processed and analysed.

5.2.3.1   *Pre-Processing*

The pre-processing options available are peak enhancement, FIR filtering, and outlier detection. The peak enhancement function attempts to normalise the amplitude, then increases R-peak amplitude relative to the rest of the signal. A Butterworth filter implementation is available to remove high frequency noise. Outlier detection on the raw signal is implemented based on a modified Hampel Filter (Davies & Gather, 1993) with a window of half the sampling rate. By default, only the peak enhancement is performed. Details are discussed in the repository's documentation (van Gent, 2018a).

5.2.3.2   *Peak Detection*

The peak detection phase attempts to accommodate amplitude variation and morphology changes of the PPG complexes by using an adaptive peak detection threshold (Figure 5.3, III), followed by outlier detection and rejection. To identify heartbeats, a moving average is calculated using a window of 0.75 seconds on both sides of each data point. The first and last 0.75 seconds of the signal are populated with the signal's mean, no moving average is generated for these sections. Regions of interest (ROI) are marked between two points of intersection where the signal amplitude is larger than the moving average (Figure 5.3, I-II), which is a standard way of detecting peaks. R-peaks are marked at the maximum of each ROI.
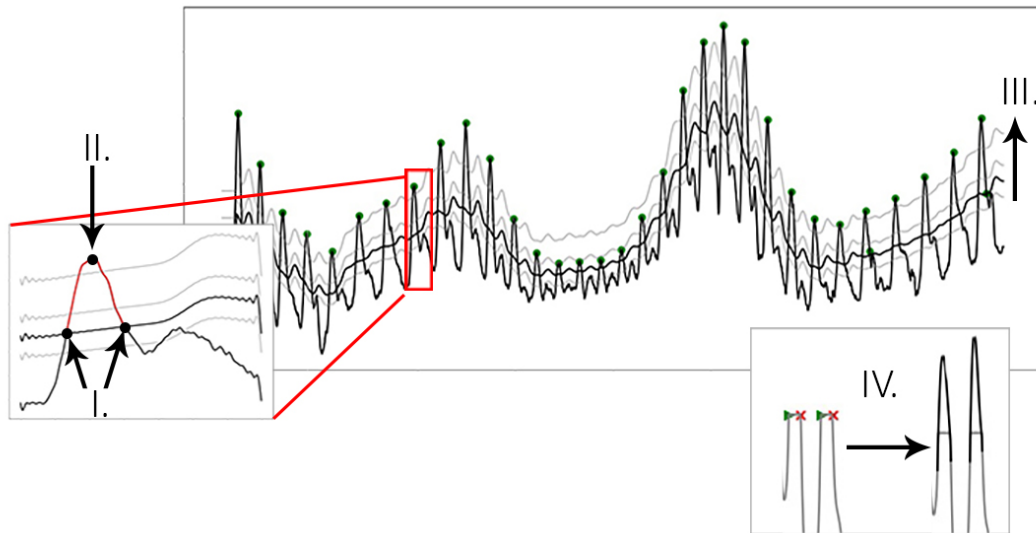
Figure 5.3 - Figure showing the process of peak extraction. A moving average is used as an intersection threshold (II). Candidate peaks are marked at the maximum between intersections (III). The moving average is adjusted stepwise to compensate for varying PPG waveform morphology (I). IV. shows the detection of the onset and end of clipping, and the result after interpolating the clipping segment.

A special case arises when the signal clips, which can happen for example when a sensor has constraints on the range of the signal it can measure, or when digitising an analog signal. The algorithm has clipping detection for R-peaks and will attempt to reconstruct the waveform by spline interpolation whenever an R-peak displays clipping. To interpolate, 100ms of data before clipping onset and 100ms of data after clipping end is used. An example of the process is shown in Figure 5.3-IV.

During the peak detection phase, the amplitude of the calculated threshold is adjusted stepwise. To find the best fit, the standard deviation between successive differences (SDSD, see also 2.2) is minimised. The instantaneous heart rate (BPM) is computed and evaluated in tandem with the SDSD. This represents a fast method of approximating the optimal peak detection threshold by exploiting the relative regularity of the heart rate signal. As shown in Figure 5.4, missing one R-peak (III.) already leads to a substantial increase in SDSD compared to the optimal fit (II.). Marking incorrect R-peaks also leads to an increase in SDSD (I.). The lowest SDSD value that is not zero, in combination with a likely BPM value, is selected as the best fit. The BPM must lie within a predetermined range (default: 40 <= BPM <= 180, range settable by user).
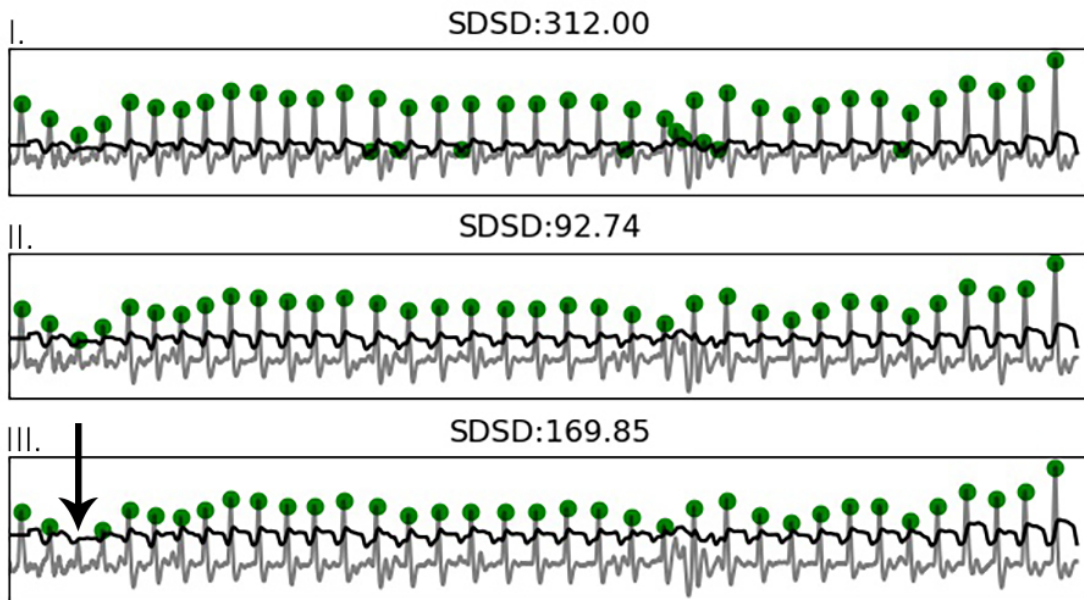
Figure 5.4 - Image showing how the dynamic threshold is fitted using SDSD. The last image (III.) shows that even missing a single beat will lead to a large increase in SDSD compared to the optimal fitting. BPM is also taken into account when fitting.

5.2.3.3   *Error Detection*

Due to the variable PPG waveform morphology, it is possible that after the initial peak fitting phase incorrectly marked R-peaks remain. Motion artefacts may be another cause of detection error. A correction is performed by thresholding the sequence of RR-intervals. R-peaks are considered low confidence if the interval created between two adjacent R-peaks deviates by more than 30% of the mean RR-interval of the analysed segment (Figure 5.5). The threshold is adaptive based on the current segment with a minimum value of 300ms. We've found this to be a good approximation for incorrect detections. If any peaks are considered incorrect detections, the array of RR-values is recomputed to only contain intervals between two high confidence R-peaks.
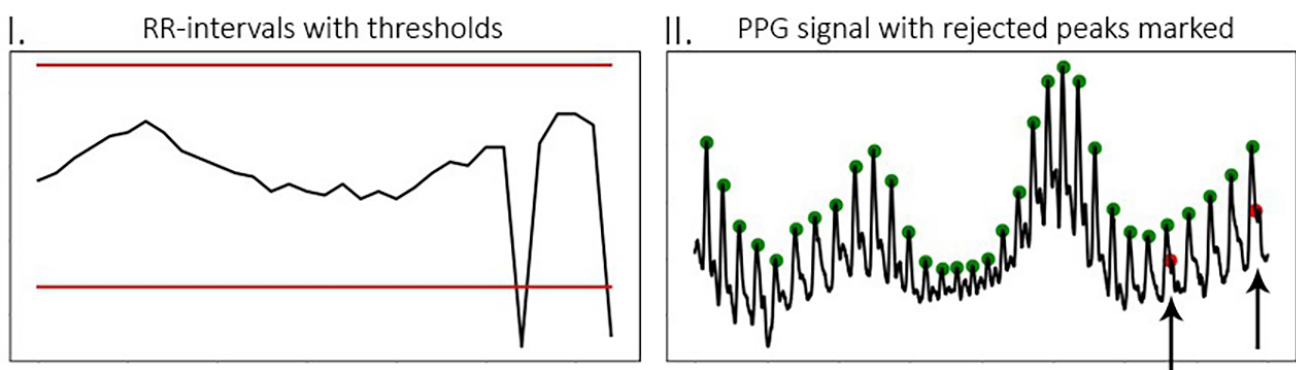
Figure 5.5 - The plotted RR-intervals with thresholds (I.), and the resulting rejected peaks (II.).

An optional error detection pass is available. Using the method, the signal is segmented into n-peak sections and each segment evaluated. Segments are marked low quality if more than a predetermined percentage of peaks are marked low confidence (default n=10, rejection percentage=30%). We found that this pattern of short segments displaying multiple rejected peaks, are often indicative of periods of poor signal/noise ratio or signal loss, such as displayed in Figure 5.6. By eliminating these short periods from the analysis, the output measures remain reliable because only RR-intervals resulting from analysable segments are used in their calculation.

The heart rate analysis package was implemented in both Python and embedded C. The following two sections describe both implementations as well as their requirements, dependencies and availability.
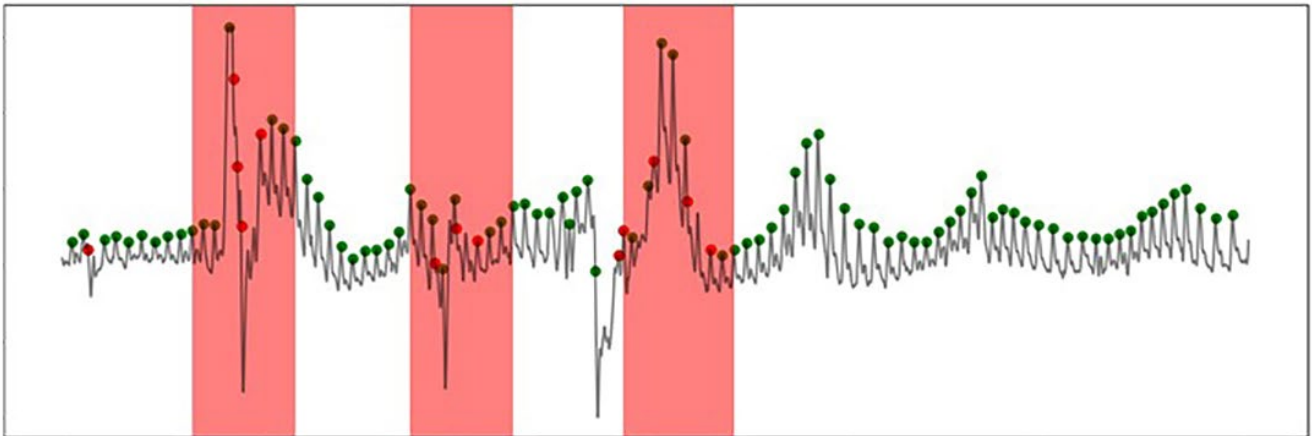
Figure 5.6 – Plot from PPG dataset with low-confidence sections marked. These are ignored in the computation of output measures.

5.2.4   Python Implementation

Python is a flexible programming language that is well suited for scientific use (Oliphant, 2007). During development the reliance on external dependencies was minimised. The package uses the following external packages:

- *Numpy* is used to handle the data, numerical computations, and the Fast Fourier Transform. For these purposes Numpy is much faster than the standard Python interpreter.
- *Scipy* is used for various filtering and interpolation tasks.
- *Matplotlib* is included to plot the results of the analysis if requested by the user.

The implementation of the functions had readability as the main aim. Pep-8 conventions were followed in code styling and function design. A quickstart and background information can be found in the documentation (van Gent, 2018b) and the code together with detailed Jupyter tutorial notebooks can be found on the GitHub (van Gent, 2017).

5.2.5   Embedded Implementation

Several C implementations have been developed to facilitate data collection and analysis in lab-based and field-based studies that utilise wearable technology. Hardware interrupt timers are

used to ensure a precise sampling rate is maintained. Most implementations contain a double switching buffer to collect the sensor data. As one of the buffers fills up, logging switches to the secondary buffer and the content of the first buffer is processed and stored. This ensures logging without interruption.

The repository contains Arduino IDE sketches for several popular boards. Wiring diagrams, and suggested PCB (printed circuit board) design files for various (wearable) applications are in development. The implementations available are briefly discussed below.

5.2.5.1   Data Logger
A data logging application is available. Users can set the desired sampling rate they wish to log. Adaptive input scaling is available (on by default), which attempts to normalise amplitude over time. This is especially useful when measuring at locations where the PPG signal is weaker (wrist, neck), or when measuring it on participants, with reduced perfusion, such as those with advanced age or a history of smoking.

5.2.5.2   Peak Finder
The peak finder implementation analyses the incoming signal real-time for peaks and returns both the peak position and RR-interval created between the current and the previous detected peak. Error detection based on the last 20 RR-intervals, as well as based on various settable parameters is available. See the documentation for more details (van Gent, 2018b).

5.2.5.3   Time Series Analysis
The time series analysis implementation is similar to the peak finder implementation, except that it calculates and outputs the time-series measurements of both heart rate and heart rate variability. It tracks detected peaks in time to computer RR-intervals and ignores intervals when there is a missing or rejected peak in between.

5.2.5.4   Full Implementation
The full implementation contains the HR and HRV online analysis. All the HR and HRV measures mentioned under 5.2.2 are derived from the signal and stored to an on-board SD card, together with the original signal. Since a full signal period is first collected, several pre-processing steps can be taken to improve signal quality prior to analysis. This makes the full implementation the most noise-resistant of the available versions. The memory and processing requirements, however, are also higher than of the other versions. This makes it less suited for long-term wearable solutions required in naturalistic studies, but very suited for environments where power is available (in-car, driving simulator, lab-based studies, bicycle with power bank) or situations of shorter measurement periods.

5.3   Quality Control

5.3.1   General Quality Control

The code development was centred around ease-of-use and reusability of functions and methods. Coding best practices were followed (Wilson et al., 2014). Throughout the development process, cyclomatic complexity (cc) was frequently calculated for all functions using the python Radon package and the Lizard[7] package. Refactoring was applied for functions

---

[7] See https://github.com/terryyin/lizard

that had a cc of over 10 [8], to ensure maintainability and readability of the code. Git was used for version control (Loeliger & McCullough, 2012) throughout the project.

Means for automatic source code validation and automated testing have been implemented. In the Python implementation, examples are available in the docstrings that double as doctests. Automated continuous integration (CI) testing is implemented through the Travis-CI platform. Code coverage, build status and supported Python versions are displayed as badges on the GitHub repository.

Several end-to-end examples are included in Jupyter notebooks on the repository, detailing how to handle various types of signals with HeartPy. Available examples deal with both good and poor quality PPG and ECG signals from various sources (sensors, electrodes, smartwatch, smart ring). The examples are designed to familiarise new users with the functionality of the package and to highlight possible use cases.

A tutorial series is available (van Gent, 2016), detailing the basics behind the Python implementation of the algorithm. Users seeking deeper understanding in the mechanics behind the algorithm can follow these.

5.3.2   Validation

HeartPy was validated on a dataset collected by PPG sensor from a previous experiment (van Gent, Melman, et al., 2018). Heart beats in the dataset were manually annotated to serve as a ground truth. The validation was performed on the set and compared to two popular available open source algorithms. Error rates showed superior performance of HeartPy on the noisy PPG data in the test set. The full validation is described in (van Gent, Farah, et al., 2018b).

5.4   Availability

5.4.1   Operating System

HeartPy has been tested to run on Python 2.7, 3.4, 3.5, 3.6 and 3.7 and above. All updates are automatically built using Travis-CI and tested. Results are dynamically displayed on the repository as badges.

Several Arduino IDE sketch files have been provided as well for different boards. These have been tested on their respective boards, and developed in the Arduino IDE version 1.8.5. They are designed to enable researchers low-cost ways of collecting heart rate data as well.

5.4.2   Additional System Requirements

The Python implementation's data handling happens in Numpy, which ensures efficient RAM usage and fast execution. The size of the input data will determine memory usage, although the requirements are low for most datasets. As an example, we loaded the first data file from the first participant in the PPG validation set included on the GitHub ('pp1_Eind_Som_C.csv'). The file represents 10:52 minutes of data sampled at 100Hz. The loaded data takes 261Kb in memory. Temporary containers created during analysis and (pre-)processing take up an additional 593Kb. This indicates very low resource requirements for most analyses.

---

[8] See http://radon.readthedocs.io/en/latest/api.html#radon.complexity.cc_rank

Requirements of the embedded hardware conform to RAM and CPU resources available on the SOC's for which the implementation has been designed. More information and absolute values are available in the documentation (van Gent, 2018a).

5.4.3   Dependencies

HeartPy is dependent on the Numpy, SciPy and Matplotlib packages. The lowest versions we've tested with HeartPy are Numpy==1.15, Scipy==1.1.0, Matplotlib==2.2.3. These versions allow functionality on Python 2.7.

The Arduino implementations depend on standard modules available in the Arduino IDE. The versions for Arduino and Teensy boards depends on the SDfat module for communication with the SD card for data storage. This module is installed in the Arduino IDE by default.

5.4.4   List of Contributors

Jonathan de Bruin has provided valuable advice and suggestions during development and testing of HeartPy and will remain active in further development.

5.4.5   Software Location of Python version

Name: Zenodo.org
　　　Persistent identifier: https://zenodo.org/badge/latestdoi/91584229
　　　Licence: GNU General Public License V3.0
　　　Publisher: Paul van Gent
　　　Version published: V1.0.0
　　　Date published: 31-07-2018

Code repository: GitHub
　　　Name: Python Heart Rate Analysis Toolkit
　　　Identifier: https://github.com/paulvangentcom/heartrate_analysis_python
　　　Licence: GNU General Public License V3.0
　　　Date published: 31-07-2018

5.4.6   Software Location of Embedded Version

Code repository: GitHub
　　　Name: Arduino Heart Rate Analysis Toolkit
　　　Identifier: https://github.com/paulvangentcom/heartrate_analysis_Arduino
　　　Licence: GNU General Public License V3.0
　　　Date published: 31-07-2018

5.5   Reuse Potential

HeartPy can be used researchers, makers, and engineers to create applications that make use of (real-time) heart rate data. The toolkit can be used in research settings both in the lab and 'in the wild'. HeartPy handles noise that is typically introduced into heart rate signals when recording outside the lab well, and contains many pre-processing options to help clean up poor quality signals. The software has been used in the past in for example lab-based simulator

contexts (van Gent et al., 2017; van Gent, Melman, et al., 2018), real world driving contexts (Stapel et al., 2017, 2018), and as a backend for a pregnancy monitoring tool (Gupta, Kumar, & Mago, 2019).

Detailed examples are available on the repository and in the documentation on handling different data types that serve to kick-start any new project based on HeartPy. These examples are available on the repository as Jupyter notebooks (https://github.com/paulvangentcom/heartrate_analysis_python/tree/master/examples).   The examples cover how to analyse PPG signals from sensors, smartwatches and smart rings (and similar devices), as well as ECG signals ranging from good to very poor quality.

HeartPy is designed to easily be integrated into existing projects. All methods are documented separately, and most can be used in isolation as well. Throughout the processing pipeline everything of interest is stored in a dict{} object, which can be accessed each step of the analysis. This facilitates integration with other projects by allowing a fine level of control over each step.

We are currently working on incorporating a GUI for use with HeartPy, which will expand the reuse potential further towards researchers without coding experience.

Funding Statement

References

Aasman, J., Mulder, G., & Mulder, L. J. M. (1987). Operator effort and the measurement of heart-rate variability. Human Factors, 29(2), 161–170. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-0023330102&partnerID=tZOtx3y1

Belin, M.-Å., Tillgren, P., & Vedung, E. (2012). Vision Zero – a road safety policy innovation. International Journal of Injury Control and Safety Promotion, 19(2), 171–179. https://doi.org/10.1080/17457300.2011.635213

Bernardi, L., Leuzzi, S., Radaelli, a, Passino, C., Johnston, J. a, & Sleight, P. (1994). Low-frequency spontaneous fluctuations of R-R interval and blood pressure in conscious humans: a baroreceptor or central phenomenon? Clinical Science (London, England : 1979), 87(6), 649–654.

Bousefsaf, F., Maaoui, C., & Pruski, a. (2014). Remote detection of mental workload changes using cardiac parameters assessed with a low-cost webcam. Computers in Biology and Medicine, 53, 1–10. https://doi.org/10.1016/j.compbiomed.2014.07.014

Brookhuis, K. A., & de Waard, D. (2010). Monitoring drivers' mental workload in driving simulators using physiological measures. Accident Analysis and Prevention, 42(3), 898–903. https://doi.org/10.1016/j.aap.2009.06.001

Danisman, T., Bilasco, I. M., Djeraba, C., & Ihaddadene, N. (2010). Drowsy driver detection system using eye blink patterns. 2010 International Conference on Machine and Web Intelligence, ICMWI 2010 - Proceedings, 230–233. https://doi.org/10.1109/ICMWI.2010.5648121

Davies, L., & Gather, U. (1993). The identification of multiple outliers. Journal of the American Statistical Association, 88(429), 782–792.

Goldberger, A. L., Amaral, L. A. N., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., … Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet - Components of a New Research Resource for Complex Physiologic Signals.

Gupta, Y., Kumar, S., & Mago, V. (2019). Pregnancy Health Monitoring System based on Biosignal Analysis. 2019 42nd International Conference on Telecommunications and Signal Processing (TSP), 664–667. https://doi.org/10.1109/tsp.2019.8769074

Kaplan, S., Guvensan, M. A., Yavuz, A. G., & Karalurt, Y. (2015). Driver Behavior Analysis for Safe Driving: A Survey. IEEE Transactions on Intelligent Transportation Systems, 16(6), 3017–3032. https://doi.org/10.1109/TITS.2015.2462084

Lewandowska, M., Ruminsky, J., Kocejko, T., & Nowak, J. (2011). Measuring Pulse Rate with a Webcam - a Non-contact Method for Evaluating Cardiac Activity. Proceedings of the Federated Conference on Computer Science and Information Systems, (January), 405–410.

Loeliger, J., & McCullough, M. (2012). Version Control with Git: Powerful Tools and Techniques for Collaborative Software Development.

Mehler, B, Reimer, B., & Wang, Y. (2011). Comparison of heart rate and heart rate variability indices in distinguishing single task driving and driving under secondary cognitive workload. Proceedings of the Sixth International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design, 590–597.

Mehler, Bruce, Reimer, B., Coughlin, J. F., & Dusek, J. a. (2010). Impact of Incremental Increases in Cognitive Workload on Physiological Arousal and Performance in Young Adult

Drivers. Transportation Research Record: Journal of the Transportation Research Board, 2138(1), 6–12. https://doi.org/10.3141/2138-02

Montano, N., Porta, A., Cogliati, C., Costantino, G., Tobaldini, E., Casali, K. R., & Iellamo, F. (2009). Heart rate variability explored in the frequency domain: A tool to investigate the link between heart and behavior. Neuroscience and Biobehavioral Reviews, 33(2), 71–80. https://doi.org/10.1016/j.neubiorev.2008.07.006

Oliphant, T. E. (2007). Python for scientific computing. Computing in Science and Engineering, 9(3), 10–20. https://doi.org/10.1109/MCSE.2007.58

Stapel, J., Mullakkal-Babu, F. A., & Happee, R. (2017). Driver Behaviour and Workload in an On-road Automated Vehicle. Proceedings of the RSS2017 Conference.

Stapel, J., Mullakkal-Babu, F. A., & Happee, R. (2018). Attentive monitoring of automated driving requires more effort than manual driving. Manuscript Submitted for Publication.

Stuiver, A., de Waard, D., Brookhuis, K. A., Dijksterhuis, C., Lewis-Evans, B., & Mulder, L. J. M. (2012). Short-term cardiovascular responses to changing task demands. International Journal of Psychophysiology : Official Journal of the International Organization of Psychophysiology, 85(2), 153–160. https://doi.org/10.1016/j.ijpsycho.2012.06.003

Sun, Y., Hu, S., Azorin-Peris, V., Kalawsky, R., & Greenwald, S. (2012). Noncontact imaging photoplethysmography to effectively access pulse rate variability. Journal of Biomedical Optics, 18(6), 061205. https://doi.org/10.1117/1.JBO.18.6.061205

van Gent, P. (2016). Analyzing a Discrete Heart Rate Signal Using Python. Retrieved from http://www.paulvangent.com/2016/03/15/analyzing-a-discrete-heart-rate-signal-using-python-part-1/

van Gent, P. (2017). Python Heart Rate Analysis Toolkit. Retrieved from GitHub Repository website: https://github.com/paulvangentcom/heartrate_analysis_python

van Gent, P. (2018a). Embedded Heart Rate Analysis Toolkit Documentation. Retrieved from https://embedded-heart-rate-analysis-toolkit.readthedocs.io

van Gent, P. (2018b). Python Heart Rate Analysis Toolkit Documentation. Retrieved from https://python-heart-rate-analysis-toolkit.readthedocs.io

van Gent, P., Farah, H., van Nes, N., & van Arem, B. (2017). Towards Real-Time, Nonintrusive Estimation of Driver Workload: A Simulator Study. Road Safety and Simulation 2017 Conference Proceedings.

van Gent, P., Farah, H., van Nes, N., & van Arem, B. (2018a). A Conceptual Model for Persuasive In-Vehicle Technology to Influence Tactical Level Driver Behavior. Manuscript Submitted for Publication.

van Gent, P., Farah, H., van Nes, N., & van Arem, B. (2018b). Heart Rate Analysis for Human Factors: Development and Validation of an Open Source Toolkit for Noisy Naturalistic Heart Rate Data. Proceedings of the 6th HUMANIST Conference, 173–178.

van Gent, P., Melman, T., Farah, H., van Nes, N., & van Arem, B. (2018). Multi-Level Driver Workload Prediction Using Machine Learning and Off-The-Shelf Sensors. Transportation Research Record: Journal of the Transportation Research Board. https://doi.org/10.1177/0361198118790372

Wilson, G., Aruliah, D. A., Brown, C. T., Chue Hong, N. P., Davis, M., Guy, R. T., … Wilson, P. (2014). Best Practices for Scientific Computing. PLoS Biology, 12(1). https://doi.org/10.1371/journal.pbio.1001745

Chapter 6.

HeartPy: A Novel Heart Rate Algorithm for the Analysis of Noisy Signals

Abstract

This chapter presents the validation of HeartPy, the development of which was described in the previous chapter (5). Heart rate data are often collected in human factors studies, including those studies into vehicle automation. Advances in open hardware platforms and off-the-shelf photoplethysmogram (PPG) sensors allow for the non-intrusive collection of heart rate data at very low cost. However, the PPG signal from these studies is often not trivial to analyse, as the PPG signal has different morphology and noise characteristics when compared to the often used but more intrusive electrocardiogram (ECG) signals, and the use of low-cost sensors often introduces extra noise into the signal. Few validated open source algorithms exist that can handle noisy PPG data well, as most available algorithms are designed for ECG data. We benchmark the performance on two types of datasets and show that HeartPy performs well.

6.1   Introduction

Vehicle automation is rapidly gaining popularity in the agendas of the automotive sector and governments. Automation promises to increase traffic flow efficiency (Hoogendoorn, van Arem, & Hoogendoorn, 2014) and free up the time of the driver for other activities. However, (semi-)autonomous vehicles below SAE level 5 will still need to interact with the driver, for example for a transition of control, or in emergency situations when the automation fails. This means that the vehicle needs to be aware of the driver's state (e.g. distraction, fatigue), because a transition of control can be dangerous when the driver is not able to take over control of the vehicle (Merat, Jamson, Lai, Daly, & Carsten, 2014), because of for example high workload or distraction. Additionally, In-Vehicle Information Systems (IVIS) that interact with the driver to provide information or advices can also benefit from knowledge about the driver's state to choose the most appropriate interface (Birrel, Young, Stanton, & Jennings, 2017; Park & Kim, 2015), content and timing of the information (van Gent, Farah, van Nes, & van Arem, 2018a).

Algorithms are being developed that can estimate the driver's state, whether this is driver workload (Solovey, Zec, Garcia Perez, Reimer, & Mehler, 2014; van Gent, Melman, Farah, van Nes, & van Arem, 2018), driver distraction (Liang, Reyes, & Lee, 2007) or a driver's interruptibility (S. Kim, Chun, & Dey, 2015). Heart rate is frequently included as an input for predicting a driver's state since it contains information about changes in (driver) workload (Mehler et al., 2012; Mehler et al., 2010), stress (Healey & Picard, 2005), and general driver state such as drowsiness (Danisman, Bilasco, Djeraba, & Ihaddadene, 2010). In addition to the benefits for autonomous syshiring ffefefefetems, many human factors studies focusing on the interaction between the driver and (semi-)autonomous vehicles also include heart rate measurements (Jamson, Merat, Carsten, & Lai, 2011; Reimer, Mehler, & Coughlin, 2016; Reimer, Mehler, Coughlin, Roy, & Dusek, 2011; Stapel, Mullakkal-Babu, & Happee, 2017).

However, capturing heart rate in the often noisy conditions of either a driving simulator or in an in-vehicle setting, and subsequently analysing the complex signals either real-time or offline, can be difficult or costly (Brookhuis & de Waard, 2010). Low-cost commercial devices are available, but these are generally designed for sporting contexts and not specifically for scientific research. Furthermore, the proprietary nature of the firmware and software used in these devices creates problems with data reliability, reproducibility of results, and integration into in-vehicle hardware for the purpose of real-time driver monitoring. This reduces the usefulness of these devices for research into (partially) self-driving vehicles and makes it nearly impossible to integrate them into actual in-vehicle systems.

One potential solution lies in the recent advances in wearable technology and open hardware platforms, such as Arduino[9] and Raspberry Pi[10]. They are open source in both hardware and software design, meaning that integration into existing systems is feasible. There is, however, a lack of open source available heart rate analysis algorithms that are validated, easy to use and able to handle noisy data from low-cost PPG sensors. Implementations of heart rate analysis algorithms described in research papers are often not available, poorly documented, or require substantial technical expertise to implement properly.

---

[9] See http://www.arduino.cc
[10] See http://www.raspberrypi.org

In a previous study we collected heart rate data with low-cost sensors to develop an affordable driver workload estimation approach (van Gent, Farah, van Nes, & van Arem, 2017; van Gent, Melman, et al., 2018). Available open source algorithms did not function well on this type of - often noisy- data and couldn't easily be integrated into a real-time system. To overcome this issue, our aim is to develop a novel algorithm that (i) functions better on this type of noisy data, and (ii) provides an easy-to-use analysis method for the collected heart rate data both, offline and real-time. We've named the developed algorithm HeartPy. For a technical overview of HeartPy, its development and its availability, please see (van Gent, Farah, van Nes, & van Arem, 2018b). The main aim of this paper is to describe the validation of HeartPy using two datasets: a noisy dataset collected in a driving simulator (van Gent et al., 2017), and an openly available medical dataset (Jager et al., 2003).

In the rest of this paper, we first describe basic properties of the heart rate signal as they relate to data collection and analysis. This is followed by a brief overview of the algorithm's functioning, discussion of our methods, results and concluding remarks.

6.1.1   Measuring Heart Rate in Naturalistic or Simulated Settings

There are two major approaches to measuring heart rate, which mainly differ in the physiological properties they measure.
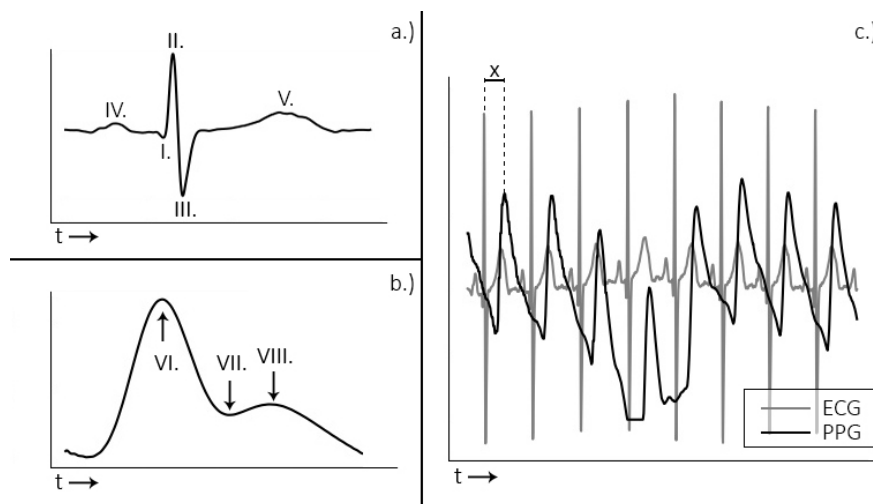
Figure 6.1 - The differences in morphology of the ECG wave (a) and PPG wave (b), and the time lag 'x' between both waves (c). The ECG (a) wave consists of most notably the Q-R-S complex (I-III). The P (IV) and T (V) waves are also marked in the plot. The PPG (b) wave consists of the systolic peak (VI), the diastolic peak (VIII) and the dichrotic notch (VII).

Electrocardiogram recordings (ECG) are collected by placing electrodes on the chest near the heart. These electrodes measure the electrical activation of the heart during each cardiac cycle. The defining feature in the ECG signal is the QRS complex (Figure 6.1a I-III). Advantages of the ECG signal are that it directly measures the heart's electrical activation and that it presents a strong QRS complex presence in the resulting signal (Figure 6.1 a). A common source of noise in ECG signals are motion artefacts resulting from sensor displacement due to participant movement. These tend to fall in the same frequency range as the QRS-complexes, which can make it difficult to filter them without deforming the QRS complex (Kirst, Glauner, & Ottenbacher, 2011). In traffic related studies, ECG recordings have been used in for example

(Brouwer & Dijksterhuis, 2015; Fallahi, Motamedzade, Heidarimoghadam, Soltanian, & Miyake, 2016; Farah et al., 2012; Miyaji, Danno, Kawanaka, & Oguri, 2008). The main disadvantage is that ECG is not easily measured unintrusively.

Photoplethysmogram (PPG) recordings offer a less invasive method of assessing the cardiac cycle. These devices employ an optical sensor to measure the changes in coloration of the skin as blood perfuses through the arteries and capillaries with each heartbeat. PPG is typically measured at the fingertip or through wrist bracelets. The PPG signal consists of a systolic peak (Figure 6.1b-VI), a dicrotic notch (6.1b-VII), and a secondary peak called a diastolic peak (6.1b-VIII). The secondary peak may be absent in some recordings or of very low amplitude. Advantages of the PPG method are that it is low cost, easy to set up, and non-invasive (Elgendi, 2012; Millasseau et al., 2000). Ways of obtaining the PPG signal contactless through cameras have been demonstrated, further reducing intrusiveness (Sun, Hu, Azorin-Peris, Kalawsky, & Greenwald, 2012). However, PPG tends to display more amplitude variation over short time-intervals (Figure 6.1c), more variation in waveform morphology, as well as contain more noise from various sources when compared to ECG measurements. This makes analysis more difficult. In the traffic domain, PPG sensors have been used by for example (Jarvis, Putze, Heger, & Schultz, 2011; van Gent, Melman, et al., 2018; Zhai & Barreto, 2006).

6.1.2   Analysing Heart Rate Data

The heart signal is often split into heart rate (HR) and heart rate variability (HRV) measures. Heart rate is a simple measure of the heart period, expressed in the beats per minute and the inter-beat interval. Heart rate variability measures describe how the heart rate signal varies over time, and can be divided into time-domain measures and frequency-domain measures (B Mehler, Reimer, & Wang, 2011; Montano et al., 2009).

When extracting heart beats from a signal, a marker is chosen that can reliably be detected at the same position on all heartbeat complexes in the signal. In the ECG the R-peak is often taken (Fig 1a-II), in the PPG signal the maximum of the Systolic wave is usually marked (Fig 6.1b-I).

Common measures expressing the HR found in the literature are the beats per minute (BPM) and the mean inter-beat interval (IBI). HRV is expressed by the median absolute deviation of intervals between heart beats (MAD), the standard deviation of intervals between heart beats (SDNN), the root mean square of successive differences between neighbouring heart beat intervals (RMSSD), the standard deviation of successive differences between neighbouring heart beat intervals (SDSD), and the proportion of differences between successive heart beats greater than 50ms and 20ms (pNN50, pNN20, resp.). HRV can also be expressed in the frequency domain, where two frequency bands are usually included: low frequency (LF, 0,04-0,15Hz), which is related to short-term blood pressure variation (Bernardi et al., 1994), and high frequency (HF, 0,16-0,5Hz), which is a reflection of breathing rate (Montano et al., 2009).

Despite the different underlying physiological constructs that are measured, a high correlation (median 0.97) between peak-peak intervals extracted from ECG and PPG signals has been reported (Selvaraj, Jaryal, Santhosh, Deepak, & Anand, 2008). This makes the PPG a valid alternative for applications that require non-intrusive heart rate measurements, given that validated analysis algorithms exist.

6.1.3   Development and Availability of HeartPy

We developed HeartPy to help analyse noisy heart rate data collected in driving settings (both simulated and on-road). The algorithm runs on desktop computers (Python) as well as wearables (embedded C) such as Arduino and Teensy boards, both offline and in real-time. The latter allows for real-time heart rate analysis in in-car settings as well as other mobile situations, such as with cyclists or pedestrians. The algorithm is available as the Python package 'HeartPy', hosted on GitHub (van Gent, 2017) and is installable through Python's 'pip' package manager as well. Documentation is available through the GitHub page. The wearable embedded C version is available on GitHub as well (van Gent, 2018), together with documentation linked there.

HeartPy was designed to be resistant to typical noise patterns (e.g. motion artefacts, momentary signal loss) of participants engaged in other tasks (driving simulator, on-road car experiment, bike experiment), to be capable of handling signals from low-cost off-the-shelf sensors, and to be user friendly.

6.1.4   Overview of the HeartPy Algorithm

HeartPy comes with various pre-processing options to clean up signals, including FIR filtering and outlier detection. This section briefly outlines the peak detection methods. Please refer to van Gent et al., (2018b) for more information on the software, its availability and its functioning.
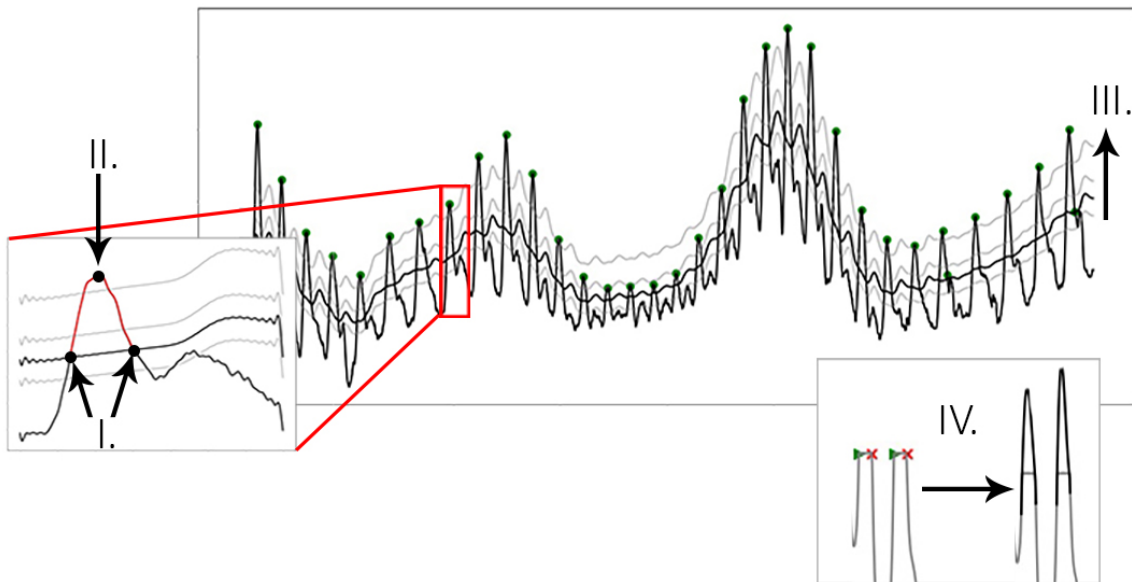
Figure 6.2 – Figure showing the process of peak extraction. A moving average is used as an intersection threshold (I). Candidate peaks are marked at the maximum between intersections (II), with optional spline interpolation available to improve position accuracy. The moving average is raised stepwise (III). IV. shows the detection of the onset and end of clipping, and the result after interpolating the clipping segment.

Peak detection uses an adaptive threshold (Figure 6.2, III) to accommodate for morphology and amplitude variation in the PPG waveform, followed by outlier detection and rejection. To identify heartbeats, a moving average is calculated using a window of 0.75 seconds on both

sides of each data point. Regions of interest (ROI) are computed between two points of intersection where the signal amplitude is larger than the moving average (Figure 6.2, I-II), which is a standard way of detecting peaks. Two methods of obtaining a peak's location are included. In the first approach, the peak position is simply taken to be the highest point in the marked ROI. Although this is a computationally low-cost operation, its accuracy depends on the sampling rate used, with a higher sampling rate resulting in more accurate results. In the second method a univariate spline is used to upsample and interpolate the ROI, which is then solved for its maximum. This requires more computation but is also more accurate, especially with lower sampling rates. Both methods are available in HeartPy, by default the fast method is used.

Signal clipping is a special case that hinders the accurate placement of a peak's position. Clipping can occur for various reasons, for example when digitising an analog signal. HeartPy detects the onset and end of clipping segments, and will attempt to reconstruct the waveform by spline interpolation, as shown in Figure 6.2-IV.

During the peak detection phase, the amplitude of the calculated threshold is adjusted stepwise (Figure 6.2-III). The best fit is determined by minimising the standard deviation of peak-peak intervals (SDNN, see also 2.2). The instantaneous heart rate (BPM) is computed and evaluated in tandem with the SDNN. This represents a fast method of approximating the optimal threshold amplitude by using the (relative) regularity of the heart rate signal. As shown in Figure 6.3, missing one peak (III.) already leads to a substantial increase in SDNN compared to the optimal fit (II.). Marking incorrect peak positions also leads to an increase in SDNN (I.). The lowest SDNN value that is not zero, in combination with a reasonable BPM value, is selected. The BPM must lie within a predetermined range (default: 40 <= BPM <= 180, range settable by user).
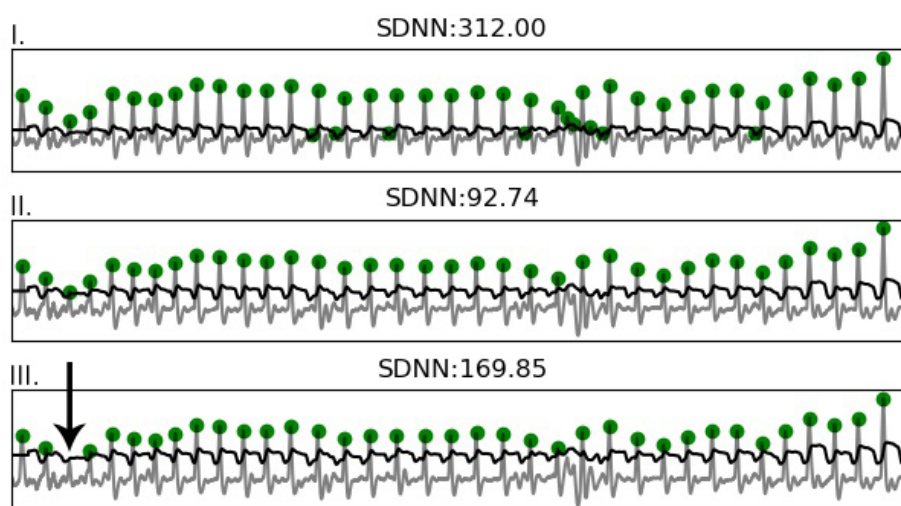
Figure 6.3 – Figure showing the effects of missing or additional peaks on the SDNN metric. The last image (III.) shows that missing a single beat will already lead to a large increase in SDNN compared to the optimal fitting. BPM is also considered when fitting.

Due to the variable PPG waveform morphology, it is possible that after the initial peak fitting phase incorrectly marked peaks remain. Motion artefacts may be another cause of detection error. A correction is performed by thresholding the sequence of peak-peak intervals. Peaks are considered low confidence if the interval created between two adjacent peaks deviates by more than 30% of the mean peak-peak interval of the analysed segment (Figure 6.4). The threshold

is adaptive based on the current segment with a minimum value of 300ms. We've found this to be a good approximation for incorrect detections. If any peaks are considered incorrect detections, the array of peak-peak intervals is recomputed to only contain intervals between two high confidence peak positions.
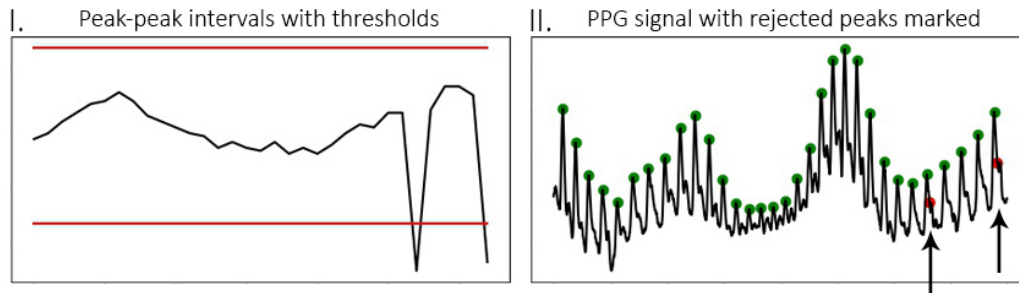
Figure 6.4 – The plotted peak-peak intervals with thresholds (I.), and the resulting rejected peaks (II.)

6.2   Methods

The algorithm was validated using two datasets from two different experiments and research domains. The first dataset used was collected with a low-cost PPG sensor in a driving simulator experiment (van Gent, Melman, et al., 2018). This dataset contains approximately 20.7 hours of PPG recordings. The second dataset is the openly available Long-Term ST Database (Jager et al., 2003), containing 86 long ECG recordings of 80 participants, with each recording being between 21 and 24 hours.

The PPG dataset was used in its entirety and split into one-minute segments. Because the ECG dataset was so large, 1.000 one-minute segments were randomly selected from the database. The peak positions in all the segments from both datasets were annotated manually and checked a second time to serve as a ground truth. These annotations are also available on the GitHub page (van Gent, 2017). For both data sets a one-minute length for the segments was used to balance both the number of peaks in each segment with the time needed to manually annotate all segments. The algorithm performance was compared to the annotated data on four variables: detected peak position, mean of the peak-peak intervals calculated over the analysed segment, beats per minute computed by the algorithm (a HR measure), and a common heart rate variability (HRV) measure: the standard deviation of successive differences (SDSD). To quantify the accuracy of the algorithm's predictions, we used the Root Mean Squared Error (RMSE), defined as:

$$Eq.1 \quad RMSE = \sqrt{\frac{\Sigma(y - \hat{y})^2}{n}}$$

Where y is the ground truth value, yˆis the value predicted by the algorithm, and n the number of comparisons. For the accuracy of the absolute peak-positions in time as compared to the annotated ground truth, we used the mean of the absolute deviations.

6.2.1   Error Types
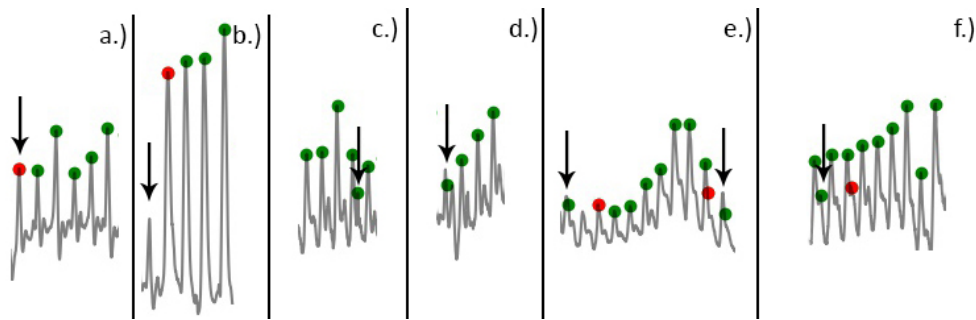
Figure 6.5 – Figure displaying the possible errors. These are: a.) 'incorrectly rejected', b.) 'missed', c.) 'incorrectly accepted'. Peaks marked on a correct QRS complex but not on its peak maximum, are also counted as 'incorrectly accepted'. This type of error is shown in d.). Other possible mistakes counted as 'incorrectly accepted' are marking a peak at a non-maximum position (e), or incorrectly marking a diastolic peak (f).

In addition to the performance comparison, the results of the one-minute segments were plotted and three types of errors annotated (shown in Figure 6.5): 'Incorrectly rejected' – meaning that a correct peak has been marked as low confidence (Figure 6.5a). 'Missed' – indicating that a peak is present but not marked (Figure 6.5b). 'Incorrectly accepted' – indicating a peak is marked where no peak is considered present by the human annotator (Figure 6.5c). Figure 6.5d, e and f show other examples that were classified as 'incorrectly accepted': cases where a peak was marked at a non-maximum position, or where a diastolic peak was marked instead of a systolic peak. The algorithm has been designed to minimise the 'incorrectly accepted' error type for reasons discussed in the next section.

6.2.2   Minimising the Correct Error Type

The algorithm was designed to minimise the 'incorrectly accepted' peak error types because this error type has the strongest effect on calculated output measures. This section illustrates why the choice was made.

Heart rate variability (HRV) measures are not robust against outliers. Marking a peak on an anomalous position affects these measures since they express the variation in the intervals between peak positions. Marking a peak at an incorrect time position creates a deviation in the length of the surrounding intervals which will strongly influence the variance in the sample. Note that the heart rate (HR) measures such as IBI and BPM are quite resistant to outliers because they use the mean of all peak-to-peak intervals in a given signal segment.

To further explain and show these effects, a bootstrapped simulation was performed. We took a manually annotated one-minute segment of PPG heart rate data and artificially introduced two types of errors:

1. "Incorrectly rejected" peaks were simulated by dropping a random n% of peaks from the signal. Measures were then calculated on intervals between peaks where no missing value occurred in between, mimicking the algorithm's behaviour of only computing peak-peak intervals between two accepted peaks.

2. "Incorrectly accepted" peaks were simulated by introducing a position error into a random n% of all peaks. The error disturbed the peak position randomly by 0.1% - 10%, meaning a random positional disturbance of between 1ms and 100ms. For each selected peak the disturbance magnitude was randomised.

Simulations were run with values of 5%, 10% and 20% for 'n'. Each simulation run was bootstrapped for 10.000 iterations to reduce the effects of the random selection process.

Results show that the effect of incorrect beat detections (displacement scenario) is significantly stronger than the effect of missing values, especially the effect on HRV measures. The effect on BPM is negligible in both the missing and displacement simulations, showing the HR measures' resilience to outliers. Effects on HRV measures are substantial. The data are displayed in table 6.1 below, and the analysis notebook is available on the Python GitHub (van Gent, 2017).

Table 6.1 –The absolute errors introduced in the simulations. The errors introduced in the HRV measures by displaced peaks are particularly evident (marked grey).

	Errors introduced in simulated scenarios (n=10.000)					
	Missing n% of peaks			Displacing n% of peaks		
	n=5%	n=10%	n=20%	n=5%	n=10%	n=20%
BPM	0.001	0.007	0.062	0.000	0.000	0.000
IBI	0.012	0.067	0.541	0.002	0.004	0.003
SDNN	0.004	0.042	0.267	2.635	5.172	10.006
SDSD	0.614	1.210	2.354	9.543	16.365	26.200
RMSSD	0.773	1.565	3.161	10.774	19.658	34.446
pNN20	0.006	0.012	0.025	0.043	0.082	0.149
pNN50	0.005	0.010	0.022	0.064	0.125	0.232

The data in the table show that the errors induced are especially large in the case of the variability measures. As discussed, this has to do with what the measures are designed to express: the variability measures express the variation in the beat-to-beat intervals. Considering that for example the RMSSD range tends to lie between roughly 20 – 55 (G. M. Kim & Woo, 2011) and in our experience rarely exceeds 125, the error of 34.446 introduced by displacing 20% of peaks is more than large enough to bury any effects of external factors on HRV. Effects on BPM (and thus IBI as well) are negligible, however, reflecting their relative insensitivity to outliers.

6.3   Results

6.3.1   PPG Data

The PPG dataset represents 20.7 hours of PPG recordings split into 1,240 one-minute segments. Due to sensor disconnects, 1.095 (18.25 hours) of the data contained a heart rate signal. The signals were recorded at the tip of the finger as participants were driving in a driving simulator.

The sensor placement did not interfere with driving. Participants were instructed to drive as they normally would in real life.

The data set was manually annotated by a human, then checked a second time to ensure accuracy. To annotate the set a custom tool was developed based on the Python's 'pyplot' plotting library. Each of the segments was visualised, and in the tool the annotator could manually mark peak locations, correct incorrectly placed peak locations and delete incorrect detections. Segments where little to no heart rate signal was present (for example due to the sensor detaching from the fingertip) were excluded. This left a total of 1.095 segments for the validation phase. A total of 89,837 peaks were detected by the algorithm. Of these, 84,845 (95.11%) were correctly accepted, and 2,977 (3.34%) were correctly rejected automatically. This indicates that for 98.45% of all detections, the algorithm correctly labelled the peak locations. 957 (1.07%) peaks were incorrectly rejected. 426 (0.48%) peaks were incorrectly accepted. A total of 632 peaks were annotated as missed. Most of the incorrectly accepted peaks occur either because a peak location was marked not at but nearby its maximum (Figure 6.5, e) which induces a minor error, or because a diastolic (secondary) peak is marked as a peak (Figure 6.5, f) which induces a larger error. Future updates of the algorithm aim to further reduce these error rates.

We compared the performance of our algorithm with an implementation of the Pan-Tompkins QRS algorithm (Pan & Tompkins, 1985), as well as with an open source algorithm called HRVAS ECGViewer[11]. The latter was chosen because it is one of the first hits when searching for open source heart rate analysis software on Google and it shows high usage statistics. It is designed for Matlab, but a standalone version is also available. The Pan Tompkins algorithm is a computationally efficient algorithm widely used in ECG analysis.

The comparison results are displayed in Table 6.2. They indicate that our algorithm significantly outperforms the other two open source algorithms on PPG data. The peak position error is 0.89 (milliseconds), indicates that the mean of the errors between the actual peaks and the predicted peaks was low compared to the other two algorithms. The resulting peak-peak intervals were also more accurate compared to the other algorithms. This is likely due to less missed and less incorrectly accepted beats in our case.

Differences in BPM error are not very large. Since the BPM uses the mean of all peak-peak intervals in a segment, it is relatively robust to a few incorrectly placed peak positions. However, effects on heart rate variability measures are large. To evaluate HRV performance we selected the Standard Deviation of Successive Differences (SDSD), which is one often used HRV measure that expresses how the intervals between the heart beats vary over time. It shows a large root mean squared error in the other two algorithms. This shows the importance of correctly identifying peak positions as well as identifying incorrectly labelled peaks, as deviations risk introducing substantial error to the output measures.

---

[11] See: https://github.com/jramshur/ECG_Viewer

Table 6.2 – Table showing how our algorithm compares to two other popular open source algorithms on key metrics.

Comparison of algorithm performance on PPG dataset (N=1,240)			
	peak location error (ms)	RMSE peak-peak intervals	RMSE BPM	RMSE SDSD
Developed algorithm	0.89	29.64	3.77	167.77
Pan-Tompkins	7.62	267.67	10.73	1060.77
HRVAS ECGViewer	4.25	171.32	4.76	364.74

All analysis data, code and results are also available on the GitHub page (van Gent, 2017) in the form of Jupyter notebooks. These can be opened and viewed directly on GitHub, or downloaded and executed using the Python 3.6 Anaconda distribution[12].

6.3.2   ECG Data

The 1.000 sections selected from the ECG dataset comprise a total of 16.67 hours of heart rate data. More information about the dataset is available in the publication of Jager et al. (2003).

The dataset was fully annotated. A total of 73.841 peaks were detected by the algorithm. Of these, 73.443 (99.46%) were correctly accepted and 190 (0.26%) were correctly rejected, representing a total of 99.72% of peaks correctly treated. 54 (0.07%) of peaks were incorrectly rejected, and 154 (0.21%) were incorrectly accepted. A total of 929 peaks were annotated as missed, meaning they were not detected by the algorithm. Note that with ECG, which has a more stable morphology, performance is significantly improved compared to PPG.

This performance was compared to the same algorithms as described above. The developed algorithm again showed superior performance on ECG data, although we were impressed with the performance of ECGViewer as well. The peak position error is lower compared to the PPG data, reflecting that the ECG waveform is easier to detect and more stable than the PPG waveform. The peak finding method we employ does not discriminate between the types of waveforms and can handle considerable morphological distortion.

The details are displayed in table 6.3 below.

Table 6.3 – Table showing how our algorithm compares to two other popular open source algorithms on key metrics.

Comparison of algorithm performance on ECG dataset (N=1,000)			
	RMSE peak location	RMSE peak-peak intervals	RMSE BPM	RMSE SDSD
Developed algorithm	0.16	6.38	0.41	221.79
Pan-Tompkins	6.84	335.25	3.07	371.38
HRVAS ECGViewer	3.64	97.34	1.88	231.96

---

[12] See: https://www.anaconda.com/download/

As with the PPG data, all analysis data, code and results are also available on the GitHub page (van Gent, 2017) in the form of Jupyter Notebooks. These can be used in conjunction with the manually annotated datasets to validate the performances on both datasets.

It must be noted that the lower performance of the other two algorithms is mainly due to a relatively small number of segments. Further analysis showed that for the developed algorithm, 98.7% of all segments had an error of 25 milliseconds or less in the computed peak-peak intervals, with 58.1% showing no error at all. For the Pan-Tompkins implementation the majority (67.5%) also showed an error of 25ms of less, with 9.5% showing no error at all. For the ECGViewer implementation 77% showed an error of less than 25ms. No segments were without error. The reason for this is likely that this implementation uses a template matching system for beat detection, which while robust to noise, also creates slight errors in the positions of detected peaks because the template rarely matches the heartrate waveform in the measured signal perfectly.

Furthermore, the ECGViewer implementation for example failed to detect any peaks in 108 segments (10.8%), likely due to noise or deviating morphology. One such example is shown in the figure 6.6 below. These segments were excluded from the calculation[13] of the performance measures from Table 6.3, so they do not negatively influence these measures.
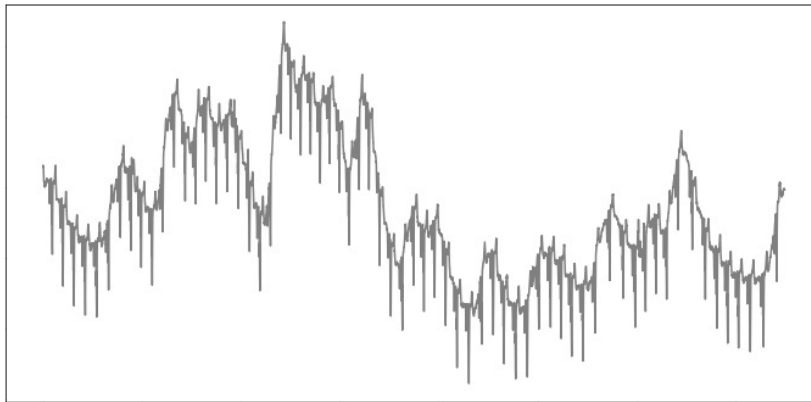
Figure 6.6 - An example of a noisy ECG recording where the ECGViewer implementation fails to detect any complexes. These segments were excluded from further analysis.

6.3.3   Additional Information Embedded in the Heart Rate Signal: Breathing Patterns

In addition to heart rate, the developed algorithm extracts breathing patterns from the collected heart rate data as well. Heart rate tends to increase during inhalation and decrease again during exhalation (Grossman & Taylor, 2007). This creates the possibility of extracting an estimate of the breathing rate from heart rate signals.

We've included a basic estimation method in the algorithm and validated it on an existing dataset (Karlen, Raman, Ansermino, & Dumont, 2013), that includes both PPG and respiratory data from patients undergoing surgery. Ground truth breathing rate was calculated from the CO_2 capnometry signal, which measures the increase in carbon dioxide concentration whenever the patient exhales. To extract breathing rate from the PPG signal the peak-peak intervals were

---

[13] Note that HeartPy can handle these edge cases since V1.2.4

upsampled, and their cycles marked. The following image visualises how the upsampled signal relates to the CO_2 capnometry signal.
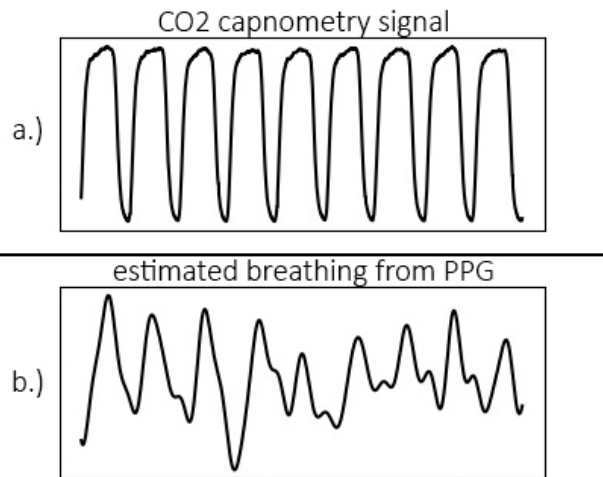
Figure 6.7 - Figure showing how the CO_2 capnometry signal (top) relates to the breathing signal extracted from the PPG.

As discussed in (Grossman & Taylor, 2007), the relationship between heart rate variability and breathing rate is not a linear one and can be impacted by factors such as medication use and physical strain. The results of our validation on this dataset reflect this nonlinear relationship. The estimates correspond to the breathing rate determined by capnometry but include detection errors. See table 6.4 for an overview of the detection errors expressed as the difference in Hertz between ground truth and calculated measures. The mean error present in the PPG estimation compared to the ground truth estimation of breathing rate, corresponds to a confidence interval of roughly 5-10%. This illustrates that the breathing rate extracted from the PPG signal should only be taken as an estimation and not an absolute value. Future improvements to the method might increase accuracy. The data files and analysis code are available on the GitHub page (van Gent, 2017).

Table 6.4 - Table showing the difference between the ground truth breathing rate and the breathing rate estimated from the PPG.

	mean	median	minimum	maximum
Error magnitude	0.028	0.015	0	0.102

6.4   Discussion and Conclusion

In this paper we have presented the validation analysis and results of a novel, robust heart rate analysis algorithm developed for use in lab settings, as well as in-vehicle and other mobile settings. The motivation to develop such an algorithm is that current available open source algorithms do not work well on noisy PPG data, are often highly technical or expensive to implement, and because low-cost commercial measurement devices offer no suitable solution for scientific purposes. The developed algorithm runs both in real-time and offline, on desktop computers and wearable (embedded) hardware. This makes it ideal for human factors studies

seeking to incorporate heart rate analysis in their design, which includes studies into how automated vehicles can obtain and maintain an awareness of the driver's state.

We have evaluated the algorithm's performance on a manually annotated PPG and an ECG data set and compared the performance to two other available algorithms. Results showed superior performance on PPG data. This reflects that the lower performance of the two other algorithms is specific to the type of data: PPG data collected in the field using low-cost sensors has quite different signal and noise properties compared to ECG data, for which many available open source algorithms are designed. Despite the higher noise rate, PPG data can be collected using less intrusive or even contactless methods, which makes it ideal for real-world driving settings. On ECG data the differences between algorithms were less pronounced.

Making automated vehicles smarter means they need to be aware of the driver's state. Heart rate is one physiological marker that allows the estimation of driver state on several levels. By offering an openly available and validated toolkit for heart rate analysis, we aim to increase the research possibilities into this field, as well as the reliability and reproducibility of results obtained.

One limitation of the present validation is that it was performed on 2.095 one-minute segments, due to the time-intensive nature of the manual annotation. The annotation was done by hand, which means minor errors can exist in the annotations. Although we believe that the reported performance is a good reflection of real-world performance due to the two different datasets used, a larger validation will add further confidence to the results. Furthermore, the included breathing rate estimation should be taken as an estimated value rather than an absolute one, since the relationship between heart rate and breathing rate is not linear. Future steps include this larger validation, as well as increasing the accuracy and functionality of the algorithm, and comparing its performance to commercially available solutions.

References

Bernardi, L., Leuzzi, S., Radaelli, a, Passino, C., Johnston, J. a, & Sleight, P. (1994). Low-frequency spontaneous fluctuations of R-R interval and blood pressure in conscious humans: a baroreceptor or central phenomenon? Clinical Science (London, England : 1979), 87(6), 649–654.

Birrel, S., Young, M., Stanton, N., & Jennings, P. (2017). Using Adaptive Interfaces to Encourage Smart Driving and Their Effect on Driver Workload, 484, 764. https://doi.org/10.1007/978-3-319-41682-3

Brookhuis, K. A., & de Waard, D. (2010). Monitoring drivers' mental workload in driving simulators using physiological measures. Accident Analysis and Prevention, 42(3), 898–903. https://doi.org/10.1016/j.aap.2009.06.001

Brouwer, A., & Dijksterhuis, C. (2015). Physiological correlates of mental effort as manipulated through lane width during simulated driving. In Proceedings of the 2015 International Conference on Affective Computing and Intelligent Interaction (ACII) (pp. 42–48).

Danisman, T., Bilasco, I. M., Djeraba, C., & Ihaddadene, N. (2010). Drowsy driver detection system using eye blink patterns. 2010 International Conference on Machine and Web Intelligence, ICMWI 2010 - Proceedings, 230–233. https://doi.org/10.1109/ICMWI.2010.5648121

Elgendi, M. (2012). On the Analysis of Fingertip Photoplethysmogram Signals. Current Cardiology Reviews, 8(1), 14–25. https://doi.org/10.2174/157340312801215782

Fallahi, M., Motamedzade, M., Heidarimoghadam, R., Soltanian, A. R., & Miyake, S. (2016). Effects of mental workload on physiological and subjective responses during traffic density monitoring: A field study. Applied Ergonomics, 52, 95–103. https://doi.org/10.1016/j.apergo.2015.07.009

Farah, H., Koutsopoulos, H. N., Saifuzzaman, M., Kölbl, R., Fuchs, S., & Bankosegger, D. (2012). Evaluation of the effect of cooperative infrastructure-to-vehicle systems on driver behavior. Transportation Research Part C: Emerging Technologies, 21(1), 42–56. https://doi.org/10.1016/j.trc.2011.08.006

Grossman, P., & Taylor, E. W. (2007). Toward understanding respiratory sinus arrhythmia: Relations to cardiac vagal tone, evolution and biobehavioral functions. Biological Psychology, 74(2), 263–285. https://doi.org/10.1016/j.biopsycho.2005.11.014

Healey, J. A., & Picard, R. W. (2005). Detecting Stress During Real-World Driving Tasks Using Physiological Sensors. IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, 6(2), 156–166.

Hoogendoorn, R., van Arem, B., & Hoogendoorn, S. (2014). Automated Driving, Traffic Flow Efficiency, and Human Factors. Transportation Research Record: Journal of the Transportation Research Board, 2422(2422), 113–120. https://doi.org/10.3141/2422-13

Jager, F., Taddei, A., Moody, G. B., Emdin, M., Antolič, G., Dorn, R., … Mark, R. G. (2003). Long-term ST database: A reference for the development and evaluation of automated ischaemia detectors and for the study of the dynamics of myocardial ischaemia. Medical and Biological Engineering and Computing, 41(2), 172–182. https://doi.org/10.1007/BF02344885

Jamson, a. H., Merat, N., Carsten, O., & Lai, F. (2011). Fully-automated driving: the road to future vehicles. Proceedings of the Sixth International Driving Symposium on Human Factors

in   Driver   Assessment,   Training   and   Vehicle   Design,   2–9.
https://doi.org/10.17077/drivingassessment.1370

Jarvis, J., Putze, F., Heger, D., & Schultz, T. (2011). Multimodal person independent
recognition of workload related biosignal patterns. Proceedings of the 13th International
Conference   on   Multimodal   Interfaces   -   ICMI   ’11,   205.
https://doi.org/10.1145/2070481.2070516

Karlen, W., Raman, S., Ansermino, J. M., & Dumont, G. A. (2013). Multiparameter respiratory
rate estimation from the photoplethysmogram. IEEE Transactions on Biomedical Engineering,
60(7), 1946–1953. https://doi.org/10.1109/TBME.2013.2246160

Kim, G. M., & Woo, J. M. (2011). Determinants for heart rate variability in a normal Korean
population.   Journal   of   Korean   Medical   Science,   26(10),   1293–1298.
https://doi.org/10.3346/jkms.2011.26.10.1293

Kim, S., Chun, J., & Dey, A. K. (2015). Sensors Know When to Interrupt You in the Car.
Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems -
CHI ’15, 487–496. https://doi.org/10.1145/2702123.2702409

Kirst, M., Glauner, B., & Ottenbacher, J. (2011). Using DWT for ECG motion artifact reduction
with noise-correlating signals. Proceedings of the Annual International Conference of the IEEE
Engineering   in   Medicine   and   Biology   Society,   EMBS,   4804–4807.
https://doi.org/10.1109/IEMBS.2011.6091190

Liang, Y., Reyes, M. L., & Lee, J. D. (2007). Real-time detection of driver cognitive distraction
using support vector machines. IEEE Transactions on Intelligent Transportation Systems, 8(2),
340–350. https://doi.org/10.1109/TITS.2007.895298

Mehler, B., Reimer, B., & Coughlin, J. F. (2012). Sensitivity of Physiological Measures for
Detecting Systematic Variations in Cognitive Demand From a Working Memory Task: An On-
Road Study Across Three Age Groups. Human Factors: The Journal of the Human Factors and
Ergonomics Society, 54(3), 396–412. https://doi.org/10.1177/0018720812442086

Mehler, B, Reimer, B., & Wang, Y. (2011). Comparison of heart rate and heart rate variability
indices in distinguishing single task driving and driving under secondary cognitive workload.
Proceedings of the Sixth International Driving Symposium on Human Factors in Driver
Assessment, Training, and Vehicle Design, 590–597.

Mehler, Bruce, Reimer, B., Coughlin, J. F., & Dusek, J. a. (2010). Impact of Incremental
Increases in Cognitive Workload on Physiological Arousal and Performance in Young Adult
Drivers. Transportation Research Record: Journal of the Transportation Research Board,
2138(1), 6–12. https://doi.org/10.3141/2138-02

Merat, N., Jamson, A. H., Lai, F. C. H., Daly, M., & Carsten, O. M. J. (2014). Transition to
manual: Driver behaviour when resuming control from a highly automated vehicle.
Transportation Research Part F: Traffic Psychology and Behaviour, 27(PB), 274–282.
https://doi.org/10.1016/j.trf.2014.09.005

Millasseau, S. C., Guigui, F. G., Kelly, R. P., Prasad, K., Cockcroft, J. R., Ritter, J. M., &
Chowienczyk, P. J. (2000). Noninvasive Assessment of the Digital Volume Pulse: Comparison
With the Peripheral Noninvasive Assessment of the Digital Volume Pulse Comparison With
the   Peripheral   Pressure   Pulse.   Hypertension,   36,   952–956.
https://doi.org/10.1161/01.HYP.36.6.952

Miyaji, M., Danno, M., Kawanaka, H., & Oguri, K. (2008). Driver’s cognitive distraction
detection using adaboost on pattern recognition basis. Proceedings of the 2008 IEEE

International Conference on Vehicular Electronics and Safety, ICVES 2008, 51–56. https://doi.org/10.1109/ICVES.2008.4640853

Montano, N., Porta, A., Cogliati, C., Costantino, G., Tobaldini, E., Casali, K. R., & Iellamo, F. (2009). Heart rate variability explored in the frequency domain: A tool to investigate the link between heart and behavior. Neuroscience and Biobehavioral Reviews, 33(2), 71–80. https://doi.org/10.1016/j.neubiorev.2008.07.006

Pan, J., & Tompkins, W. J. (1985). A simple real-time QRS detection algorithm. IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, BME-32(3), 230–236. https://doi.org/10.1109/IEMBS.1996.647473

Park, H. S., & Kim, K. H. (2015). Adaptive multimodal in-vehicle information system for safe driving. ETRI Journal, 37(3), 626–636. https://doi.org/10.4218/etrij.15.0114.1104

Reimer, B., Mehler, B., & Coughlin, J. F. (2016). Reductions in self-reported stress and anticipatory heart rate with the use of a semi-automated parallel parking system. Applied Ergonomics, 52, 120–127. https://doi.org/10.1016/j.apergo.2015.07.008

Reimer, B., Mehler, B., Coughlin, J. F., Roy, N., & Dusek, J. a. (2011). The impact of a naturalistic hands-free cellular phone task on heart rate and simulated driving performance in two age groups. Transportation Research Part F: Traffic Psychology and Behaviour, 14(1), 13–25. https://doi.org/10.1016/j.trf.2010.09.002

Selvaraj, N., Jaryal, A., Santhosh, J., Deepak, K. K., & Anand, S. (2008). Assessment of heart rate variability derived from finger-tip photoplethysmography as compared to electrocardiography. Journal of Medical Engineering and Technology, 32(6), 479–484. https://doi.org/10.1080/03091900701781317

Solovey, E. T., Zec, M., Garcia Perez, E. A., Reimer, B., & Mehler, B. (2014). Classifying driver workload using physiological and driving performance data. Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems - CHI '14, 4057–4066. https://doi.org/10.1145/2556288.2557068

Stapel, J., Mullakkal-Babu, F. A., & Happee, R. (2017). Driver Behaviour and Workload in an On-road Automated Vehicle. In Proceedings of the RSS2017 Conference.

Sun, Y., Hu, S., Azorin-Peris, V., Kalawsky, R., & Greenwald, S. (2012). Noncontact imaging photoplethysmography to effectively access pulse rate variability. Journal of Biomedical Optics, 18(6), 061205. https://doi.org/10.1117/1.JBO.18.6.061205

van Gent, P. (2017). Python Heart Rate Analysis Toolkit. Retrieved from https://github.com/paulvangentcom/heartrate_analysis_python

van Gent, P. (2018). Arduino Heart Rate Analysis Toolkit. Retrieved from https://github.com/paulvangentcom/heartrate_analysis_Arduino

van Gent, P., Farah, H., van Nes, N., & van Arem, B. (2017). Towards Real-Time, Nonintrusive Estimation of Driver Workload: A Simulator Study. In Road Safety and Simulation 2017 Conference Proceedings.

van Gent, P., Farah, H., van Nes, N., & van Arem, B. (2018a). A Conceptual Model for Persuasive In-Vehicle Technology to Influence Tactical Level Driver Behaviour. Submitted to Transportation Research Part F, 1–20.

van Gent, P., Farah, H., van Nes, N., & van Arem, B. (2018b). Analysing Noisy Driver Physiology Real-Time Using Off-the-Shelf Sensors: Heart rate analysis software from the Taking the Fast Lane Project. Submitted to Journal of Open Research Software.

van Gent, P., Melman, T., Farah, H., van Nes, N., & van Arem, B. (2018). Multi-Level Driver Workload Prediction Using Machine Learning and Off-The-Shelf Sensors. Transportation Research Record: Journal of the Transportation Research Board. https://doi.org/10.1177/0361198118790372

Zhai, J., & Barreto, A. (2006). Stress Recognition Using Non-invasive Technology. Proceedings of the 19th International Florida Artificial Intelligence Research Society Conference (FLAIRS)., 395–400.

Chapter 7.

The Persuasive Automobile: Design and Evaluation of a Persuasive Lane-Specific Advice Human Machine Interface

Abstract

Traffic congestion is a major societal challenge. By advising drivers on the optimal lane to drive, traffic flow can be improved and congestion reduced. This chapter presents the development of a lane-specific Human Machine Interface (HMI) designed to deliver lane-specific advices to the driver. It builds on the conceptual model from chapter 3 by determining and evaluating both message content and message modality from the information transfer level, as well as persuasive techniques as descibed in the system level. The challenge is to persuade drivers to follow an advice that is beneficial to the traffic situation, but may not be immediately beneficial to the drivers themselves. To solve this challenge a persuasive lane-specific advice system was developed and tested. This chapter describes the design process of the persuasive system, followed by two questionnaire studies and a simulator study. In the simulator study two types of persuasion were tested: gamified and socially cooperative persuasion. Participants drove on two separate days, with a web-portal intervention being shown to the treatment groups between the two days. Those in the treatment groups followed significantly more advices (117 and 111) than those in the control group (89). No significant differences were visible between competitive and cooperative groups. The differences between groups only emerged on the second day, indicating the intervention was the likely cause of the effect.

7.1   Introduction

7.1.1   Background

The effects of congestion on both the economy and individuals are large. Aside from annoyance and time loss, congestion is a source of higher emissions (Zhang, Batterman, & Dion, 2011) and negatively impacts safety. The benefits of reducing congestion are obviously large. Driver-assistance systems that can help reduce congestion and improve flow are for example connected cruise control (Schakel, Arem, & Netten, 2010), or a congestion assistant (Van Driel & Van Arem, 2010) which, based on simulation experiments, would reduce travel-time delay by 30% even at a 10% penetration rate.

Recent technological advancements add to the possibilities by enabling vehicles to detect the specific lane they are driving on based on low-cost precise point positioning GPS receivers (V. L. Knoop, De Bakker, Tiberius, & Van Arem, 2013; Victor L. Knoop, De Bakker, Tiberius, & Van Arem, 2017). This makes traffic control on the individual level possible by advising drivers on a specific lane they can take (Schakel & Van Arem, 2014; Yao, Knoop, & van Arem, 2017). Such an advice system needs to be safe as well as persuasive, in order for it to be successful (van Gent, Farah, van Nes, & van Arem, 2019). The next question then becomes how to make such a system persuasive and safe. To determine this, we investigate and describe the development of a persuasive lane-specific advice Human Machine Interface (HMI) in this paper.

The rest of section 1 introduces the literature background for the study. Section 2 reports the methods and results of two questionnaire studies that were performed to determine the type of auditory chime used to alert the driver to an advice, the location of the interface, and whether to provide context for the advice (reason for advice and feedback on behaviour). In section 3 we develop the persuasive advices and a web-portal for the simulator study based on the results of the questionnaire studies from section 2. Section 4 describes the methods used in the simulator study to evaluate the effectiveness of the persuasive lane change advice. Section 5 describes the results of the simulator study, and in section 6 and 7 the results are discussed, and conclusions drawn.

7.1.2   Objectives

We are working on an in-car system with the goal of reducing congestion through lane-specific advices. This will be achieved by stimulating a better distribution of traffic over the available lanes on a multi-lane highway through lane-specific advices. The challenge is to persuade drivers to follow non-compulsory advices that are in the benefit of all drivers on a given road segment, but not necessarily in the benefit of individual drivers (Risto, 2014). Some drivers may, for example, be asked to move to a slower lane in order to maintain a balanced traffic system.

The main objective of this study is to find a way to persuade drivers to comply with these voluntary lane-specific advice messages, using methods from the field of persuasive technology (Fogg, 2003; Oinas-Kukkonen, 2013; van Gent et al., 2019). To achieve this, we develop a multimodal (auditory, visual) interface to convey lane-specific requests to the driver. This leads to the following sub-goals: to design an auditory and visual signal, to determine whether to

provide context for the advice to the driver (reason for advice, feedback on behaviour), and to define the safest location for the interface. This paper describes the design process of the interface in two iterative steps, and the evaluation of the lane-specific advice HMI's effects in a driving simulator.

7.1.3   Techniques for Driver Persuasion

Our aim is to stimulate drivers to follow lane-specific advice messages, without enforcing compliance. Gamification has been used to change behaviour in people (Hamari, Koivisto, & Sarsa, 2014). Video games are designed to create environments that motivate people to display certain behaviours over others, often to win the game. Gamification is about applying those game design elements that elicit different behaviour patterns to non-game contexts (Deterding, Dixon, Khaled, & Nacke, 2011). Such elements include challenges, leader boards and achievements (Hamari et al., 2014). In driving contexts gamification has been used for example to encourage eco-driving behaviour (Ecker, Holzer, Broy, & Butz, 2011; Nousias et al., 2019; Steinberger, Proppe, Schroeter, & Alt, 2016), and to encourage safer driving behaviour (Bahadoor & Hosein, 2016; Shi, Lee, Kurczak, & Lee, 2012). Other ways of achieving behavioural change include methods from persuasive technology (Fogg, 2003; Hutchison & Mitchell, 2008) and behavioural economics (Avineri, 2011; Cialdini, 2006; Kahneman, 2013).

The different approaches are unified in the Persuasive Systems Design (PSD) model, which takes concepts from the different persuasive fields and brings them together into a single model (Oinas-Kukkonen & Harjumaa, 2008, 2009). The PSD specifies that a system can be made more persuasive by offering support to the user in various categories: primary task support, dialogue support, system credibility support and social support (Oinas-Kukkonen & Harjumaa, 2008; van Gent et al., 2019).

Persuading different people in different situations might require different approaches, and there are indications that not every person is equally susceptible to being persuaded, at least from studies on health-based persuasive applications (Kaptein, Lacroix, & Saini, 2010) and gaming settings (Orji, Mandryk, & Vassileva, 2015; Orji, Vassileva, & Mandryk, 2014). This provides a challenge because we need to maximise persuasive potential while not creating a personalised solution for every driver, which would needlessly complicate the design. Orji et al. (Orji et al., 2014) provide a possible solution. The authors investigated persuasive effectiveness on a range of 'gamer personalities' in 1.108 gamers. The personality types they used were derived from a neurobiological study into gamer personalities called BrainHEX (Nacke, Bateman, & Mandryk, 2011). The personality types found (seeker, survivor, daredevil, mastermind, conqueror, socialiser, achiever) had, as expected, stronger relations with gaming and cannot readily be translated to the driving environment. However, a set of persuasive techniques were found that worked well across all the different personality types. These are **competition** and **comparison**, which fit in the "social support" component from the PSD model (Oinas-Kukkonen & Harjumaa, 2008). **Self-monitoring** and **suggestion**, respectively from "dialogue support" and "primary task support" in the PSD, were found to be effective across the different personality types. Interestingly, **praise** and **rewards** did not have a strong effect in this study, contrary to what others have reported. This may be in line with what is reported by (Scott, Pereira, & Oakley, 2012), where the effectiveness of feedback combined especially with emotionally expressive avatars did not always work well, especially when negative emotions on avatars were combined with negative text messages.

7.1.4 Defining Message Modality

Aside from persuading a driver, the modality that is used to convey any type of information to a driver is of major importance, as humans have limited information processing capacity. Dangerous and even life-threatening situations may occur when overloading a driver (de Waard, 1996; Fuller, 2005; Young, Brookhuis, Wickens, & Hancock, 2015), or when distracting a driver with an advice at the wrong moment (Horberry, Anderson, Regan, Triggs, & Brown, 2006; Reyes & Lee, 2004).

Visual interfaces have the advantage of having high information bandwidth and being self-paced. However, many visual interfaces require the driver to take their eyes off the road. Taking eyes off the road has been shown to have serious consequences for lane-keeping ability (Peng, Boyle, & Hallmark, 2013), and may cause drivers to miss safety-critical events on the road.

Heads-Up Displays (HUD) have been put forward as a means of reducing the negative aspects of visual displays in cars. However, HUDs have some problems as well related to both psychological and biological processes. The 'looked-but-failed-to-see' problem (Herslund & Jørgensen, 2003) is an example. This occurs when an object (like a pedestrian, cyclist, or other car) is within the field of view of a driver, but is not perceived. This seems to be a cognitive problem rather than a sensory one, where the object is visible on the retina but not consciously registered by the driver. HUDs might exacerbate this issue by adding an additional stimulus to the driver's field of view. In other words: even if the driver's eyes are on the road, that does not mean the driver's attention is on the road. In this regard it is important to keep visually presented information brief and easily understood, for example by making stimuli similar to their real-world counterparts. This reduces cognitive distance (Kim & Dey, 2015), which is defined as the ease of transforming digital information to a task at hand (Kim & Dey, 2009). An example of an advice with a short cognitive distance is a lane change request that displays the current lane configuration, the ego vehicle on its current lane, and an arrow or instruction pointing to the lane to which the driver needs to move. This way the driver does not need to expend much cognitive processing on understanding the advice, but can instead focus on the requested behaviour.

Biological processes might also interfere with driving. For example, the eye has a so-called resting focus (or 'dark focus'), which is the focal distance of the eye when the iris is relaxed. Typically, this is between 0.5-2.0 meters. Stimuli placed in this distance can draw a particularly dominant accommodation response from the eye (Edgar, 2007). This was originally called the Mandelbaum effect and it is especially prevalent when visibility conditions are poor (Owens, 1979). This might create issues with HUDs in certain weather conditions, which needs to be considered when designing an HMI for on-road use, for example by not having the HUD be always on, and to be sensitive to contexts by reducing its saliency when visibility conditions are poor.

Multimodal interfaces have been proposed to reduce the negative aspects of using a single modality, especially in complex environments (Sarter, 2006). From a theoretical perspective this works by reducing load on a single modality and allowing drivers to better spread work over their available mental resources (Wickens, 2002). Spreading information over multiple modalities has been shown to induce lower workloads (Y C Liu, 2001) and better reaction times (Ho, Kingdom, & Reed, 2007) in participants.

Based on these benefits for workload and reaction times, in this study we chose to design for a multimodal display, where the advice is visually presented and announced by an auditory

chime. The chime is used to alert the driver whenever an advice is available, as described by for example (Sarter, 2006). This way the driver can focus on the road and only has to look at the display whenever an advice is available.

7.1.5   Making it Personal

Avatars are representations of a virtual character. They are more effective than textual information in eliciting a human-like interaction between system and driver (Scott et al., 2012). Scott et al. showed that adding emotional expressions increased persuasive effectiveness and trustworthiness of a system. Avatars have been used in gamified driving contexts such as Driving Miss Daisy (Shi et al., 2012), which helps improve driving skills by providing a virtual passenger that occasionally comments on driving style. To facilitate more human communication, we developed an avatar based off a freely available clipart from www.clipartroo.com.

Fig. 7.1 **Design of the avatar, showing its happy state (left) and sad state (right).**

The avatar (Figure 7.1) had a happy and an unhappy state depending on how drivers would react to advices. We chose a stylized avatar, so it resembled a car rather than a human. The choice was based on work by Verberne et al (Verberne, Ham, & Midden, 2012), who showed that trust in an in-car system improved if drivers perceived it as sharing their driving goals. By styling the avatar like a car that was happy when congestion was avoided, we aimed to visualize that the driver's goal of reaching a destination without congestion was shared by their car. This stylizing is unlikely to change participant's response to the avatar, as Bailenson et al (Bailenson, Blascovich, Beall, & Loomis, 2001) for example demonstrated people tend to respond to avatars in a natural way as if they are human, even if they are highly stylized and don't resemble humans at all.

7.1.6   Using a Driving Simulator for HMI Research

The driving simulator is a powerful tool to investigate human behaviour in a controlled setting where traffic and weather conditions can be standardised (Carsten & Jamson, 2011). In the context of our study, a simulator offers an environment where our novel HMI design can be safely tested without the danger of distracting a participant in real traffic.

Wang et al. (Wang et al., 2010) have shown that medium fidelity driving simulators can be used effectively to evaluate in-vehicle information interfaces, which our proposed persuasive HMI is, although care must be taken to ensure no confounding variables are introduced (Engen, Lervåg, & Moen, 2009).

7.2   Developing the Persuasive Interface – Two Questionnaire Studies

Prior to performing our simulator experiment we needed to define several important aspects. These include the type of auditory alert used to announce the advice, the location of the advice, and whether to provide a reason for the advice or feedback on the performed behaviour. If the advice is unclear, the alert not salient enough, or if the system is considered annoying, it is

unlikely drivers will follow advices or continue using the system (Risto & Martens, 2013; van Gent et al., 2019). Two questionnaire studies were performed. The first questionnaire study is described in section 2.1 and 2.2, and investigates whether to precede the advice by an auditory chime, and if so, which chime. The second questionnaire is described in section 2.3 and 2.4. It uses the chime determined in the first questionnaire, and investigates where the advice should be located based on driver preferences (central console, HUD, or near speedometer), and whether to provide a reason for the given advice or feedback on driver behaviour.

7.2.1   Determining the Auditory Alert Chime – Methods

To determine which auditory chime to use to alert drivers to an available advice, we performed a questionnaire study. The aim was to select a chime that sounded friendly (to not irritate the driver), could alert the driver, and that was not judged to be distracting.

A range of auditory alert chimes were designed using Apple's Logic Pro digital audio workstation, and the Omnipshere digital synthesizer. The chimes were designed around the C Major tonality, which has an open and warm character. 15 chime types were generated in total. Where applicable, variations in rhythm and variations in pitch were generated per chime type. This gave a total of 53 possible alert sounds. We reduced these possibilities by making a subjective pre-selection of seven auditory alerts.

The questionnaire was distributed through Google Forms. In the questionnaire participants were informed about the goals of our proposed lane-specific advice HMI, and subsequently presented with the seven selected auditory chimes. After each chime they were asked for their impression regarding the alert, specifically if it was: informative, intrusive, friendly, distracting, annoying, easy to miss, and urgent. Each item was rated on a 7-point scale, ranging from disagree completely (-3), to neutral (0), to agree completely (3). Participants were recruited by an advert on social media (LinkedIn, Twitter), and through a recruitment e-mail to several departments at Delft University of Technology.

7.2.2   Determining the Auditory Alert Type – Analysis and Results

20 participants took part in the auditory chime questionnaire. 7 participants were female, 13 male. All participants were frequent drivers. 14 participants indicated driving at most 1000 km per month, and the remaining 6 participants between 1000 and 2000 km per month. This range is close to the Dutch national average for private cars of 13.000 km on a yearly basis (CBS, 2019).

Questionnaire data were analysed using a principal component analysis (PCA), a method that transforms a set of observations into uncorrelated variables as described for example in (Wold, Esbensen, & Geladi, 1987). This way we can find underlying constructs shared by different questions in a questionnaire.

The result of the PCA was visualised in a scree plot, which displayed a distinctive 'knee' at a two-component solution, which together explained 81.97% of all variance in the data set. The factor loadings for the two-component solution are displayed in table 7.1. We removed factor loadings smaller than 0.2.

Table 7.1 - PCA loadings on first two components.

Label	Component 1	Component 2
Informative	-	-0.779
Intrusive	-0.423	-
Friendly	0.377	-0.297
Distracting	-0.341	-
Annoying	-0.507	0.278
Easy to miss	0.378	0.422
Urgent	-0.395	-

The first component loads negatively on intrusiveness, distraction potential, annoyance and urgency, while loading positively on friendliness and being easy to miss. It seems to reflect a general 'likeability' of the chime. The chime being easy to miss is likely inversely related to its potential to be intrusive, distracting and annoying. The second component loads strongly negative on informativeness and on friendliness, while loading positively on being easy to miss and annoyance. This component seems to indicate that the alert is unclear: it is rated low on being informative, and high on being easy to miss and annoyance. It seems likely that an unclear message during driving would lead to annoyance.

The loadings of each of the seven chimes on the two components are displayed below in table 7.2. We selected chime #1, which loads strongly on the first component ('likeability') and not on the second component ('unclear').

Table 7.2 – Loadings of each chime on the two main PCA components.

Chime number	Component	Component 2
1	1.499	0.063
2	0.224	-0.194
3	-1.409	-0.060
4	-1.018	0.192
5	0.432	0.178
6	0.903	-0.058
7	-0.243	-0.017

7.2.3 Determining the Interface and Message Characteristics – Methods

After choosing the alert chime, we needed to determine the driver preferences regarding the implementation details of the lane-specific advice HMI, thus a second questionnaire study was performed. The questionnaire consisted of three parts. Most questions were answered on the same seven-point scale as the previous questionnaire (-3 – completely disagree, 0 – neutral, 3 – completely agree).

In the first part of the questionnaire participants were presented with three videos (figure 7.2), each showing the same lane-specific advice but in a different location: central console (1), heads-up display (2), and on the speedometer (3). After viewing each video, participants answered on a 7-point scale whether they noticed the advice quickly, if it was distracting, if they were used to looking at the specific location, if they felt they had to take their eyes off the road too long, if they felt safe looking at the specific location, if the location was convenient, and if they thought they would miss the advice easily at this location.
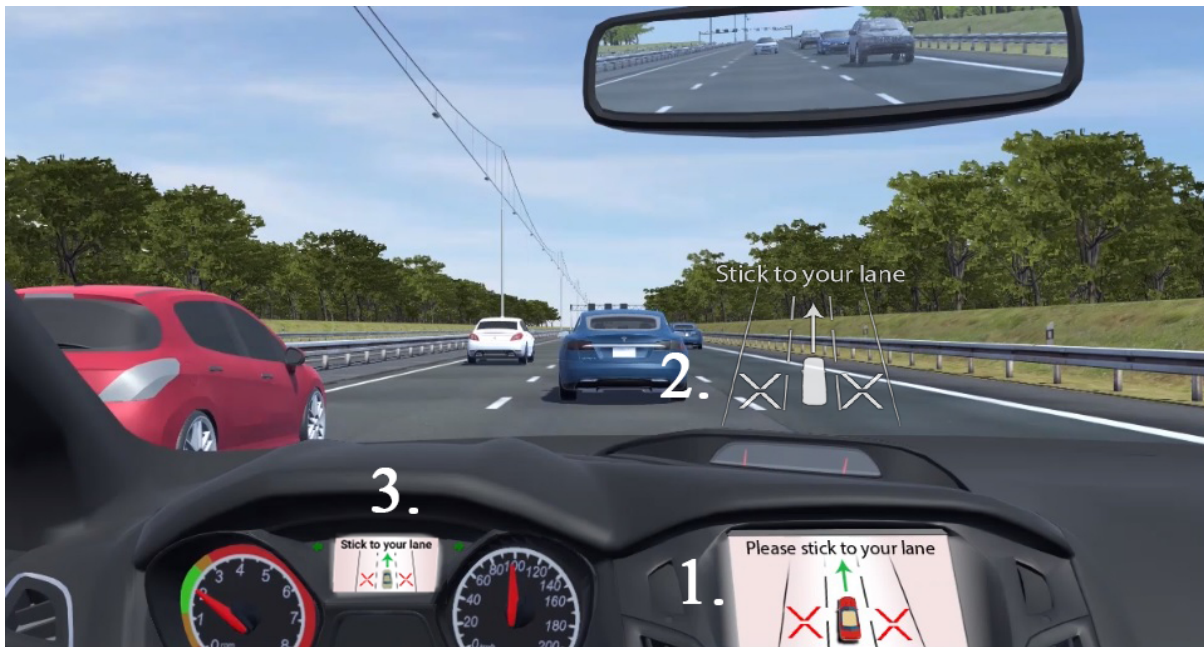
Fig. 7.2 Screenshot from one of the three videos shown to participants. The video shows regular traffic on a typical Dutch highway, with the active advice shown on the central console (1), the HUD (2), and near speedometer displays (3).

In the second part participants were presented with a full screen video of the same advice (figure 7.3), but with included audio and haptic feedback. This section served to test responses to the selected audio chime from the previous questionnaire, and to test whether to include haptic feedback in the steering wheel as well. Since no actual steering wheel would be available while filling in the questionnaire, the haptic feedback was displayed on a steering wheel below the advice visualisation as shown in figure 7.3 a.) and accompanied by a vibration sound. If vibration occurred on a particular side of the steering wheel, the vibration audio was only played through the corresponding stereo channel.

The last section of the questionnaire examined the context needed for the advice, specifically whether to provide the reason for the advice and feedback on driver behaviour. We know from earlier research (Risto & Martens, 2013) that if drivers do not perceive the reason for an advice, they are less inclined to follow it. Providing feedback can also support the formation of habits which are a main factor in making persuasive effects last over time (Lally & Gardner, 2013). Participants were shown an example video of an advice preceded by a message displaying the reason for the advice (figure 7.3 b.), and a message after the advice displaying feedback about their behaviour that consisted of the avatar thanking them or encouraging them to do better next time (figure 7.3 c.)). After this, they answered several questions about how it would impact their understanding of the advice, their likelihood of following it, and their perceived safety.
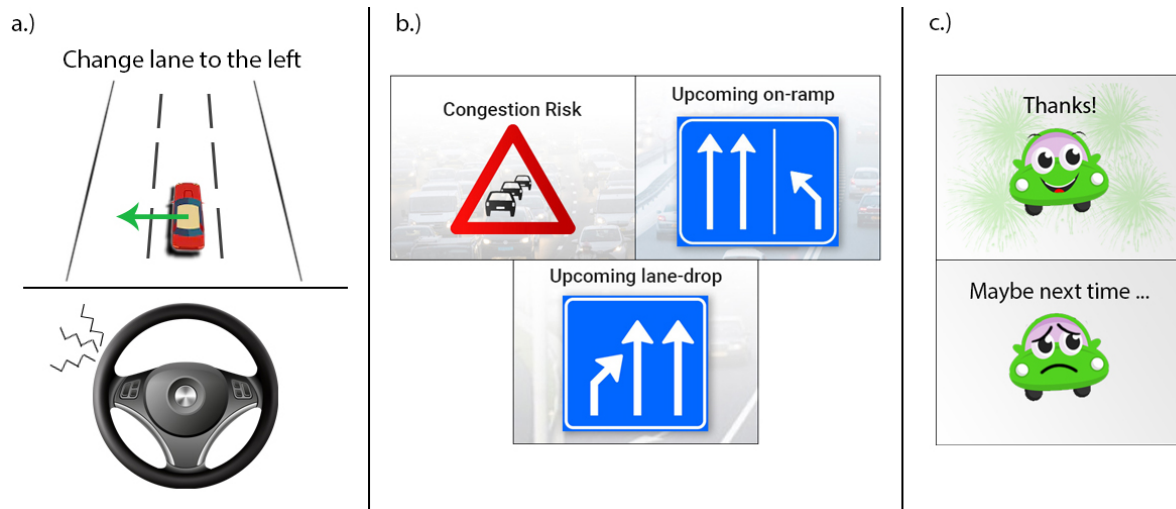
Fig. 7.3 Haptic feedback was shown on the steering wheel a.), and an audio message was played through the corresponding stereo channel. Image b.) shows possible reasons for an advice. Image c.) shows the two avatar states.

7.2.4 Determining the Interface and Message Characteristics – Analysis and Results

34 participants filled in the questionnaire. 2 did not complete the questionnaire and were excluded from the analysis. That left 32 participants in total. 23 were male, 9 female. 23 participants owned a car and 9 did not (no statistically significant correlation with gender, r = 0.227, p = 0.211). 18 indicated driving a maximum of 1.000 km per month, 6 drove 1.000-2.000 km, 6 drove 2.000-5.000 km, one drove over 5000 km a month and one participant indicated they didn't know their monthly mileage.

Overall, participants had a slight preference for the HUD (15, 46.88%), over the central console display (11, 34.38%), and the speedometer display (6, 18.74%). Answers to the questions were analysed using a series of repeated measures t-tests. Due to the number of comparisons run on the data a Bonferroni correction was applied which put the alpha used at p = 0.0023. A single value was significant. This was for the question where participants indicated they were more used to looking at their speedometer than a HUD (t = -3.503, p = .001). Since few cars are equipped with HUDs while all cars have speedometers, this information was obviously not informative or beneficial for choosing a location.

We chose to select the HUD combined with an auditory chime based on both its advantages offered as described in the literature (section 1.4), and based on the trend that slightly more participants preferred that location. The results indicated that drivers, at least in their self-reported answers, show little differences in preference, perceived safety, and perceived ease of the different locations tested. This runs contrary to earlier research where participants had a strong preference for the HUD, likely based on a novelty effect (Yung Ching Liu & Wen, 2004). Perhaps now, nearly 15 years later, participants are more used to these systems despite them not being widely available in cars yet.

Most drivers (25) indicated they liked having the audio chime available to alert them to whenever an advice becomes available, although 16 of the 25 indicated they would like to have the option to turn the chime off. We used a 7-point scale that ran from -3 (disagree completely), to 0 (neutral), to 3 (agree completely). Results are displayed in table 7.3-a. On average,

participants felt the chime helped them know an advice had become available, was appropriate, was not annoying, was not unnecessary, and would not startle them. Participants indicated it would not help them keep their eyes on the road, nor would it help them understand the advice. The latter was expected, as the chime was designed to alert drivers and did not vary based on the type of advice. The fact that participants indicated it would not help them keep their eyes on the road might be because the chime would prompt them to look at the interface. This again raises the importance of taking the driving context into consideration when choosing to communicate to the driver using in-car technology (van Gent et al., 2019).

Table 7.3 – Summary of results of the second questionnaire.

a. The audio chime:	Mean (SD)	b. Presenting the reason for the advice:	Mean (SD)
tells me when an advice is available	1.625 (1.516)	helps me understand the advice	1.500 (1.455)
is appropriate	1.063 (1.600)	makes it more likely I will follow the advice	1.667 (1.655)
is annoying	-0.969 (1.741)	is unsafe	-1.167 (1.485)
helps me understand the message	-0.438 (1.519)	is necessary	0.200 (1.759)
is unnecessary	-1.313 (1.570)	is confusing	-1.100 (1.620)
helps me keep my eyes on the road	-0.594 (1.599)	is distracting	-0.167 (1.593)
might startle me	-0.781 (1.653)		
		c. Giving me feedback about the results of my behaviour:	
		motivates me to follow the next advice	-0.333 (2.055)
		is unsafe	-0.867 (1.688)
		is unnecessary	0.567 (1.909)
		is distracting	0.033 (1.906)

The questionnaire also inquired into whether haptic feedback in the steering wheel would be preferred to signal a new advice. Three types of vibrations were presented to the 32 participants (left side, right side, both sides). The vibrations were positively evaluated in only 14 cases (14.58%). In the 42 cases the vibrations were disliked (43.75%), and in 38 (39.58%) cases the vibrations were evaluated positively if there was a way to turn them off. In 2 (2.09%) cases no evaluation was recorded.

We also inquired about whether to provide context for the advice, meaning whether to precede it with the reason for the advice and conclude it with feedback about the performed behaviour. Questions were again answered on a 7-point scale from -3 (disagree completely), to 0 (neutral), to 3 (agree completely), and results are shown in table 7.3-b. Participants indicated providing the reason prior to the advice helped them understand the advice better. Providing the reason before the advice also made it more likely they would follow the advice, did not feel unsafe, and was not confusing. On average participants were neutral about the necessity of providing the reason and whether it would be distracting. This neutral rating on necessity is remarkable, since participants indicated that providing the reason for the advice would help understand the advice and would make it more likely an advice will be followed.

Table 3-c displays results regarding providing feedback about the consequences of (not) following an advice. Providing the feedback was perceived as safe, and somewhat necessary. Participants were neutral about whether the feedback would motivate to follow more advices, or whether it would be distracting. The latter is likely because the participants lacked hands-on experience with the advices and as such were unsure about the effects of receiving the feedback.

7.2.5   Interface and Persuasive Message Characteristics – Conclusion

In this section we described the two questionnaires that were distributed. The goal of the questionnaires was to find driver preferences among the modalities used for the advice, its location, and how to best present the advice.

Results showed that participants preferred having an audio-visual multimodal interface where the advice was preceded by an auditory chime, and the advice displayed through a HUD. Adding haptic feedback was generally disliked, especially when the option to turn the vibrations off would be unavailable, we therefore chose to avoid using haptic feedback in our simulator study. Providing context will help participants understand when an advice is available and make it more likely that the advice will be followed. Participants were more divided on whether to provide feedback on their behaviour. We chose to include both in our simulator study to observe the effects.

The next section describes the development of the persuasive lane-specific advice HMI.

7.3   Developing Persuasive Advice Based on Driver Preferences

7.3.1   Lane-Specific Advice

Based on the results from the questionnaire studies, we developed persuasive advices that were preceded by the reason for the advice, and followed by feedback on the driver behaviour. Three types of advices were developed, two persuasive variants and one control. Advices for all conditions followed the same basic design of a diagram of the road with the ego vehicle displayed on the current lane as displayed in figure 7.4. The reasons for the advice were based on standard signage in use on Dutch motorways, so as to be quickly recognisable by participants. The reasons used in the experiment were congestion, and a lane-drop where the right lane would drop off. This type of lane-drop may occur when an incident has happened on the right lane, when there are road works, or where the rush-hour lane terminates.
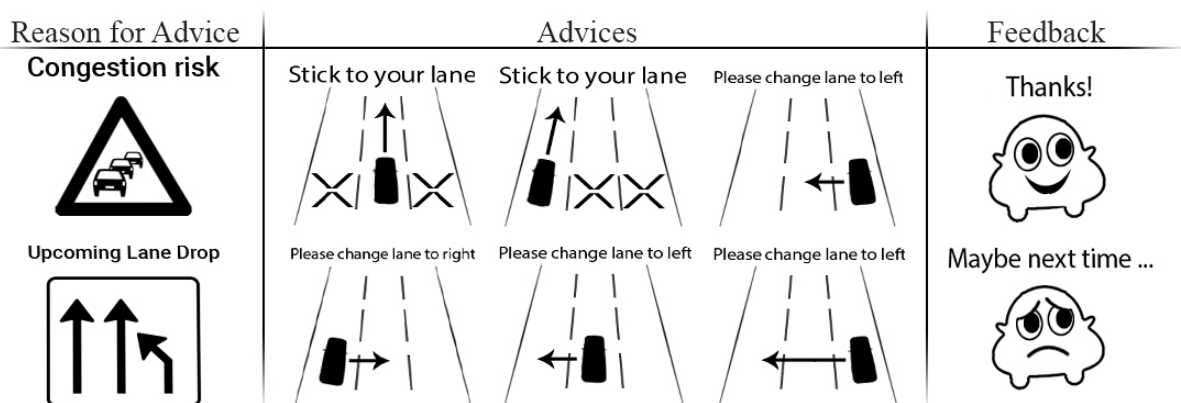
Fig. 7.4 Overview of the interfaces used in the simulator study. The figure shows possible advice reasons (left), lane-specific advices (middle), and feedback to the driver (right).

We split the gamified group into two conditions to be able to incorporate both competition and comparison from the study by Orji et al (Orji et al., 2014) as discussed in 1-C. Among the persuasive advices were a competitive type and a cooperative type. Variations for the competitive and cooperative group are displayed in Figure 7.5. In the competitive group the number of points to be earned is clearly displayed below an advice, and in the cooperative group the percentage of other drivers following their advice is displayed. Participants were informed that the number of drivers following their advice included those adhering to 'stick to your lane' advices.
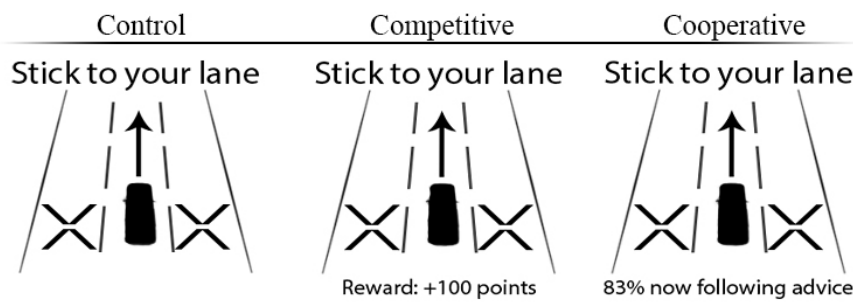
Fig. 7.5 Example of the three types of advices given in the simulator, in case of requiring the driver to stick to their lane. The competitive variant displayed the amount of points the driver could earn, the cooperative version displayed how many other drivers on the same road segment were following their advice.

7.3.2   Going Online: A Web-Portal Intervention

To limit effects on workload while driving, we chose to keep the advices simple and add a web-portal for both intervention groups. In this web-portal, drivers could at their own pace review their performance parameters. These included a page with information on their latest trips, as well as a page with the progress made to the next achievement. Aside from an insight into their performance, the web-portal gave participants an extended interaction with the avatar, whose emotion and comments changed depending on how well the participants had performed during their first driving session. The avatar's two emotional states are shown in Figure 7.1 and the full range of responses are shown in Table 7.4. The web-portal is shown in Fig 7.6.

Table 7.4 Table showing conditions and avatar comments.

Advices followed	Home page	Latest trip page
0 out of 4 *(sad avatar)*	Too bad [Name], can you do better next time?	You followed 0 out of 4 advices. That's too bad, can you try following more next time?
1, 2, 3 out of 4 *(happy avatar)*	[Name] you are doing good! Can you do better next time?	You followed [1, 2, 3] out of 4 advices. That's cool! Can you try following more next time?
4 out of 4 *(happy avatar)*	[Name] you are doing great! Can you keep it up?	You followed 4 out of 4 advices. Wow! Can you do that again next time?

The web-portal had a competitive and a cooperative variant. In both versions the avatar gave feedback to the driver depending on what part of the interface the participants clicked. Both versions also showed the participant's name, score, latest trip summary and next achievement. The information on the latest trip was dependent on the performance of the participants in their first driving session. The points required to unlock the next achievement were also based on performance during the drive, but scaled so that it was always attainable by following more advices on the second day than on the first, or an equal number of advices if all were followed the first day.

The competitive version had a leader board showing the participant's relative position to others. Like the upcoming achievement, the position on the leader board was also fixed for all participants. First place was always attainable by following more advices on the second day than on the first, or an equal number if all advices were followed the first day. The cooperative version of the web-portal showed the number of other drivers on the road that followed their advice while the participants were driving, including 'stick to your lane' advices.
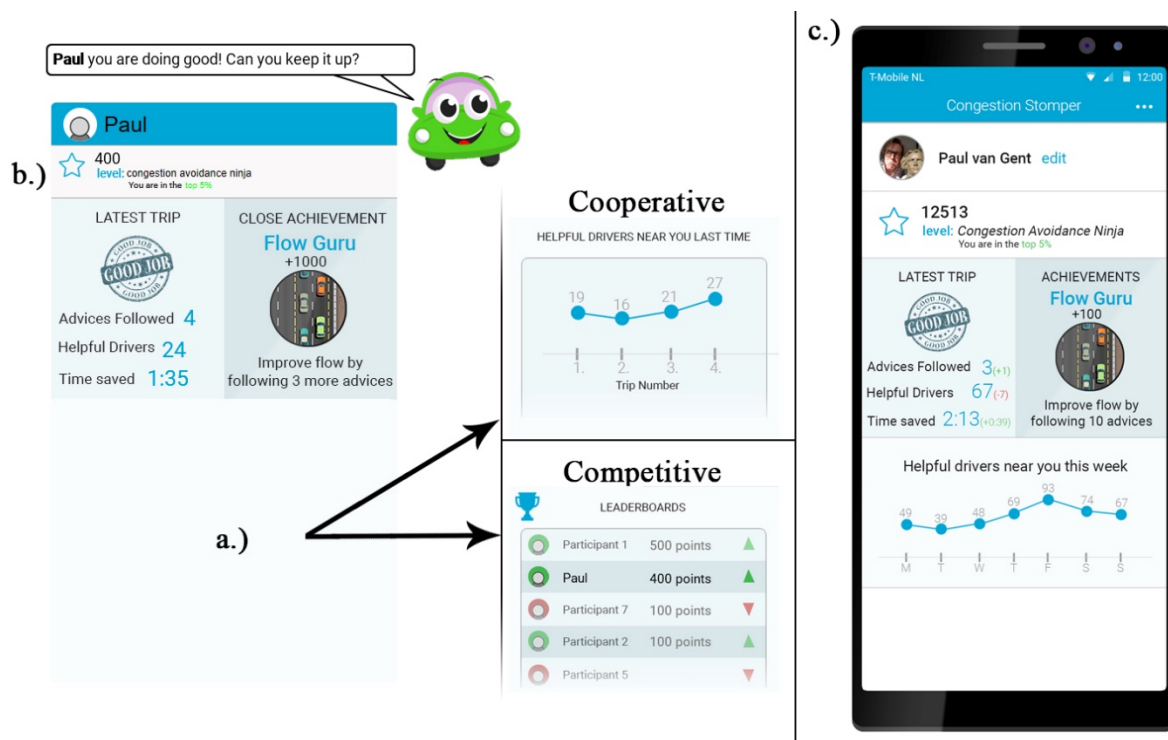
Fig. 7.6 The web-portal that was developed. Part a.) was variable. More information on the rank of the driver (i.e. 'top 5%' as shown in b.)) was only shown in the gamified condition. Part c.) shows an example of the cooperative portal as it looked when viewed on a phone.

7.4 Simulator Experiment

This section discusses the equipment used in the experiment, the scenarios developed and the procedure that was followed while collecting the data. In the simulator study we chose to use a persuasive approach that combines the mentioned techniques from the PSD that were found to work well across different personalities (Oinas-Kukkonen & Harjumaa, 2008; Orji et al., 2014; van Gent et al., 2019). To include the competition and comparison elements we decided to split the experiment into three groups: a competitive group where drivers could earn points and compete through a leader board, a cooperative group where drivers had real-time insight into how many other drivers followed their advices, and a control group. To incorporate the self-monitoring and suggestion without distracting the drivers we chose to implement a web-portal where drivers could review their performance (see figure 7.4, section 3.2). Praise and rewards were implemented using an avatar (see figure 7.2, figure 7.1, section 1.5), which we hoped would be instrumental in forming habits, which are a main factor in making persuasive effects last over time (Lally & Gardner, 2013). In this context it is a form of "dialogue support" and "primary task support" from the PSD (Oinas-Kukkonen & Harjumaa, 2008) and our theoretical framework (van Gent et al., 2019).

7.4.1 Equipment

A medium-fidelity driving simulator was used to perform the experiment. It consisted of three 4K (resolution 4096 * 2160 pixels) displays mounted on top of a dashboard mock-up. It provided participants with roughly 180-degree vision of their virtual surroundings. Fanatec

steering wheel and pedals were used along with custom key-based ignition and blinker controls were used. The simulation was run on the Unity3D game engine on a Windows 10 desktop pc.

Car kinematics were logged in Unity3D on the simulator pc. Participant responses and video recordings were logged on a Windows 10 laptop computer situated behind the participant out of their view, as not to be distracting.

7.4.2   Scenarios

We developed a congestion scenario and a lane-drop scenario. In the congested scenario, participants encountered a traffic jam after driving for several minutes. In the lane-drop scenario, participants encountered a lane-drop after the same amount of time had passed. We varied whether the reason for the advices was visible to the participants. In two scenarios the reason for the advice was visible ('congruent' scenarios), and in two others the reason for the advice was not visible ('incongruent' scenarios). For example, in a congruent lane-drop scenario participants encountered signage indicating an upcoming lane-drop together with a lane-specific advice, whereas in the incongruent version the signage and lane-drop were not encountered but the advice was given nonetheless. The same was true for the congested scenario; in the congruent version the overhead matrix signs indicated a reduced speed limit and a congested section was encountered, whereas in the incongruent version the traffic jam was too far ahead to be visible and no signage was active, but the advice was given. This gave a total of four scenarios. The type of advice was either non-persuasive (control group), competitive (competitive group), or cooperative (coop group) in nature.

Advices were developed as described in section 3.1. During the drive the advice was projected on a Heads-Up Display (HUD) in the centre of the car window. The choice for a HUD was made based on the questionnaire research and relevant literature, as described in section 1.5. and 2.4. The HUD was made semi-translucent, so it would not occlude any vital information from participants. In the competitive variant, the number of points to be earned was displayed below the advice, in the cooperative variant this was the percentage of other drivers currently following their advice.

Each scenario started on a highway-side parking lot. Participants had to start the vehicle, navigate off the parking area and merge onto the highway. After approximately two minutes participants were given an advice on the car's HUD. This advice was preceded by an alert that specified the reason for the advice (figure 7.4, left). The advice (figure 7.4, middle) was active for approximately 1.5 minutes and was lane-sensitive, meaning that the advice (change left, change right, stick to lane) updated real-time based on the lane participants were driving. After the advice period ended, feedback (figure 7.4, right) was displayed based on whether participants followed the advice. Traffic was programmed to drive defensively and give way to participants whenever they turned on their blinker or started a lane change. This was done to eliminate the situations where participants could not change lane due to other traffic as much as possible, so that we could observe the effects of the advices. It is also in line with our design goals of only generating an advice when the driver has the opportunity to follow it and when it is safe to do so (van Gent et al., 2019).

In the congestion scenario, participants were advised to either change to the middle lane of the three-lane highway, or stick to the middle lane if they were already driving there. In the congruent scenario the matrix signs above the highway were switched on and displayed a dynamic speed limit of 80 km/h. Congestion was visible in the distance when the advice was

given, and participants approached slow moving (15-20km/h) traffic while the advice was active. In the incongruent scenario, traffic was driving with a regular speed limit of 130km/h, the dynamic speed limit signs were off, and no congestion was encountered by participants.

In the lane-drop scenario, participants were advised to move to the leftmost lane in anticipation of the righter most lane dropping off. In the congruent scenario, signs announcing the lane-drop were posted at 1 km, 300 meters, at the start of the weaving section, and near the end of the weaving section, as specified by Dutch traffic regulations. In the incongruent scenario no signage was visible and no lane-drop was encountered by participants.

7.4.3 Competitive and Cooperative Interventions

Advices in the competitive version of the scenarios displayed the amount of points that participants could earn by following it. In the cooperative scenario the percentage of drivers currently following their advice was displayed alongside the advice. In the control group no extra information was displayed. See figure 7.5 for a visualisation of all three variations.

Participants were recruited to drive on two separate days, and in between both days those in the competitive and cooperative groups received a link to the web-portal. The two versions of the web-portal that showed the same general information but emphasized different aspects. The competitive version accentuated the amount of points earned, and participants could view their position relative to other participants through a leader board. Unbeknownst to the participants the web-portal placed every participant as second. The point-gap between them and the first position could in all cases be closed by following more advices on the second day. The cooperative version of the web-portal emphasized how many of the other drivers on the road followed advices. These data were fabricated and showed an upward trend of more drivers following advices recently.

For both groups the portal showed the travel time saved, advices followed, and their next achievement. The avatar communicated their performance and encouraged them to either keep up good performance when all advices were followed the first day, or encouraged participants to follow more advices the second day if they did not follow all advices during the first day. The avatar also communicated relevant details about their performance when they clicked the different parts of the site. Both web-portal versions are displayed in figure 7.6.

7.4.4 Procedure

A pilot study was performed to test the equipment, scenarios and experimental procedure. The hardware functioned properly, and participants had no trouble performing the tasks.

Approval for the experiment was obtained from the TU Delft ethics committee. Participants could apply for the experiment through e-mail, after which they received a copy of the informed consent and were allowed to ask any questions. During the first session participants were seated in the simulator and had a second opportunity to ask questions about the informed consent or procedure, and signed the document when all questions were answered. A familiarization scenario was first started. This scenario had no traffic and no advices so that participants could drive at their own pace and get used to the simulator. Once participants indicated they felt comfortable driving the car, the experiment started.

Prior to starting the experiment, participants received a written instruction. The document asked participants to drive as they would in everyday life and emphasized there was no desired behaviour. Rather, participants were made aware of the fact that, just as with a real-life in-car system, it is unknown what the accuracy of the given advice is. In the competitive group, participants were told they could earn points by following the advice and that the potential rewards would be displayed with the advice message. Those in the cooperative group were instructed that the system was a cooperative system that only worked when most of the people on the road followed the advices, and that the number of computer-controlled cars that 'chose' to follow their advice would be displayed real-time on the advice as well.

Participants were randomly assigned to control, competitive or cooperative groups and drove the four scenarios in a randomized order. At the end of the session participants filled in the van der Laan scale (van der Laan, Heino, & de Waard, 1997), a short questionnaire that measures perceived usefulness and satisfaction with advanced in-car systems.

Those in the competitive or cooperative group received an e-mail with a link to the web-portal after the first day, where they could view their performance in a personalised version of the portal. On the second day participants drove the same scenarios as the first day, again in a randomized order. At the end of the second day the van der Laan scale was filled in again.

During the familiarization drive and between scenarios, participants were asked for signs of discomfort and/or simulator sickness, and asked to indicate it the moment they experienced any discomfort.

7.5   Results

7.5.1   Participant Demographics

A total of 55 participants took part in the experiment. One participant dropped out due to simulator sickness. 24 (44.4%) of participants were female, 30 (55.6%) male, with an average age of 36.19 years (SD: 10.75). The participants were assigned randomly to conditions (control, competitive, cooperative) with 18 participants per condition.

All participants held a valid driver's license and drove regularly. 30 (55.6%) of participants drove at most 1,000 km per month, 14 (25.9%) between 1,000 and 2,000 km, 8 (14.8%) between 2,000 and 5,000 km, and 1 participant (1.85%) over 5000 km per month. One participant (1.85%) didn't know how many kilometres they drove every month.

28 participants (51.85%) indicated they regularly used a navigation device in while driving, 21 (38.89%) sometimes, and 5 (9.25%) rarely to never used a navigation device while driving.
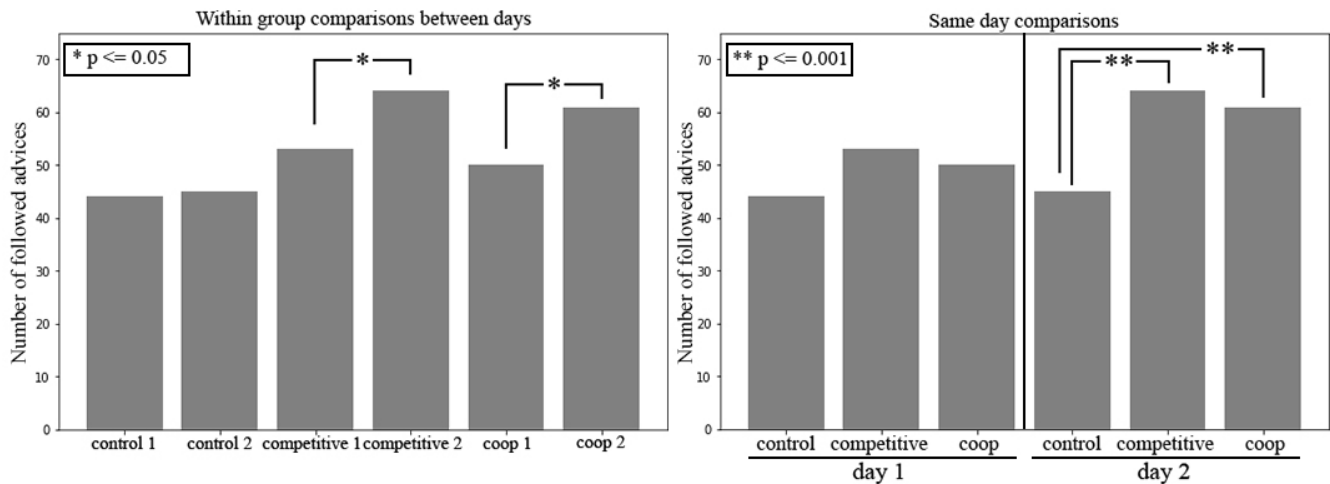
Figure 7.7 – a.) shows comparisons within groups between driving day 1 and 2. The right-hand plot b.) shows comparisons between days. Significant differences are marked with *.

7.5.2 Persuasive Effectiveness of Interventions

First, we analysed the total advices followed by each group. Levene's test for equality of variances indicated the assumption of equality of variances was violated, so instead of a T-test we used the Mann-Whitney U Rank Test, which does not assume equality of variances. With each result we give the test statistic 'U' and significance level 'p'. Out of 144 advices, participants in the control group followed 89 (61.81%) advices, in the competitive group 117 (79.17%) advices, and in the cooperative group 111 (77.08%) of advices. The difference between control and competitive groups was statistically significant ($U = 8352$, $p < .001$), as well as between the control and cooperative group ($U = 8784$, $p = .002$). The difference between competitive and cooperative groups was not statistically significant ($U = 9936$, $p = .193$). This indicates that both interventions were more effective than the control group, but there were no clear differences between them in effectiveness.

Second, we analysed the effects of the intervention given between both driving days. We used a Wilcoxon-Pratt Signed-Rank test, suitable for dependent (non-normal within-participant) data, to test the number of followed advices on the first and second day. With each result we give the test statistic 'Z' and significance level 'p'. The control group followed 44 advices on the first day and 45 advices on the second day, a difference that was not statistically significant ($Z = 720$, $p = .841$). Participants in the competitive group followed 53 advices on the first day and 64 on the second day, which was statistically significant ($Z = 252$, $p = .012$). Those in the cooperative group followed 50 advices on the first day and 61 on the second day, which was also statistically significant ($Z = 252$, $p = .012$). This indicated that after exposure to the web-interface, participants followed significantly more advices, and that the difference was not attributable to repeated exposure to the advices as the control group showed no significant difference. Results are visualised in Figure 7.7 a.).

Lastly, we analysed the differences between groups on the same days. Again, the assumption of equal variances was violated, so a Mann-Whitney U Test was used. With each result we give the test statistic 'U' and significance level 'p'. Each group was given a total of 72 advices per day, on both days. On the first day, participants in the control group followed 44, those in the competitive group 53, and those in the cooperative group 50 advices. The difference between

control and competitive and control and cooperative groups was not significant (U = 2268, p = .056, U = 2376, p = 0.148, respectively), and the difference between competitive and cooperative was not statistically significant either (U = 2484, p = .291). On the second day, those in the control group followed 50, those in the competitive group 64, and those in the cooperative group 61 advices. The differences between control and competitive and between control and cooperative were statistically significant (U = 1908, p < .001, U = 2016, p = .001, respectively), but the results between cooperative and competitive were not (U = 2484, p = .232). This indicates the effectiveness of the intervention: the first day no significant differences between the groups were observable, but on the second day differences emerged, with those in the competitive and cooperative groups following significantly more advices than those in the control group. Results are visualised in figure 7.7 b.).

Surprisingly, we found no statistically significant relation between whether or not the reason for the advices was visible (congruent vs incongruent) to the driver (t = .377, p = .706), which runs contrary to what has been observed before (Risto & Martens, 2013). It is possible this discrepancy results from participants driving in a simulator rather than in the real world.

Using a t-test (test statistic '*t*', significance level '*p*'), no statistically significant difference was found between advices followed and the lane-drop or the congestion advices on the first day (t = 1.963, p = .052), the second day (t = .364, p = .717), or both days combined (t = 1.634, p = .103). Furthermore, no statistically significant correlation was found between advices followed and gender (r = -0.150, p = 0.279), age (r = -0.072, p = 0.603), or average kilometres travelled per month (r = 0.139, p = 0.312).

7.5.3   Types of Advices and Behaviour

Participants were free to drive as they normally would. This meant that the types of advices given (change lane, stick to lane) were determined dynamically based on participant driving behaviour. Because this might skew results, we analysed the link between the types of advices given and the behaviour of participants as well.

In total 427 advices were given to participants during the experiment. 87 (20.37%) advices required drivers to stay in their lane, 229 (53.63%) advices required drivers to move one lane left or right, and 111 (26.00%) advices required drivers to move two lanes. No significant correlation existed between the choice to follow or not follow an advice and the number of lanes the driver had to change (r = 0.004, p = 0.941). The same held for within-group correlations for all groups: control (r = -0.034, p = 0.690), competitive (r = -0.060, p = 0.478), and cooperative (r = 0.160, p = 0.056). This ran contrary to our expectations. We expected stick-to-your-lane advices to be complied to more often, as these require less effort from the driver to follow compared to advices requiring a lane change.

7.5.4   Perceived Usefulness and Driver Satisfaction

The van der Laan scale (van der Laan et al., 1997) was used to assess both the perceived usefulness of, and the participants' satisfactions with, the lane-specific advice HMI. Both the perceived usefulness and the satisfaction scales range from -2 (low usefulness, low satisfaction), 0 (neutral usefulness, neutral satisfaction), to 2 (high usefulness, high satisfaction). The assumption of normality was not violated so the data was analysed using the appropriate t-tests depending on whether dependent or independent data were being analysed. Analysis follows the same pattern as in the previous section.

First, we analysed the differences between the groups. Perceived usefulness for the control group was 0.589, for the competitive group 1.072, and 1.006 for the cooperative group. The differences were statistically significant between control and competitive groups (t = -3.531, p = .001), and between control and cooperative groups (t = -3.277, p = .002), but not between competitive and cooperative groups (t = 0.427, p = .672). Satisfaction was 0.472 for the control group, 1.014 for the competitive group, and 0.931 for the cooperative group. The differences were statistically significant between control and competitive (t = -2.949, p = .006), and between control and cooperative (t = -2.692, p = .011), but not between competitive and cooperative groups (t = 0.477, p = .637). This indicates that in general, the competitive and cooperative advices were evaluated as more useful, and participants were more satisfied with them compared to the control group. We note that participants only had two short driving sessions to receive advices and become familiar with them. It is likely that satisfaction and perceived usefulness will increase or decrease over time as participants get more experienced with the advices and their effects.

Table 7.5 - The usefulness and satisfaction scores of the lane-specific advices, both the total as well as split by day.

	Both days		
	control	competitive	cooperative
Usefulness	0.589	1.072	1.006
Satisfaction	0.472	1.014	0.931

	Day 1			Day 2		
	control	competitive	cooperative	control	competitive	cooperative
Usefulness	0.578	1.033	0.900	0.600	1.111	1.111
Satisfaction	0.486	0.931	0.847	0.458	1.097	1.014

Second, we analysed the effects between driving days in the same groups. Data were analysed using a paired-samples t-test. Within the control group the perceived usefulness on day 1 was 0.578 and 0.600 on day 2, which was not statistically significant (t = 0.163, p = .872). Satisfaction was 0.486 on day 1 and 0.458 on day 2, which was not statistically significant (t = 0.243, p = .811). Within the competitive group the perceived usefulness was 1.033 on day 1 and 1.111 on day 2, which was not statistically significant (t = -0.999, p = .331). Satisfaction was 0.931 on day 1 and 1.097 on day 2, which was not statistically significant (t = -1.531, p = .144). In the cooperative group the perceived usefulness on day 1 was 0.900 and 1.111 on day 2, which was not statistically significant (t = -1.769, p = .095). Satisfaction was 0.847 on day 1 and 1.014 on day 2, a difference that was not statistically significant (t = -1.531, p = 0.144). This indicates no effects of the web portal on either perceived usefulness of, or satisfaction with the HMI, as there is no difference after receiving the intervention, and no significant increase between both driving days. Drivers did not need the web-portal to see the usefulness of the HMI or evaluate it as satisfying to use.

Lastly, we analysed the group differences on the same days. On the first day, perceived usefulness differed significantly between control (0.577) and competitive (1.033) groups (t = -2.888, p = .007) and between control and cooperative (0.900) groups (t = -2.240, p = .03), but not between competitive and cooperative groups (t = 0.777, p = .442). Satisfaction differed significantly between control (0.486) and competitive (0.931) groups (t = -2.031, p = .050), and between control and cooperative (0.847) groups (t = -2.097, p = .044), but not between competitive and cooperative groups (t = 0.436, p = .665). On the second day the same patterns

were present, with perceived usefulness differing between control (0.600) and competitive (1.111) groups (t = -3.239, p = .003), and between control and cooperative (1.111) groups (t = -3.053, p = .004), but not between competitive and cooperative groups (t = 0, p = 1.000). Satisfaction differed between control (0.458) and competitive (1.097) groups (t = -3.571, p = .001), and between control and cooperative (1.014) groups (t = -2.753, p = .009), but not between competitive and cooperative groups (t = 0.435, p = .666). The differences between the groups remained stable over time, confirming that the web-portal intervention did not seem to contribute significantly to overall usefulness of satisfaction scores.

7.6   Conclusion

In this paper we outlined the development of persuasive advices for a lane-specific advice HMI, with the goal of reducing congestion.

During the driving experiment participants drove the same scenarios on two different days. Those in the competitive group could earn points by following advices, those in the cooperative group could see how many others were following an active advice, and those in the control group only received an advice. Those in the competitive and cooperative groups viewed a web-portal in between both sessions where they could review their performance and were encouraged by an avatar. Results showed that, on a group level, the competitive and cooperative groups followed significantly more advices in total. Secondly, after exposure to the persuasive web-portal, those in the competitive and cooperative groups followed significantly more advices on the second day than on the first, which indicates the intervention's effectiveness. Finally, the differences between groups only emerged on the second day, meaning there was no significant behavioural difference between the groups prior to the intervention, but there was a significant difference after the intervention. This indicates the effectiveness of the persuasive intervention over the control group, but shows no clear distinction between the competitive or the cooperative approach to say which is more effective.

Based on the van der Laan scale, perceived usefulness and satisfaction were higher for both persuasive groups compared to the control group, but not between them. Over time there were no significant within-group changes between both driving days, although there was a slight upward trend in perceived usefulness for all groups, as well as for satisfaction in both treatment groups but not the control group. Differences between groups were also stable over time, with the cooperative and competitive HMI's being perceived as more useful. We interpret this as meaning the web-portal interface had no significant effect on overall perceived usefulness or satisfaction, but that both persuasive interventions were perceived as more useful and satisfying in use.

7.7   Discussion

Persuading drivers to follow a message that may not be in their personal benefit is a complex issue. The significant effects on driver willingness to follow advices are important in light of newly developed lane-specific (cooperative) advice systems. These systems only work to improve flow if drivers follow the advices generated, however drivers may be unwilling to do so until they see that doing so will benefit them (Risto & Martens, 2013, 2014). This creates a catch-22 situation where deployment of such a system may fail because for it to work drivers need to follow the advices, but drivers will not follow the advices until they see the that system works. Using persuasive advices in such a system creates an added incentive for drivers to follow the advices, which may boost the amount of advices followed, subsequently leading to

drivers observing benefits from the system which further reinforces willingness to follow lane-specific advices. This way the persuasive aspects are employed mainly in the early phases when rolling out a lane-specific or cooperative system. This overcomes a major limitation of such persuasive interventions, which is that persuasive effectiveness may reduce over time (Farzan et al., 2008a, 2008b), by stimulating the formation of habits. This is a key factor in making persuasive effects last over time (Lally & Gardner, 2013).

Based on what we discussed in this paper, when implementing persuasive in-car advice systems we recommend spreading information over multiple modalities to reduce impact on driver workload (Y C Liu, 2001), to keep the eyes-off-road time to a minimum (Peng et al., 2013), and to manage driver workload by timing messages to appropriate moments (van Gent et al., 2019). Using an avatar that shared driving goals with the driver, and a web-portal that gave insight into participant performance had a positive effect on driver willingness to comply with persuasive messages. While in this paper we describe the choices for and development of a visual advice combined with an auditory alert, an avatar and a web-portal, the approach taken for such systems is dependent on the required behaviours and the type of advice given.

When implementing a lane-specific advice system such as the one described in this paper, the accuracy of the given advices is of paramount importance. If the information is inaccurate, trust in the system erodes over time (Fox & Boehm-Davis, 1998) and participants might stop following advices altogether. This also includes situations where a driver may not be able to judge whether the information is trustworthy or not (Risto & Martens, 2013). Any such system, therefore, must ensure its advices are correct, and that information about the reason for the advice is visible to the driver.

7.7.1 Limitations

The present work consists of two questionnaires and a simulator study. Although all possible care was taken to make the generated videos and simulator scenarios as realistic as possible, differences between simulator and real-world driving do exist. Our study shows significant effects of gamification on driver persuasion to follow advices. However, in real-world driving other factors like time-pressure, driver mood, weather conditions or the behaviour of other drivers might influence driver willingness to follow an advice, among other factors. When using a driving simulator in research, its validity is usually relative rather than absolute (Carsten & Jamson, 2011), meaning that behavioural effects found translate to the real world, but that effect magnitudes might differ. Wang et al. (Wang et al., 2010) performed an evaluation on using medium fidelity simulators to test in-car interfaces, and found that the effects of in-car interfaces can be effectively investigated using medium fidelity driving simulators.

The two questionnaires were based on 20 and 34 respondents, and the simulator study on 54. Self-selection bias may be present, since we put out adverts for all study steps and participants were free to apply themselves. Although the sample size is adequate for the analyses performed, as is often the case a larger sample size will make the results more generalizable. This is especially since, although the sample is diverse, it still consists mainly of Caucasian Europeans. Results may differ among ethnicities.

Lastly, since we developed the interface for a specific goal during the design phase, it is conceivable that different persuasive goals, or different environments in which the persuasive intervention is applied, will lead to different HMI requirements. This means that for different application domains, the HMI discussed in this paper needs to be validated.

7.7.2   Next Steps and Recommendations

Following the mentioned limitation of potential differences between simulated and real-world driving, as a next step we recommend an on-road trial to evaluate the persuasive HMI in real-world driving conditions. Ideally such a study would take place in a naturalistic driving setting over a longer period. This will give insight into how persuasive advice following might change over time.

A second recommendation relates to our theoretical model on driver persuasion (van Gent et al., 2019). To improve safety and effectiveness of the advices we suggested to time them to a moment where the driver' workload is low. This can be achieved by integrating the persuasive HMI with for example a workload estimator (van Gent, Melman, Farah, van Nes, & van Arem, 2018) to make the interface adaptive (Birrel, Young, Stanton, & Jennings, 2017).

Third, the motivations for following an advice as offered are different between the gamified condition, where participants could earn points, and the cooperative condition, where participants mainly had a social motivation to follow advices. We know from research that different personalities are sensitive to different types of persuasion (Kaptein, Markopoulos, Ruyter, & Aarts, 2009). Investigating this in the context of persuasive in-car advice is an interesting avenue for future research.

Lastly, in the present study only two advice contexts were tested: congestion ahead, lane-drop ahead. More reasons for giving an advice exist, such as road works, an accident, or adverse weather conditions. Although we found no statistically significant differences in numbers of advices followed between the congestion and lane-drop scenarios, it may still be that drivers show different compliance rates to different advice contexts. This should be examined in a future study.

Regarding recommendations for applying persuasive systems to in-car settings in practice, based on what we discussed in the paper and on the results, we recommend that:
- An app or web-portal is combined with the in-car HMI, to reduce information clutter on the in-car HMI, and for the drivers to review their progress at their own pace.
- An avatar is used to encourage drivers. The avatar should share the driver's goals.
- Auditory or haptic feedback have the option to be turned off.
- The visual HMI is only on when it needs to be.
- If an HMI is used, it is best to reduce salience (e.g. increase transparency or reduce brightness) or not use the HMI at all during conditions of poor visibility, such as fog, heavy rain, or darkness. This is to prevent dangerous situations related to the Mandelbaum effect.

References

Avineri, E. (2011). Applying behavioural economics in the design of travel information systems. 43rd Universities Transport Study Group Conference, 1–12. https://doi.org/10.1017/CBO9781107415324.004

Bahadoor, K., & Hosein, P. (2016). Application for the detection of dangerous driving and an associated gamification framework. Proceedings - 2016 4th International Conference on Future Internet of Things and Cloud Workshops, W-FiCloud 2016, 276–281. https://doi.org/10.1109/W-FiCloud.2016.63

Bailenson, J. N., Blascovich, J., Beall, A. C., & Loomis, J. M. (2001). Equilibrium theory revisited: Mutual gaze and personal space in virtual environments. Presence: Teleoperators and Virtual Environments, 10(6), 583–598. https://doi.org/10.1162/105474601753272844

Birrel, S., Young, M., Stanton, N., & Jennings, P. (2017). Using Adaptive Interfaces to Encourage Smart Driving and Their Effect on Driver Workload. 484, 764. https://doi.org/10.1007/978-3-319-41682-3

Carsten, O., & Jamson, A. H. (2011). Driving simulators as research tools in traffic psychology. In Handbook of Traffic Psychology. https://doi.org/10.1016/B978-0-12-381984-0.10007-4

CBS. (2019). Personenautoverkeer. Retrieved from https://www.cbs.nl/nl-nl/maatschappij/verkeer-en-vervoer/transport-en-mobiliteit/mobiliteit/verkeer/categorie-verkeer/personenautoverkeer

Cialdini, R. B. (2006). Influence: The Psychology of Persuasion. https://doi.org/10.1017/CBO9781107415324.004

de Waard, D. (1996). The Measurement of Drivers ' Mental Workload. https://doi.org/10.1016/j.apergo.2003.11.009

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From Game Design Elements to Gamefulness: Defining "Gamification." Proceedings of the 15th International Academic MindTrek Conference on Envisioning Future Media Environments - MindTrek '11, 9–11. https://doi.org/10.1145/2181037.2181040

Ecker, R., Holzer, P., Broy, V., & Butz, A. (2011). Ecochallenge: a race for efficiency. Proceedings of the 13th International Conference on Multimodal Interfaces - ICMI '11, 13, 91–94. https://doi.org/10.1145/2037373.2037389

Edgar, G. K. (2007). Accommodation, cognition, and virtual image displays: A review of the literature. Displays, 28(2), 45–59. https://doi.org/10.1016/j.displa.2007.04.009

Engen, T., Lervåg, L. E., & Moen, T. (2009). Evaluation Of IVIS/ADAS using driving simulators. 16th ITS World Congress, 1–10.

Farzan, R., Dimicco, J. M., Millen, D. R., Brownholtz, B., Geyer, W., & Dugan, C. (2008a). Results from Deploying a Participation Incentive Mechanism within the Enterpris. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'08), 563–572. https://doi.org/10.1145/1357054.1357145

Farzan, R., Dimicco, J. M., Millen, D. R., Brownholtz, B., Geyer, W., & Dugan, C. (2008b). When the experiment is over : Deploying an incentive system to all the users. Proceedings of the AISB 2008 Symposium on Persuasive Technology, 33–38. https://doi.org/10.1145/2145204.2145362

Fogg, B. J. (2003). Persuasive Technology: Using computers to change what we think and do. Elsevier Science & Technology.

Fox, J. E., & Boehm-Davis, D. A. (1998). Effects of Age and Congestion Information Accuracy of Advanced Traveler Information Systems on User Trust and Compliance. Transportation Research Record, 1621(1), 43–49. https://doi.org/10.3141/1621-06

Fuller, R. (2005). Towards a general theory of driver behaviour. Accident; Analysis and Prevention, 37(3), 461–472. https://doi.org/10.1016/j.aap.2004.11.003

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work? - A literature review of empirical studies on gamification. Proceedings of the Annual Hawaii International Conference on System Sciences, 3025–3034. https://doi.org/10.1109/HICSS.2014.377

Herslund, M.-B., & Jørgensen, N. O. (2003). Looked-but-failed-to-see-errors in traffic. 35(May 2001), 885–891. https://doi.org/10.1016/S0001-4575(02)00095-7

Ho, C., Kingdom, U., & Reed, N. (2007). Multisensory In-Car Warning Signals for Collision Avoidance. 49(6), 1107–1114. https://doi.org/10.1518/001872007X249965.

Horberry, T., Anderson, J., Regan, M. a., Triggs, T. J., & Brown, J. (2006). Driver distraction: The effects of concurrent in-vehicle tasks, road environment complexity and age on driving performance. Accident Analysis and Prevention, 38(1), 185–191. https://doi.org/10.1016/j.aap.2005.09.007

Hutchison, D., & Mitchell, J. C. (2008). Persuasive Technology. Persuasive, 5033(December), 1–5. https://doi.org/10.1007/978-3-540-68504-3

Kahneman, D. (2013). Thinking Fast and Slow. In Journal of Chemical Information and Modeling. https://doi.org/10.1017/CBO9781107415324.004

Kaptein, M., Lacroix, J., & Saini, P. (2010). Individual differences in persuadability in the health promotion domain. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 6137 LNCS, 94–105. https://doi.org/10.1007/978-3-642-13226-1_11

Kaptein, M., Markopoulos, P., Ruyter, B. De, & Aarts, E. (2009). Can You Be Persuaded? Individual Differences in Susceptibility to Persuasion. Human-Computer Interaction–INTERACT, 115–118.

Kim, S., & Dey, A. K. (2009). Simulated augmented reality windshield display as a cognitive mapping aid for elder driver navigation. Chi, 9. https://doi.org/10.1145/1518701.1518724

Kim, S., & Dey, A. K. (2015). Augmenting human senses to improve the user experience in cars: applying augmented reality and haptics approaches to reduce cognitive distances. Multimedia Tools and Applications. https://doi.org/10.1007/s11042-015-2712-4

Knoop, V. L., De Bakker, P. F., Tiberius, C. C. J. M., & Van Arem, B. (2013). Single Frequency Precise Point Positioning: Obtaining a map accurate to lane-level. IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, (Itsc), 2255–2261. https://doi.org/10.1109/ITSC.2013.6728563

Knoop, Victor L., De Bakker, P. F., Tiberius, C. C. J. M., & Van Arem, B. (2017). Lane Determination with GPS Precise Point Positioning. IEEE Transactions on Intelligent Transportation Systems, 18(9), 2503–2513. https://doi.org/10.1109/TITS.2016.2632751

Lally, P., & Gardner, B. (2013). Promoting habit formation. Health Psychology Review, 7, S137–S158. https://doi.org/10.1080/17437199.2011.603640

Liu, Y C. (2001). Comparative study of the effects of auditory, visual and multimodality displays on drivers' performance in advanced traveller information systems. Ergonomics, 44(4), 425–442. https://doi.org/10.1080/00140130010011369

Liu, Yung Ching, & Wen, M. H. (2004). Comparison of head-up display (HUD) vs. head-down display (HDD): Driving performance of commercial vehicle operators in Taiwan. International Journal of Human Computer Studies, 61(5), 679–697. https://doi.org/10.1016/j.ijhcs.2004.06.002

Nacke, L. E. ., Bateman, C. ., & Mandryk, R. L. . (2011). BrainHex: Preliminary results from a neurobiological gamer typology survey. In Entertainment Computing - ICEC 2011: (pp. 288–293). https://doi.org/10.1007/978-3-642-24500-8_31

Nousias, S., Tselios, C., Bitzas, D., Amaxilatis, D., Montesa, J., Lalos, A. S., … Chatzigiannakis, I. (2019). Exploiting Gamification to Improve Eco-driving Behaviour: The GamECAR Approach. Electronic Notes in Theoretical Computer Science, 343, 103–116. https://doi.org/10.1016/j.entcs.2019.04.013

Oinas-Kukkonen, H. (2013). A foundation for the study of behavior change support systems. Personal and Ubiquitous Computing, 17(6), 1223–1235. https://doi.org/10.1007/s00779-012-0591-5

Oinas-Kukkonen, H., & Harjumaa, M. (2008). A Systematic Framework for Designing and Evaluating Persuasive Systems. Persuasive Technology, Third International Conference, PERSUASIVE 2008, Oulu, Finland, June 4-6, 2008. Proceedings, 5033, 164–176. https://doi.org/10.1007/978-3-540-68504-3

Oinas-Kukkonen, H., & Harjumaa, M. (2009). Persuasive systems design: Key issues, process model, and system features. Communications of the Association for Information Systems, 24(1), 485–500.

Orji, R., Mandryk, R. L., & Vassileva, J. (2015). Gender, Age, and Responsiveness to Cialdini's Persuasion Strategies. In Persuasive Technology, Third International Conference, PERSUASIVE 2008, Oulu, Finland, June 4-6, 2008. Proceedings (Vol. 9072, pp. 147–159). https://doi.org/10.1007/978-3-319-20306-5

Orji, R., Vassileva, J., & Mandryk, R. L. (2014). Modeling the efficacy of persuasive strategies for different gamer types in serious games for health. User Modeling and User-Adapted Interaction, 24(5), 453–498. https://doi.org/10.1007/s11257-014-9149-8

Owens, D. A. (1979). The Mandelbaum effect: evidence for an accommodative bias toward intermediate viewing distances. Journal of the Optical Society of America, 69(5), 646–652. https://doi.org/10.1364/JOSA.69.000646

Peng, Y., Boyle, L. N., & Hallmark, S. L. (2013). Driver's lane keeping ability with eyes off road: Insights from a naturalistic study. Accident Analysis and Prevention, 50, 628–634. https://doi.org/10.1016/j.aap.2012.06.013

Reyes, M. L., & Lee, J. D. (2004). The Influence of IVIS Distractions on Tactical and Control Levels of Driving Performance. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 48(September 2004), 2369–2373. https://doi.org/10.1177/154193120404801935

Risto, M. (2014). Cooperative In-Vehicle Advice. University of Twente.

Risto, M., & Martens, M. H. (2013). Factors Influencing Compliance to Tactical Driver Advice: An Assessment Using a Think-Aloud Protocol. Proceedings of the 16th International IEEE

Annual Conference on Intelligent Transportation Systems (ITSC 2013), 1923–1928. https://doi.org/10.1109/ITSC.2013.6728510

Risto, M., & Martens, M. H. (2014). Assessing driver's ability to estimate compliance rates to in-car, advisory driver support. European Transport Research Review, 6(3), 287–294. https://doi.org/10.1007/s12544-013-0129-6

Sarter, N. B. (2006). Multimodal information presentation: Design guidance and research challenges. International Journal of Industrial Ergonomics, 36(5), 439–445. https://doi.org/10.1016/j.ergon.2006.01.007

Schakel, W. J., Arem, B. Van, & Netten, B. D. (2010). Effects of Cooperative Adaptive Cruise Control on Traffic Flow Stability. Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference On, (Idm), 759–764. https://doi.org/10.1109/ITSC.2010.5625133

Schakel, W. J., & Van Arem, B. (2014). Improving traffic flow efficiency by in-car advice on lane, speed, and Headway. IEEE Transactions on Intelligent Transportation Systems, 15(4), 1597–1606. https://doi.org/10.1109/TITS.2014.2303577

Scott, M., Pereira, L., & Oakley, I. (2012). Show Me or Tell Me: Designing Avatars for Feedback. Interacting with Computers, 27(4), 458–469. https://doi.org/10.1093/iwc/iwu008

Shi, C., Lee, H. J., Kurczak, J., & Lee, A. (2012). Routine Driving Infotainment App: Gamification of Performance Driving. Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI'12), 181–183.

Steinberger, F., Proppe, P., Schroeter, R., & Alt, F. (2016). CoastMaster: An ambient speedometer to gamify safe driving. Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, 83–90. https://doi.org/10.1145/3003715.3005412

van der Laan, J. D., Heino, A., & de Waard, D. (1997). A Simple Procedure for the Assessment of Acceptance of Advanced Transport Telematics. Transportation Research Part C: Emerging Technologies, 5(1), 1–10.

Van Driel, C., & Van Arem, B. (2010). The Impact of a Congestion Assistant on Traffic Flow Efficiency and Safety in Congested Traffic Caused by a Lane Drop. Journal of Intelligent Transportation Systems, 14(4), 197–208. https://doi.org/10.1080/15472450.2010.516226

van Gent, P., Farah, H., van Nes, N., & van Arem, B. (2019). A conceptual model for persuasive in-vehicle technology to influence tactical level driver behaviour. Transportation Research Part F: Traffic Psychology and Behaviour, 60, 202–216. https://doi.org/10.1016/j.trf.2018.10.004

van Gent, P., Melman, T., Farah, H., van Nes, N., & van Arem, B. (2018). Multi-Level Driver Workload Prediction Using Machine Learning and Off-The-Shelf Sensors. Transportation Research Record: Journal of the Transportation Research Board. https://doi.org/10.1177/0361198118790372

Verberne, F., Ham, J., & Midden, C. (2012). Trusting Automation Technology for Safer Roads: The Effect of Shared Driving Goals. Persuasive Technology, 57, 57–60.

Wang, Y., Mehler, B., Reimer, B., Lammers, V., D'Ambrosio, L. A., & Coughlin, J. F. (2010). The validity of driving simulation for assessing differences between in-vehicle informational interfaces: A comparison with field testing. Ergonomics, 53(3), 404–420. https://doi.org/10.1080/00140130903464358

Wickens, C. D. (2002). Multiple resources and performance prediction. Theoretical Issues in Ergonomics Science, 3(2), 159–177. https://doi.org/10.1080/14639220210123806

Wold, S., Esbensen, K., & Geladi, P. (1987). Principal Component Analysis. Chemometrics and Intelligent Laboratory Systems, 2(1–3), 37–52. https://doi.org/10.1016/0169-7439(87)80084-9

Yao, S., Knoop, V. L., & van Arem, B. (2017). Optimizing traffic flow efficiency by controlling lane changes: Collective, group, and user optima. Transportation Research Record, 2622(1), 96–104. https://doi.org/10.3141/2622-09

Young, M. S., Brookhuis, K. A., Wickens, C. D., & Hancock, P. A. (2015). State of science: mental workload in ergonomics. In Ergonomics (Vol. 58, pp. 1–17). https://doi.org/10.1080/00140139.2014.956151

Zhang, K., Batterman, S., & Dion, F. (2011). Vehicle emissions in congestion: Comparison of work zone, rush hour and free-flow conditions. Atmospheric Environment, 45(11), 1929–1939. https://doi.org/10.1016/j.atmosenv.2011.01.030

Chapter 8.

Findings, Discussion, and Conclusions

This dissertation investigated how to persuade drivers to follow lane-specific advices in dense traffic situations, for the purpose of optimizing the traffic distribution on the different lanes of a motorway. This final chapter overviews the main findings first (8.1). This is followed by a discussion of the main findings (8.2), the methodological choices made (8.3), recommendations for future research (8.4), and recommendations for practice (8.5).

8.1 Main Findings

This section summarises the main findings and contributions made to science and practice by the research embedded in this thesis.

8.1.1 A Conceptual Model for Persuasive In-Vehicle Technology to Influence Tactical Level Driver Behaviour

The conceptual model was developed to persuade drivers to change their tactical driving behaviour. Targeting short-term behavioural responses, such as those at the tactical level, has been shown to lead to more effective persuasion compared to persuading someone to change longer term behavioural patterns.

The behavioural component in the conceptual model is based on the theory of planned behaviour. This model was chosen to represent driver behaviour after a review of existing models for behaviour, based on its capacity to explain both short-term behavioural responses and the longer-term effects related to social and attitudinal factors. Especially the term

perceived behavioural control was of importance, since it represents the control an individual perceives to have over their behaviour and is thus strongly related to persuasive effectiveness and safety: if a driver has no confidence in their capacity to follow an advice, they will not do so.

The conceptual model describes the interaction between the driver and the persuasive system by defining aspects of the information transfer between them. It is divided into three layers: the system level, the information transfer level, and the driver level. At the system level the persuasive strategy is planned, and it is decided whether it is executed based on the evaluation of the safety criteria such as driver workload, distraction, and potential unsafe situations. These safety criteria should act as a decision filter on whether to continue or not. The interaction between the driver and the persuasive advice system is captured in the information transfer level, where the content, modality and timing of the information transfer is of importance. The driver level describes driver behaviour based on the theory of planned behaviour and extends this theory with effects from workload, indirect behavioural effects, driver characteristics, and driver safety.

8.1.2 Multi-Level Driver Workload Prediction Using Machine Learning and Off-the-Shelf Sensors

Communicating with the driver can cause unsafe situations if the state of the driver is unknown. The risk can be reduced by estimating the driver state in real-time and timing the communication to moments where the driver still has spare capacity to perceive, process and if necessary, act upon given advices or information.

The studies captured heart rate, skin response, blink rate from each participant together with kinematic vehicle data from the simulated car. Data were pre-processed to extract useful features such as heart rate variability measures, which were used to train a support vector machine regressor and a random forest regressor. Workload data was collected using a self-report scale. Based on the collected data a workload prediction model was developed and shown to perform well with individual or group-based models, but was less suited to generalizing to unknown drivers.

8.1.3 Analysing Noisy Driver Phsyiology Real-Time Using Off-the-Shelf Sensors: Heart Rate Analysis Software

During the development of the workload prediction model we developed an open source heart rate analysis algorithm. Especially heart rate variability (HRV) analysis requires robust and accurate peak detections because HRV measures are sensitive to outliers, and existing software performed unsatisfactorily. The problem is compounded by the need to capture heart rate data while participants are driving, which increases the noisiness of the signal due to sensor movements, skin deformations, and muscle activity. Noisy signals in turn make the analysis more difficult. For that purpose, an analysis algorithm was developed and implemented in both Python and embedded C.

The software was developed with low cost off-the-shelf sensors in mind, which can introduce extra noise. This choice was made to allow researchers, regardless of budgetary constraints, to collect and analyse heart rate data using this toolbox. The development of this toolbox was split into three parts: pre-processing tools, peak detection, and heart rate analysis.

The implemented pre-processing tools in the toolbox help to prepare the collected signals for analysis. Several finite impulse response (FIR) filters are available to reduce specific frequency bands from the signal. This is useful for example to remove the characteristic 50 or 60 Hz 'power mains hum' (the frequency of AC power often leaks into the data through interference). Peak enhancement tools for ECG and baseline wander removal tools have been developed as well, along with some further improvements such as colour-blind support and nonlinear analysis tools [14].

8.1.4   HeartPy: A Novel Heart Rate Algorithm for the Analysis of Noisy Signals

The algorithm was implemented in Python under the name HeartPy and is freely available through GitHub and the Python Package Index (pip). Several implementations have also been developed for (wearable) embedded environments such as the Arduino platform, Teensy platform, and most 8-bit RISC (Reduced Instruction Set Computer) and ARM (Advanced RISC Machine) chipsets. These implementations allow for basic signal collection, basic peak finding, and full analysis similar to what HeartPy does but within the constraints of the more limited RISC architecture. Together with the Python package this enables researchers to use low-cost heart rate collection and analysis methods in their research.

The performance of HeartPy was quantified and compared to two other available methods of heart rate analysis (pan-tompkins (Pan & Tompkins, 1985), ECGViewer (Ramshur, 2010) and the results showed that Heartpy performs equally well or better on both PPG and ECG datasets. A method of extracting breathing rate from the PPG signal was validated a publicly available dataset as well and shown it can be used as a good approximation of actual breathing rate.

8.1.5   The Persuasive Automobile: Design and Evaluation of a Persuasive Lane-Specific Advice Human Machine Interface

To design the HMI an iterative approach was used with two sequential questionnaire studies. These served to determine optimal advice location, modalities through which to transmit the information, and the information flow of the advice. An avatar was designed to give limited feedback to a driver after completing an advice.

The persuasive HMI was tested in two driving simulator scenarios: a congestion scenario and a lane-drop scenario. Advices were either baseline, gamified, or cooperative in nature. More advices were followed in the gamified (79.17%) and cooperative groups (77.08%) compared to the baseline (61.81%). Between sessions the treatment groups were exposed to a web-portal that displayed their progress. After this in the second day, the treatment groups showed a significant increase in the number of advices followed compared to the first day. On day 2 those in the control group followed: 45 advices (+2.3%), those in the gamified: 64 (+20.8%), and those in the cooperative group: 61 (+15.1%) advices. Significant differences in advice following behaviour did not exist between the groups on the first day, but only emerged on the second day. This indicates the web-portal is an effective intervention, as it shows that the groups behaved similarly on the first day, but not on the second day. No significant difference was observed between the two treatment groups, meaning that the cooperative and competitive persuasion were equally effective.

---

[14] These are not mentioned in chapter 4 but are available on the GitHub repository mentioned.

Perceived usefulness and satisfaction showed higher perceived usefulness and satisfaction with the persuasive HMI for both treatment groups compared to the control group, but again no significant difference between the treatment groups. No significant changes regarding perceived usefulness and satisfaction occurred between both driving days and differences between groups were also stable over time. This indicates that there seems to be no effect of the web-portal on perceived usefulness or satisfaction, but that the persuasive interventions were perceived to be more useful and satisfying to use than the control version.

8.2   Discussion of Main Findings

The main research question for this thesis, as posed in chapter 2, was: *How can we persuade a driver to follow a lane-specific advice without enforcing behaviour?*

This question was divided into three sub-questions, being:

1. How to communicate with the driver? Fundamental requirements for a persuasive system to be effective and safe.
2. When to communicate with the driver? Timing messages to low workload periods is safer and more likely to persuade.
3. What to communicate with the driver? Design of a persuasive HMI system.

This section discusses how the research presented in this thesis helps answer the sub-questions and the main research question.

8.2.1   How To Communicate With the Driver?

The context for this thesis was a lane-specific advice system that would allow for reduced congestion by advising drivers to spread out over the available road space. The first question posed was how to get a message across to the driver. This implies knowing what communication strategy to follow, what modality to use to deliver the information, and what external factors to take into account as well.

Chapter 3 developed a conceptual model for driver persuasion targeting tactical level behaviours. Managing demands on the driver is crucial from a safety perspective, as demands exceeding driver capability can lead to loss-of-control or collision situation. In the theoretical model, the tactical level was selected because it contains mostly skill-based and rule-based behaviours from Rasmussen's taxonomy, behaviours which are less demanding on the driver than the more complex knowledge-based behaviours. On top of selecting the tactical level as less likely to increase workload by much, communications with the driver should be timed to periods when the driver has spare capacity (e.g. low workload) as well. The communication strategy is thus: target tactical level driver behaviours and wait for a period of low workload before communictating with the driver. Alternatively, when dealing with time-critical messages such as lane-specific advice, rather than waiting for a period of low workload for one driver, it is of course equally possible to select drivers with the lowest workload from a larger pool of drivers as well.

Chapter 7 described a combination of questionnaires and a simulator study. Based on the information bandwidth and speed of information uptake, the visual modality was selected as a primary information carrier of communications to the driver. This would lead to the lowest increase in workload and was most preferred by drivers. In the questionnaires, drivers preferred

the Heads-Up Display (HUD) over a central console- or speedometer-based display, and preferred that messages be announced by an auditory chime. The HUD in combination with the chime would allow them to keep their eyes on the road. Drivers indicated they would at all times like control to turn the chime off. Haptic feedback in the steering wheel was not a suitable replacement for the chime, as it was seen as distracting and annoying by drivers.

In conclusion, the research presented in this thesis shows that communicating lane-specific advice to a driver is best done by:
- requesting a tactical level behaviour from the driver.
- targeting communications to periods of low driver workload.
- using the visual modality.
- having messages announced by an auditory chime, which the driver can turn off if desired.

8.2.2   When To Communicate With the Driver?

Based on the theoretical model described in chapter 3, the best moment to initiate communication with a driver is when their workload is low and their perceived behavioural control is high. This ensures driver safety because requiring low workload prior to communicating with the driver means the chance that the driver will become overloaded is small. Targeting low workload periods also increases the success of any persuasive attempt because both perceived behavioural control and ability to perform the requested behaviour are high in such a situation.

Measuring driver workload is a complex topic, and results from the field as to what physiological signals can be used to measure it have been diverse and sometimes contradictory. We hypothesized that one reason for the discrepancies is the frequent use of heart rate variability (HRV) as a workload proxy. HRV as a set of metrics is highly sensitive to outliers in the annotation of individual heart beats from the heart rate signal, which is usually done through automated software. Especially for low-cost PPG sensors, accurate analysis software was not available open source. This prompted the development of HeartPy as an accurate heart rate analysis toolkit capable of handling these types of noisy PPG signals. Chapters 5 and 6 describe the development and validation of HeartPy, and showed good performance on diverse datasets.

HeartPy was used to analyse heart rate signals from two driving simulator studies, which showed that in normal driving conditions workload of a driver can be determined using ML methods. The main constraint was that the models could not robustly generalise to unknown drivers in regular driving conditions, likely because workload was never very high. Group-based or individualised models could predict workload well. To test this assumption a second study was performed that induced high workload in drivers with a difficult forced-pace lane keeping task. In this study, models were much more capable of generalising to unknown drivers, although performance was still not excellent. A larger and more diverse dataset, along with more powerful ML models, will likely improve on this performance gap.

Together, these chapters show not just that workload measurement in drivers is possible by using group-based or individualised models, but also provide the tools to automatically annotate and analyse noisy PPG and ECG heart rate signals collected in the field and with low-cost devices.

8.2.3   What To Communicate With the Driver?

The last subquestion was what content should be communicated to the driver to ensure a high adherence to lane-specific advices.

Persuading drivers to follow messages that are in the collective benefit, but may not be in their personal benefit, is tricky. Based on the development of the theoretical model, various persuasive techniques were identified fom the Persuasive Systems Design (PSD) model that can help make a message more enticing and likely to be followed. In this thesis gamification and socially cooperative approaches were chosen, as they each incorporate different persuasive elements as outlined in the Persuasive Systems Design (PSD) model, as well as influence the 'social norms' component from the Theory of Planned Behaviour (TPB).

When implementing persuasive in-car advice systems it is recommended to spread information over multiple modalities to reduce impact on driver workload (for example use an auditory alert chime to announce a visual message), to keep the eyes-off-road time to a minimum, and to manage driver workload by timing messages to appropriate moments. Using an avatar that shared driving goals with the driver coupled with a web-portal that gave insight into their own performance, had a positive effect on driver willingness to comply and actual compliance with persuasive messages. A simulator study confirmed that both the gamified and cooperative advices, combined with the avatar and web-portal, increased the advices followed significantly compared to only giving drivers an advice. Drivers indicated that presenting the reason for the advice prior the actual advice would lead to a higher compliance to the advices, but that feedback on the results of their behaviour might have a detrimental effect.

8.2.4   Tying It Into the Main Research Question

The main research question was *How can we persuade a driver to follow a lane-specific advice without enforcing behaviour?*

The conceptual model outlined in this thesis can help designers of such systems decide on which strategy to employ and how to approach a driver. Of especial interest is workload; when persuading a driver it is important to time advices to periods when drivers can accommodate the advices safely, as well as are capable of following up on them. In terms used in this thesis, this means timing messages to periods of low workload, high perceived behavioural control, and high motivation. To determine driver workload on-line, this thesis outlined a ML-based approach that performed well. The tools for analysing heart rate real-time developed and discussed in this thesis should enable researchers and practitioners to get started quickly.

Both gamification and socially cooperative approaches worked significantly better to persuade drivers to follow advices than simply providing an advice. The application of a web-portal where drivers could follow their progress between driving sessions helped to increase motivation and persuasive power. When deciding on either approach for an in-car advice system, important considerations are which approach is more suited to the use case and how long the system is envisaged to run. Gamified solutions tend to perform well for shorter durations, as once users reach their goals it is easy to lose motivation. Playing the game is only fun for so long.

8.3   Methodological Limitations

This section discusses the methodological decisions that were made, and the advantages and drawbacks this brings with it including the research limitations.

8.3.1   The Nature of Driving Simulator Studies

The method of experimentation employed throughout the research is the use of a fixed-base driving simulator. The advantage of a driving simulator is that it offers the experimenter full control of the environment and eliminates changes in factors that otherwise are difficult to control in reality, for example varying weather conditions, transient traffic events, and variation in time of day lighting conditions. Using a driving simulator also enables safety-critical research, since little to no risk is posed to the participants and virtual crashes are harmless. Simulator studies are logistically easier and thus can also be employed on a larger scale and offer the possibility for larger sample sizes.

The main drawback of a driving simulator experiment is that the participants are not exposed to a real driving environment. A driving simulator has relative validity rather than absolute validity (Carsten & Jamson, 2011), meaning that the behavioural patterns observed will likely transfer to real world conditions, but their magnitude may differ. Despite remarkable advances in computer graphics and the open sourcing of several rendering engines such as Unity3D and the Unreal Engine, the participant will remain aware that they are not driving a real car. Aside from the graphics not being completely photorealistic, participants often remarked it felt different to drive a car in the simulator when compared to a real car. This is because in a fixed-base driving simulator motion cues are missing (Kaptein, Theeuwes, & van der Horst, 1996). Whereas in a real driving environment accelerating and decelerating both laterally (steering) and longitudinally (accelerating/braking) will create forces acting on the body that act as motion cues and aid in vehicle control. Such motion cues are missing when driving in a fixed-base driving simulator. Moving base simulators exist that can create these motion cues, but these suffer from realism issues as well, and are prohibitively expensive to acquire and run. Although only two participants in the three driving simulator studies suffered from this, the possibility for simulator sickness is another potential risk that can lead to drop-out of participants.

Despite the previous limitations, a driving simulator was used in the studies because of ethical concerns regarding safety. One part of the research focused on inducing various levels of workload in drivers including high workload, which for obvious reasons has the potential to endanger the safety of someone driving in real traffic. The second part of my research determined how to persuade drivers using an in-car HMI. While this study could have been performed in real traffic, it was chosen to use a driving simulator because it allowed testing a larger sample size and give more advices in total. This would give a clearer view of any possible effects of the persuasive intervention as compared to real traffic, where varying traffic and weather conditions might confound any effects present. However, the usage of a driving simulator may also have affected the behavioural responses of participants as was discussed in chapter 7.

8.3.2   Measuring Workload of Individual Drivers is Possible, Generalising if Less Straightforward

The workload models developed in this thesis could predict workload well and generalize to unknown individuals relatively well in the lane-keeping task (high workload scenario) but not

in the simulated regular driving conditions. In regular driving conditions, individual and group-level models worked well, but generalizing to unknown drivers did not give good results. The reason for this may be as simple as that workload induced in the regular simulated driving conditions was not high enough to create distinct patterns that generalize across individuals. In other words, it might be that physiological responses to extreme levels of workload are similar across individuals, but responses to small variations differ between individuals.

However, a possible reason for this discrepancy is that workload cannot be captured as a singular concept. Indeed, the capacity to control a car relies not only on the mental workload of the driver, but it is likely also influenced by other states such as fatigue, momentary physical distraction, mental distraction, boredom, transient secondary tasks such as controlling infotainment of navigation systems, and the general driving context. These states might all interact with the driving task to influence workload levels, yet can still be distinct enough from workload that they may not be captured in one aggregate measure.

An added difficulty with predicting an aggregate measure like workload in practice is that more factors are involved in real world driving compared to the lab. This means that the variance of the data used to predict workload will be larger. Models built in lab settings might have a hard time accurately generalizing to the more noisy real-world conditions. Those built on real-world driving sets may suffer from not being able to fit all variance, or from the collected dataset not being exhaustive enough. Future research can take this into account by ensuring data used to build models is representative not just of the research question being answered, but also of the practical environments in which any predictive inference will be run.

8.3.3   On Using Machine Learning Methods

Machine learning (ML) methods were used to fit models and predict workload. The main disadvantage often associated with ML approaches is their black box nature: something is learned, the prediction is verified, but not much is known about exactly what the model learns. Recently, the calls for transparent and easily interpretable ML models have intensified (Adadi & Berrada, 2018; Samek, Wiegand, & Müller, 2017) and work on 'explanatory artificial intelligence' (XAI) has expanded. These efforts are necessary, especially amidst recent reports of, for example, medical deep learning models not generalizing well between different brands of the same devices (Zech et al., 2018).

This issue of generalizability is one that was also raised in the workload study and is one that is in my opinion insufficiently discussed in applied ML literature. ML approaches (of which especially in Deep Learning (DL)) can achieve very high accuracy but at the same time function only within the precise parameters of the training data provided. If data at runtime varies from what is used at training time, the predictions are at risk of losing their validity. This problem was minimized by by collecting data on four separate days, with the goal of collecting a varied data set.

The strength of ML methods is their capacity to fit vast amounts of data to desired output mappings or learn a range of behaviours without a priori ideas of relations present in the data set. As research moves into the area of big data, it becomes more difficult for traditional statistical modelling to result in comprehensive models. Simply put, the human mind is too limited to understand and map all relations in a complicated modern data set (Breiman, 2001). The main advantage of ML methods is that they can. This is the main reason to employ this

method in the workload prediction study, and also because the work on workload prediction using more traditional methods has so far resulted in mixed results.

8.4 Recommendations for Scientific Research

The conclusions of the work in this dissertation are subject to certain limitations. These are related to the methods of data collection, analysis, sample of participants used, and the experimental designs chosen. Future research can address some of the shortcomings and build on the conclusions presented.

8.4.1 On-Road Trials of Persuasive HMI

I developed a persuasive HMI and tested its effectiveness in a simulator study. As mentioned, the main advantage of a driving simulator is full control over the scenarios, but the main drawback is that it differs from real driving. This that the results from a driving simulator have relative validity, rather than absolute validity.

It was found that the persuasive variants, which included cooperative and competitive elements, an avatar, and a web-portal, were more effective than the control condition which only included the advices and the avatar. An open question remains how well these results will generalize to real-world driving conditions. Since a decent sample size of 54 participants was obtained, that collectively received 432 advices, no radically different results are expected in real-world driving settings. However, differences in effect size can be expected. In normal driving conditions other factors potentially confound persuasive effects, such as being in a hurry, emotional state, fatigue, as well as general driving styles which may not have manifested in the driving simulator. On-road tests are required to further validate the potential of the proposed persuasive approach.

Another recommendation relates to the types of advices given. A congestion warning and a lane-drop warning were tested. Other reasons for advices are possible, such as changing weather conditions, road works, or accidents. While no difference was found in the number of advices followed between the two tested contexts, it may be that different situations show different compliance patterns. A future study can examine this in more detail.

A final point relates to the long-term effectiveness of persuasive technology. As drivers become familiar with the system and its features, they may become less sensitive to certain persuasive aspects such as competition or rewards. Further research is required to quantify long-term effectiveness of persuasive technology, not only within driving context but outside of it as well.

8.4.2 Predict More Than Workload

The chapter on driver workload discussed two studies. Data was collected in typical driving conditions, and artificial lane keeping conditions, both in the driving simulator. The latter was a much more demanding driving task and induced higher workload levels.

In both studies predicting workload with individual models (trained on individual level data) and group-based models (trained on group level data) worked well. This indicates there is enough variance in the data that is tied to driver workload, and the models had sufficient capacity to capture this. However, trying to generalise the results to unknown drivers proved challenging in especially in the realistic driving conditions. In the generalization case only the

forced-pace study (higher levels of workload) provided useful results, but with an important caveat: it was observed that in most cases, the correct workload pattern was predicted over time, but a significant offset error was present.

I hypothesize the reason for this is that workload cannot be viewed as a singular construct. While workload may arise from the difference between the cognitive resources required to control the vehicle and the cognitive resources that the driver has available, the ability to control the vehicle is dependent on many factors beyond this. Transient distractions like operating a navigation or infotainment device may take both the eyes and mind of the driver off the road and thus create a window for dangerous situations to arise. Similarly, talking to a passenger or being involved in a call, even if hands-free, can decrease a driver's capacity to deal with safety-critical situations that may arise. Driver emotional states like anger from a recent conversation, or bodily states like fatigue, can hamper vehicle control and likely modulate a driver's workload response to external events. This is important in the context of using driver state prediction models to time communication between in-vehicle systems and the driver. The difficulty in generalizing the predictions to unknown drivers in real-world driving conditions, may have been because workload was predicted as an aggregate measure, rather than through its possible subcomponents.

Another source for the poor generalising performance of the models could be that self-reported workload measures were used as ground truth input for the models. It was chosen to do so because from a persuasive perspective, self-reported (perceived) workload is what you want to predict: when deciding whether to follow a persuasive advice or not, it matters how a driver perceives their own workload level, not what it really is. However, because this relatively indirect workload metric was chosen as ground truth, a disconnect may exist between the perceived workload and the actual workload, and subsequently the physiological responses of the driver to the workload levels. Future research can work to quantify the relation between self-reported mental load measures and actual task loads put on a driver.

I recommend future studies work towards classifying and predicting sub-components important to workload, such as transient distractions, eyes-off-road moments, fatigue, boredom, the driving context, or other factors that can influence a driver's capacity to adequately control the vehicle. These studies can build upon existing attempts to predict these sub-components, and work towards unifying them into a fast classifier that can be deployed in in-car settings. Additionally, when planning to use self-reported measures, the relationship between self-reported workload and actual task loads needs to be examined further.

8.4.3 Validity of Machine Learning Models: Generalisability, Robustness, and Available Data Sets

It was observed in many applied ML papers in the transportation domain and psychophysiological domain, that insufficient attention was given to how the trained models will be applied in practice and what performance to expect under variable settings. Segmenting a dataset into train/validation/test sets is a robust way of reducing overfitting and of estimating some generalization performance of models. A problem with relying on this approach is that all three segmented sets are not independent of each other: they tend to still come from the same underlying distribution. Factors like camera type used, lighting or weather conditions (with visual data), to the type of sensors used for most data, will have an effect. For example a recent critical overview found that for chest x-ray pneumonia detection, deep learning networks failed to generalize well between different something innocuous like different brands of x-ray devices

(Zech et al., 2018). Such inherent issues are dangerous as these problems are not detected with the standard evaluation methods.

I recommend that research focuses on defining not only model performance on training/vallidation/test sets, but also critically reflects on issues such as sensitivity of trained models to differences between training data and real-world input data. One way to work towards this is to open source the data sets and annotations that are generated and used in research whenever possible. This increases the available data variance which allows for more robustness of the developed models.

8.5   Recommendations for Practice

Most of the research presented in this dissertation can be viewed as 'building blocks' for the development and construction of smarter in-car systems. This section outlines some recommendations for using them in practice.

8.5.1   Persuasive HMI Usage in Practice

The research presented in this dissertation has led to several recommendations for using persuasive HMI in practice. When persuading drivers, context matters. Drivers are not likely to follow advices if they cannot understand or observe the reasons for it. Provide the reason for why an advice is given, especially when drivers cannot observe for themselves why an advice may be given.

Not just when persuading a driver, but when communicating with a driver in any way, workload matters. Using a workload monitor as a filter to determine when to communicate to the driver, as proposed in this dissertation, is one way to provide safe interaction. However, it was discussed that predicting workload as an aggregate measure might not be reliable. Splitting workload into relevant subcomponents as discussed in 8.3.2 might be a solution. Often it may not be required to predict general workload.

Regarding using persuasive systems to change driver behaviour: it is recommended to not plan on using persuasion long-term, especially gamification. Persuasive effects reduce over time as people get used to a system and subsequently become bored with (or blind to) its features. Competing with other drivers is fun, but for how long? Where persuasion can however be very effective, is in the type of system described in this dissertation. For the proposed lane-specific system to function many drivers need to follow its advices from the start. To stimulate compliance to advices in early stages it was proposed to use persuasive methods. Once drivers observe the benefits a system gives, it no longer matters if persuasive effectiveness drops off. Drivers will have an incentive to keep following advices simply because it benefits them.

8.5.2   Initiate Communication with Drivers Only at Appropriate (Safe) Moments

As in-car technology moves towards more information and assistive systems the information density in the car increases. From the point of view of the driver, who is also performing the already taxing task of vehicular control and navigation, an increase in information density can create a more demanding and thus less safe and less comfortable driving experience.

Aside from the information content that is being transmitted to the driver, its timing is most important. Distracting a driver at the wrong time can be dangerous in addition to having the

potential of leading to irritation with the driver. Communicating with a driver requires knowledge of both the environment, as factors like traffic density or the distance to the car in front may increase risks dynamically, as well as knowledge of the driver state, as difficult driving conditions, preoccupation with a conversation, or distraction caused by operating in-car infotainment may also pose extra risk factors.

8.5.3   Workload Prediction in Practice

While the goal in practice can often be to predict workload, in many cases another often easier to measure factor is required. For example, finding whether a driver is alert and capable of vehicle control may be better detected by measuring eyes-on-road, distraction from in-car devices, and vehicle kinematics, rather than an aggregate measure such as workload.

Collecting and analysing representative data sets, for example from naturalistic driving studies, can help reduce this issue. When considering workload prediction in practice, it is recommended to determine relevant subcomponents as discussed in 8.3.2 and decide on which facets are most important for a given feature. The diverse results from scientific work over the past decades seems to indicate that general workload cannot be predicted accurately when driving, nor is it necessary to predict it in many cases.

Physiological data take a special place in monitoring drivers. They are often collected to predict workload or related metrics, and in practice various trade-offs have to be made compared to the lab. Whereas in the lab the participant movement can be more restricted, types of sensors used more intrusive, and the tasks a driver is asked to perform are more delineated, driving in the real world is more dynamic, and a driver might at any time engage in secondary asks as well. This dissertation has contributed by developing a robust heart rate analysis algorithm and making it available open source.

8.5.4   Machine Learning Ready for Practice: Speed Versus Accuracy

One important consideration in developing ML models that are ready for practical usage is the consideration of resources required. This problem is especially prominent in Deep Learning, a sub-field of Machine Learning that generally deals with much larger model architectures. Developing a model that accurately performs a task with good accuracy and performance parameters is great, but where will it be deployed?

While it may be tempting to use a standard model architecture like a ResNet50 or an Xception network and keep working with that, even despite their relatively low requirements relative to their performance (Bianco, Cadene, Celona, & Napoletano, 2018), they still require about 5 and 8 gFLOPs, and require at least 740 and 1003 MB of memory, respectively. A challenge to future researchers is to take this one step further and find the minimum requirements for acceptable model performance, either by reducing the complexity of existing models, or by working towards custom architectures. In practice a given architecture that reaches 98% predictive accuracy but requires 300ms inference time may still be less desirable than one reaching 94% predictive accuracy that requires an inference time of 50ms. This constraint is especially true for embedded environments such as in in-car systems. I encourage researchers to work towards optimized network architectures within the scope of their research, to speed up adoption of the research in practice.

While platforms like Nvidia Jetson and more recently Google's Edge TPU[15] offer impressive embedded performance at low power levels (2-7.5 Watts), when for example working on driver state prediction as described in 8.4.2 it may be necessary to run multiple models in parallel to detect fatigue, distraction, eyes-off-road, phone calls, and other factors related to driver state. By ensuring accurate models that also have relatively low computational complexity, practical use of these models can be increased.

8.5.5   Keep an Eye on Where Your Machine Learning Models Will Run

Some of the ML recommendations for future research also apply to recommendations for practice, such as a fair quantification of generalizing power of the models used. Realizing what the distribution of real-world data is and how this compares to the more limited data sets used to train the networks can help in improving model robustness. Aside from collecting a more representative set, data augmentation techniques can help close the gap between training data and real-world data.

Another potential problem comes from so-called adversarial attacks. ML (and especially DL) networks are specialists capable of only performing within the boundaries of what they have learned. Adversarial attacks aim to make small changes to model inputs to confuse classification and prediction networks. Oftentimes these changes are imperceptible to humans. For example, by adding a small amount of noise to images, classifiers can give false predictions with very high confidence. In some areas this is potentially life threatening. Healthcare is vulnerable (Finlayson, Chung, Kohane, & Beam, 2018) with patient lives being at stake. Transportation is not immune either. By giving examples never encountered by a road signage classifier, such as company logos or slightly manipulated traffic signs, a study showed that road signs were reliably misclassified with very high confidence (Sitawarin, Bhagoji, Mosenia, Chiang, & Mittal, 2018). Even simple misses, such as classifying a 30 km/h area as an 80 km/h one, can be disastrous in semi or fully autonomous cars. This underlines the importance of redundancy and time investment in ways of making the models more robust (Madry, Makelov, Schmidt, Tsipras, & Vladu, 2018; Prakash, Moran, Garber, Dilillo, & Storer, 2018).

---

[15] Available through Coral.ai

References

Adadi, A., & Berrada, M. (2018). Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). IEEE Access, 6, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

Bianco, S., Cadene, R., Celona, L., & Napoletano, P. (2018). Benchmark analysis of representative deep neural network architectures. IEEE Access, 6, 64270–64277. https://doi.org/10.1109/ACCESS.2018.2877890

Breiman, L. (2001). Statistical Modeling: The Two Cultures. Statistical Science, 16(3), 199–215. https://doi.org/10.2307/2676681

Carsten, O., & Jamson, A. H. (2011). Driving simulators as research tools in traffic psychology. In Handbook of Traffic Psychology. https://doi.org/10.1016/B978-0-12-381984-0.10007-4

Finlayson, S. G., Chung, H. W., Kohane, I. S., & Beam, A. L. (2018). Adversarial Attacks Against Medical Deep Learning Systems. Retrieved from http://arxiv.org/abs/1804.05296

Jager, F., Taddei, A., Moody, G. B., Emdin, M., Antolič, G., Dorn, R., … Mark, R. G. (2003). Long-term ST database: A reference for the development and evaluation of automated ischaemia detectors and for the study of the dynamics of myocardial ischaemia. Medical and Biological Engineering and Computing, 41(2), 172–182. https://doi.org/10.1007/BF02344885

Kaptein, N. A., Theeuwes, J., & van der Horst, R. (1996). Driving Simulator Validity: Some Considerations. Transportation Research Record, 1550(1), 30–36. https://doi.org/10.1177/0361198196155000105

Madry, A., Makelov, A., Schmidt, L., Tsipras, D., & Vladu, A. (2018). Towards deep learning models resistant to adversarial attacks. 6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings, 1–28.

Pan, J., & Tompkins, W. J. (1985). A simple real-time QRS detection algorithm. IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, BME-32(3), 230–236. https://doi.org/10.1109/IEMBS.1996.647473

Prakash, A., Moran, N., Garber, S., Dilillo, A., & Storer, J. (2018). Deflecting Adversarial Attacks with Pixel Deflection. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 8571–8580. https://doi.org/10.1109/CVPR.2018.00894

Ramshur, J. (2010). Design, Evaluation, and Application of Heart Rate Variability Analysis Software (HRVAS). https://doi.org/10.13140/RG.2.2.33667.81444

Samek, W., Wiegand, T., & Müller, K.-R. (2017). Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models. ArXiv Preprinte. Retrieved from http://arxiv.org/abs/1708.08296

Sitawarin, C., Bhagoji, A. N., Mosenia, A., Chiang, M., & Mittal, P. (2018). DARTS: Deceiving Autonomous Cars with Toxic Signs. Retrieved from http://arxiv.org/abs/1802.06430

van der Laan, J. D., Heino, A., & de Waard, D. (1997). A Simple Procedure for the Assessment of Acceptance of Advanced Transport Telematics. Transportation Research Part C: Emerging Technologies, 5(1), 1–10.

Zech, J. R., Badgeley, M. A., Liu, M., Costa, A. B., Titano, J. J., & Oermann, E. K. (2018). Variable generalization performance of a deep learning model to detect pneumonia in chest

radiographs: A cross-sectional study. PLoS Medicine, 15(11), 1–17. https://doi.org/10.1371/journal.pmed.1002683

TRAIL Thesis Series

The following list contains the most recent dissertations in the TRAIL Thesis Series. For a complete overview of more than 275 titles see the TRAIL website: www.rsTRAIL.nl.

The TRAIL Thesis Series is a series of the Netherlands TRAIL Research School on transport, infrastructure and logistics.

Gent, P. van, *Your Car Knows Best,* T2021/20, July 2021, TRAIL Thesis Series, the Netherlands

Wang, Y., *Modeling Human Spatial Behavior through Big Mobility Data,* T2021/19, June 2021, TRAIL Thesis Series, the Netherlands

Coevering, P. van de, *The Interplay between Land Use, Travel Behaviour and Attitudes: a quest for causality,* T2021/18, June 2021, TRAIL Thesis Series, the Netherlands

Landman, R., *Operational Control Solutions for Traffic Management on a Network Level,* T2021/17, June 2021, TRAIL Thesis Series, the Netherlands

Zomer, L.-B., *Unravelling Urban Wayfinding: Studies on the development of spatial knowledge, activity patterns, and route dynamics of cyclists,* T2021/16, May 2021, TRAIL Thesis Series, the Netherlands

Núñez Velasco, J.P., *Should I Stop or Should I Cross? Interactions between vulnerable road users and automated vehicles,* T2021/15, May 2021, TRAIL Thesis Series, the Netherlands

Duivenvoorden, K., *Speed Up to Safe Interactions: The effects of intersection design and road users' behaviour on the interaction between cyclists and car drivers,* T2021/14, April 2021, TRAIL Thesis Series, the Netherlands

Nagalur Subraveti, H.H.S., *Lane-Specific Traffic Flow Control,* T2021/13, March 2021, TRAIL Thesis Series, the Netherlands

Beirigo, B.A., *Dynamic Fleet Management for Autonomous Vehicles: Learning- and optimization-based strategies,* T2021/12, March 2021, TRAIL Thesis Series, the Netherlands

Zhang, B., *Taking Back the Wheel: Transition of control from automated cars and trucks to manual driving,* T2021/11, February 2021, TRAIL Thesis Series, the Netherlands

Boelhouwer, A., *Exploring, Developing and Evaluating In-Car HMI to Support Appropriate use of Automated Cars,* T2021/10, January 2021, TRAIL Thesis Series, the Netherlands

Li, X., *Development of an Integrity Analytical Model to Predict the Wet Collapse Pressure of Flexible Risers,* T2021/9, February 2021, TRAIL Thesis Series, the Netherlands

Li, Z., *Surface Crack Growth in Metallic Pipes Reinforced with Composite Repair System,* T2021/8, January 2021, TRAIL Thesis Series, the Netherlands

Gavriilidou, A., *Cyclists in Motion: From data collection to behavioural models,* T2021/7, February 2021, TRAIL Thesis Series, the Netherlands

Methorst, R., *Exploring the Pedestrians Realm: An overview of insights needed for developing a generative system approach to walkability,* T2021/6, February 2021, TRAIL Thesis Series, the Netherlands

Walker, F., *To Trust or Not to Trust? Assessment and calibration of driver trust in automated vehicles,* T2021/5, February 2021, TRAIL Thesis Series, the Netherlands

Schneider, F., *Spatial Activity-travel Patterns of Cyclists,* T2021/4, February 2021, TRAIL Thesis Series, the Netherlands

Madadi, B., *Design and Optimization of Road Networks for Automated Vehicles,* T2021/3, January 2021, TRAIL Thesis Series, the Netherlands

Krabbenborg, L.D.M., *Tradable Credits for Congestion Management: support/reject?,* T2021/2, January 2021, TRAIL Thesis Series, the Netherlands

Castelein, B., *Accommodating Cold Logistics Chains in Seaport Clusters: The development of the reefer container market and its implications for logistics and policy,* T2021/1, January 2021, TRAIL Thesis Series, the Netherlands

Huang, B., *The Influence of Positive Interventions on Cycling*, T2020/20, December 2020, TRAIL Thesis Series, the Netherlands

Xiao, L., *Cooperative Adaptive Cruise Control Vehicles on Highways: Modelling and Traffic Flow Characteristics,* T2020/19, December 2020, TRAIL Thesis Series, the Netherlands

Polinder, G.J., *New Models and Applications for Railway Timetabling,* T2020/18, December 2020, TRAIL Thesis Series, the Netherlands

Scharpff, J.C.D., *Collective Decision Making trough Self-regulation*, T2020/17, November 2020, TRAIL Thesis Series, the Netherlands

Guo, W. *Optimization of Synchromodal Matching Platforms under Uncertainties,* T2020/16, November 2020, TRAIL Thesis Series, the Netherlands

Narayan, J., *Design and Analysis of On-Demand Mobility Systems,* T2020/15, October 2020, TRAIL Thesis Series, the Netherlands

Gong, X., *Using Social Media to Characterise Crowds in City Events for Crowd Management,* T2020/14, September 2020, TRAIL Thesis Series, the Netherlands

Rijal, A., *Managing External Temporal Constraints in Manual Warehouses,* T2020/13, September 2020, TRAIL Thesis Series, the Netherlands

Alonso González, M.J., *Demand for Urban Pooled On-Demand Services: Attitudes, preferences and usage,* T2020/12, July 2020, TRAIL Thesis Series, the Netherlands