



A Bayesian reference model for visual time-sharing behaviour in manual and automated naturalistic driving

Downloaded from: <https://research.chalmers.se>, 2020-04-24 15:20 UTC

Citation for the original published paper (version of record):

Morando, A., Victor, T., Dozza, M. (2019)

A Bayesian reference model for visual time-sharing behaviour in manual and automated naturalistic driving

IEEE Transactions on Intelligent Transportation Systems, 21(2): 803-814

<http://dx.doi.org/10.1109/TITS.2019.2900436>

N.B. When citing this work, cite the original published paper.

©2019 IEEE. Personal use of this material is permitted.

However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

A Bayesian reference model for visual time-sharing behaviour in manual and automated naturalistic driving

Alberto Morando, Trent Victor, and Marco Dozza

Abstract—Visual time-sharing (VTS) behavior characterizes an inattentive driver. Because inattention has been identified as the major contributing factor in traffic crashes, understanding the relation between VTS and crash risk could help reduce crash risk through the development of inattention countermeasures. The aims of this study are 1) to develop a reference model of VTS behavior and 2) reveal if vehicle automation influences VTS behavior. The reference model was based on naturalistic eye-tracking data. VTS sequences were extracted from routine driving data (including manual and automated driving). We used Bayesian Generalized Linear Mixed Models for a range of on- and off-path glance-based metrics. Each parameter was estimated with a probability distribution and summarized with credible intervals containing the model parameters with 95% probability. The reference model corroborates previous findings from driving simulator experiments and on-road studies, but also captures the characteristics of on-path and off-path glance behavior in greater detail. The model demonstrated that 1) there was minimal change in VTS behavior due to automation, and 2) the percentage of time that glances fell on-path (PRC) was greater for all routine driving (~80%) than for VTS sequences (~50%). The PRC was the only metric that was sensitive to VTS, but it did not differentiate between manual and automated driving. Our model, by describing a measure of inattention (VTS behavior), can be used in future driver models to improve the computer simulations used to design ADASs and evaluate their safety benefits. Additionally, the model could serve as a detailed reference for inattention guidelines.

Index Terms—ADAS, attention, eye tracker, glance distribution, vehicle automation, visual behavior.

I. INTRODUCTION

DRIVER inattention is the major contributing factor in traffic crashes [1]–[3]. Inattention is defined here as misdirected attention [4, p. 34], occurring “when the demands of activities currently critical for safe driving are not matched

This research was financially supported by the European Marie Curie ITN project Human Factors of Automated driving (HFAuto, PITN-GA-2013-605817) and by VINNOVA (Swedish governmental agency for innovation) as part of the project Quantitative Driver Behavior Modelling for Active Safety Assessment Expansion (QUADRAE).

A. Morando is with the Division of Vehicle Safety at the Department of Mechanics and Maritime Sciences, Chalmers University of Technology, 412 96 Göteborg, Sweden (e-mail: alberto.morando@chalmers.se).

T. Victor is with Volvo Cars Safety Center (Volvo Cars Corporation), 405 31 Göteborg, Sweden (e-mail: trent.victor@volvocars.com). He is also with the Division of Vehicle Safety at the Department of Mechanics and Maritime Sciences, Chalmers University of Technology, 412 96 Göteborg, Sweden.

M. Dozza is with the Division of Vehicle Safety at the Department of Mechanics and Maritime Sciences, Chalmers University of Technology, 412 96 Göteborg, Sweden (e-mail: marco.dozza@chalmers.se).

due to the allocation of resources to other safety-critical or non-critical activities”. When a driver’s attention is misdirected towards non-critical activities it is usually known as distraction. An inattentive driver often switches visual attention back and forth between the forward path and another location; this behavior is called visual time-sharing (VTS) [5]–[8].

Previous research has shown that VTS increases crash risk because 1) frequent and inappropriate off-path glances increase the uncertainty of the driving situation [1], [3], [9]–[13] and 2) short on-path glances are not long enough to make up for information decay or to uptake enough information to predict a critical situation [11]–[13].

Inattention can have many causes, both internal and external to the vehicle [4], [14]. The traditional approach to investigating inattention and VTS has been the semantic categorization of activities performed while driving. For example, in the 100-car study [15], about 60 categories of secondary tasks were identified (Appendix D [1]). Unfortunately, this approach is time-consuming, relies on subjective judgment, and it can only be performed post hoc [15]. Moreover, task-based methods do not consider that the crash risk of a low-demanding task performed frequently—or for an extended time—may be comparable to that of a more demanding task performed less often [16].

An alternative approach is to investigate inattention and VTS based on glance behavior, independent of the task performed [5], [6]. It is unclear whether unsafe behavior is dependent on glance characteristics and independent of task type [3], but a glance-based approach is justified because eye movements are a strong indicator of where attention is directed [17]. Glance-based methods also have practical advantages: 1) they enable the programmatic analysis of long time series; 2) they are not limited to any specific category of tasks; and 3) they enable the real-time, unobtrusive collection of eye-tracking data.

Glance-based methods have been used to develop ADASs that counteract inattention via feedback and warnings [18], [19]. Such driver-state monitoring systems are relatively immature, but some solutions are already on the market [20]. Glance-based methods have also been used to develop guidelines to minimize the inattention caused by in-vehicle interfaces (e.g., the National Highway Traffic Safety Administration (NHTSA) distraction guidelines [16]).

However, to be effective, an inattention countermeasure requires a reference model (a set of metrics and target values)

that captures abnormal driver states, applies to real world driving, is statistically robust, and incorporates the latest vehicle automation. Such a reference model is currently lacking. In fact, there are many unanswered questions about the reference values currently available [16], [18], [21]: Which glance metrics are the most sensitive indicators of driver inattention? Are the results from experiments based on commanded tasks ecologically valid? Is the reference model robust with respect to individual driver characteristics? Do the results in manual driving transfer to automated driving?

The development of a reference model for driver behavior requires better statistical methods, especially for analyzing real-world/naturalistic datasets; they are large, but they are also noisy, sparse, and unbalanced due to the lack of experimental control (e.g., the proportion of time the driver spent in a certain driving condition). In this paper, we applied a full Bayesian data analysis that addresses these challenges, producing results that can be used in a reference model.

Bayesian methods are gaining traction in other fields but are under-represented in the engineering and human-factors field. Bayesian data analysis has many advantages compared to traditional frequentist statistics and null hypothesis testing [22], [23]: it provides a more intuitive interpretation of the results in terms of probabilities; it easily accommodates data that are not normally distributed; it focuses on the estimation of the magnitude of the effects and the quantification of uncertainty of the estimation—not on the dichotomous rejection of a null hypothesis [24]; therefore it allows the use of results in simulation and quantifies more precisely the difference between driving conditions; finally it encourages comparisons of replication and results because the results can be used by other researchers to carry out additional analysis, or to apply the results to subsequent studies.

The aim of this paper is to develop a Bayesian reference model for VTS behavior in real-world naturalistic routine driving, in both manual and automated driving (and reveal if vehicle automation influences VTS behavior). Automated driving is intended here as adaptive cruise control and lane keeping aid active (ACC+LKA). This model should allow a comprehensive assessment of VTS behavior and enable the quantification of the differences in visual behavior between manual and automated driving. This model could also be used for driver modeling, safety assessments and ADAS development. Moreover, it could serve as reference for developing inattention guidelines.

II. METHODS

A. Data source

The data used in this study are from the *EyesOnRoad* naturalistic field operational test [25], [26]. The dataset is described in greater detail in an earlier paper by the same authors [27], which will henceforth be referred to as “the previous paper”. In short, data were collected from ten Volvo cars (2014 V60 model) in the region of Västra Götaland (Sweden) from December 2014 to September 2015. Most of the data were collected in Gothenburg, the second-largest city in Sweden. The cars were equipped with the automated systems ACC+LKA. Drivers were free to drive their vehicle and to use the ADASs installed in the car as they wished.

Vehicle data were continuously collected at 60 Hz from the controller area network (CAN) bus. Eye movements were recorded at 60 Hz by an eye-tracking system that automatically classified glances as either on-path or off-path [25]. A glance is defined as the transition of the eyes to an area of interest (e.g., the forward path) followed by one or more continuous fixations within that area, until the eyes move to another area (e.g., off path) (ISO 15007-1:2014). Classification of glances as on- and off-path is routinely done [1], [3]. Our previous paper’s analysis of this dataset [27] showed that the eye-tracker was reliable and robust in real-world driving scenarios. Our previous paper also provided the algorithm we used to remove artifacts and unrealistic glances from the dataset and the filtering criteria we used to collect routine (i.e., non-safety-critical) driving segments on straight roads (rural roads and highways).

From this preliminary collection of driving segments, we selected the ones that involved open-road driving (i.e., range to other vehicles greater than 50 m) in daylight (inferred from the current time and date and the vehicle’s GPS position) when the vehicle speed was above 60 km/h. The datasets for the other conditions (e.g., car following and night driving) were too sparse and unbalanced to provide meaningful results. The segments were grouped into manual driving or ACC+LKA driving, depending on whether the assistance systems were turned on and operational or not. A total of 1770 VTS sequences were available for the analysis (296 in manual driving, 1474 in ACC+LKA driving). The VTS sequences included 16 unique drivers: the manual driving group had 7 drivers (2 males, 5 females) with an average age of 53 years (SD 7.5 years), the ACC+LKA driving group had 11 drivers (8 males, 3 females) with an average age of 48 years (SD 13 years), and two drivers (1 male and 1 female) were common to both groups.

B. VTS sequence extraction

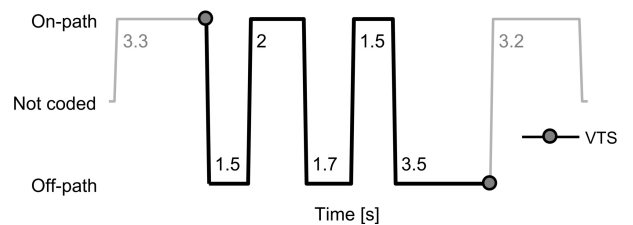


Figure 1. Example of a visual time-sharing (VTS) sequence detected by the algorithm proposed by Victor et al. [6]. The thicker, darker line represents the time series of on- and off-path glances that belong to the VTS sequence. The thinner, lighter line represents the remainder glances in the driving segment. Each glance in the graph has been labeled with its duration in seconds.

VTS sequences were detected with a rule-based algorithm based on the glance location coded by the eye tracker (Fig. 1). The algorithm was originally proposed by Victor et al. [6] and recently applied, in a modified version, by Ahlström and Kircher [5]. Each driving segment could contain from zero to multiple VTS sequences. To code a sequence as VTS, the algorithm requires that [6]:

- the driver is looking on-path for at least 3 s between VTS sequences;

- there are at least 3 off-path glances in the VTS sequence;
- off-path glances in the VTS sequence are in the range $[0, \infty)$ s;
- on-path glances in the VTS sequence are in the range $(0, 3]$ s.

C. Metrics

The glance-based metrics used to quantify VTS sequences were:

- the on- and off-path glance distributions, together with the 50th (median), 85th, 95th, and 99th percentiles.
- the proportion of on-path glances shorter than 1 s ($PGDon \leq 1$) and of off-path glances longer than 2 s ($PGDoff \geq 2$). $PGDon \leq 1$ and $PGDoff \geq 2$ quantify information uptake deficiencies. ($PGDoff \geq 2$ is comparable to percentage of extended duration glances in ISO 15007-1:2014.)
- the proportion of time that glances fall within the on-path area (percent road center: PRC) [28]. PRC combines glance duration and frequency and has demonstrated sensitivity to changes in driving demand and context [27]–[29]. (PRC is equivalent to percent time on the area of interest and the on-path inverse of the percentage of eyes off road time (PEORT) defined in ISO 15007-1:2014.)
- the total duration in seconds of the VTS sequence (total task time: TTT), computed as the sum of the duration of the on- and off-path glances (thicker, darker line in Fig. 1). TTT is a metric for quantifying task engagement and difficulty.

Two additional metrics were evaluated, but are not presented due to space constraints: total glance time on-path ($TGTON$) and off-path ($TGTOff$), which are the sum of the on- and off-path glances in the VTS sequence, respectively. ($TGTOff$ is also known as total eyes off-path time in ISO 15007-1:2014.) The general trend of $TGTON$ and $TGTOff$ can be estimated as $TGTON = TTT \cdot PRC$ and $TGTOff = TTT \cdot (1 - PRC)$.

D. Statistical analysis

We defined a series of generalized linear mixed models (GLMMs: also known as hierarchical or multilevel models [30]) to model the individual drivers and the overall tendencies of the group [31]. We fit one Bayesian GLMM for each driving condition (manual or ACC+LKA driving) and glance metric. Because only two drivers were common between driving conditions, we considered the two driving groups as independent. In general, the likelihood function is defined as follows:

$$y \sim f(g(\mu), \epsilon) \quad (1)$$

with $\mu = \mathbf{X}\beta + \mathbf{Z}\gamma$

where y is the vector of the observations of size $(n_{observations} \times 1)$, \mathbf{X} is the design matrix of size $(n_{observations} \times 1)$ corresponding to the parameter β of the group-level effect, \mathbf{Z} is the design matrix of size $(n_{observations} \times n_{drivers})$ corresponding to the vector γ of the effects of the individual driver in the group of size

$(n_{drivers} \times 1)$, and ϵ is the family-specific parameter (the unexplained variability), assumed to be homogeneous across observations. In this form, $\mathbf{Z}\gamma$ is the deviation of the drivers from to the overall group tendency $\mathbf{X}\beta$. Depending on the distribution $f(\cdot)$ of the observations, the appropriate link function $g(\cdot)$ was selected. For all models we placed vague priors over the parameters. The models were parametrized in a non-centered way to improve sampling and eliminate divergences [32], [33].

1) Models:

a) *Off-path glance distribution:* Off-path glances are log-normally distributed [27]. Hence, we fit a mixed log-normal model (2), $f(\cdot) := \log\mathcal{N}$, $g(\cdot) := \text{identity}$:

$$y \sim \log\mathcal{N}(\mu, \sigma) \quad (2)$$

We placed vague priors on each parameter: $\beta \sim \mathcal{N}(0, 2.5)$, $\gamma \sim \mathcal{N}(0, \sigma_z)$, $\sigma_z \sim \text{half}\mathcal{N}(1)$, and $\sigma \sim \text{half}\mathcal{N}(2.5)$. The percentiles and $PGDoff \geq 2$ were estimated based on 1000 draws from the log-normal distribution defined by the combination of β and σ in the Markov Chain Monte Carlo (MCMC) trace.

b) *On-path glance distribution:* On-path glances are distributed as an inverse-gaussian [27]. Hence, we fit a mixed inverse-gaussian model (3), $f(\cdot) := \mathcal{IG}$, $g(\cdot) := \text{logit}$. Because of the algorithm used (presented in II-B), the on-path glance distribution was truncated at 3 s and required a link function to constrain μ (mean of the inverse-gaussian distribution) within the range $[0, 3]$.

$$y \sim \mathcal{IG}(\mu, \lambda) \quad (3)$$

with $\mu = 3 \cdot \text{logistic}(\mathbf{X}\beta + \mathbf{Z}\gamma)$
and $y \in [0, 3]$

We placed vague priors on each parameter: $\beta \sim \mathcal{N}(0, 5)$, $\gamma \sim \mathcal{N}(0, \sigma_z)$, $\sigma_z \sim \text{half}\mathcal{N}(1)$, and $\lambda \sim \text{half}\mathcal{N}(3)$. The percentiles and $PGDon \leq 1$ were estimated based on 1000 draws from the truncated inverse normal distribution defined by the combination of β and λ in the MCMC trace.

c) *Percent road center (PRC):* PRC can be represented as a proportion bounded by the unit interval $[0, 1]$. Hence, we fit a mixed beta model (4), $f(\cdot) := \mathcal{B}$, $g(\cdot) := \text{logit}$. The beta distribution was parametrized in terms of mean μ ($0 \leq \mu \leq 1$) and precision ϕ ($\phi \geq 0$) [34]:

$$y \sim \mathcal{B}(\mu\phi, (1 - \mu)\phi) \quad (4)$$

with $\mu = \text{logistic}(\mathbf{X}\beta + \mathbf{Z}\gamma)$

We placed vague priors on each parameter: $\beta \sim \mathcal{N}(0, 10)$, $\gamma \sim \mathcal{N}(0, \sigma_z)$, $\sigma_z \sim \text{half}\mathcal{N}(1)$, and $\phi \sim \text{HalfCauchy}(5)$.

d) *Total task time (TTT):* For TTT we fit a mixed log-normal model 5, as for the off-path glance distribution.

$$y \sim \log\mathcal{N}(\mu, \sigma) \quad (5)$$

We placed vague priors on each parameter: $\beta \sim \mathcal{N}(0, 25)$, $\gamma \sim \mathcal{N}(0, \sigma_z)$, $\sigma_z \sim \text{half}\mathcal{N}(1)$, and $\sigma \sim \text{half}\mathcal{N}(5)$.

2) *Bayesian inference*: The data were analyzed using Python (ver. 3.6) and the probabilistic programming library PyMC3 (ver. 3.4.1) [35]. Four MCMC chains were used; 5000 samples were drawn from the posterior distribution for each chain (1000 were used for tuning the sampler and then discarded, and 1000 were discarded as burn-in) with the No-U-Turn Sampler (NUTS) [36]. The model convergence was evaluated qualitatively by inspecting the graph of the traces and quantitatively by evaluating both the diagnostics included in PyMC3 and the statistics \hat{R} [37]—which should be close to 1 for convergence. The goodness of fit for each model was then assessed by comparing the posterior predictive distribution against the empirical data, to verify that the model could generate credible observations (graphs supplied as supplemental material [38]).

The result of the Bayesian analysis is the joint posterior distribution over the parameters of the fitted model. The joint posterior distribution was summarized by the 95% highest posterior-density (HPD) interval of the parameters' marginal distributions: that is, the interval includes the range of values that have 95% probability, given the prior distribution and the data [22], [23], [39].

3) *Group comparison*: Comparing groups by basing the estimate of a null effect solely on whether the null value is contained in the HPD is discouraged (for a discussion see [24]). An alternative decision rule for accepting or rejecting null values is based on the HPD interval and on the region of practical equivalence (ROPE: the range of values sufficiently close to the null value to be considered equivalent for practical purposes) [24]. The rule states that if the HPD interval falls entirely outside the ROPE, the null value is rejected; if the HPD interval is entirely contained within the ROPE, the null value is accepted (this means that the HPD could exclude zero, yet the difference is negligible); otherwise, one should remain undecided [24].

There is no unique way to set the limits for the ROPE, however, because a given difference may be trivial—producing values that are still practically equivalent—or meaningful, depending on the context of current knowledge and the practical real-world effect of the results [24], [40]. The great advantage of the Bayesian method over the frequentist one is that it provides evidence that readers can interpret for themselves with different criteria (e.g., different ROPE limits), based on specific knowledge about the importance of the findings—“which might change through time as risks are reassessed and as theories are refined” [24, p. 276]. Because one aim of our study was to model drivers' VTS behavior, we focused on providing reference values useful for modeling. We estimated the differences in the parameters underlying the model in manual and ACC+LKA driving, but we leave any practical significance that can be inferred from our results to be assessed by the readers.

III. RESULTS

A. Off-path glance distribution

a) *Glance distribution*: Table I summarizes the MCMC trace statistics for the fitted log-normal model defined in (2) for

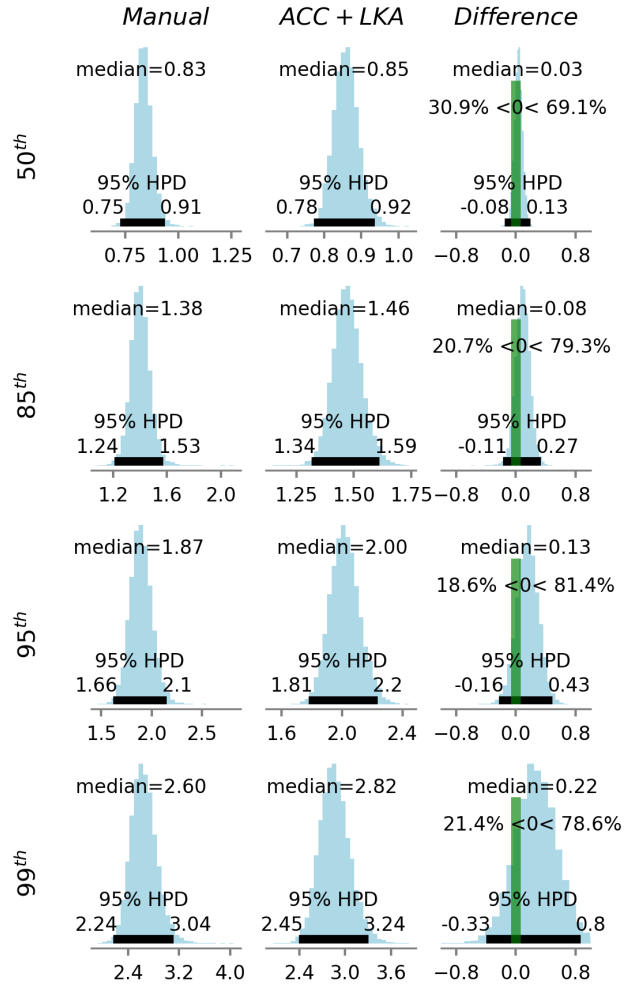


Figure 2. Marginal posterior distributions of the group-level off-path glance percentiles (grouped in rows) in manual and ACC+LKA driving. Each histogram is annotated with its central tendency and the highest posterior density (HPD) interval (thick horizontal black line). The right-most column shows the distributions of the differences in percentile values between the two groups (i.e., the value for ACC+LKA driving minus the value for manual driving). The thick vertical green line marks the null value, and the percentages indicate the proportion of posterior samples below and above the null value. The plots are in the style of Kruschke [41].

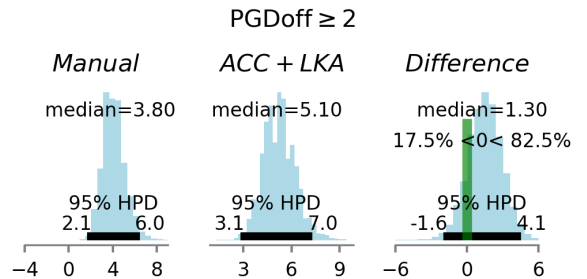


Figure 3. Marginal posterior distributions of the group-level percentage of off-path glances exceeding 2 s ($\text{PGDoff} \geq 2$) in manual and ACC+LKA driving. Each histogram is annotated with its central tendency and the highest posterior density (HPD) interval (shown as a thick horizontal black line). The right column shows the distribution of the difference in $\text{PGDoff} \geq 2$ values between the two groups (i.e., the value for ACC+LKA driving minus the value for manual driving). The thick vertical green line marks the null value, and the percentages indicate the proportion of posterior samples below and above the null value. The plots are in the style of Kruschke [41].

Table I

SUMMARY STATISTICS OF THE MARKOV CHAIN MONTE CARLO (MCMC) TRACE FOR THE PARAMETERS OF THE OFF-PATH GLANCE MODEL, INCLUDING THE DIFFERENCE Δ BETWEEN THE TWO GROUPS (I.E., THE VALUE FOR AUTOMATED DRIVING MINUS THE VALUE FOR MANUAL DRIVING). THE TABLE ALSO PROVIDES A TRANSFORMATION OF THE PARAMETERS TO EASE THE COMPARISON WITH THE RESULTS IN THE LITERATURE. THE MEDIAN (Md) GLANCE DURATION IS COMPUTED AS $\exp(\beta)$; THE MEAN (M) GLANCE DURATION IS COMPUTED AS $\exp(\beta + \sigma^2/2)$.

Parameter	Mean	Sd	Median	HPD	
				2.5%	97.5%
<i>Group level effect</i>					
β_{manual}	-0.19	0.05	-0.19	-0.29	-0.10
$\beta_{ACC+LKA}$	-0.16	0.04	-0.16	-0.23	-0.09
$\Delta(\beta)$	0.03	0.06	0.03	-0.09	0.14
<i>Individual driver effect</i>					
Md_{manual}	0.83	0.04	0.83	0.75	0.90
$Md_{ACC+LKA}$	0.85	0.03	0.85	0.79	0.92
$\Delta(Md)$	0.02	0.05	0.02	-0.07	0.12
<i>Family specific</i>					
M_{manual}	0.94	0.04	0.94	0.85	1.02
$M_{ACC+LKA}$	0.98	0.03	0.98	0.91	1.05
$\Delta(M)$	0.04	0.06	0.04	-0.07	0.15
<i>Individual driver effect</i>					
$\sigma_{z,manual}$	0.11	0.05	0.10	0.04	0.21
$\sigma_{z,ACC+LKA}$	0.11	0.04	0.10	0.05	0.18
<i>Family specific</i>					
σ_{manual}	0.50	0.01	0.50	0.48	0.52
$\sigma_{ACC+LKA}$	0.52	0.00	0.52	0.51	0.53
$\Delta(\sigma)$	0.02	0.01	0.02	0.00	0.04

both driving conditions. At the group level, we found a slight tendency towards higher median off-path glance duration in ACC+LKA than in manual driving. To obtain the group-level off-path glance distribution for both driving conditions, we considered only the parameters β and σ in Table I: we set $\mu = \beta$ in (2). A more intuitive interpretation of the parameter β is to take the exponential transformation, $\exp(\beta)$, that indicates the group-level median of the distribution. Another way to interpret β is to compute $\exp(\beta + \sigma^2/2)$ to obtain the group-level mean of the distribution (off-path glances are often reported in the literature as mean values, even though the mean is not a representative measure of central tendency for skewed distributions). The tendency towards higher median off-path glance duration in ACC+LKA compared to manual driving is revealed by the fact that, although the HPD interval for the difference in median values between driving conditions ($\Delta(Md)$ in Table I) contains the null value, it tends towards positive values (about 70% of the marginal posterior distribution is above zero).

In another finding at the group level, we also estimated higher variance and skewness in the off-path glance distribution in ACC+LKA compared to manual driving. The HPD interval for the difference in σ between driving conditions (Table I) barely includes the null value, and nearly all the marginal posterior distribution (about 98%) falls above zero (the parameter σ , for a fixed μ , is proportional to the skewness and variance of the log-normal distribution).

At the driver level, we found that the between-driver vari-

ability was similar in manual and ACC+LKA driving. In fact, there was little difference in the value for σ_z (Table I), which is the standard deviation of the normal distribution with zero mean, estimates the drivers' variations from the group-level off-path glance tendency.

b) Percentiles: At the group level, we found a tendency towards higher off-path glance percentile values in ACC+LKA compared to manual driving (Fig. 2). The figure shows the distribution of the estimated percentile values in both driving conditions. Their pairwise difference is shown in the right column—which can be interpreted as the Bayesian version of the shift function [42]. An estimation of the difference in percentile values between driving conditions reveals that the null value (thick green vertical line) falls within the HPD interval (black horizontal line). The tendency towards higher percentile values in ACC+LKA compared to manual driving is revealed by the fact that, in general, the distribution of the difference between driving conditions tends towards positive values for all four percentile intervals.

c) PGDoff ≥ 2 : At the group level, we found a tendency towards higher PGDoff ≥ 2 values in ACC+LKA compared to manual driving (Fig. 2). Although the HPD interval of the differences in PGDoff ≥ 2 values between driving conditions (black horizontal line in right graph) contains the null value (thick green vertical line in the right graph), it tends towards positive values (about 81% of the marginal posterior distribution of the difference falls above zero).

B. On-path glance distribution

Table II

SUMMARY STATISTICS OF THE MARKOV CHAIN MONTE CARLO (MCMC) TRACE FOR THE PARAMETERS OF THE ON-PATH GLANCE MODEL, INCLUDING THE DIFFERENCE Δ BETWEEN THE TWO GROUPS (I.E., THE VALUE FOR AUTOMATED DRIVING MINUS THE VALUE FOR MANUAL DRIVING). THE TABLE ALSO PROVIDES A TRANSFORMATION OF THE PARAMETERS TO EASE THE COMPARISON WITH THE RESULTS IN THE LITERATURE. THE MEAN (M) GLANCE DURATION IS COMPUTED AS $3 \cdot \text{logistic}(\beta)$

Parameter	Mean	Sd	Median	HPD	
				2.5%	97.5%
<i>Group level effect</i>					
β_{manual}	2.62	1.95	2.09	0.12	6.67
$\beta_{ACC+LKA}$	6.40	2.64	5.93	2.14	11.58
$\Delta(\beta)$	3.78	3.29	3.58	-2.43	10.97
<i>Individual driver effect</i>					
M_{manual}	2.61	0.33	2.67	1.97	3.00
$M_{ACC+LKA}$	2.97	0.06	2.99	2.86	3.00
$\Delta(M)$	0.36	0.33	0.30	-0.11	1.06
<i>Family specific</i>					
λ_{manual}	1.26	0.61	1.22	0.00	2.33
$\lambda_{ACC+LKA}$	0.75	0.56	0.64	0.00	1.80
<i>Family specific</i>					
λ_{manual}	1.64	0.08	1.63	1.48	1.80
$\lambda_{ACC+LKA}$	1.47	0.03	1.47	1.41	1.52
$\Delta(\lambda)$	-0.17	0.08	-0.17	-0.34	-0.01

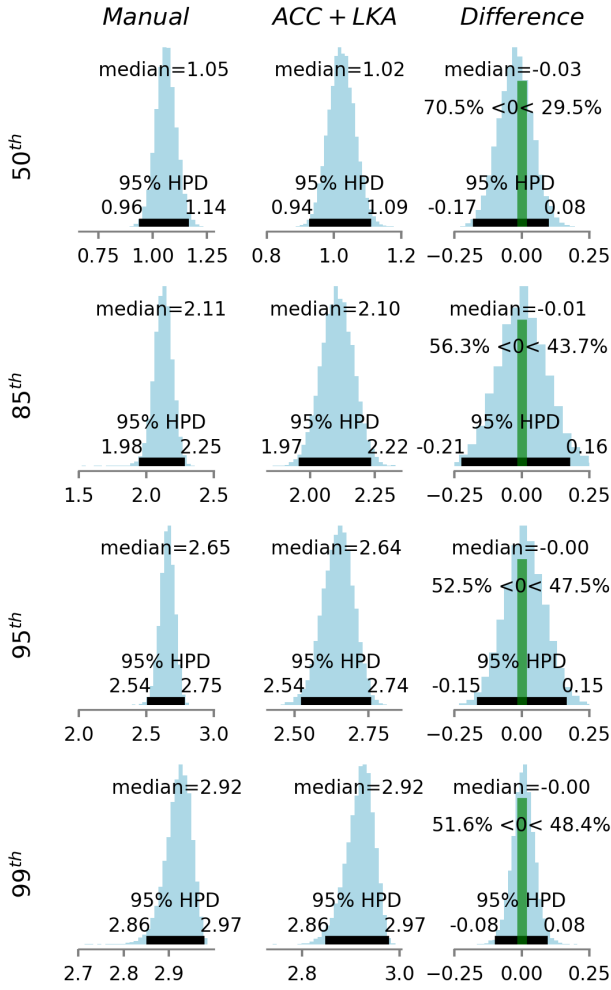


Figure 4. Marginal posterior distributions of the group-level on-path glance percentiles in manual and ACC+LKA driving. Each histogram is annotated with its central tendency and the highest posterior density (HPD) interval (thick horizontal black line). The right column shows the distribution of the difference in percentile values between the two groups (i.e., the value for ACC+LKA driving minus the value for manual driving). The thick vertical green line marks the null value, and the percentages indicate the proportion of posterior samples below and above the null value. The plots are in the style of Kruschke [41].

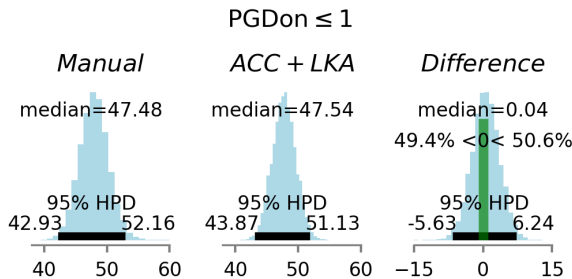


Figure 5. Marginal posterior distributions of the group-level percentage of on-path glances below 1 s (PGDon \leq 1) in manual and ACC+LKA driving. Each histogram is annotated with its central tendency and the highest posterior density (HPD) interval (thick horizontal black line). The right column shows the distribution of the difference in PGDon \leq 1 values between the two groups (i.e., the value for ACC+LKA driving minus the value for manual driving). The thick vertical green line marks the null value, and the percentages indicate the proportion of posterior samples below and above the null value. The plots are in the style of Kruschke [41].

a) *Glance distribution*: Table II gives the summary statistics of the MCMC trace for the fitted inverse-gaussian model defined in (3) for both driving conditions.

At the group level, we identified two noteworthy findings: first, a tendency towards higher mean on-path glance duration in ACC+LKA compared to manual driving. To obtain the group-level mean on-path glance distribution for both driving conditions we considered only the parameters β and λ in Table II: we set $\mu = 3 \cdot \text{logistic}(\beta)$ in (3). Although the HPD interval for the difference in driving condition means ($\Delta(M)$ in Table II) contains the null value, it tends towards positive values (about 92% of the marginal posterior in above zero). The difference between the mean on-path glance durations in ACC+LKA and in manual driving can be up to 1 s (Table II). Note, however, that the logistic link function in (3) quickly saturates to its upper limit, which causes the inverse-gaussian distribution to change slightly for $\beta > 3$, making the differences indistinguishable for values of $\beta > 6$. Hence, some differences in β between driving conditions—caused by values of $\beta > 6$ in the HPD interval (Table II)—may be negligible.

In the second finding at the group level, we estimated higher skewness and variance in the on-path glance distribution for ACC+LKA compared to manual driving. In fact, the HPD interval for the difference in λ between driving conditions (Table II) barely includes the null value, and nearly all of the posterior distribution (about 98%) falls below zero. (The parameter λ , for a fixed μ , is inversely proportional to the skewness and the variance of the inverse-gaussian distribution.)

At the driver level, we found larger between-driver variability in manual driving than in ACC+LKA driving. This difference is represented by σ_z (the standard deviation of the normal distribution with zero mean that estimates the driver's variation from the group-level tendency), which was larger in manual driving.

b) *Percentiles*: At the group level, we found a minimal difference between the percentiles' values in manual and ACC+LKA driving (Fig. 4). As mentioned in the results for the off-path glances, the pairwise difference in percentile values shown in the right column of Fig. 4 can be interpreted as the Bayesian version of the shift function [42]. An estimation of the differences in percentile values between driving conditions reveals that the null value (thick green vertical line) generally falls in the middle of the HPD interval (thick black horizontal line). The exception is the 50th percentile, for which there may be a tendency towards higher percentile values in manual compared to ACC+LKA driving (about 71% of the difference distribution falls below zero).

c) *PGDon \leq 1*: At the group level, we found a slight tendency towards higher PGDon \leq 1 values in manual driving (Fig. 5). In fact, the HPD interval of the differences in PGDon \leq 1 values between driving conditions (thick black horizontal line in right graph) contains the null value (thick green vertical line in the right graph), but it tends slightly towards negative values (about 60% of the difference distribution falls below zero).

C. PRC

Table III gives the summary statistics of the MCMC trace for the fitted beta model defined in (4) for both driving

Table III

SUMMARY STATISTICS OF THE MARKOV CHAIN MONTE CARLO (MCMC) TRACE FOR THE PARAMETERS OF THE PERCENT ROAD CENTER (PRC) MODEL, INCLUDING THE DIFFERENCE Δ BETWEEN THE TWO GROUPS (I.E., THE VALUE FOR AUTOMATED DRIVING MINUS THE VALUE FOR MANUAL DRIVING). THE TABLE ALSO PROVIDES A TRANSFORMATION OF THE PARAMETERS TO EASE THE COMPARISON WITH THE RESULTS IN THE LITERATURE. THE MEAN (M) PRC IS COMPUTED AS $logistic(\beta)$.

Parameter	Mean	Sd	Median	HPD	
				2.5%	97.5%
<i>Group level effect</i>					
β_{manual}	-0.05	0.08	-0.05	-0.22	0.11
$\beta_{ACC+LKA}$	-0.07	0.04	-0.07	-0.15	0.02
$\Delta(\beta)$	-0.02	0.09	-0.02	-0.20	0.17
<i>Individual driver effect</i>					
M_{manual}	0.49	0.02	0.49	0.45	0.53
$M_{ACC+LKA}$	0.48	0.01	0.48	0.46	0.51
$\Delta(M)$	0.00	0.02	0.00	-0.05	0.04
<i>Family specific</i>					
$\sigma_{z,manual}$	0.18	0.08	0.17	0.05	0.35
$\sigma_{z,ACC+LKA}$	0.12	0.04	0.11	0.05	0.19
<i>Family specific</i>					
ϕ_{manual}	14.99	1.21	14.95	12.66	17.35
$\phi_{ACC+LKA}$	15.03	0.54	15.02	14.01	16.12
$\Delta(\phi)$	0.03	1.32	0.06	-2.59	2.54

conditions. At the group level, we found that the mean PRC values in ACC+LKA and manual driving were similar at about 50%. To obtain the group-level PRC distribution for both driving conditions we considered only the parameters β and ϕ in Table III: we set $\mu = logistic(\beta)$ in (4). The transformation $logistic(\beta)$ indicates the group-level mean of the distribution. The HPD interval for the difference in mean value between driving conditions ($\Delta(M)$ in Table III) reveals that the difference around the null value could be up to about 4%. In a second finding at the group level, we also estimated that the variance in the PRC distribution was similar in ACC+LKA and manual driving. In fact, the HPD interval for the difference in ϕ between driving conditions (Table III) is centered at the null value; however, the magnitude of the difference in ϕ could be about 2.5 in absolute value (the parameter ϕ is inversely proportional to the variance of the beta distribution). At the driver level, we found that the between-driver variability was slightly higher in manual than in ACC+LKA driving. This is revealed by the value of σ_z in Table III (the standard deviation of the normal distribution with zero mean that estimates the drivers' variation from the group-level off-path glance tendency), which is slightly larger in manual driving.

D. TTT

Table IV gives the summary statistics of the MCMC trace for the fitted log-normal TTT model defined in (5) for both driving conditions. At the group level, two results stand out: the first is a slight tendency towards higher median TTT duration in ACC+LKA compared to manual driving. To obtain the

Table IV

SUMMARY STATISTICS OF THE MARKOV CHAIN MONTE CARLO (MCMC) TRACE FOR THE PARAMETERS OF THE TOTAL TASK TIME (TTT) MODEL, INCLUDING THE DIFFERENCE Δ BETWEEN THE TWO GROUPS (I.E., THE VALUE FOR AUTOMATED DRIVING MINUS THE VALUE FOR MANUAL DRIVING). THE TABLE ALSO PROVIDES A TRANSFORMATION OF THE PARAMETERS TO EASE THE COMPARISON WITH THE RESULTS IN THE LITERATURE. THE MEDIAN (Md) TTT IS COMPUTED AS $\exp(\beta)$; THE MEAN (M) TTT IS COMPUTED AS $\exp(\beta + \sigma^2/2)$.

Parameter	Mean	Sd	Median	HPD	
				2.5%	97.5%
<i>Group level effect</i>					
β_{manual}	2.00	0.04	2.00	1.93	2.08
$\beta_{ACC+LKA}$	2.03	0.03	2.03	1.97	2.09
$\Delta(\beta)$	0.03	0.05	0.03	-0.07	0.12
<i>Individual driver effect</i>					
Md_{manual}	7.41	0.28	7.41	6.90	7.98
$Md_{ACC+LKA}$	7.60	0.23	7.60	7.13	8.04
$\Delta(Md)$	0.19	0.36	0.19	-0.53	0.88
<i>Family specific</i>					
M_{manual}	8.35	0.32	8.34	7.73	8.98
$M_{ACC+LKA}$	8.48	0.26	8.48	7.97	8.98
$\Delta(M)$	0.13	0.41	0.13	-0.68	0.94
<i>Individual driver effect</i>					
$\sigma_{z,manual}$	0.04	0.04	0.03	0.00	0.12
$\sigma_{z,ACC+LKA}$	0.07	0.03	0.07	0.02	0.14
<i>Family specific</i>					
σ_{manual}	0.49	0.02	0.49	0.45	0.53
$\sigma_{ACC+LKA}$	0.47	0.01	0.47	0.45	0.48
$\Delta(\sigma)$	-0.02	0.02	-0.02	-0.07	0.02

group-level TTT distributions for both driving conditions, we considered only the parameters β and σ in Table IV: we set $\mu = \beta$ in (5). As previously mentioned, the exponential transformation of the parameter β , $\exp(\beta)$, indicates the group-level median of the distribution. The slight tendency towards higher median TTT in ACC+LKA driving is revealed by the fact that the HPD interval for the difference in median value between driving conditions ($\Delta(Md)$ in Table IV) tends towards positive values (about 71% of the marginal posterior distribution is above zero)—although it also contains the null value. However, the magnitude of the difference is below 1 s (Table IV), which may be negligible. The second result at the group level was a higher variance and skewness in the TTT distribution in manual compared to ACC+LKA driving. In fact, although the HPD interval for the difference in σ between driving conditions (Table IV) includes the null value, most of the marginal posterior distribution (about 83%) falls below zero (the parameter σ , for a fixed μ , is proportional to the skewness and variance of the log-normal distribution). At the driver level, we found that the between-driver variability for TTT was similar in manual and ACC+LKA driving. In fact, there were only minor differences in the values for σ_z (Table IV). As mentioned before, σ_z is the standard deviation of the normal distribution with zero mean that estimates the drivers' variation from the group-level TTT distribution.

IV. DISCUSSION

The aim of this paper was to develop a reference model (metrics and target values) for VTS behavior and reveal if vehicle automation (ACC+LKA) influences VTS behavior in real-world driving. The key assets of this paper were 1) the large naturalistic dataset of eye-tracking data (which was discussed and validated in the previous paper [27]), including driving with low-level automation and 2) the Bayesian analysis for driver behavior modeling. To our knowledge, this is the first comprehensive study of VTS in real-world driving.

Naturalistic driving data provide an opportunity to study driver behavior at the highest level of ecological validity, albeit at the cost of other challenges for data analysis: naturalistic data are large, noisy, sparse, and unbalanced. We addressed these challenges by using a rule-based algorithm for extracting VTS sequences and applying the Bayesian method.

The programmatic approach to extracting VTS sequences allowed us to quickly process a large amount of data. Manual video reduction would not have been feasible. The algorithm is similar to the one recently used by Ahlström and Kircher [5], but it differs in two respects. First, the length of the on-path glances before and after the VTS sequence is set to 3 s instead of 4 s. The threshold of 3 s is supported by Tivesten and Dozza’s results [29], in which glance behavior was coded during a phone-related task. (The paper [29] does not report the on-path glance distribution but we received this information as additional glance data provided by the authors of the paper. All driving-related glances were considered on-path glances [29].) It was found that 95% of the on-path glances were shorter than 3 s; hence, glances longer than 3 s may be less likely to be associated with a VTS sequence. Second, there need to be at least three off-path glances instead of only one. Although the value of three glances is arbitrary, we argue that a single glance off-path is not a meaningful VTS sequence.

We applied a Bayesian data analysis, which is a step forward with respect to traditional statistical approaches in the engineering and human factors field. As mentioned in the introduction, Bayesian statistics has many advantages for statistical modeling and inference over frequentist statistics. These advantages have practical implications for modeling drivers’ behavior.

First, with the Bayesian method we estimated the parameters underlying the distribution of the data that comprise the generative model. For example, we not only computed the mean/median of the observations for the off-path glance distribution, but we also computed the parameters μ and σ of the log-normal distribution that described the observations. Results in the literature tend to be scattered because some studies report the mean, others the median, etc. In contrast, the generative model allows the computation of any descriptive statistics of choice, which simplifies the comparison between studies. For example, we could use the parameters of the distribution directly for computing the median ($\exp(\mu)$) or the mean ($\exp(\mu + \sigma^2/2)$) of the log-normal distribution. Or we could compute any other statistics over samples drawn from the log-normal distribution defined by a combination of μ and σ (as we did for the percentiles). The generative model could

have been estimated via maximum likelihood estimation (as in, for example, the previous paper [27]). However, the Bayesian method yields the full distribution of the parameters—not only a point estimate (i.e., the maximum likelihood estimator). The distribution of the parameter’s values reveals its central tendency, along with the uncertainty in the estimation; therefore it enables accurate and robust models for computer simulations. For example, in counterfactual simulations (e.g., see [43]), we can generate many credible glance distributions from the set of values in Table I. (In this case, we should use the samples from the MCMC trace, supplied as supplemental material [38], which carries more information on the correlation between parameters than the marginal distribution does.)

Second, with the Bayesian method, statistical inferences about model parameters are expressed as probability statements based on the posterior distribution. In this paper, we refrained from null-hypothesis testing, because statistical significance may not be practically meaningful [44]. Instead, we gained richer insights into the data by quantifying the magnitude of each effect, its tendency, and its uncertainty. This approach facilitates re-evaluating the results in future, in light of any new or refined human factor theories about the challenges related to emerging vehicle technologies. For example, let’s assume a difference of about 1% in $\text{PGDoff} \geq 2$ between ACC+LKA and manual driving negligible, based on the current state of knowledge; if future research should find that such a difference has practical safety consequences (e.g., a measurably longer response time to threats), then the results in this paper are still valid—but they would need to be re-discussed.

Finally, the Bayesian GLMMs accommodate sources of heterogeneity typical of naturalistic datasets while delivering robust estimations of the overall trend of the data and of between-driver variability. In fact, Bayesian GLMMs estimate group and individual drivers’ parameters simultaneously [31]; estimating parameters for a single driver is informed by all the other drivers in the group. In situations where the data are scarce, the estimate is pulled towards the group level tendency [31]. The results from the GLMMs for the glance-based metrics in both manual and ACC+LKA driving could be the reference model (metrics and target values) for safety assessment, driver modeling, development of ADAS, and guidelines design.

In the following sections we will compare our findings to those available in the literature. Unfortunately, a literature search revealed little quantitative information on visual behavior in VTS. We also found little information on the effects of ACC+LKA on VTS.

A. Off-path glance distribution

a) *Manual driving:* At the group level ($\mu = \beta$ in (2)), the HPD intervals for the parameters β and σ , underlying the off-path glance distribution (Table I), contain the reference values in routine naturalistic driving found in the previous paper [27], $\mu = -0.21$ and $\sigma = 0.50$. The HPD intervals for the 50th and 95th percentiles (Fig. 2) also contain the reference values of, respectively, 0.80 s and 1.90 s, found in the previous paper [27]. In their 2017 field experiment, Lee, Roberts, Reimer, and Mehler found that the average off-path

glance was 1.02 s for the radio-tuning tasks, and 0.91 s for the navigation-entry task [45]; both estimates fall within our HPD interval (Table I). The graphs in Victor, Harbluk, and Engström’s 2005 paper [28] indicate, for a synthetic visual secondary task, a mean off-path glance between 0.95 s and 1.5 s [28], depending on the task difficulty; furthermore, in general, the values were higher in simulated driving than in on-road driving. These values agree with our HPD interval, except for the most challenging task, which caused longer mean off-path glances. Field experiments by Seaman et al. (2016) found a mean off-path glance duration of about 1 s for radio tasks [46], which agrees with our estimated HPD interval.

The HPD interval for $\text{PGDoff} \geq 2$ (Fig. 3) contains the reference value of about 4% in routine naturalistic driving found in the previous paper [27], and the median $\text{PGDoff} \geq 2$ of about 3.4% found for naturalistic phone-related tasks [29]. However, when these phone-related tasks were classified into dialing, texting, or reading, then the median $\text{PGDoff} \geq 2$ for both dialing and texting was higher than our estimation (8.7% and 9.3% respectively). The average $\text{PGDoff} \geq 2$ in another work [28] increased from about 2.5% to 30% as the task difficulty increased in simulated driving [28] (no data were reported for the on-road experiment). These estimates are only partly included in the HPD interval, except for the most challenging task—which required many long off-path glances. The 2017 field experiment [45] by Lee, Roberts, Reimer, and Mehler revealed an average $\text{PGDoff} \geq 2$ of about 5% for the radio-tuning task [45], which is included in the estimated HPD interval (Table I). In the same paper, for a navigation entry task, the average $\text{PGDoff} \geq 2$ was about 1.8%, which is slightly lower than our estimation. Seaman et al.’s field experiments revealed a mean $\text{PGDoff} \geq 2$ between 3% and 4% for a radio-related tasks [46], which agrees with our estimation. Finally, the NHTSA guidelines recommend that “For at least 21 of the 24 test participants, no more than 15% (rounded up) of the total number of eye glances away from the forward road scene have durations of greater than 2 s while performing the testable task one time” [16, p. 24888]. Unfortunately, this recommendation cannot easily be applied to our results. However, if we assume that, at the aggregate level, $\text{PGDoff} \geq 2$ should be less than 15%, then our results comply with the suggested upper limit: we found that the upper limit of the HDP for $\text{PGDoff} \geq 2$ was about 4% in manual driving and about 5% in ACC+LKA driving.

In summary, 1) there was only a minimal difference in off-path visual behavior between routine driving and VTS; 2) the results are congruent with previous studies on visual behavior while performing secondary tasks, in both simulated and on-road driving (the exception being some highly challenging tasks in simulator experiments [28], which may be uncommon in real-world driving).

b) ACC+LKA driving: At the group level ($\mu = \beta$ in (3)) the HPD intervals for the parameters β and σ , underlying the group-level off-path glance distribution (Table I), contain the reference values in routine driving found in the previous paper [27], $\mu = -0.20$ and $\sigma = 0.51$ [27]. Moreover, the values for the 50th and 95th percentiles (Fig. 2), as well as for $\text{PGDoff} \geq 2$ (Fig. 3), also agree with the results from that paper [27]. There was minimal difference in off-path glance behavior

in VTS between manual and ACC+LKA driving. The tendency, however, was towards higher values for all metrics related to the off-path glance behavior (glance distribution, percentiles, and $\text{PGDoff} \geq 2$) in ACC+LKA.

B. On-path glance distribution

In our previous study [27] we reported that few studies have investigated on-path glance behavior. There is much more qualitative and quantitative information about off-path glances. Recent studies, however, have started to acknowledge the relevance of on-path glance behavior for evaluating safety and crash risk [13].

At the group level ($\mu = \beta$ in (3)), the HPD intervals for the parameters β and λ , underlying the group-level on-path glance distribution (Table II), contain the values in routine driving from the previous paper [27], both in manual driving ($\mu = 3.56$, $\lambda = 1.50$) and in ACC+LKA driving ($\mu = 4.08$). The exception is higher values in the HPD interval for the parameter λ than the value of 1.59 reported in the previous paper [27], which may indicate higher skewness in the glance distribution of VTS compared to that of routine driving.

When we compared manual and ACC+LKA driving in VTS, we found 1) shorter mean on-path glances in manual driving (the difference reached 1 s in the HPD interval) and 2) higher between-driver variability in on-path glance duration for manual driving. However, it is possible that, since the on-path glance distribution was truncated at 3 s, the results may be conflated. In a skewed distribution, the mean rarely describes the distribution accurately. In fact, in our results the difference in percentiles (a more robust measure of location) reveals only a slight tendency towards higher median on-path glances in ACC+LKA driving. This finding suggests that the mean is misleading because the distribution is truncated and non-normal.

The HPD interval for $\text{PGDon} \leq 1$ (Fig. 3) contains values that are congruent with what we found in routine naturalistic driving, after we truncated the on-path glance distribution from the previous paper [27] at 3 s to make the models comparable. The truncated on-path glance distribution in routine driving had a $\text{PGDon} \leq 1$ of 48% in manual driving and 46% in ACC+LKA driving.

In summary, we found minimal differences in on-path visual glance behavior both between routine driving and VTS and between manual and ACC+LKA driving in VTS.

C. PRC

At the group level, the HPD interval for the average PRC ($\mu = \text{logistic}(\beta)$ in Table III) contains values that are considerably lower than those of routine driving, in both manual and ACC+LKA driving [27]. In routine driving, the median PRC in daylight, without a lead vehicle in front, was about 85% in manual driving and about 79% in ACC+LKA driving [27]. Interestingly, while in the previous paper [27] the effect of vehicle automation was evident in the PRC values, the results in this study indicate that the effects of automation on PRC during VTS were minimal. The 2005 study by Victor, Harbluk, and Engström reported that PRC decreased from about 70/85%

in baseline driving to a value around 40% when the driver engaged in visual secondary tasks, in both simulated and on-path driving (the reduction in PRC value was correlated with task difficulty) [28]. In general, the results from the current paper agree, except for some values that are slightly lower than our estimation. For phone-related tasks, Tivesten and Dozza described a median PRC of 26% [29], which is considerably lower than our estimation. The median PRC value found in their paper was barely affected by the type of phone task or driving condition (speed and other traffic) [29]. During VTS, we found that the difference in mean PRC between manual and ACC+LKA driving was about 5% (in both directions).

In summary, because VTS involves frequent off-path glances and short on-path glances, a lower value for PRC compared to routine driving is not surprising. However, the results suggest that PRC may be the metric able to discriminate between VTS and routing driving (as previously noted [28]), without being too sensitive to the type of task being executed by the driver.

D. TTT

At the group level, the HPD interval for the median TTT in manual driving ($\Delta(Md)$ in Table IV) contains values that are lower than the median TTT for naturalistic phone-related tasks (the median TTT for phone-dialing was 10.4s, for phone-reading was 11.2s, and for phone-texting was 42.2s [29]). However, because only VTS longer than 3s were included, the results may be biased towards high TTT values. Conversely, the results from a field experiment revealed a median TTT of 2.2s when interacting with the infotainment system [47], a value below our estimation. The median TTT for various secondary tasks included in the NEST database [48] was found to be 1.7s for non-safety critical events, and 6s for events involving a crash [46]; both are congruent with our estimated HPD interval. A series of field experiments [46] found, for tasks involving interactions with the center stack, an average TTT of 3.8s in non-safety critical driving (which is lower than our estimation in Table IV), and 7.7s in crash events (which falls inside our estimation in Table IV). In that work, VTS sequences closer than 5s were merged together, so the results may be biased towards high mean TTT.

In summary, the results in the literature on TTT seem to only partly agree with ours. One reason might be that TTT is sensitive to the task performed, and it does not generalize well. Moreover, the way different studies have defined TTT may also influence the estimation. Unfortunately, a literature search revealed no information on the effect of ACC+LKA on TTT. Our results, however, suggest that there was some difference between the TTT distribution in manual and ACC+LKA driving (we found a slight tendency towards higher median TTT in ACC+LKA compared to manual driving).

E. Limitations and future directions

In this paper, we did not investigate the effect of driving context on VTS behavior (e.g., car-following situations and the effect of illumination) although it is known that drivers' visual response—and secondary task engagement—depends on the driving context situation [27], [29], [49]. The reason was that

there were too few VTS segments in driving situations other than open road driving in daylight. This limitation, however, suggests that when the demand of driving increases (e.g., because of a lead vehicle in front or reduced visibility at night) drivers may be less prone to visual time-sharing. Further research is needed to understand how the VTS reference model can be tuned according to different driving situations.

Another limitation is that the eye-tracker did not provide information about the off-path areas of interest nor the glance eccentricity (i.e., the radial angle between the forward path and the glance location). For example, off-path glances towards the mirrors may have different safety implications compared to glances towards a secondary, distracting task. Moreover, as visual detection performance generally deteriorates towards the retinal periphery, the ability to detect threats and objects on the road may degrade with increasing visual eccentricity [7], [50]–[53]. Further studies, which take glance eccentricity and location into account, will need to be undertaken to improve the reliability of glance-based methods for studying inattention.

V. CONCLUSIONS

We proposed a novel Bayesian reference model for VTS behavior in manual and automated driving. The model is built upon naturalistic eye-tracking data. In general, the model agrees with previous results from simulator and on-road studies on VTS but captures the characteristics and differences in glance-based metrics in greater detail. We found that, in general, 1) the effect of automation on VTS is minimal, and 2) the difference between VTS and routine naturalistic driving is negligible, except for the PRC metric: during VTS there was a lower proportion of time that glances fell within the on-path area (PRC of about 50%) compared to routine driving. The PRC metric may be useful for its sensitivity at distinguishing VTS behavior from routine driving in both manual and automated driving. The model can inform the development of driver models to be used in computer simulations for designing and evaluating the safety benefit of ADASs. Finally, the model could also serve as a detailed reference for developing inattention guidelines.

VI. ACKNOWLEDGMENTS

We thank the colleagues at the Crash Analysis and Prevention group for their comments and suggestions on this paper; Carol Flannagan for her advice on Bayesian data analysis; Emma Tivesten for supplying additional data from her paper; the PyMC3 community (<https://discourse.pymc.io>) for the support; Kristina Mayberry for language revisions; the three anonymous reviewers. Additionally, we thank Volvo Cars and Autoliv for data collection in the EyesOnRoad project. The current work was performed at Volvo Car Safety Center and at SAFER, the vehicle and traffic safety center at Chalmers University of Technology (Gothenburg, Sweden).

REFERENCES

- [1] S. G. Klauer, T. A. Dingus, V. L. Neale, J. D. Sudweeks, and D. J. Ramsey, "The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data," Virginia

- Tech Transportation Institute (VTTI), 3500 Transportation Research Plaza, Blacksburg, VA 24061, United States, Report, 2006. [Online]. Available: <https://vtechworks.lib.vt.edu/bitstream/handle/10919/55090/DriverInattention.pdf>
- [2] S. Singh, "Critical reasons for crashes investigated in the national motor vehicle crash causation survey. (traffic safety facts crash stats)," National Highway Traffic Safety Administration (NHTSA), 1200 New Jersey Avenue, SE, Washington, DC 20590, United States, Report DOT HS 812 115, 2015. [Online]. Available: <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812115>
 - [3] T. Victor, M. Dozza, J. Bärghman, C.-N. Boda, J. Engström, C. Flannagan, J. D. Lee, and G. Markkula, "Shrp2 - analysis of naturalistic driving study data: Safer glances, driver inattention and crash risk," Transportation Research Board and National Academies of Sciences, Engineering, and Medicine, 500 Fifth St. N.W., Washington, DC 20001, United States, Report, 2015.
 - [4] J. Engström, C. A. Monk, R. J. Hanowski, W. J. Horrey, J. D. Lee, D. V. McGehee, M. Regan, A. Stevens, E. Traube, M. Tuukkanen, T. Victor, and C. Y. D. Yang, "A conceptual framework and taxonomy for understanding and categorizing driver inattention," United States Department of Transportation and the European Commission Directorate General for Communication Networks, Content and Technology, 1200 New Jersey Avenue, SE, Washington, DC 20590, United States, Report, 2013, Driver Distraction & Human Machine Interaction Working Group, United States and European Union Bilateral Intelligent Transportation Systems Technical Task Force.
 - [5] C. Ahlstrom and K. Kircher, "A generalized method to extract visual time-sharing sequences from naturalistic driving data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 11, pp. 2929–2938, Nov 2017.
 - [6] T. Victor, C. Ahlström, E. Steinmetz, A. Rydström, J. L. Cano, C. Blåberg, and D. Sandberg, "Semifot task report - visual behavior analysis," SAFER Vehicle and Traffic Safety Centre at Chalmers, Lindholmospiren 3A, Göteborg, SE-417 56, Sweden, Report, 2009.
 - [7] T. Victor, J. Engström, and J. L. Harbluk, *Distraction Assessment Methods Based on Visual Behavior and Event Detection*. CRC Press, 2008, pp. 135–165. [Online]. Available: <http://dx.doi.org/10.1201/9781420007497.ch10>
 - [8] W. Wierwille, "Visual and manual demands of in-car controls and displays," *Automotive ergonomics*, 1993.
 - [9] W. Horrey and C. Wickens, "In-vehicle glance duration: distributions, tails, and model of crash risk," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2018, no. 1, pp. 22–28, 2007.
 - [10] Y. Liang, J. D. Lee, and W. J. Horrey, "A looming crisis: The distribution of off-road glance duration in moments leading up to crashes/near-crashes in naturalistic driving," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 58, no. 1, pp. 2102–2106, 2014.
 - [11] G. Markkula, "Modeling driver control behavior in both routine and near-accident driving," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 58, 2014, pp. 879–883.
 - [12] J. W. Senders, A. B. Kristofferson, W. H. Levison, C. W. Dietrich, and J. L. Ward, "The attentional demand of automobile driving," *Highway research record*, no. 195, 1967.
 - [13] B. D. Seppelt, S. Seaman, J. Lee, L. S. Angell, B. Mehler, and B. Reimer, "Glass half-full: On-road glance metrics differentiate crashes from near-crashes in the 100-car data," *Accid Anal Prev*, vol. 107, pp. 48–62, 2017. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/28787612>
 - [14] L. Tijerina, *Driver distraction and road safety*. LAWYERS & JUDGES PUB, 2015.
 - [15] T. A. Dingus, S. G. Klauer, V. L. Neale, A. Petersen, S. E. Lee, J. D. Sudweeks, M. A. Perez, J. Hankey, D. J. Ramsey, and S. Gupta, "The 100-car naturalistic driving study, phase ii-results of the 100-car field experiment," Virginia Tech Transportation Institute (VTTI), 3500 Transportation Research Plaza, Blacksburg, VA 24061, United States, Report, 2006.
 - [16] Department of Transportation, "Visual-manual nhtsa driver distraction guidelines for in-vehicle electronic devices (docket no. nhtsa-2010-0053)," National Highway Traffic Safety Administration (NHTSA), 1200 New Jersey Avenue, SE, Washington, DC 20590, United States, Report Federal Register vol. 78, no. 81, pp. 24818–24890, 2013. [Online]. Available: <https://www.gpo.gov/fdsys/pkg/FR-2013-04-26/pdf/2013-09883.pdf>
 - [17] M. Corbetta, E. Akbudak, T. E. Conturo, A. Z. Snyder, J. M. Ollinger, H. A. Drury, M. R. Linenweber, S. E. Petersen, M. E. Raichle, D. C. Van Essen, and G. L. Shulman, "A common network of functional areas for attention and eye movements," *Neuron*, vol. 21, no. 4, pp. 761–773, 1998. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0896627300805930>
 - [18] J. D. Lee, J. Moeckli, T. L. Brown, S. C. Roberts, C. Schwarz, L. Yekhshatyan, E. Nadler, Y. Liang, T. Victor, and D. Marshall, "Distraction detection and mitigation through driver feedback," National Highway Traffic Safety Administration, 1200 New Jersey Avenue, SE, Washington, DC 20590, United States, Report DOT HS 811 547A, 2013.
 - [19] J. D. Lee, J. Moeckli, T. L. Brown, S. C. Roberts, C. Schwarz, L. Yekhshatyan, E. Nadler, Y. Liang, T. Victor, D. Marshall, and C. Davis, "Traffic safety facts: Distraction detection algorithm evaluation," National Highway Traffic Safety Administration (NHTSA), 1200 New Jersey Avenue, SE, Washington, DC 20590, United States, Report DOT HS 811 548, 2013. [Online]. Available: <https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/811548.pdf>
 - [20] 2017-04-10 2017. [Online]. Available: <http://media.gm.com/media/us/en/cadillac/news.detail.html/content/Pages/news/us/en/2017/apr/0410-superccruise.html>
 - [21] R. Broström, P. Bengtsson, and M. L. Aust, "Individual glance strategies and their effect on the nhtsa visual manual distraction test," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 36, pp. 83–91, 2016. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1369847815001680>https://ac.els-cdn.com/S1369847815001680/1-s2.0-S1369847815001680-main.pdf?_tid=d687081b-d7d2-4825-982a-695f770cb09e&acdnat=1530004243_fe3e79232d1fda5c3adc7a383b9878f4
 - [22] J. K. Kruschke and T. M. Liddell, "Bayesian data analysis for newcomers," *Psychonomic Bulletin & Review*, pp. 1–23, 2017. [Online]. Available: <http://dx.doi.org/10.3758/s13423-017-1272-1>
 - [23] J. K. Kruschke and T. M. Liddell, "The bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a bayesian perspective," *Psychonomic Bulletin & Review*, pp. 1–29, 2017. [Online]. Available: <http://dx.doi.org/10.3758/s13423-016-1221-4>
 - [24] J. K. Kruschke, "Rejecting or accepting parameter values in bayesian estimation," *Advances in Methods and Practices in Psychological Science*, vol. 1, no. 2, pp. 270–280, 2018. [Online]. Available: <http://journals.sagepub.com/doi/abs/10.1177/2515245918771304>
 - [25] J. Karlsson, C. Apoy, H. Lind, S. Dombrovskis, M. Axestål, and M. Johansson, "Eyesonroad - an anti-distraction field operational test," Vinnova FFI - Vehicle and Traffic Safety Program, Mäster Samuelsgatan 56, Stockholm, SE-101 58, Sweden, Report, 2016. [Online]. Available: https://www.vinnova.se/contentassets/5cb2a63a302b4e859e642ff20ea86550/2013-01303_en.pdf
 - [26] B. Nilsson, "Driver attentiveness detection method and device," U.S. Patent 9 834 221, 2017.
 - [27] A. Morando, T. Victor, and M. Dozza, "A reference model for driver attention in automation: Glance behavior changes during lateral and longitudinal assistance," *IEEE Transactions on Intelligent Transportation Systems*, 2018.
 - [28] T. Victor, J. L. Harbluk, and J. A. Engström, "Sensitivity of eye-movement measures to in-vehicle task difficulty," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 8, no. 2, pp. 167–190, 2005. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1369847805000161>
 - [29] E. Tivesten and M. Dozza, "Driving context and visual-manual phone tasks influence glance behavior in naturalistic driving," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 26, no. PA, pp. 258–272, 2014.
 - [30] T. Sorensen and S. Vasishth, "Bayesian linear mixed models using stan: A tutorial for psychologists, linguists, and cognitive scientists," *arXiv preprint arXiv:1506.06201*, 2015.
 - [31] J. K. Kruschke and W. Vanpaemel, "Bayesian estimation in hierarchical models," *The Oxford handbook of computational and mathematical psychology*, pp. 279–299, 2015.
 - [32] M. Betancourt and M. Girolami, "Hamiltonian monte carlo for hierarchical models," *Current trends in Bayesian methodology with applications*, vol. 79, p. 30, 2015.
 - [33] O. Papaspiliopoulos, G. O. Roberts, and M. Sköld, "A general framework for the parametrization of hierarchical models," *Statistical Science*, pp. 59–73, 2007.
 - [34] J. I. Figueroa-Zúñiga, R. B. Arellano-Valle, and S. L. Ferrari, "Mixed beta regression: A bayesian perspective," *Computational Statistics & Data Analysis*, vol. 61, pp. 137–147, 2013.
 - [35] J. Salvatier, T. V. Wiecki, and C. Fonnesbeck, "Probabilistic programming in Python using PyMC3," *PeerJ Computer Science*, vol. 2, p. e55, 2016. [Online]. Available: <https://doi.org/10.7717/peerj-cs.55>

- [36] M. D. Hoffman and A. Gelman, "The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1593–1623, 2014.
- [37] A. Gelman and D. B. Rubin, "Inference from iterative simulation using multiple sequences," *Statist. Sci.*, vol. 7, no. 4, pp. 457–472, 1992. [Online]. Available: <https://projecteuclid.org:443/euclid.ss/1177011136>
- [38] "Supplemental material," https://github.com/ruvigroup/Paper_VTS_2018_supplemental, accessed: 2018-10-10.
- [39] J. K. Kruschke, "Bayesian estimation supersedes the t test," *Journal of Experimental Psychology: General*, vol. 142, no. 2, p. 573, 2013.
- [40] P. D. Ellis, *Interpreting effects*, ser. The essential guide to effect sizes : statistical power, meta-analysis, and the interpretation of research results. Cambridge: Cambridge University Press, 2016.
- [41] J. Kruschke, *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*, 2nd ed. Academic Press, 2014.
- [42] G. A. Rousselet, C. R. Pernet, and R. R. Wilcox, "Beyond differences in means: robust graphical methods to compare two groups in neuroscience," *European Journal of Neuroscience*, vol. 46, no. 2, pp. 1738–1748, 2017. [Online]. Available: <http://dx.doi.org/10.1111/ejn.13610>
- [43] J. Bärgerman, V. Lisovskaja, T. Victor, C. Flannagan, and M. Dozza, "How does glance behavior influence crash and injury risk? a 'what-if' counterfactual simulation using crashes and near-crashes from shrp2," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 35, no. Supplement C, pp. 152–169, 2015. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S136984781500162X>
- [44] P. D. Ellis, *The essential guide to effect sizes : statistical power, meta-analysis, and the interpretation of research results*. Cambridge: Cambridge University Press, 2016.
- [45] J. Lee, S. C. Roberts, B. Reimer, and B. Mehler, "Does order matter? investigating the effect of sequence on glance duration during on-road driving," *PLOS ONE*, vol. 12, no. 2, pp. 1–19, 02 2017. [Online]. Available: <https://doi.org/10.1371/journal.pone.0171730>
- [46] S. Seaman, J. Lee, L. Angell, B. Mehler, B. Seppelt, and B. Reimer, "Exploring generalizability of field experiment radio tasks with naturalistic driving data: A comparison with SHRP2 NEST," in *Adjunct Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. New York, NY, United States: ACM, 2016, pp. 111–116. [Online]. Available: <http://doi.acm.org/10.1145/3004323.3004350>
- [47] M. A. Perez, L. S. Angell, and J. M. Hankey, "Assessment of naturalistic use patterns of advanced infotainment systems," *Human factors*, vol. 57, no. 4, pp. 674–688, 2015.
- [48] J. M. Owens, L. Angell, J. M. Hankey, J. Foley, and K. Ebe, "Creation of the naturalistic engagement in secondary tasks (nest) distracted driving dataset," *Journal of Safety Research*, vol. 54, pp. 11–116, 2015. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S002243751500050X>
- [49] E. Tivesten and M. Dozza, "Driving context influences drivers' decision to engage in visual–manual phone tasks: Evidence from a naturalistic driving study," *Journal of Safety Research*, vol. 53, pp. 87 – 96, 2015. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0022437515000225>
- [50] A. Morando, T. Victor, and M. Dozza, "Drivers anticipate lead-vehicle conflicts during automated longitudinal control: Sensory cues capture driver attention and promote appropriate and timely responses," *Accident Analysis & Prevention*, vol. 97, pp. 206–219, 2016.
- [51] H. Summala, T. Nieminen, and M. Punto, "Maintaining lane position with peripheral vision during in-vehicle tasks," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 38, no. 3, pp. 442–451, 1996.
- [52] H. Summala, D. Lamble, and M. Laakso, "Driving experience and perception of the lead car's braking when looking at in-car targets," *Accident Analysis & Prevention*, vol. 30, no. 4, pp. 401–407, 1998.
- [53] D. Lamble, M. Laakso, and H. Summala, "Detection thresholds in car following situations and peripheral vision: Implications for positioning of visually demanding in-car displays," *Ergonomics*, vol. 42, no. 6, pp. 807–815, 1999.



Alberto Morando received his M.Sc. in Mechatronic engineering from University of Trento (Trento, Italy) in 2014. He is currently a PhD student at Chalmers University of Technology (Gothenburg, Sweden) in Human factors of automated driving. His research includes the development of new methods for glance analysis and driver modelling. He was a Marie Curie Fellow (Early Stage Researcher) in the HF Auto ITN project.



Trent Victor received the Ph.D. degree in psychology from Uppsala University, Uppsala, Sweden, in 2005. He is currently Senior Technical Leader Crash Avoidance at Volvo Cars Safety Centre, Adjunct Professor at Chalmers University of Technology (affiliated with the SAFER Vehicle and Traffic Safety Center) in Gothenburg, Sweden, and Adjunct Professor at University of Iowa. At Volvo Cars Safety Centre he provides leadership safety-related issues in the development of crash avoidance systems, autonomous vehicles, safety impact analyses, and human factors. He has over 22 years of work experience in design and evaluation in the field of intelligent vehicle systems but his deepest technical knowledge is within driver attention and visual behavior.



Marco Dozza received the M.E. degree from the University of Bologna (Bologna, Italy) in 2002 and the Ph.D. degree in bioengineering from the University of Bologna, in collaboration with Oregon Health & Science University (Portland, OR, USA) in 2007. After graduation, he worked as a System Developer for over two years with Volvo Technology, a research and innovation company inside the Volvo group. Since 2009, he has been at Chalmers University of Technology (Gothenburg, Sweden) where he is an Associate Professor. Marco Dozza is Examiner for the course Active Safety in the Master's Programme for Automotive Engineering. He is also affiliated with the SAFER Vehicle and Traffic Safety Center, where he leads several projects on traffic safety.