A comparison of glance coding approaches for driver attention assessment

Katja Kircher*a, Christer Ahlströmab

The Swedish National Road and Transport Research Institute (VTI), Linköping, Sweden
b Department of Biomedical Engineering, Linköping University, Linköping, Sweden

ABSTRACT

Eye tracking is a common tool to assess drivers’ attentional state, either in real-time with the goal to prevent incidents, or offline, to understand underlying processes. While seemingly objective, eye tracking data can be coded and interpreted in different ways, which can have substantial effects on the results. The objective of this paper is to highlight and discuss the possibilities and limitations of three different approaches to code glance data: the direction-based encoding, the target-based approach, and the purpose-based approach. The direction-based coding scheme describes glances relative to the direction of travel. The target-based approach classifies the glance targets into different categories. The purpose-based approach needs additional layers of information to deduce the reason for the glance. Data from a field study with 23 participants driving an instrumented vehicle on an urban route was used to illustrate differences between the three methods. The results showed that the coding approach clearly affected the interpretation of the measured glance data. A purely target-based approach is limited by its inability to account for spare visual capacity and that also the absence of a target constitutes valuable information, and a purely direction-based approach does not account for the need to scan areas located off forward. The purpose-based approach requires contextual information that can be cumbersome to integrate. Regardless of the approach used, additional layers of contextual information increase understanding and interpretability, potentially at the cost of increased complexity. The three approaches are suitable for different contexts and their feasibility also depends on the availability of additional data. A key message is that context awareness improves the accuracy of driver attention monitoring and inattention identification.

1. Introduction

Occupant state monitoring, including driver distraction detection, has become more relevant in recent years. From a regulatory perspective there is a desire to prevent negative effects of non-driving related activities (NDRAs). Advanced driver distraction warning systems will therefore be mandated in all new vehicle types via the General Safety Regulations (GSR) of the European Union from 2024. Similarly, Euro NCAP’s safe driving assessment protocol now includes requirements for how long and how often a driver can look away from the forward road before being warned, with specific detection requirements for mobile phone use (Euro NCAP, 2022). From an automation perspective, assisted driving tempts drivers to engage in NDRAs for extended periods of time (Noble et al., 2021). Driver monitoring solutions are thus needed to ensure an appropriate level of situation awareness that matches the present automation level. Eye tracking is a central ingredient in all these examples, where it is often assumed that visual attention can be estimated from eye movement data alone.

There are, however, several limitations to this assumption. First, it must be acknowledged that eye tracking data neither provide a direct overt measure of visual attention (Deubel & Schneider, 1996), nor do they reveal what information the brain cognitively processes during a glance (Viviani, 1990). Second, despite the capabilities of peripheral vision in visual information acquisition (Gong...
It is seldom taken into consideration. Third, not all foveated information is processed (Herslund & Jørgensen, 2003; Pammer et al., 2018; White & Caird, 2010). And fourth, eye tracking data do not provide an easy way to determine whether the sampled information was relevant, necessary, and sufficient for the driver in the current situation (Wolfe et al., 2020).

In addition to real-time driver monitoring, offline processing of visual behaviour can contribute to basic research aiming at understanding driver attention and behaviour better. For this purpose, additional information, including such that is only available in hindsight, can be applied. Here, we consider both aspects.

Different strategies for gaze data interpretation have been suggested in the literature (Ahlström, Kircher, et al., 2021). These strategies can be summarized as the direction-based approach (where are you looking), the target-based approach (what objects are you looking at), and the purpose-based approach (why are you looking where you are looking). The three approaches have different strengths and weaknesses when applied to real-time monitoring or basic research. To illustrate the differences between the three approaches, an example is provided where a driver intends to go straight in an intersection (Fig. 1). The driver glances to the right in the direction of a bicyclist leaving the intersection on the main road.

1.1. Glance coding from a driver monitoring perspective

In the direction-based approach, the gaze is typically coded as directed forward, left, right, behind, etc. Remote eye trackers without a coupled scene camera usually provide a gaze vector indicating a direction without context. In the example in Fig. 1, this vector would be coded as a glance to the right. In a driver monitoring setting, the gaze direction is often used to determine if a driver is looking away from forward too often or for too long (e.g., Victor et al., 2005). In environments where much of the relevant information is located off-forward, such as in urban areas, this may lead to many false distraction detections. Different solutions have been presented to adapt attention monitoring algorithms to the surroundings. The demand on forward glances could be relaxed (Han et al., 2023; Kujala et al., 2016), the size of the forward region could be adapted based on the surroundings (Ahlström et al., 2011), and the forward region could be complemented with additional “allowed” target areas (Ahlström, Georgoulas, et al., 2021). In practice, this can be implemented by fusing eye tracking data with data from digital maps and proximity sensors (Bickerdt et al., 2021). However, such solutions are yet uncommon in driver monitoring.

Several implications follow when moving from one relevant region (i.e., forward) to multiple target areas. In the example in Fig. 1, two target areas have been added in addition to forward, representing traffic streams that must be yielded to. Note that the target area is the location where traffic with right-of-way may appear, and not the traffic itself. Also note that the additional target areas are only present in the vicinity of the intersection. Each target area is therefore associated with a zone within which the driver must sample information from the target area. Any traffic present in a target area while the ego driver, that is, the driver under scrutiny, is in the associated zone is considered relevant. Any traffic outside the pre-defined target areas is considered irrelevant. Both the target areas and the zones are pre-defined based on knowledge about the infrastructure, traffic rules that apply, and the driver’s intended manoeuvre. From a distraction monitoring perspective, bringing in multiple target areas shifts the focus from detecting when drivers have looked away too much towards detecting when drivers have not considered all target areas.

In the target-based approach, glance targets are coded and categorised depending on the target type. This requires a scene camera with accompanying gaze direction data. Examples of coding categories include “car”, “bicyclist”, “traffic sign”, or “mobile phone”. Manual coding of scene videos with a gaze overlay is the most common approach, but automated solutions based on computer vision are emerging (Panetta et al., 2020). The glance to the right in Fig. 1 would be coded as “bicyclist”. Whether a target is relevant for the driving task or not is typically determined directly via the target type. A glance towards a car would be categorised as relevant while a glance to a billboard or to a mobile phone would be categorised as irrelevant. In reality, situational dependencies rather than its type.
dictate the relevance of an object. For example, the bicyclist in Fig. 1 is not relevant given the driver’s intention to go straight. However, the same cyclist would have been relevant if travelling in the opposite direction. For driver monitoring applications, the target-based approach is therefore mostly used to identify glances to specific objects such as mobile phones, and then not via eye tracking but rather via driver activity recognition (Xing et al., 2019).

1.2. Glance coding to understand attention and behaviour

The role of vision is to provide relevant information for making decisions to achieve behavioural goals (Hayhoe, 2017). Since vision and action are so tightly coupled, the conclusions that can be drawn from eye tracking data without contextual data is limited. Just the knowledge that a glance is made in a driving setting provides context since it is then known that the forward direction is aligned with the direction of travel. Adding more context enables further insights about the observed behaviour, where each layer of additional context gradually allows for more detailed interpretations of the measured glance behaviour. One such layer could contribute map data including road layout and priority rules. Another layer could add data on sight distance and speed limits. Further layers may represent the planned route and consequently the upcoming manoeuvres, or data on surrounding road users and other objects, and the relevance of these people and objects for the ego driver. More qualitative layers such as thinking aloud protocols, where the drivers verbally describe what they are doing and what they are expecting from each action, are also valuable. Each of these contextual layers strengthens the eye tracking data, regardless of whether they are coded as glances in certain directions, as foveated objects, or as glances towards various target areas. Some of these layers are possible to add in real time, whereas others, especially information about the actual development of the situation, can only be added in hindsight and offline.

Direction data can be used to create heat maps, describe basic glance patterns, or investigate glance distributions to various directions, potentially linked to external data layers. As described above, the addition of road layout, traffic rules and intended direction allows the definition of target areas, and from that follows the possibility of estimating the purpose of the glance. Hazard perception studies are a special case where the target areas containing a latent hazard should be observed while the driver is in the so-called launch zone (Krishnan et al., 2019). In this setting, the purpose of potential glances in that direction is given by the latent hazard, whether it materialises or not. The objective of purpose-based coding is to infer the purpose of a glance in a general case, where the target areas are pre-defined based on attentional requirements, but without staging them in the experimental design. If the purpose of a glance can be deducted, it may allow conclusions about the driver’s understanding of the situation. This, in turn, could help when developing support systems or improving infrastructure layout.

Target-based coding is typically only done in certain situations of interest, such as when investigating glances in an intersection (Vetturi et al., 2020), or when analysing glances during the execution of a certain task (Ahlström & Kircher, 2017). The context, and to some extent also the purpose, of observed glances is then provided, and constrained, by the situation of interest. The purpose of a sequence of glances to an in-vehicle information system is simply the desire to interact with the system, and the purpose of in-between glances to other objects is to maintain situational awareness.

A limitation with target-based coding schemes is that only foveated objects are included in the data representation. This makes it difficult to know whether any relevant objects were not looked at by the driver. For such analyses, it would be necessary to include a listing of all available objects in the vicinity of the vehicle in the data representation and to assess their relevance. Accounting for exposure is a related issue. For example, the share of glances to cars is not very informative without knowing the number of cars that could have been looked at. With background information about the prevalence of various targets in combination with additional context, it can be assessed whether certain target types take precedence over others, and whether this may be linked to driver characteristics or other variables.

Returning to the rightward glance towards the bicyclist in our example in Fig. 1, given the contextual data we can glean from the picture, it is likely that the purpose of the glance is to check for traffic with right-of-way. With no such traffic present, the salient bicyclist happens to be foveated, even though the bicyclist is not relevant for the driver’s upcoming manoeuvre. A purpose-based interpretation of the rightward glance would therefore be that the driver checks for traffic in the target area to the right rather than looking specifically at the bicyclist.

1.3. Objective

The objective of this paper is to provide a side-by-side comparison of the direction-based, target-based, and purpose-based coding schemes of glance data. The coding schemes were implemented with the purpose to understand driver attention and behaviour based on eye tracking, but the results are also interpreted from a driver monitoring perspective where feasibility and real-time aspects are considered. The three methods are applied to an on-road dataset including multiple participants driving in an urban environment in an instrumented vehicle. We demonstrate how the different methods require different amounts of additional contextual data before they become interpretable, and how this may lead to different conclusions in terms of attention and distraction while driving. Finally, we investigate and discuss the challenges and research gaps in relation to the three approaches, especially in terms of benefits gained by coupling eye tracking data to contextual information like infrastructure, rules, and other traffic.

2. Method

The original purpose of the data collection was to investigate differences in attentional requirements for car drivers and cyclists in the same traffic environment (Ihlström et al., 2021; Kircher & Ahlström, 2020; Nygårdh et al., 2020). In the present publication the
focus is on car driving. Details related to the cyclist part of the experiment have been omitted. This research complied with the tenets of the Declaration of Helsinki and was approved by the Regional Ethical Committee in Linköping (Dnr 2017/107–31). Informed consent was obtained from each participant.

Based on their answers to an online questionnaire, 23 participants (39 ± 14 years old, 12 females) were recruited from 382 respondents. Inclusion criteria were the possession of a driving licence and being familiar with driving a car in town, not needing more than ± 4 dioptres vision correction (eye tracker requirements) and being able and willing to participate in the study. Technical problems with the eye tracker led to the exclusion of all data from one participant.

After giving informed consent, the participant was equipped with a head-mounted eye tracker (SMI glasses 2.0, SensoMotoric Instruments, Teltow, Germany), which was calibrated using a three-point procedure. The experimental vehicle was a Volvo V60 equipped with a data logger (Video VBOX Pro, RaceLogic, Buckingham, UK) recording speed, position, and video of the view forward, rearward and to the right side of the driver. The route went through the town centre of Linköping, Sweden, and near-centre living areas. Traffic density varied with time of day. The participant was shown a map of the route and received additional navigation instructions from the experimenter in the rear passenger seat. The instruction was to drive as one normally would.

Due to the resource-intensive nature of the data reduction, a sub-set of the available data was selected as follows (totalling 56 min 22 s). Four intersections and the link roads leading up to and away from them were chosen. Applying the MiRA-theory (Kircher & Ahlström, 2017) during on-site visits, we determined the required target areas that needed to be sampled visually (based on traffic rules and infrastructure layout), and the associated physical zones within which the sampling would need to take place (based on sight distance). In other words, the “zone” is the road segment within which the driver should collect visual information from the “target area” (see Fig. 2). In a yield-controlled intersection, for example, the required “target areas” would be to the left and to the right of the intersection on the main road, where cross traffic can be expected, and the “zone” would be the stretch of road beginning where potential traffic on the main road can be seen and ending before the intersection is entered. When all required target areas and their corresponding zones had been identified, all zones were selected where one of four target area combinations needed to be sampled concurrently (Table 1). Each passing of such a zone combination was considered an “event”. The whole dataset consisted of 552 events.

Each event was coded in a direction-based, target-based, and purpose-based fashion, as determined by the three data reduction schemes presented in Table 2. Coding was done using the Noldus Observer 14 software (Wageningen, the Netherlands), which implies manual video annotations. After coming to consensus about the three coding schemes using a sub-set of the data in a group session with four coders, the schemes were fixed, whereupon one person proceeded with coding the remaining data. A glance is defined as a sequence of gaze data points longer than 100 ms directed towards an area or object.

As outlined in the introduction, relevance and context information are integral parts of the purpose-based glance data representation. In a naturalistic setup, relevant targets and zones come and go as one drives along a road, but here, only the geopositioned regions of interest from Table 1 were encoded. As all three methods were applied to the same regions, this means that all three encodings can be interpreted within this context. Results from all methods are therefore interpreted as a function of the two situational variables that originally come from the purpose-based data encoding: zone combination (forward, forward + zebra, forward + right, and forward + left + right), and encounter level (no/irrelevant traffic, traffic not on collision course, traffic on potential collision course, and traffic on collision course). Analyses of variance were computed for categories pertaining to various classifications of inattention, using “event” as research unit.

3. Results

Fig. 3 shows the time share of glances resulting from the three data encoding methods. All zone combinations that correspond to one of the four selected groups from Table 1 are subsumed and presented together.

The distribution of glance directions for the different zone combinations clearly reflects the direction of where the required target areas are located. If multiple concurrent target areas requiring glances in several directions are present, these directions are clearly visible in the glance direction distribution. The percentage of dwell time to forward decreases the more attentional requirements there are outside of the forward area. An analysis of variance showed that the share of forward glances within each event differs significantly.
Table 1
Description and explanation of the zone combinations for which data was reduced (see also Fig. 2).

<table>
<thead>
<tr>
<th>Zones with target areas to …</th>
<th>#</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>forward</td>
<td>312</td>
<td>Required sampling in the direction of travel only. Typical on link roads.</td>
</tr>
<tr>
<td>forward, zebra</td>
<td>174</td>
<td>Additionally, the presence or absence of pedestrians wanting to cross must be ascertained before the zebra crossing is entered.</td>
</tr>
<tr>
<td>forward, right</td>
<td>22</td>
<td>In addition to forward, the road coming from the right needs to be checked for traffic before the crossing is entered. E.g. when crossing a one-way street or in intersections with priority from the right.</td>
</tr>
<tr>
<td>forward, right, left</td>
<td>44</td>
<td>In addition to forward, potential traffic coming from the right and left needs to be checked for, e.g. in an intersection with a stop- or yield-sign.</td>
</tr>
</tbody>
</table>

Table 2
Description of manually encoded data.

<table>
<thead>
<tr>
<th>Data selection</th>
<th>Data coding</th>
</tr>
</thead>
<tbody>
<tr>
<td>glance coding</td>
<td>gaze direction relative to own trajectory, approximately in 90-degree-segments: forward; left; right; behind (including mirrors); down (dashboard etc.)</td>
</tr>
<tr>
<td></td>
<td>glance target: motorised traffic; bicycles; pedestrians; traffic signs and traffic lights; zebra crossings; no target (usually forward roadway with no specific targets present); other (off forward: billboards, trees, in-car targets, nothing in particular)</td>
</tr>
<tr>
<td></td>
<td>purpose: traffic check (first glance off forward to target area or other area where traffic can appear, regardless of traffic presence); monitor traffic on collision course; monitor traffic on potential collision course; monitor traffic not on collision course; monitor irrelevant traffic; information collection (e.g. from traffic signs, the speedometer, traffic light); other (billboards, trees, in-car targets); default (all glances to forward that do not seem to have a specific purpose)</td>
</tr>
<tr>
<td>target areas and associated zones (as described above)</td>
<td>location and size of each zone and target area</td>
</tr>
<tr>
<td>all road users who are reasonably close</td>
<td>presence/absence of other road users: traffic on collision course (the participant must act to avoid a collision); potential collision course (the participant may have to act to avoid a collision depending on how the situation unfolds); not on collision course (no collision will occur if both continue on their trajectory); irrelevant traffic (road users present but outside the predefined target areas, no need to take any action). Note that a collision course event does not have to be critical, it just means that someone must act, for example by slowing down.</td>
</tr>
<tr>
<td>road furniture that is reasonably close</td>
<td>presence/absence of road furniture: traffic signs; traffic lights; zebra crossings</td>
</tr>
</tbody>
</table>

Fig. 3. Share in dwell time for different categories depending on the data reduction approach (rows), for different coinciding target areas within one zone (columns).
between zone combination types (F(3, 548) = 74.8; p < .001). With the zone combination information readily available, it is easy to explain the differences in the four pie charts. For example, the large share of glances to the right in the forward + right and forward + left + right conditions come from the need to monitor the road to the right. This illustrates the limitations of eye tracking data and the necessity of context information when interpreting glance behaviour.

From a driver monitoring perspective, quantification of off-forward glances has its limitations, especially outside motorway environments. In environments with several target areas in multiple directions, such as the urban setting used here, a distraction detection algorithm should rather monitor whether the driver has looked at all relevant target areas. For the present dataset we found that all forward, left, and right target areas necessitating a glance away from the forward direction were foveated at least once by each participant within the associated zone. It was harder to ascertain whether all zebra crossings were sampled. In 43 cases the participants looked directly at the zebra crossing, while in 72 unclear cases it is likely that peripheral vision was used. No incidents where pedestrians were not given priority occurred.

Context information is needed when interpreting glance targets, too. The distribution of foveated targets differs substantially between the different zone combinations. For forward + right it is more common to look at motor vehicles, whereas cyclists receive more glances in forward + right + left. Given the coding approach, in which only foveated targets are coded, no exposure information is available. That means, it is not known how many road users of each type are present in the two zone types, and whether or not they are on a critical path relative to the participant’s intended direction of travel. Further, glances in the “no target” category, mostly consisting of glances to the forward roadway but without glancing at anything in particular, are more frequent in events with fewer target areas (F(3, 547) = 26.5, p < 0.001). At the same time, the category “other”, which encompasses glances that were not directed forward and did not target other road users or road furniture, has the largest glance share for forward + left + right and the smallest for forward only (F(3, 547) = 30.8, p < 0.001). Reduced scanning in the forward direction (included in the “no target” category), and increased scanning towards the sides (included in the “other” category), is a finding that speaks to our intuition. Again, the extra contextual information is needed to make sense of the coded glance data. For the target-based approach, if no targets are present in a relevant target area, this will otherwise typically be coded as a glance to a non-specified target (here “other”).

The purpose-based coding scheme assumes that the first glance towards a target area within the corresponding zone is to check for presence of traffic in that area. The glance is coded as such independent of whether there is any traffic or not. Fig. 3 reveals that a higher number of different target areas leads to more traffic checks (F(3, 548) = 60.5; p < .001). It is also connected to a smaller share of default-glances forward (F(3, 548) = 20.5; p < .001). The share of glances devoted to monitoring traffic also increases with the number of target areas. The share of glances to targets not belonging to any of the previous categories in the purpose-based coding scheme (“other”) lies at five per cent or below for all zone combinations, indicating that aimless glances away from forward are rare.

In addition to analysing events by zone combination, it is also possible to analyse them by “encounter type”, that is, whether other traffic is present and how it is interacting with the participant (see Table 2). The most critical encounter type in each event determines the group membership, for example, if both irrelevant traffic and relevant traffic on potential collision course is present, the event will be categorised as the latter (Fig. 4).

![Fig. 4. Glance target and purpose depending on traffic encounter type.](attachment:glance_target_purpose_traffic_encounter.png)
Regardless of encounter type, the main glance direction is forward, as expected. The share of glances to the right is somewhat larger when traffic on collision course is present. The potential collision and the collision course events typically occur in intersections with pedestrians or bicyclists present, which explains the larger share of glances to the right. The share of downwards glances is most common when no or only irrelevant traffic is present, indicating that spare visual capacity is available (this can also be seen by the larger share of “default” glances in the purpose-based encoding). As before with the zone combinations, without the encounter type information in combination with information about where such encounters typically occur, it would not have been possible to explain the differences between the four charts.

“No target”, mostly representing glances to the forward roadway, is the main glance target when no or only irrelevant traffic is present. In these situations, irrelevant traffic draws a share of around 15% of the dwell time, but more glances go to other targets off the forward roadway. The largest share of glances to traffic is found when there is a potential collision course, that is, when it has yet to be determined whether action is required on the side of the participant or not. The target-based approach does not include information about the prevalence of different road user types in the different events, so it is unknown whether the glance share distribution simply reflects the prevalence or whether it is related to encounter type. As hypothesized above, the target category “zebra crossing” is more common in the collision course events, indicating that such events mostly occur in intersections with pedestrians and cyclists present.

The purpose-based data reduction shows that when no relevant traffic is present, around half of the glance share consists of “default” glances, and the share of “default” glances decreases with increasing criticality of encounter type. With no relevant traffic present, glances checking for traffic lead to a confirmation of the absence of traffic in the target area. “Irrelevant traffic” is glanced at for around a quarter of the time. The purpose-based approach includes information about the presence and encounter level of all surrounding road users. Each road user classified as “relevant” (i.e., inside of a target area) was foveated at least once. In contrast, overall, only about half of the irrelevant traffic (road users outside of target areas) was glanced at. When relevant traffic was present concurrently, only around a quarter of the irrelevant traffic was glanced at. The combined share of glances devoted to checking for traffic and monitoring traffic increased with encounter criticality. The reason why the drivers are not looking at the collision course targets more is that these events typically occur in more complex situations where multiple areas/targets are present, and where all these areas/targets must be monitored/checked to avoid risky encounters. This indicates that the drivers do not solely focus on the most urgent object, but also keeps track of additional relevant targets or areas. Additionally, traffic on collision course is likely to be discerned easily, whereas judging whether a road user is on collision course or not may require more glances.

4. Discussion

The aim of this study was to provide a side-by-side comparison of three different eye tracking data encoding procedures and to explore their potential for both real-time and offline applications depending on which additional data layers are available. This methodological approach was supported and partially illustrated via a dataset from real traffic.

4.1. Understanding glance tracking

In the context of driver attention and distraction, eye tracking is a widely used method, as visual behaviour is seen as a cornerstone for traffic-related information sampling. Here, we investigate how the three coding approaches lend themselves to inattention assessment in that context and which issues may arise.

In a context-blind setting, the three coding approaches led to rather different assessments of the drivers’ likely level of inattention. All target areas were checked for traffic, albeit with some uncertainty regarding the zebra crossings, implying that the purpose-based approach did not detect insufficient attention for the present study. Monitoring attention based on glances in the forward direction, or by the amount of “other” target glances, gives the impression that drivers are less attentive in more complex scenarios. This counterintuitive result emphasises the need to interpret glance data in relation to the context where the glances were made, at least in more complex environments.

Given the controlled setting of the experiment, where the participants did not use their phones or engaged in other activities not related to driving, it is probable that the “other” category in the target-based coding mainly consisted of targets that did not fit the strictly traffic-related categories from Table 2. If a narrower criterion for which targets qualified as “distractors” is chosen, the percentage of inattention occurrences would decrease. This very reasoning shows that inattention levels vary with the definition of the target categories as relevant or not relevant for driving, challenging the ostensible objectivity of glance target coding. The situation is different when determining relevance based on zones and target areas. Relevant target areas are identified a priori, based on external features like infrastructure, traffic regulations and sight distances. The zones and the target areas provide the context that classifies road users within the target areas as relevant and objects outside of the target areas as not relevant. Thus, the same road user can be relevant or irrelevant to look at, depending on the situation.

The caveat is that traffic regulations often include a component that requires road users to be wary of others’ potential mistakes. For simplicity and clarity, the representation used here does not consider this. To be attentive within this framework, it is enough to glance at least once towards the identified relevant target areas. Possible incidents that occur, because another road user does not act according to the rules cannot be ascribed to inattention on the part of the ego driver.

The difficulties in ascertaining whether the zebra-crossings were monitored by the drivers illustrates a problem arising from requiring foveation. For target areas in the general forward direction, peripheral vision may be enough. On the other hand, it is possible that a single glance towards a target area may not necessarily be enough to assess the situation sufficiently. Larger datasets with a controlled variation of specific features in the infrastructure, regulations and the presence or absence of relevant and irrelevant
traffic would improve the identification of situation specific minimum glance requirements. However, the number of interacting factors that can play a role is large. Also, in many situations different visual strategies can be used to fulfil the requirements, which complicates the process of identifying the least common denominator.

We would argue that drivers taking part in a controlled study of rather short duration with an experimenter sitting next to them, on an urban course, are unlikely to be distracted for a substantial amount of time. They might still not sample all relevant areas, as reported by Kaya et al. (2021). In their field study, drivers frequently neglected looking over their shoulder in right-signalised turns, which would indicate inattention according to the purpose-based approach as it is used here. In the present study no turning manoeuvres were included.

The purely direction- and target-based approaches neither actively incorporate spare visual capacity nor the need to scan areas located off forward. Also, the target-based approach cannot reflect that also the absence of a target constitutes valuable information. The concept of spare visual capacity in driving is not new and there is strong evidence that spare capacity is available under various circumstances, varying between situations and people (Liu et al., 2020; Senders et al., 1967; Underwood, 2007). In the purpose-based coding scheme, spare capacity is explicitly reflected in the “default glance” category. The decreasing share of default glances with increasing task complexity relates well to previous research showing that experienced drivers need less time to attend to all relevant objects (Bos et al., 2015; Underwood et al., 2002). Looking at irrelevant traffic, other objects outside of the car, and even engaging in NDRAs, would not in itself lead to a classification as inattentive, as long as the required target areas are sampled sufficiently. For the direction-based approach, the situation with a larger amount of target areas may provide an approximation for the share of glances needed to forward, which can give an indication of the percentage of spare capacity available in the situations with fewer requirements. In the target-based approach, the other-category comes closest to reflect spare capacity. The purpose-based glance classification used here is based on the MiRA-theory (Kircher & Ahlström, 2017), but other categorisation methods are thinkable. Wolfe et al. (2020) describe and discuss the processes of visual information acquisition in traffic, which could be a good starting point for a conceptual framework. The aspect of external demands, that is, the situation-specific requirements on attention must be incorporated.

4.2. Glance metrics

Glance frequency and duration are common metrics to describe gaze behaviour (Crundall & Underwood, 2011; Holmqvist et al., 2011), and they also serve as input to driver monitoring algorithms (Ahlström, Georgoulas, et al., 2021; Ahlström et al., 2011; Han et al., 2023; Kujala et al., 2016; Victor et al., 2005). The underlying assumptions are that a higher number of glances to a certain object or area enables more information extraction from that area, and that information decays over time, such that longer glances away from a certain object or area make it more difficult to anticipate how information has changed in that area. The share of glances is another commonly used metric which combines frequency and duration. The share of glances is useful in a limited time window spanning a certain situation of interest, such as the zone combinations used here. Such metrics can be used for all investigated eye tracking encodings.

Another type of metric summarises whether a driver is “sufficiently” attentive, and that relevant information has been sampled “enough”. For the purpose-based approach, the criterion applied for “sufficient attention” was at least one glance (with a minimum duration) to each relevant target area, with the so far not well operationalised assumption that peripheral perception can be enough in certain cases. Ringhand et al. (2022) showed that glance behaviour does not only vary with zones and target areas, but also with traffic and task engagement, indicating that the one-glance criterion is not sufficient. Uncertainty and information decay (Clark, 2013; Kujala et al., 2023), taking the presence, type, trajectory, and speed of other road users into account, is likely to require certain glance durations, frequencies or potentially a minimum gap between consecutive glances to the same object. This would need to be explored in future research. Purpose-based coding could be of help here to understand the underlying processes.

4.3. Real-time driver monitoring

Real-time driver monitoring usually aims at ascertaining that a driver is attentive enough to keep driving or to take over from automation, or at detecting occasions of distraction and potentially warn the driver.

An advantage with the direction-based approach is its simplicity. Setting up a distraction detection system that monitors glances away from the forward direction, or downward glances to the car’s interior, is straightforward. However, the observed behaviour is difficult to interpret without adding contextual information. Therefore, such systems are most useful for real-time monitoring in motorway-like environments where relevant information mainly resides in the forward region. In more complex environments, such as in urban areas, the glance direction data should be complemented with contextual data allowing for at least a rudimentary implementation of relevant target areas, making it possible for an attention monitoring system to verify that all relevant directions have been covered in a timely manner. The direction coding of the present dataset clearly shows, unsurprisingly, that road layout in combination with traffic rules plays a big role in where drivers direct their glance.

A purely target-based encoding provides information about what the driver is looking at. For real-time applications its benefits appear to be limited. Glances to a phone, billboards or other objects could be identified and potentially warned about, and it would also be possible to detect the absence of glances to mirrors, traffic lights, etc., given their presence is known from additional data sources. The target coding of the present data is not very useful for real-time monitoring, even when context information about road layout and traffic encounter type is available. This is due to the lack of knowledge about target relevance and the presence or absence of road users or objects that were not foveated.

To integrate purpose-based coding into real-time monitoring, a better understanding is needed of how dynamic features like the
presence or absence of other road users influence information sampling. The present data indicate that there is a relationship, which could be explored to improve attention assessment. The identified purpose changes logically both with relevant target areas and with the traffic encounter type at hand. The requirements on additional data layers to be fused with the eye tracking data in real-time are high, though.

Note that information about the present situation, the driver’s intentions, and the relevance of the foveated direction/target/area, strengthens all three coding approaches. To avoid hindsight-bias when determining intentions and relevance, the only information considered should be from the past, the present, and the likely (but not the actual) future development of the situation. This can be compared with a situationally aware driver who is able to perceive and interpret the situation to anticipate upcoming events (Endsley, 1995). Predicting a driver’s intentions with a long enough prediction horizon is not yet possible, but having access to the future travel path from a navigation system would provide a useful estimate of intent which can be used in a real-time driver monitoring system.

5. Methodological considerations and limitations

All data encodings were done by a single person. For the relative comparison of different coding schemes done here, this is not a major limitation. However, a formal content analysis, where the glance share results are treated as absolute, representative, and generalisable, would require both an intra-rater analysis and a larger dataset with a controlled variation of specific features in the infrastructure, regulations and the presence or absence of relevant and irrelevant traffic. Such analyses are much needed but are out of scope of the present project.

To increase repeatability and reliability, it is important to be as specific as possible in the coding instructions regardless of which approach is used. However, this needs to be balanced with functionality or common sense. The latter may sound counterintuitive when the goal is to strive for objectivity, but the crux lies in the apparent precision of eye trackers, which can be deceptive for several reasons. The first issue is data correctness, which can be compromised in different ways, such as an offset of the foveal point. Even if the foveation is correct, the indicated target does not need to be the sampled information (Vater et al., 2020). A typical example is a driver glancing over the shoulder to check for the presence of a vehicle. In many cases the foveation target seems to be meaningless, but the presence or absence of a vehicle in the blind spot can be confirmed with peripheral vision. Therefore, the driver does not need to turn the head and gaze all the way around, and therefore the actual field of interest is not foveated. Similarly, when target areas are located roughly in the direction of the forward roadway, for example zebra crossings, a driver may confirm the absence of pedestrians via peripheral vision, which could be interpreted as inattention, no gaze to the sides of the zebra crossing is registered. Related evidence shows that people avoid obstacles without foveating them (Harms et al., 2019; Hyman et al., 2014). A similar mechanism may lead to an object being foveated, but not attended to. This can occur when the foveated object indeed was the actual gaze target, but was not processed enough, or when the actual sampling area lies along the same gaze vector, but on a different distance (Kim & Gabbard, 2022).

Permanent features such as traffic signs may be familiar to the driver from previous trips. The information content is thus already known even if the feature is not foveated. Again, this may be deduced from behaviour, but cannot be extracted directly from glance behaviour. A range of studies on change-blindness confirms this by showing that drivers act on information that was present in the past before it was changed (Charlton & Starkey, 2012; Charlton & Starkey, 2013; Martens, 2018; Wolfe & Horowitz, 2017).

Due to the extensive data encoding efforts, the data material used in this study is limited. This means that the reported glance shares can have an idiosyncratic component that is related to the preconditions found at the specific study locations. However, the main goal here was not to investigate glance share distributions for representative infrastructure or traffic occurrences, but to illustrate the differences resulting from the three eye-tracking data coding approaches. Even though the same raw data were used in all three approaches, the different coding schemes and aggregation procedures influenced the representation of gaze shares as well as the interpretation in terms of driver inattention.

6. Conclusions

While seemingly objective, the results obtained from eye tracking depend strongly on the data analysis approach and the amount of contextual information that is used when interpreting the data. Regardless of whether glance data are coded with the direction-based, target-based, or purpose-based approach, additional information about the present situation, the driver’s intentions, and the relevance of the foveated direction/target/area is needed to explain the observed glance behaviour.

A limitation with the target-based approach is that it equates distraction with glancing at targets not relevant for driving. This is problematic since the concept of spare capacity is not accommodated. Coding glances towards relevant target areas (that may, or may not, include different targets such as cars or traffic lights) opens for analyses of whether all relevant targets have been sampled, which subsequently accommodates for spare capacity. The latter requires context information when setting up the target areas.

It is well known that context, not only in the shape of infrastructure, traffic and rules as shown here, but also as determined by the level of vehicle automation and user characteristics, strongly influence sampling strategies. This should be considered when monitoring driver inattention and detecting driver distraction. In a driver monitoring setting, a good starting point would be to automatically add static target areas and the corresponding zones based on information from digital maps, possibly in combination with object recognition to establish whether any traffic is present within the target areas.
K. Kircher and C. Ahlström

Transportation Research Part F: Psychology and Behaviour 100 (2024) 243–253

CRediT authorship contribution statement

K. Kircher: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Writing – original draft, Writing – review & editing. Christfer Ahlström: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgements

This work was supported by the Swedish Strategic Vehicle Research and Innovation Programme (FFI; grant number 2017-05526).

References


Han, S., Guo, F., & Krause, C. (2023). In-depth Analysis of Crash Risk Associated with Eyes-off-road Duration by Road Control Type and Intersection Type (National Surface Transportation Safety Center for Excellence (NSTSCE), Issue Report # 23-UM-122). http://hdl.handle.net/10919/113544.


252