

# On the necessity of adaptive eye movement classification in conditionally automated driving scenarios

Christian Braunagel\*, David Geisler, Wolfgang Stolzmann  
Camera Systems  
Daimler AG  
Germany

Wolfgang Rosenstiel, Enkelejda Kasneci  
Computer Engineering  
University of Tübingen  
Germany

## Abstract

Algorithms for eye movement classification are separated into threshold-based and probabilistic methods. While the parameters of static threshold-based algorithms usually need to be chosen for the particular task (task-individual), the probabilistic methods were introduced to meet the challenge of adjusting automatically to multiple individuals with different viewing behaviors (inter-individual). In the context of conditionally automated driving, especially while the driver is performing various secondary tasks, these two requirements of task- and inter-individuality fuse to an even greater challenge. This paper shows how the combination of task- and inter-individual differences influences the viewing behavior of a driver during conditionally automated drives and that state-of-the-art algorithms are not able to sufficiently adapt to these variances. To approach this challenge, an extended version of a Bayesian online learning algorithm is introduced, which is not only able to adapt its parameters to upcoming variances in the viewing behavior, but also has real-time capability and lower computational overhead. The proposed approach is applied to a large-scale driving simulator study with 74 subjects performing secondary tasks while driving in an automated setting. The results show that the eye movement behavior of drivers performing different secondary tasks varies significantly while remaining approximately consistent for idle drivers. Furthermore, the data shows that only a few of the parameters used for describing the eye movement behavior are responsible for these significant variations indicating that it is not necessary to learn all parameters in an online-fashion.

**Keywords:** Automated analysis methods, Eye movements and cognition, Machine learning methods and algorithms

**Concepts:** •Computing methodologies → Activity recognition and understanding; Mixture modeling; Model verification and validation;

## 1 Introduction

Of the six primary eye movement types, namely fixation, saccade, smooth pursuit, vergence, optokinetic and vestibulo-ocular reflex, saccades and fixations occur most frequently in daily life [Leigh and Zee 2015]. In brief, fixations keep the gaze stable on a stationary target while saccades enable the eye to switch between differ-

ent fixation targets [Holmqvist et al. 2011]. Because the automated classification of these two eye movement types is a crucial topic for many applications such as the diagnosis of pathologies [Isotola et al. 2009], human-machine interaction (HMI) [Barea et al. 2003; Gandhi et al. 2010], human activity recognition [Bulling et al. 2011; Banerjee et al. 2014], and hazard perception during driving [Kasneci et al. 2015], the literature provides various methods for this purpose [Salvucci and Goldberg 2000; Tafaj et al. 2012]. An especially complex environment for eye movement classification algorithms is represented by conditionally automated driving scenarios [International 2014]. This automation level is characterized by an autonomous driving function able to take over the driving task and the responsibility for a specific time interval. However, such systems will experience situations, which are difficult to handle and, therefore, the driver is requested to take over in a limited amount of time. While the vehicle is in control, the driver is enabled to perform various secondary tasks like reading, writing a mail, or just relaxing. By switching between such tasks, the driver shows task-individual eye movements such as large saccades while scanning the environment or small saccades while reading. Furthermore, due to different drivers or different driver seat settings, inter-individual viewing behavior occurs. Therefore, a robust and reliable detection of eye movements is indispensable for in-vehicle systems based on eye movement detection such as driver-activity recognition [Braunagel et al. 2015]. Such systems are necessary for classifying the driver's current take-over readiness.

However, current algorithms for eye movement classification are suited and evaluated for either task- or inter-individual differences in the viewing behavior. A thorough examination of how significant the effect of combined task- and inter-individual differences on the viewing behavior is as well as an evaluation of the performance of existing methods under these circumstances does not exist up to now. Furthermore, to the best of the author's knowledge, there exists no literature on the viewing behavior and the individual eye movements of the driver in conditionally automated driving situations. Hence, in this paper we investigate how the task- and inter-individual differences in conditionally automated drives influence the eye movements. Furthermore, we introduce a novel approach for an adaptive eye movement classification based on a Bayesian Mixture Model [Tafaj et al. 2012] and suitable for common rapid control prototyping (RCP) and hardware in the loop (HIL) tools in the vehicle.

## 2 Related Work

The automated classification of fixations and saccades is a well-studied topic resulting in a variety of algorithms using different ways for detecting these basic eye movements. Despite several related work in this context, two questions are still open: (1) how significantly is the viewing behavior influenced by individual differences among various tasks and (2) to which extent the parameters of such classification algorithms need to be adaptively adjusted. This work focusses on the inter- and task-individual differences in the viewing behavior during conditionally automated driving while performing secondary tasks.

\*e-mail: christian.braunagel@daimler.com

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org. © 2016 ACM. ETRA 2016, March 14 - 17, 2016, Charleston, SC, USA ISBN: 978-1-4503-4125-7/16/03 DOI: <http://dx.doi.org/10.1145/2857491.2857529>

For a better overview of the different classification methods, a new taxonomy consisting of a minimal set of three spatial and two temporal criteria was first introduced in [Salvucci and Goldberg 2000]. Furthermore, by means of five different representative algorithms, the particular advantages and disadvantages of the chosen categories were evaluated. One of these categories describes velocity-based algorithms, which separate saccades and fixations by analyzing the velocity of the sequential data points. As representatives, an adaptive as well as a threshold-based method is introduced, but there is no further distinction of these types in the taxonomy. Instead, it is suggested that fixed thresholds are usually sufficient for the classification, since the velocity profiles are assumed to be physiologically stable [Salvucci and Goldberg 2000].

However, how this fixed threshold has to be chosen depends on the respective task. As an example, with a fixed velocity threshold of  $15^\circ s^{-1}$  the trajectories of self-paced saccades were examined in [Erkelens and Vogels 1995], while in [Sen and Megaw 1984] a threshold value of  $20^\circ s^{-1}$  is used to detect effects on saccades while working on visual display units. A summary of further settings for this threshold is given in [Rötting 2001].

For an appropriate threshold selection for eye movement classification algorithms, a logic based on novel standardized scores was discussed in [Komogortsev et al. 2010]. Furthermore, how significantly the classification performance of well-known algorithms is influenced by different choices of the threshold parameter was analyzed in detail. However, the calculation of these metrics is based on the assumption that the stimuli were specified by the examiner. In case of real-world eye-tracking data, a ground truth for the evaluation of eye movement classification algorithms is typically provided by a tedious and time-consuming manual classification. Note that all settings summarized in [Rötting 2001] and the two examples mentioned above were applied to eye-tracking data recorded in a static lab environment, too.

Driving scenarios can doubtlessly be considered as far more dynamic environments than lab environments. Therefore, an additional distinction of algorithms for eye movement classification in threshold-based and probabilistic methods was given in [Kasneji et al. 2014] and their applicability to non-automated driving scenarios was discussed. It was stated that threshold-based methods, which make use of empirically adjusted thresholds, are not applicable to highly dynamic scenarios and the strongly physically- and physiologically-dependent viewing behavior. Yet no further references or details on this statement were given. Further, two state-of-the-art probabilistic methods, namely Bayesian Mixture Model (BMM) [Tafaj et al. 2012] and Hidden Markov Model (HMM) [Salvucci and Anderson 1998], were compared in terms of their applicability to online driving scenarios, revealing a superior classification performance of the BMM over the HMM. BMM considers the Euclidean distances of sequential data points for its classification and is based on the assumption that these distances, either describing a fixation or a saccade, are Gaussian distributed. The parameters of the applied Gaussian Mixture Model (GMM) will be updated with every new classified distance in an online-fashion. Note that if a constant sample time is considered, the distances can be seen as velocities.

Studies which specifically examined individual viewing behavior, typically in the field of psychology, focus on the eye movements during specific tasks. For example, in [Castelhano and Henderson 2008] the authors found inter-individual differences in the saccadic amplitudes during the scanning process of images, while the intra-individual saccadic amplitudes were stable. Moreover, there is a considerable amount of eye movement studies in reading describing differences in the viewing behavior among various types of readers. A comprehensive overview on the above work is given

in [Rayner 1998]. All these findings indicate that there is a significant individual difference in the eye movement parameters among individuals performing the same task, while the existence of various threshold settings for different tasks suggests a non-negligible task-individual difference. The scenario of conditionally automated driving exposes further challenges to the eye movement classification algorithms, since both task- and inter-individual differences occur at the same time. There are plenty of possible secondary tasks, which can be performed in conditionally automated driving scenarios and among which the driver can switch frequently. Examples for possible secondary tasks are reading news, writing emails, watching a movie, or just relaxing and observing the environment. Even the level of automation can change between conditionally automated and non-automated route sections, so that the driver needs to take over or hand over the control of the vehicle. All these varying task-individual conditions can influence the eye movement behavior of the driver. Furthermore, inter-individual differences need to be considered due to the possibility of multiple various drivers per vehicle and drive. Since the eye movement behavior of various drivers can react individually for the different conditions, the task- and inter-individual differences intensify each other.

The contributions of this paper are two-fold. On the one hand, it is shown that state-of-the-art classification algorithms such as the BMM cannot sufficiently adapt their parameters to eye-tracking data which frequently changes its underlying distribution. This is a crucial factor since a driver often switches between different secondary tasks in the conditionally automated driving scenarios. Therefore, an enhanced version of the BMM is introduced with several advantages towards the previous method such as real-time capability, easier implementation geared towards embedded architectures, and an improved adaptation. On the other hand, a thorough examination is performed of the task- and inter-individual differences of the eye movement behavior in case of conditionally automated driving situations, where the driving situations are characterized by frequent changes between automated and manual driving sections and among various secondary tasks. This is done using the novel classification algorithm to describe the currently underlying distributions of the saccade and fixation velocities as well as occurring changes in these distributions to avoid manual classification.

### 3 MERCY: Moving Estimation Classification

Conditionally automated driving enables the driver to perform secondary tasks while the automated driving function takes over control of the vehicle. As outlined in Section 2, the literature suggests that by varying the task the resulting eye movement behavior varies as well. Hence, a classification algorithm needs to be able to adapt its internal parameters to classify fixations and saccades sufficiently. To what extend current classification methods satisfy these requirements will be investigated in this section by means of the BMM [Tafaj et al. 2012], resulting in a **Moving Estimation Classification** or MERCY, an enhanced version of the BMM.

#### 3.1 Evaluation of BMM

Based on the same assumption as in [Tafaj et al. 2012] that the underlying process generating the velocities of fixations and saccades can be described by a GMM, the probability density function  $p(\|v_i\|)$  of the model is given by

$$p_f(\|v_i\|) = \pi_f N(\|v_i\|, \mu_f, \beta_f) \quad (1)$$

$$p_s(\|v_i\|) = \pi_s N(\|v_i\|, \mu_s, \beta_s) \quad (2)$$

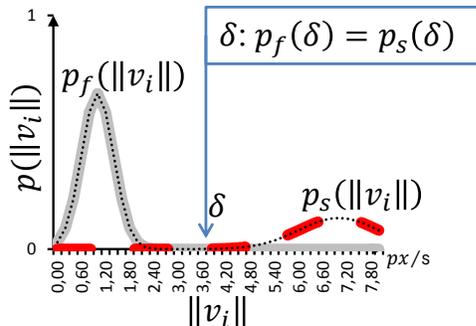
$$p(\|v_i\|) = p_f(\|v_i\|) + p_s(\|v_i\|) \quad (3)$$

where  $v_i$  is the measured velocity between the two sequential data points with the index  $i-1$  and  $i$ , the parameters  $\mu$  and  $\beta$  describe the

mean resp. the variance of the Gaussian distribution,  $\pi$  describes the mixture parameter, and the indices  $f$  and  $s$  denote the components of the fixations or the saccades. The norm  $\|\cdot\|$  represents the Euclidean distance. The classification process can be seen as the determination of the intersection  $\delta$  of the two probability density functions  $p_f$  and  $p_s$  and consequently boils down to the estimation of the means, variances, and mixture components denoted by the parameter set

$$\Theta_k = \{\mu_k, \beta_k, \pi_k\} \quad \text{where } k = \{f, s\}. \quad (4)$$

An artificially generated example of a GMM is shown for illustration in Figure 1.



**Figure 1:** The three probability density functions  $p_f$  (grey),  $p_s$  (red), and  $p$  (dotted black).

To determine the parameter sets  $\Theta_f$  and  $\Theta_s$ , Variational Message Passing (VMP) as implemented by Infer.NET<sup>1</sup> was used [Winn and Bishop 2005]. Since VMP is an advanced approximation technique for applying variational inference to Bayesian Networks, the time required as well as the complexity for implementing such a framework in order that it can be run online on common RCP and HIL tools in the vehicle are still enormous. Furthermore, the VMP algorithm is realized as an iterative method converging in terms of a lower bound [Winn and Bishop 2005]. For this iterative approach, it cannot be determined a priori how many iterations need to be performed, which can be problematic in terms of real-time applications.

Artificially generated data is used to evaluate the ability of the BMM to adapt its parameters to frequently changing eye movement behavior. The reason for this approach is to provide a ground truth of the GMM and its parameter sets, which facilitates a simulation of the frequent changes of the underlying process and the evaluation in total. For creating the GMM, the MATLAB class *gmdistribution*<sup>2</sup> was used, which is able to generate random numbers of the specified mixture model. In total, 25000 random data samples were generated with different pre-defined parameter sets for a first evaluation. After 5000 samples the parameter set was changed for the first time and another 5000 data samples were generated. In Figure 3, this procedure was repeated four times, before 5000 random data samples were generated by a step-by-step changing model, resulting in a continuously decreasing threshold at the end of the figure. Table 1 specifies which parameters were used and varied for the different intervals each containing the 5000 samples. The BMM was trained with the first 1000 data samples, before starting the online adaption, which explains the gap at the beginning of the BMM

<sup>1</sup><http://research.microsoft.com/en-us/um/cambridge/projects/infernet/>

<sup>2</sup><http://de.mathworks.com/help/stats/index.html>

Parameter set of the x-th interval					
interval	1	2	3	4	5
$\mu_1$	1	1	1	1.1	$1.1 - i \frac{1}{e^9}$
$\mu_2$	200	220	210	207	207
$\beta_1$	0.1	0.1	0.14	0.33	0.33
$\beta_2$	400	400	400	404	$404 - i \frac{1}{e^3}$
$\pi_1$	0.8	0.5	0.6	0.7	0.7
$\pi_2$	0.2	0.5	0.4	0.3	0.3

**Table 1:** Parameter sets of the different 5000 samples large intervals shown in Figure 3. The varied values between two sequential steps are highlighted by a blue shaded background. The parameter  $i$  in the last column represents the  $i$ -th iteration, since these values were varied for every iteration.

plot. The initial parameters as well as the variations of these values in Table 1 were chosen to provide a meaningful GMM according to preliminary studies, while still providing distinctly separable data for a moderate classification task.

As shown in Figure 3, the BMM based on the velocity distributions adapts poorly to the frequently changing GMM. It shows among others the calculated intersection point of the artificial GMM and the intersection of the estimated GMM of the BMM. After the training phase, the estimated intersection point of the BMM differs from the actual intersection, but is slowly approaching it. The reason for this slow behavior is that the generated data samples are weighted lower the later they are given to the algorithm. Hence, shortly after the training phase new data samples already have little to no effect on the parameters of the BMM. Consequently, for significantly emerging differences such as at sample 15,000 in the fourth interval, the BMM is overwhelmed by the adaption process. Section 4.2 shows that the individual differences due to performing different tasks are significantly larger than for the artificially generated data at this point. Hence, an even worse estimation of the mixture model in case of conditionally automated driving data is expected.

### 3.2 A Novel Approach for Improved Adaptability

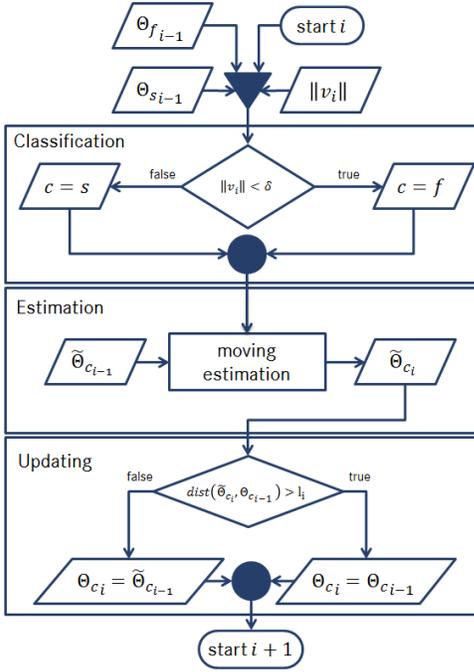
This section introduces MERCY, a novel approach for an improved estimation of the parameters of GMMs and suitable for implementation on common RCP and HIL tools in the vehicle. The architecture of this approach is illustrated in Figure 2 and can be separated into three iterative steps: estimation, updating, and classification. These steps are performed in each iteration, requiring the parameter sets  $\Theta_f$  and  $\Theta_s$  of the previous round and the current measured velocity.

#### Classification

Classification is performed in the same way as in the BMM algorithm by comparing the current velocity  $\|v_i\|$  to the intersection point  $\delta$ . If the velocity is smaller than the intersection point, i.e. it lies on the left side of the intersection, it will be marked as fixation otherwise as a saccade. After the classification, the algorithm is able to update one of the two distributions depending on the belonging of the current velocity.

#### Estimation

The main idea behind MERCY is to estimate the parameter sets  $\Theta_f$  and  $\Theta_s$  by means of sample mean and sample variance. These estimations of the means, variances, and mixture components can



**Figure 2:** Architecture of the novel algorithm MERCY.

easily be reformulated into a recursive form. Furthermore, to prevent the estimation from converging and that new data samples will be considered with decreasing weight, the recursive formulas can be provided with a weighting factor  $\omega$ , which can be interpreted as the size of a moving window. Choosing a small  $\omega$  leads to very dynamic behavior of the estimation, but increases the influence of outliers on the estimation. On the other hand, choosing a large  $\omega$  results in idle behavior which adapts slowly to the changing conditions. Given the simplifying assumption that the velocities  $v_1, v_2, \dots$  are realizations of the random variable  $\mathcal{V}$  generated by an independent and identically distributed process, the recursive equation for the weighted sample mean is defined as

$$\mu_{k_{n+1}} = \frac{\omega \mu_{k_n} + v_{n+1}}{\omega + 1} \quad (5)$$

and the recursive equation of the weighted sample variance is defined as

$$\beta_{k_{n+1}} = \frac{(\omega - 1)\beta_{k_n} + (V_{n+1} - E[V_{n+1}])^2}{\omega}. \quad (6)$$

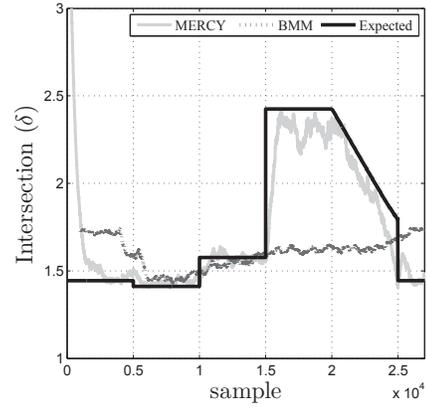
Note that the estimation of the variance depends on the estimation of the mean of the same round. The estimation of the mixture components  $\pi_f$  and  $\pi_s$ , which describe the ratio between the number of data samples classified as fixations and saccades, is realized by means of a weighted counter given by

$$\pi_{k_{n+1}} = \frac{\omega \pi_{k_n} + 1}{\omega + 1}. \quad (7)$$

In comparison to the estimation of the sample mean and sample variance, both parameters  $\pi_k$  can be updated in every round of the algorithm independent of the classification result.

## Updating

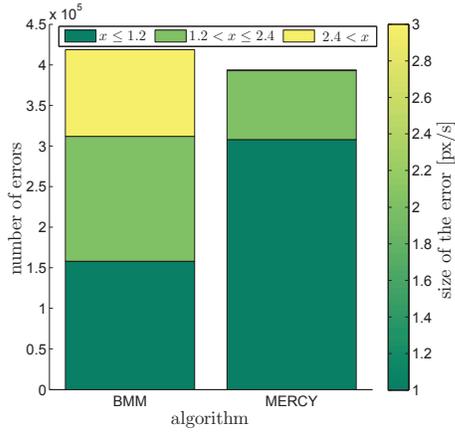
As shown in Figure 2, there are two pairs of parameter sets  $\Theta_k$  and  $\tilde{\Theta}_k$ . While the parameter sets  $\Theta_k$  describe the actual parameters used for the classification,  $\tilde{\Theta}_k$  depict the currently estimated parameters of the GMM with regard to the new data samples. If these parameter sets diverge more than a pre-defined threshold  $l$ , the currently estimated parameters  $\tilde{\Theta}_k$  will be used as new parameter set  $\Theta_k$  for the classification in the next round. As long as the threshold is set to  $l = 0$ , the actual model parameters will be updated with every new data sample, which is in general the proper approach. Nevertheless, this separation into two parameter sets was considered as a possible additional analysis of the task- and inter-individual differences. Since the whole algorithm has a constant complexity  $O(1)$ , this method is suitable for most real-time applications.



**Figure 3:** The three plots show the artificially generated threshold (black solid line), the estimated threshold by the BMM (dotted gray line), and MERCY (solid light gray line).

MERCY is applied to the same artificial data samples as the BMM in Section 3.1 and the estimation is plotted as a light gray line in Figure 3. Although the initialization values were chosen with an offset of 0.5 in the means and variances, resulting in a starting position of the estimated intersection point at  $4 \text{ px/s}$ , the algorithm adapts as fast as the BMM to the artificial model. However, MERCY is more accurate than the BMM up to iteration 15,000. In contrast to the BMM, MERCY is still able to detect and adapt to the changing distribution in the fourth interval, but the error between the actual intersection and the estimation increases slightly, due to the lack of a sufficient number of data samples. The performance of MERCY exceeds the performance of the BMM for larger steps in intersection point  $\delta$  and MERCY adapts appropriately even in the fifth interval with the continuously decreasing intersection.

A large-scale data set of half a million data samples, generated by randomly changing parameters of the artificial GMM was created in the same way as in the exemplary plot of Figure 3. The parameters were varied randomly every 10,000 samples so that every parameter, e.g. the mean  $\mu_f$ , was set to a value of the interval defined by the initial value and the radius, e.g.  $[\mu_f - \mu_f/2, \mu_f + \mu_f/2]$ . Furthermore, every 50,000 samples, MERCY was reset to the initial values and the parameter of the BMM were determined by an additional training phase. To compare the performance of both algorithms, the absolute error between the actual intersection point and the estimated points was calculated. The result is shown as a stacked bar diagram in Figure 4. For the plot, the intervals of the training phases of the BMM and all absolute errors smaller than



**Figure 4:** Two stacked bars illustrate the absolute error between each algorithm and the actual intersection point of the artificial GMM. The three stacks per bar represent three classes of error sizes.

0.1  $px/s$  were not considered. In addition, one round of 50,000 samples was discarded because the BMM was not able to calculate a meaningful initialization of the model in the training phase. The calculated error was separated into three different error classes, dividing them into small errors  $\leq 1.2 px/s$ , medium size errors between  $1.2 px/s < x \leq 2.4 px/s$ , and the class of the large errors with  $2.4 px/s < x$ . The stacked bar of MERCY shows no errors for the large errors class, since there are too few to be visible in the plot. There are 25,000 errors smaller than 0.1  $px/s$  resulting in a decreased bar height compared to the bar of the BMM. The right bar can be coarsely divided into one quarter of medium size errors and three quarters of small size errors. In contrast, the bar of the BMM can be divided into three nearly equal stacks of the different error classes. As suggested by the example in Figure 3, MERCY adapts considerably better to the given data samples than the BMM, providing fewer and smaller errors in terms of the intersection point. All errors of every parameter of  $\Theta_f$  and  $\Theta_s$  affect the estimation of the intersection point, which therefore can be seen as the worst-case scenario for the estimation. In summary, despite the simple implementation, the introduced approach provides an improved adaptability for the classification of eye movements during frequently changing viewing behavior and is suited for real-time applications due to the complexity in the order of  $O(1)$ .

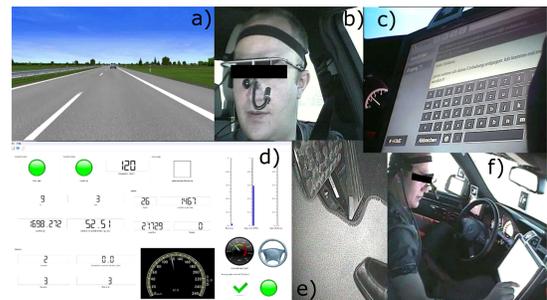
## 4 Task- and Inter-Individual Differences

Analyzing the eye movement behavior is a challenging task since there is no ground truth available for the wanted underlying distribution of fixations or saccades and the labeling effort for empirical studies is time-consuming and error-prone. MERCY, a novel approach of classifying saccades and fixations by means of a GMM, was introduced in the preceding section and shown to be capable of adapting to occurring variations in the parameters of the original mixture model. This adaptability is subsequently used for the evaluation of eye-tracking data from a driving simulator study to detect changes in the eye movement behavior of the drivers.

### 4.1 Experimental Setting

For the examination of the task- and inter-individual differences in conditionally automated driving scenarios, a driving simulator study was performed in the Mercedes-Benz moving-base Driving Simulator. This moving-base simulator realizes a  $360^\circ$  projection of the traffic situation, realistic wind and engine sound effects, as

well as acceleration forces in all directions. The test subjects drove in a conditionally automated setting on a typical German highway for about 35 minutes at 120  $km/h$  in a detailed driver's cabin of a Mercedes-Benz W212 E-Class. At the beginning of every drive, an introductory route section was simulated to introduce the participants to the driving simulator and to the automated driving system by driving manually or with the activated conditionally automated driving function without performing any secondary task for about one minute in each case. Furthermore, four take-over situations were set along the route forcing the driver to take over the control of the vehicle within 2.5 to 4 seconds. A touch screen, shown in Figure 5c) and 5f), was mounted in front of the center console so that the driver could perform the secondary tasks. On the touch screen a graphical user interface was providing the upcoming tasks which the test subject had to select manually. However, the selection process was not considered for the evaluation of the eye movement data. The set of secondary tasks included watching a movie, reading news, writing an email, listening to music, and being idle i.e. the driver was asked to not perform any task. For convenience, the manual driving sections are considered as additional tasks and the abbreviations *manual driving*, *idle*, *mail*, *music*, *read*, and *video* are used in the following to refer to the secondary tasks of *driving manually*, *being idle*, *writing an email*, *listening to music*, *reading news*, and *watching a short movie*. The eye movements of the test subjects were recorded by means of a monocular, mobile Dikablis eye tracker, shown in Figure 5b), measuring at a sampling rate of 25 Hz. Note that according to the Nyquist-Shannon theorem an accurate measuring of the peak velocities and the velocity profiles of the saccades is not feasible with such a low sampling rate. However, since MERCY only considers the point-to-point velocities the low sampling rate is sufficient. 85 test subjects participated in this study, divided into an experimental group of 74 test subjects performing secondary tasks and into a control group of 11 test subjects performing only the idle task. Introducing this second group of subjects allowed examining the driver's visual behavior in the context of automated driving, where the driver was not involved in a secondary task. Inside the cabin, four video cameras were mounted at different angles allowing the permanent observation of the driver's face, the footwell, the steering wheel area, and the touch screen. The video image of the four cameras, a synchronized image of the current simulation view in front of the driver cabin, as well as an interface for monitoring purpose of various vehicle and route information were merged and recorded (see Figure 5).



**Figure 5:** Recorded video of the different camera perspectives, the simulated scenario in front of the vehicle, and of the GUI for monitoring purpose.

### 4.2 Evaluation

For the following evaluation only 74<sup>3</sup> of the initial 85 experiments could be used due to missing signals from the eye tracker for six

<sup>3</sup>41 males/33 females, mean age of 39 years (range 20-60, SD=10)

**Table 2:** Summary of the classification results.

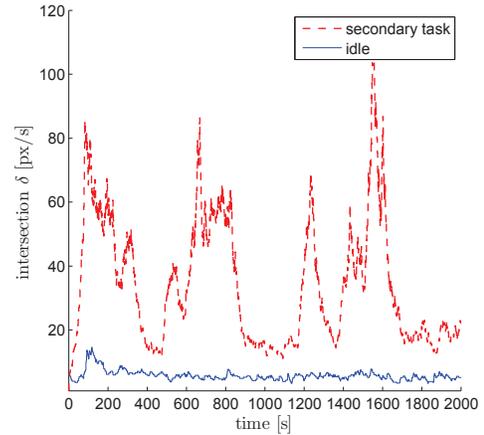
Algorithms	Recall	Precision	F1 score
<b>I-DT</b>	0.73	0.66	0.69
<b>BMM</b>	0.86	0.67	0.75
<b>MERCY</b>	0.91	0.75	0.82

subjects and erroneous simulations such as traffic freezes for five subjects. In total, the eye movement data set included 35.5 hours of recorded eye tracking data separated into 1.5 hours of manual and 34 hours of conditionally automated driving. First, accumulated eye movement behavior of the experimental versus the control group was investigated for significant differences. For this purpose, MERCY was applied to the eye-tracking data of each driver, since the adaptability of this method proved to be convenient for describing mixture models and their variations. The reliability of MERCY depends mainly on the choice of the initial parameters of the GMM. Hence, these initial parameters should at least be in the same range as the average parameters over all drivers and situations. Therefore, random segments of a pre-defined size were extracted from randomly chosen simulator drives and used as input for the Expectation Maximization algorithm. The estimated parameters by means of the Expectation Maximization algorithm were averaged, resulting in the initial values  $\Omega_{f,init} = \{0.55, 1.13, 0.90\}$  and  $\Omega_{s,init} = \{30.09, 3792.20, 0.10\}$ . As window size,  $\omega = 10$  seconds was chosen such that the algorithm could react to current changes in the eye movement behavior within a short period of time and, at the same time, the algorithm would not generate high-frequency oscillations. The threshold  $l$  was set to  $l = 0$  so that the parameters were updated in every iteration.

Before analyzing the eye movement behavior, a detailed evaluation of MERCY in comparison to the BMM and the dispersion-based algorithm *I-DT* was performed regarding their capability to distinguish between fixation and saccade points. In total, eight data sets of six different subjects performing the mentioned secondary tasks consisting of 6623 fixation points and 1384 saccade points were manually labelled by two raters. The duration and dispersion thresholds of the *I-DT* were set to the fixed values of 100 *ms* and 15 *px* in terms of the unit of the eye camera. As shown in Table 2 MERCY achieved the highest results for all three metrics of the applied algorithms. The BMM showed a high recall value, since it is sensitive to even small point-to-point velocities. However, this sensitivity leads to an increased false negative rate and, therefore, to the low precision on the labelled data set. The threshold-based algorithm *I-DT* showed the lowest results for all three metrics, which indicates the disadvantage of the fixed threshold versus the adaptive ones.

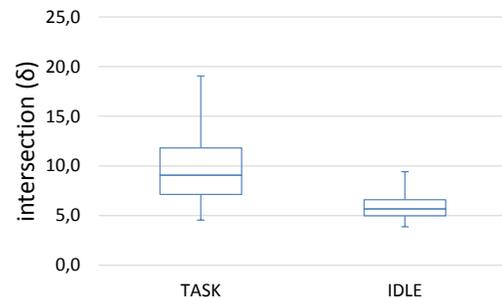
First evidence for an existing difference in the eye movement behavior between conditionally automated driving scenarios with and without performing secondary tasks is given just by looking at two examples of the curve shape of the intersection  $\delta$  in Figure 6. While the blue solid plot, representing the intersection point of one of the idle drivers, seems to be stable and shows only high-frequency noise, the red dashed plot of one of the drivers performing the secondary tasks shows huge drifts over the whole experiment. These drifts could be the result of the task-individual eye movement behavior, which would be a strong evidence for the authors' hypothesis that frequent changes in the performing task generate a significantly varying eye movement behavior. In addition, the huge differences of up to 90 *px/s* in the amplitudes as well as the steep gradients of the shown drifts require an even higher adaptability of eye movement classification algorithms than the artificially generated data. The vertical offset of the two curves can be interpreted as

inter-individual difference due to variations of the individual viewing behavior or due to the setting of the measuring system, e.g. decreased distance of the camera to the eye.



**Figure 6:** Exemplary plots for the behavior of the intersection point of a driver of the experimental and control group.

To analyze of the intersection behavior over all subjects, Figure 7 shows the boxplots of  $\delta$ , averaged over the whole test duration of every subject. For the plot, possible outliers were removed by considering only the inner 95% of the data samples. Applying the one-sample Kolmogorov-Smirnov test to the estimated data of the intersection point, it was indicated that the data is not normally distributed. The difference of the eye movement behavior between the experimental group and the control group can be seen straightaway in Figure 7, since there is no overlapping of the interquartile ranges, including median, first and third quartile, of the two boxplots. This first impression is underpinned by the Wilcoxon rank-sum test and the Hedges'  $g$  measure, implying that the difference is significant ( $p = 0.002$ ,  $z = 2.99$ ) and of practical relevance ( $g = 0.711$ ). Despite the increased value in the location parameters, the left boxplot shows an increased interquartile and whisker range. These findings illustrate the significant difference in the estimated intersection point between both groups and therefore suggest that the variations in eye movement behavior are considerably greater for the drivers performing secondary tasks than for drivers without any tasks.



**Figure 7:** Boxplots of the averaged estimated intersection point  $\delta$  of MERCY while performing secondary tasks versus being idle. The boxplots show the inner 95% of the data, excluding in this way the lowest and the largest 2.5% of the data due to outliers.

To identify the parameters of the estimated GMM, which are varying the most during the conditionally automated driving scenario

and which differ between the idle and busy driver, Table 3 compares the means, medians, minimums, maximums, and variances of the parameter sets  $\Theta_f$  and  $\Theta_s$  of the experimental and control group.

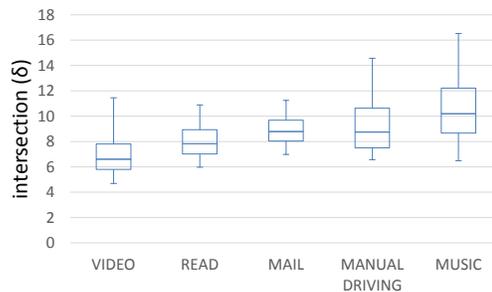
		mean	med	var	min	max
Task	$\mu_f$	0.73	0.69	0.08	0.26	1.96
	$\mu_s$	55.42	52.35	337.87	21.71	89.04
	$\beta_f$	2.26	1.96	2.51	0.55	9.73
	$\beta_s$	3604	3733	241763	1763	3999
	$\pi_f$	0.91	0.92	0.001	0.73	0.98
	$\pi_s$	0.09	0.08	0.001	0.02	0.27
		mean	med	var	min	max
Idle	$\mu_f$	0.53	0.50	0.03	0.23	1.46
	$\mu_s$	50.35	46.41	262.7	21.47	80.60
	$\beta_f$	1.32	1.15	0.65	0.49	6.00
	$\beta_s$	3622	3761	179901	1825	3999
	$\pi_f$	0.91	0.92	0.001	0.71	0.97
	$\pi_s$	0.09	0.08	0.001	0.03	0.29

**Table 3:** Statistical values of the estimated GMM divided into mean, median (med), variance (var), minimum (min), and maximum (max).

It can be seen that the parameters  $\mu_f$ ,  $\pi_f$  and  $\pi_s$  for both groups of subjects have such low variances that these parameters probably do not require learning and adapting to them at all. Especially the a priori probabilities  $\pi_f$  and  $\pi_s$  imply a constant ratio of 1/9 between saccades and fixations over the whole experiment and all statistical measures are nearly identical for both groups. The average velocity of the fixations  $\mu_f$  is not exactly zero as expected, due to measurement inaccuracies or smaller eye movements as the nystagmus<sup>4</sup>. Nevertheless, as long as such "disturbances" are kept as small as possible, there will be no significant variation in this parameter. The size of the relative variances as well as the ranges from the minimum to the maximum value of the remaining parameters  $\mu_s$ ,  $\beta_f$ , and  $\beta_s$  indicate that these values vary the most overall measured data. Note that due to the flat and wide distribution of the saccades, the influence of  $\mu_s$  on the intersection point and hence on the classification is low. In summary, for the given assumption of a GMM describing the process of generating saccades and fixations, it would be sufficient to only learn the parameter  $\beta_s$  and  $\beta_f$ , describing the variance of the distribution of the saccades resp. the fixations, since the remaining parameters of the mixture model can be considered as constant or their influence on the classification performance is vanishingly low. If the values of Table 3 are compared between the control and experimental group, an increased variance can be observed while performing secondary tasks as it occurred for the intersection point. This finding confirms the hypothesis suggesting high variations in the eye movement behavior due to task-individual differences in the same way as the evaluation of the estimated parameter  $\delta$  above.

To explain which secondary tasks cause the variation in the eye movement behavior during conditionally automated driving, Figure 8 shows a boxplot of the estimated intersection point of all performed secondary tasks and of the manual driving sections. The interquartile range of the boxplots of the tasks *video*, *mail*, and *read* are similar to the range of the idle task, but with an increased average of the estimated intersection point. These small variances probably result from the fact that all three tasks were performed on the touch screen built in the cabin. Thus, most eye movements were performed in a narrow field of view. Obviously, they cannot be the sole explanation of the increased variations in the eye movement

<sup>4</sup>Rhythmic, oscillating, and involuntary movements of the eyeball [Benjamin 1997].



**Figure 8:** Quantitative comparison of the behavior of the estimated intersection point during the different secondary tasks and while driving manually over all subjects.

behavior during the performing of the secondary tasks. In contrast, the *music* task reveals a larger variation of the eye tracking data than the idle task in Figure 7, although a similar viewing behavior of both tasks is expected. A possible explanation for this larger variation could be the gazes of the driver on the touch screen, since the display was not turned off during the *music* task and, therefore, still could attract the attention of the driver. Another explanation would be the more active scanning behavior of the driver of the environment between the usual tasks performed on the touch screen, which force the driver to focus the attention on the display and not to observe extensively the environment. An additional interesting point to mention is the high variation of the intersection point of the manual driving scenarios. This result indicates that for non-automated driving scenarios on a typical german highway, significant variations in the viewing behavior occur which need to be taken into account for a robust eye movement classification.

In summary, the tasks can be separated into two groups regarding their variation of the intersection  $\delta$ : one group consisting of the *music* and *manual driving* task, showing large variations, and in a second group, composed of the remaining tasks *read*, *video*, and *mail*, depicting small variations. Since these two groups alternate frequently in conditionally automated driving scenarios, the eye movement behavior switches between tasks of small and larger variations, leading to the higher variation of the viewing behavior while performing secondary tasks compared to while being idle.

## 5 Future Work

On the one hand, additional developments of MERCY are advisable, since the primary reason for introducing MERCY was not to provide a classification method, capable of a superior classification on all conditions, but a method to detect robustly changes in the eye movement behavior. Since this method uses sample mean and variance estimators, a reliable estimation of the variance first requires a good estimation of the sample mean. That means that in case of sudden changes in the eye movement behavior, the variance is estimated insufficiently as long as the mean has not approximated the actual mean causing an overshooting of the intersection parameter. The error of the estimation of the sample mean affects the estimation of the variance in a quadratic manner. A possible solution could be a correction function depending on the gradient of the sample mean. Another issue is given by the fact that MERCY is updating only the parameter set  $\Theta_f$  or  $\Theta_s$  of the estimated GMM belonging to the current classification result. In case of a large overlap area of the two Gaussian distributions, e.g. in case of poor initialization values, the incorrect parameters are often updated. Since the total error of the falsely classified data samples can be estimated, this error should be considered for the estimation of the parameters of the

model in terms of error minimization. In this way both parameter sets  $\Theta_f$  and  $\Theta_s$  can be updated in every iteration.

On the other hand, further evaluations of the eye movement behavior are necessary, especially in the case of different automation levels. As the results of this work indicate, even in case of non-assisted driving scenarios, variations of the eye movement behavior need to be considered, rendering static threshold-based eye movement classification algorithms impractical. Since the manual driving route section of the conducted experiment only had a duration of 1 minute, more data, preferably of real-driving scenarios, need to be considered to underpin these findings. Furthermore, the evaluation of the eye movement behavior in different driving environments such as urban, cross-country, or a highway environment, could provide additional information about the necessary adaptability of eye movement classification methods.

## 6 Conclusion

In this work, the necessity of adaptive eye movement classification methods during conditionally automated driving scenarios was thoroughly examined. The results, based on the evaluation of a large-scale driving simulator study conducted in a high-end moving-base simulator, distinctly indicate the increased challenge for the eye movement classification while performing various secondary tasks. The task-individual difference was shown to be significant between the viewing behavior of subjects performing secondary tasks and idle subjects, both driving in a conditionally automated setting. The findings suggest that the eye movement behavior during changing tasks is constantly varying and, therefore, the threshold for the classification between saccades and fixations is varying, too. These frequent changes in the eye movement behavior justify the need of classification algorithms with an increased adaptability. For this purpose, an extended approach of a probabilistic state-of-the-art algorithm, called MERCY, was introduced and its performance of adaptability was evaluated and compared to a method based on a Gaussian mixture model. MERCY not only showed a lower error rate with regard to the adaption to half a million randomly generated data samples, but also comes with the benefit of a smaller computational overhead and a constant time complexity, convenient for the application in real-time scenarios and suitable for implementation on common RCP and HIL tools in the vehicle.

## References

BANERJEE, A., DATTA, S., KONAR, A., TIBAREWALA, D., AND RAMADOSS, J. 2014. Cognitive activity recognition based on electrooculogram analysis. In *Advanced Computing, Networking and Informatics-Volume 1*. Springer, 637–644.

BAREA, R., BOQUETE, L., BERGASA, L. M., LÓPEZ, E., AND MAZO, M. 2003. Electro-oculographic guidance of a wheelchair using eye movements codification. *The International Journal of Robotics Research* 22, 7-8, 641–652.

BENJAMIN, M. 1997. Miller-keane encyclopedia and dictionary of medicine, nursing and allied health. *Philadelphia: Saunders*.

BRAUNAGEL, C., KASNECI, E., STOLZMANN, W., AND ROSENSTIEL, W. 2015. Driver-activity recognition in the context of conditionally autonomous driving. In *18th International IEEE Conference on Intelligent Transportation Systems (ITSC 2015)*.

BULLING, A., WARD, J. A., GELLERSEN, H., AND TROSTER, G. 2011. Eye movement analysis for activity recognition using electrooculography. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33, 4, 741–753.

CASTELHANO, M. S., AND HENDERSON, J. M. 2008. Stable individual differences across images in human saccadic eye movements. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale* 62, 1, 1.

ERKELENS, C. J., AND VOGELS, I. M. 1995. The initial direction and landing position of saccades. *Studies in Visual Information Processing* 6, 133–144.

GANDHI, T., TRIKHA, M., SANTHOSH, J., AND ANAND, S. 2010. Development of an expert multitask gadget controlled by voluntary eye movements. *Expert Systems with Applications* 37, 6, 4204–4211.

HOLMQVIST, K., NYSTRÖM, M., ANDERSSON, R., DEWHURST, R., JARODZKA, H., AND VAN DE WEIJER, J. 2011. *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.

INTERNATIONAL, S., 2014. J3016: Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems.

ISOTALO, E., HEIKKI, A., AND ILMARI, P. 2009. Oculomotor findings mimicking a cerebellar disorder and postural control in severe meniere's disease. *Auris Nasus Larynx* 36, 1, 36–41.

KASNECI, E., KASNECI, G., KÜBLER, T. C., AND ROSENSTIEL, W. 2014. The applicability of probabilistic methods to the online recognition of fixations and saccades in dynamic scenes. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, ACM, 323–326.

KASNECI, E., KASNECI, G., KÜBLER, T. C., AND ROSENSTIEL, W. 2015. Online recognition of fixations, saccades, and smooth pursuits for automated analysis of traffic hazard perception. In *Artificial Neural Networks*. Springer, 411–434.

KOMOGORTSEV, O. V., JAYARATHNA, S., KOH, D. H., AND GOWDA, S. M. 2010. Qualitative and quantitative scoring and evaluation of the eye movement classification algorithms. In *Proceedings of the 2010 Symposium on eye-tracking research & applications*, ACM, 65–68.

LEIGH, R. J., AND ZEE, D. S. 2015. *The neurology of eye movements*. Oxford University Press.

RAYNER, K. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological bulletin* 124, 3, 372.

RÖTTING, M. 2001. *Parametersystematik der Augen-und Blickbewegungen für arbeitswissenschaftliche Untersuchungen*. Shaker.

SALVUCCI, D. D., AND ANDERSON, J. R. 1998. Tracing eye movement protocols with cognitive process models.

SALVUCCI, D. D., AND GOLDBERG, J. H. 2000. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications*, ACM, 71–78.

SEN, T., AND MEGAW, T. 1984. The effects of task variables and prolonged performance on saccadic eye movement parameters. *Advances in Psychology* 22, 103–111.

TAF AJ, E., KASNECI, G., ROSENSTIEL, W., AND BOGDAN, M. 2012. Bayesian online clustering of eye movement data. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, ACM, 285–288.

WINN, J. M., AND BISHOP, C. M. 2005. Variational message passing. In *Journal of Machine Learning Research*, 661–694.