Sensing Time Effectiveness for Fitness to Drive Evaluation in Neurological Patients

Nadica Miljković¹, Jaka Sodnik²

1: University of Belgrade – School of Electrical Engineering, Bulevar kralja Aleksandra 73, 11000 Belgrade, Serbia
2: Faculty of Electrical Engineering, University of Ljubljana, Tržaška cesta 25, 1000 Ljubljana, Slovenia

Abstract

We present a method to automatically calculate sensing time (ST) from the eye tracker data in subjects with neurological impairment using a driving simulator. ST presents the time interval for a person to notice the stimulus from its first occurrence. Precisely, we measured the time since the children started to cross the street until the drivers directed their look to the children. In comparison to the commonly used reaction time, ST does not require additional neuro-muscular responses such as braking and presents unique information on the sensory function. From 108 neurological patients recruited for the study, the analysis of ST was performed in overall 56 patients to assess fit-, unfit-, and conditionally-fit-to-drive patients. The results showed that the proposed method based on the YOLO (You Only Look Once) object detector is efficient for computing STs from the eye tracker data in neurological patients. We obtained discriminative results for fit-to-drive patients by application of Tukey’s Honest Significant Difference post hoc test ($p < 0.01$), while no difference was observed between conditionally-fit and unfit-to-drive groups ($p = 0.542$). Moreover, we show that time-to-collision (TTC), initial gaze distance (IGD) from pedestrians, and speed at the hazard onset did not influence the result, while the only significant interaction is among fitness, IGD, and TTC on ST. Although the proposed method can be applied to assess fitness to drive, we provide directions for future driving simulation-based evaluation and propose processing workflow to secure reliable ST calculation in other domains such as psychology, neuroscience, marketing, etc.

Keywords: eye tracker, driving simulation, neurology patients, fitness to drive, object detection, sensing time.
1 Introduction

Evaluation of the ability to drive in neurological patients is a very challenging task for medical specialists. On the one hand, revoking driver's license can drastically worsen patient’s quality of life leading to social isolation, depression, and even deterioration of physical health (Carr et al. 2019; Korner-Bitensky et al., 1994; Thompson et al., 2018). Yet, on the other hand, therapists may face potential legal liability if an individual deemed fit-to-drive is involved in an accident or jeopardizes public road safety in some other way (Korner-Bitensky et al., 1994). Consequently, decision-making on a patient's fitness to drive must be performed with exceptional caution. To make this unpleasant situation even more unpleasant, medical professional cannot count on the reliable guidelines to evaluate fitness to drive in neurological patients as current protocols vary among countries and even among centers (Carr et al. 2019; Korner-Bitensky et al., 1994; Motnikar et al., 2020). A survey of physicians on practices around driving ability assessment in patients with nonepileptic seizures revealed that only 18% of physicians felt self-assured of their own decisions while the vast majority of 93.1% expressed the urge for evidence-based reporting guidelines (Farooq et al., 2018). This state of affairs poses an important question: Which method(s) to use for evaluating fitness to drive by medical professionals? Evidently, well judged decision-making consists of dedicated guidelines, as well as plausible scientific and clinical evidence (Korner-Bitensky et al., 1994).

For appropriate individual assessment, standard clinical, neurophysiological, and functional tests must be complemented with observations in a driving environment (Schanke & Sundet, 2000; Thompson et al., 2018). Indeed, proper driving evaluation for assessment of cognitive and visual functions and their complex interactions should be performed in real-life settings, as well as in response to hazards in a driving environment (Thompson et al., 2018). However, it would be way too dangerous for on-road tests to incorporate hazardous situations, especially in neurological patients (Cizman Staba et al., 2020; Edwards et al., 2003; Jurecki & Stańczyk, 2018; Motnikar et al., 2020; Olson & Sivak, 1986). Luckily, driving simulators can replace on-road tests due to their obvious safety and proven efficacy in assessing fitness to drive in neurological patients and general population (Cizman Staba et al., 2020; Frittelli et al., 2009; Motnikar et al., 2020). Furthermore, driving simulators contribute to the driving tests repeatability by providing an effective evaluation strategy for assessing driver’s behavior.
particularly related to the risky scenarios (Ciceri et al., 2013; Fisher et al., 2007; Frittelli et al., 2009; Jurecki & Stańczyk, 2018).

We would argue that driving simulator repeatability is exceptionally important for assessing established parameters in highway engineering and road traffic safety such as Perception-Response (PR) time. The determined norms of PR time that correspond to the reaction time in psychological literature are 2.5 s in the U.S. and 2 s in Europe. PR time is defined as the time from the first appearance of an obstacle to the initiation of braking and presents the most important parameter associated with the road accidents especially in critical scenarios (e.g., vehicle-pedestrian collision). (Chrysler et al., 2015; Edwards et al., 2003; Green, 2000; Jurecki & Stańczyk, 2018; Olson & Sivak, 1986)

In an insightful review and meta-analysis of driving assessment of patients with Parkinson disease (Thompson et al., 2018) it was concluded that crash rate of collisions with pedestrians and PR time (e.g., to red lights) are, among others, commonly used outcomes. Furthermore, PR time was identified as the most promising parameter for discerning the fitness to drive (Green, 2000; Motnikar et al., 2020; Schanke & Sundet, 2000), especially as a result of driving simulator complex scenery in comparison to the simple Alertness tests (Cizman Staba et al., 2020; Edwards et al., 2003; Motnikar et al., 2020). Also, patients diagnosed with Alzheimer’s disease, traumatic brain injury, and multiple sclerosis, have declined driving performance as a result of, among other factors, increased PR time (Frittelli et al., 2009; Jovanović, 2021; Schanke & Sundet, 2000; Schultheis et al., 2001).

In 2019, D’Addario & Donmez proposed an interesting approach to study the effect of cognitive distraction in healthy individuals by introducing the subcomponents of the PR time (D’Addario & Donmez, 2019). By manual analysis of the eye tracker data, D’Addario & Donmez divided the PR time into SL (Saccade Latency), PT (Processing Time), and MT (Movement Time). Their study emphasized that the cognitive distraction influenced vastly SLs related to the simulated road hazards. SL is defined as the time interval from the hazard appearance to the start of the first eye movement towards the hazard. In our study, we use slightly different definition for ST (Sensing Time) parameter as ST describes the time difference between the first appearance of the target object (i.e., pedestrian) in the scene to the first user’s gaze on that object following procedure introduced by (Ciceri et al., 2013). Our goal was to test whether subject noticed the object rather than whether the eye movement was just directed
towards or for how long it was fixed at the object of interest. To the best of our knowledge, no previous study has proposed the sensing time in different groups of neurological patients for assessing driving ability. We were keen on exploring the usability of this insightful ST\textsuperscript{1} parameter inspired by the parameters from (Ciceri et al., 2013; D’Addario & Donmez, 2019) for evaluation of fitness to drive in neurological patients. Aspects of this work have been described and tested in Master Thesis (Jovanović, 2021).

1.1 The Aim of the Study and Contributions

Our main aim is to evaluate the effectiveness of ST parameter computed from the eye tracker data in a driving simulator for discerning among three groups of neurological patients: fit-, unfit-, and conditionally-fit-to-drive. We contribute to the current body of knowledge in subsequent ways:

- We present a methodology to calculate sensing time parameter by adoption of the promising open-source state-of-the-art YOLO (You Only Look Once) object detector. Software code (with sample eye tracking video) is freely available via on GitHub Internet hosting platform with released Zenodo DOI (Miljković & Sodnik, 2022a). On top of that, we provide a table with relevant parameters with open software code for statistical analysis (Miljković & Sodnik, 2022b).
- We give a workflow for automated detection of the sensing parameter from the eye tracker video in pedestrian hazard simulations. Also, we explain encounters, as well as potential ups and downs of the proposed approach, mainly in regards to the exacerbated noise in the eye tracker videos and to the YOLO accuracy.
- We test whether sensing time is a valuable parameter for discerning among three groups of neurological patients and discuss its possible application in assessment procedures involving driving simulators or similar complex scenarios.

Altogether, we advocate that the method presented in this paper and tested for the eye tracker videos in a heterogeneous group of neurological patients could be employed in studies of

\textsuperscript{1} We coined „sensing time“ term since it explains naturally the foundation for its calculation following the perception-response time as our role model. However, in (Ciceri et al., 2013) the same parameter was termed time to first fixation, while similar parameters describing intervals from the object appearance to the fixation were termed saccade latency and perception time (D’Addario & Donmez, 2019). For details, please see the main text and for other variable names such as saccadic reaction time we recommend a remarkable guideline in (Holmqvist et al., 2022).
cognitive attention in related research fields such as experimental and cognitive psychology, architecture, neurosciences, etc.

2 Materials and Methods

The eye tracker data were recorded in 108 neurological patients. All patients participated in a standard procedure for license revalidation at the University Rehabilitation Institute Soča in the Republic of Slovenia. At the facility, patients underwent clinical, neurophysiological, functional, and on-road evaluation where the following scores were assigned to each patient: fit-, conditionally-fit- (fit-to-drive 30 km around patient’s residence), and unfit-to-drive. After initial recruitment, due to the unavailability of the data, assessment was performed in 91 subjects (24 females and 67 males) with age range from 18 to 89 years (49.88 mean and standard deviation of 17.15 years). Overall, 33 subjects were diagnosed with traumatic brain injury, 35 with non-traumatic acquired brain injury, and 21 with neurodegenerative diseases (14 with multiple sclerosis, 4 with Parkinson’s disease, and three with other). The remaining two patients were diagnosed with Guillain-Barré syndrome and epilepsy. The study included subsidiary testing of patients’ ability to drive in a motion-based driving simulator produced by Nervtech Ltd. (Ljubljana, Slovenia). All patients signed Informed Consents in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) and the study was approved by the Medical Ethics Committee at the University Rehabilitation Institute Soča. (Motnikar et al., 2020)

The videos available from the Tobii Pro Glasses 2 (Tobii Llc., Stockholm, Sweden) with sampling frequency of 50 Hz were used in order to automatically calculate ST parameters and to manually determine whether an actual collision occurred. We used only one scenario for ST calculation as it incorporates common pedestrian collision. The scenery comprised rural area where participants were expected to brake abruptly in response to two children running out behind a stopped bus (Fig. 1) (Cizman Staba et al., 2020). Other scenarios lacked the essential “surprise factor” that the selected scenario had with the children running behind the bus, so ST was calculated only for the selected collision.

Scene comprised vehicle moving in a straight direction towards the bus station at a speed that should be less than or equal to 50 km/h. As the vehicle approaches, children that were hidden behind the parked bus on the right side of the road start to run across the street. The driver
is expected to perceive the danger and decelerate. Road consists of two lanes in opposite
directions and without any traffic lights. On the right side of the road two traffic signs indicated
crosswalk ahead and children crossing road sign. Weather conditions were suitable for driving
(sunny weather without rain/snow/fog, see Fig. 1).

![Sample snapshots from the Nervtech simulator for the selected collision scene with children running behind the parked bus.](image)

**Figure 1.** Sample snapshots from the Nervtech simulator for the selected collision scene with children running behind the parked bus.

### 2.1 Video Analysis and Feature Extraction

All processing steps were performed in Python v3.8 (Python Software Foundation, Delaware, USA). We used Visual Studio Code (Microsoft, Redmond, Washington USA) as an IDE (Integrated Development Environment) for video analysis on computer with Linux Ubuntu 20.04 OS (Operational System) with 8 GB RAM (Random Access Memory) and Intel Pentium 3556U with 1.7 GHz. Apart from standard libraries for data processing such as numpy (Oliphant,
2006) and scipy (Virtanen et al., 2020), a specialized module (OpenCV (Bradski & Kaehler, 2008)) for video editing and analysis was used. YOLO v5s object identifier (Jocher et al., 2020; Redmon & Farhadi, 2018; Redmon et al., 2016) was used for object detection due to its proven outstanding possibilities to annotate objects and its respective probabilities with relatively high accuracy. Compared to the other detection systems, YOLO makes more localization errors, but is less likely to predict false positives on background which is extremely important in traffic situations (Redmon et al., 2016). To test the YOLO accuracy in pedestrian detection and to ensure reliable and correct ST calculation, manual checks of already trained YOLO classifiers were conducted and possible ST errors due to inappropriate YOLO localization were manually corrected.

Figure 2, Sample snapshots from the YOLO detector with the eye tracker gaze for pedestrian detection with enlarged details in a random simulator scene. a) A snapshot of subject’s gaze (red circle) outside of Region of Interest (ROI) marked by the yellow rectangle and determined by the YOLO detector as “person” b) Another snapshot of the subject’s gaze that matches ROI with the pedestrian on the right pavement.

The selected scenario was segmented from the eye tracker video with OpenCV tools and then each video segment comprising collision with pedestrians was converted to image sequences i.e., frames. Further, each frame was fed to the YOLO object detector to localize ROI
(Region of Interest)\(^2\) i.e., persons in an image. Sample output snapshots from YOLO detector are given in Fig. 2. YOLO output is a rectangle with annotated ROI in the relevant image. For illustration of YOLO output and its cross section with the eye tracker gaze (red circle in Fig. 2) we used a sample scene in Fig. 2 that does not correspond to the selected hazard scene with the children running behind the bus in Fig. 1.

To calculate ST, we subtracted the timestamps \(t_2\) and \(t_1\), where \(t_2\) presented the timestamp when driver noticed children. Namely, \(t_2\) corresponds to the timestamp when the driver’s gaze falls within the detection frame in a manner shown in Fig. 2b). We determined \(t_1\) as the time when children began to run across the road. Precisely, \(t_1\) presents the time when the YOLO detector indentified children appearance in the video frame. Complete video analysis workflow block diagram for ST computation is presented in Fig. 3 and freely available in (Miljković & Sodnik, 2022a).

Overall 35 eye tracker videos were excluded from the study due to the following conditions that prevented detection of the sensing parameter: tilted glasses (8 patients were excluded), missing data (11), due to the anticipatory pedestrian perception (2), altered glasses positioning (4), as eye gaze could not be determined either due to the gaze information loss or frozen gaze for the entire segment (7), due to the pauses in recording that prevented further analysis of the particular scene (2), and in one case the participant almost ran over the kids as (s)he did not look at pedestrians at all, so the ST could not be calculated (1). Remaining valid 56 recordings were analyzed and they consisted of 20 fit-, 17 conditionally-fit, and 19 unfit-to-drive patients, presenting a relatively balanced group distribution for further analysis.

As proposed in (D’Addario & Donmez, 2019), we also extracted additional features to test their influence on ST. Auxiliary features are: (1) the participants’ speed at hazard onset, (2) Initial Gaze Distance (IGD), and (3) Time-To-Collision (TTC). The children’s position at the collision onset was estimated using YOLO object detector as the center of the ROI rectangle with \((X, Y)\) coordinates in horizontal and vertical directions in pixels. Overall frame resolution of the eye-tracker video was 960 × 540 pixels. Then, the distance between children’s position and initial gaze coordinates was determined as Euclidian distance in pixels and rounded to the nearest integer value. The TTC parameter was calculated as the ratio between the distance of the ego

\(^2\) We use ROI abbreviation for Region of Interest being common term in object detection studies. However, AOI (Area of Interest) is also routinely used in eye-tracker related research (Holmqvist et al., 2022).
vehicle to the pedestrians and its speed. The distance between the ego vehicle and pedestrians was recorded automatically by the simulator. We added speed, IGD, and TTC as covariates to test their impact on statistical significance of ST parameter for differentiation of three groups of patients.

![Diagram](image)

Figure 3, Block diagram of automated video processing workflow with YOLO (You Only Look Once) object detector for calculation of ST (Sensing Time) parameter. ROI stands for Region of Interest. For more details, please see main text.

### 2.2 Statistical Analysis

For statistical analysis we used R programming language v4.1.2 (Team, 2021) in R Studio environment (R Studio, Inc., Boston, USA) with dplyr package (Wickham et al., 2015). As STs did not conform to the assumption of normal distribution, we used Welch’s ANOVA for our analysis following the approach applied also by Motnikar et al. on the same patient sample (Motnikar et al., 2020). We also used Tukey’s Honest Significant Difference (HSD) post hoc test with confidence level of 0.95 to create a set of confidence intervals on the means differences.

Similarly as in (Motnikar et al., 2020) STs deviating for more than three interquartile ranges from the third quartile were considered outliers and removed from the analysis. Moreover, valid STs were considered to be below 500 ms, as longer reaction times were previously considered as a “miss” - drivers not paying attention to the road and reacting too late or not at all (Cizman Staba et al., 2020; Edwards et al., 2003). Altogether, three observations were lost. Additionally, a general linear model (multi-factor ANOVA) was also performed to see if fitness to drive, speed, IGD, and TTC explain variation in patients’ ST with rigor significance level of 0.001 to
compensate for non-homogeneity of ST across groups of neurological patients. We used Shapiro-Wilk’s test to check the normality of model residuals. Also, Pearson’s product moment cross-correlation coefficients between ST and speed, IGD, and TTC were calculated to test whether speed, IGD, and TTC are linearly related to ST. Table with all relevant parameters and R code are available in (Miljković & Sodnik, 2022b).

We assessed measurement uncertainty type B for ST calculation, as well as combined measurement uncertainty according to the Guide to the Expression of Uncertainty in Measurement (BIPM et al., 1993) from the temporal resolution determined by the sampling frequency of the eye tracker video under the assumption of the uniform distribution.

3 Results

In Fig. 4 box plots for ST parameters for three groups of neurological patients are presented.

Figure 4, Box plot of STs (Sensing Times) for three groups of patients. Fit, unfit, and cond fit stand for fit-, unfit-, and conditionally-fit-to-drive groups of patients, respectively.

The results of Tukey’s HSD post hoc test with confidence level of 0.95 are presented in Table 1. Welch’s ANOVA for mean comparison among groups resulted in $F = 15.935$ and $p = 0.000$. 
Tukey’s honest significant difference post hoc test for ST parameter from Table 1 revealed that fit- patient group was significantly different from unfit- \( (p = 0.001) \) and conditionally-fit-to-drive \( (p = 0.000) \) groups. However, no significant differences were attained between unfit- and conditionally-fit-to-drive patient groups \( (p = 0.542) \). Clear distinction of interval for differences of the observed means for statistically significant results is shown in Table 1 corresponds to the ST box plots in Fig. 4.

Table 1. Results of Tukey’s honest significant difference post hoc test with confidence level set at 0.95. Differences in the observed means, lwr (lower end point of the interval), and upr (upper end point of the interval) are presented as integers. In Bold statistically significant results are highlighted. We report 7 decimals for adjusted \( p \) for more precise comparison.

| /                  | difference [ms] | lwr [ms]  | upr [ms]  | adjusted \( p \) |
|--------------------|-----------------|-----------|-----------|------------------|
| Fit vs. conditionally-fit | -130            | -199      | -61       | **0.0001074**    |
| Unfit vs. conditionally-fit | -31            | -101      | 39        | 0.5421314        |
| Unfit vs. fit        | 99              | 35        | 163       | **0.0012709**    |

To test the reliability of the performed statistical analysis, we repeated statistical tests with included outliers and concluded that statistically significant discrimination among groups did not change \( (p = 0.000 \) and \( p = 0.037 \) for significant differences, while non-significant had \( p = 0.268 \)).

Summary statistics for the STs, speeds, IGDs, and TTCs for three groups of patients are presented in Table 2. For three recordings, participants were not focused on driving activity and consequently the STs were relatively large, so we excluded them as outliers according to the criterion described in the previous section \( (720 \text{ ms}, 1420 \text{ ms}, \text{ and } 1060 \text{ ms}) \). Therefore, we did not incorporate outliers in summaries in Table 2. All excluded ST outliers belonged to patients from the conditionally-fit-to-drive patient group.
Table 2. Summary statistics of STs (Sensing Times) for three groups of neurological patients (fit-, conditionally-fit-, and unfit-to-drive). Min and max stand for minimal and maximal, respectively. SD stands for Standard Deviation, IGD for Initial Gaze Distance, and TTC to Time-To-Collision. For calculating SDs, we used Bessel’s correction. Please, see text for more details.

| Fitness to drive | ST [ms]       | speed [km/h] | IGZ [pixels] | TTC [s] |
|------------------|---------------|--------------|--------------|---------|
|                  | Range median | mean (SD)    | Median       | Range median | mean (SD) | Median | Range median | mean (SD) | median |
| Fit              | 120-320      | 163 (47)     | 150          | 7.4-46.9     | 19.5 (10.5) | 15     | 0.7-10.1 | 3.3 (2.6) | 2.9     |
| Conditionally fit| 140-460      | 293 (95)     | 290          | 5.5-34.8     | 19.1 (11.4) | 15.3   | 0.5-9.6 | 3.9 (3.5) | 2.7     |
| Unfit            | 100-440      | 262 (99)     | 240          | 3.3-41.1     | 20.3 (10.9) | 19.6   | 0.8-19.6 | 4.2 (4.7) | 2.4     |
Measurement uncertainty type B for ST is 5.77 ms for temporal resolution of 20 ms determined from the eye tracker video sampling frequency (50 Hz). As standard deviations of STs were much larger STs (> 47 ms from Table 2) than the measurement uncertainty type B, we discarded measurement uncertainty type B from further analysis.

A general linear model was run to observe also fitness to drive, speed, IGD and TTC on ST. While controlling for other parameters, speed ($p = 0.224$, $F = 1.588$), IGD ($p = 0.875$, $F = 0.025$), and TTC ($p = 0.015$, $F = 7.296$) do not significantly predict patient’s ST, except for the fitness to drive ($p = 0.000$, $F = 25.079$). Shapiro-Wilk’s normality test confirmed our assumption that residuals follow normal distribution ($p = 0.601$). We find significant interaction among fitness, IGD, and TTC ($p = 0.001$, $F = 10.697$) on ST. Cross-correlation coefficients revealed no significant correlation of speed (-0.105, $p = 0.503$), IGD (-0.004, $p = 0.979$), and TTC (0.290, $p = 0.059$) with ST.

All in all, for 20 recordings we applied manual correction of automatically produced results for ST which is 35.7% of all analyzed recordings. Manual checks were performed in all cases of YOLO detection of ROI and its intersection with the gaze as illustrated in Fig. 2. Whenever, YOLO failed to correctly annotate the pedestrian ROI, visual observation was performed of the gaze and pedestrian ROI junction to manually determine the corresponding frame when YOLO did not succeed to annotate ROI faithfully.

Crash rates showed that only one subject dimmed conditionally fit-to-drive crashed into the pedestrians (with TTC = 1.7 s and speed 47 km/h). The results for this subject were not included in the final analysis due to the eye tracker tilt. Overall, 6 out of the analyzed 56 subjects (10.7%) almost crashed into simulated pedestrians i.e., avoided crash, but had to abruptly stop in immediate proximity to children. Two of them were unfit-, two conditionally fit-, and two even fit-to-drive. Four of these patients had very low TTC of either 0.9 or 0.8 s, while one unfit-to-drive patient had TTC of 4.7 s, and for one we could not determine TTC due to the eye tracker noise.

4 Discussion

Sensing time proved its usability in discerning fit-to-drive patient group and at the same time did not capture the difference between conditionally-fit- and unfit-to-drive patients. Although unexpected, this is rather interesting outcome, as previous study on the same group of patients failed to find differences between fit- and conditionally-fit-to-drive groups of neurological patients (Motnikar et al., 2020). We would argue that these two
findings are complementary as conditionally-fit-to-drive patients are indeed at borderline between fit- and unfit-to-drive groups. Exploration of possible application of machine learning algorithms for investigating more in-depth this multiplex interaction in larger and more diverse samples would be an attractive future approach. In what follows, we thoroughly discuss obtained results and their possible implications.

4.1 Prospects of YOLO-based Eye Tracker Video Analysis

We sought to develop an automated procedure for evaluating ST from the eye tracker videos by application of general purpose pretrained YOLO object detector. However, due to the YOLO errors, we had to introduce manual corrections in ~36% (20 out of the selected 56 recordings). Future efforts towards automated procedures may be focused on the customization of openly available YOLO detector.

Since its introduction in 2016 by Redmond et al. (Redmon et al., 2016) YOLO reached much attention in the scientific and technical communities. In addition to welcoming advantage of YOLO open-source license, the main YOLO supremacy is its speed (from 45 up to the 155 frames per second for real-time employment). For the exact localization, YOLO can struggle to reach high accuracy with error rate of up to 34.5%, especially in cases of smaller objects in a scene (Redmon et al., 2016), though some of the previous flaws were corrected in v3 (Redmon & Farhadi, 2018), this challenge remains in v5 as well (Amin & Arby 2022). This error rate is in line with our results of ~36% of inaccurate ROI recognition. Still, we argue that the fidelity of the driving simulation may also play a significant role in pedestrian detection by YOLO identifier (Redmon et al., 2016). YOLO is a general purpose detector and for pedestrian detection some improvements of initial pretrained detector have already been proposed. For example Valiati and Menotti in 2019 introduced weak semantic segmentation in the learning phase to enhance YOLO performance for pedestrian detection (Valiati & Menotti, 2019). More examples include tiny-YOLO improvements proposed by Zhang et al. 2019 (Yi et al., 2019) and improvement of network layers by Lan et al. 2018 for pedestrian detection (Lan et al., 2018). Hence, future approach may include adaptations to the existing YOLO versions for more accurate pedestrian detection, especially in the simulation. Although our optimistically proposed approach of applying non-extended YOLO detector for automated ST calculation (Fig. 3) had relatively high error rate, the proposed approach of semi-automatic detection is at least one step ahead and less time consuming of existing hand-operated procedure (D’Addario & Donmez, 2019).
Conclusively, we believe that enlargement of the input parameters of the pretrained YOLO would be a reasonable way to improve its performance for detection of simulated pedestrians. Therefore, reliable open detection benchmark data for driving simulators are required. On one side, this is especially exciting direction as novel YOLO versions are focused on automated vehicles and can incorporate complex human-machine interactions by introducing parameters such as ST. On the other side, real-time assessment in simulators can speed up the decision-making on fitness to drive in neurological patients with immediate evaluation in complex environments. Until then, the proposed YOLO-based automated approach (Fig. 2) with manual inspection presents a relatively fast and reliable technique to determine ST.

4.2 Eye Tracker Video Analysis: Ups and Downs

Although 108 patients initially agreed to participate in the study and 91 entered eye tracker video analysis, we had a very high data loss due to the unreliability and unsteadiness of the eye-tracker glasses. In a previous study on the same sample (Motnikar et al., 2020), assessed reaction times in response to the on-screen instruction to come to a full stop were calculated for 54 subjects due to the “technical error”. Here, we were able to calculate STs in 56 patients being just two more than reported in (Motnikar et al., 2020). Apparently, the eye tracker data quality was assessed in a similar manner. Moreover, this sample size proved sufficient to achieve comparable statistical power with the previous study on the same sample.

Previous results on Tobii device data quality by Niehorster et al. (Niehorster et al., 2020) showed that signal was relatively stable, but not robust, while subjects spoke, made facial expressions, and moved the eye tracker device. Our participants were seated in a controlled environment, but they were not constrained from commenting or making facial expressions and therefore the data quality may be partly affected. Although some manufacturers provide corrective procedures and post-analysis to compensate for noise, eye tracker devices can still move in respect to the subject's head causing the “slippage” that can drastically deteriorate the signal quality regardless of the type of the eye tracker device (Niehorster et al., 2020). The Nervtech driving simulator is motion-based device and the incorporated haptic feedback may have added to the “slippage” of the eye tracking setup in our case. Therefore, we were especially cautious in interpreting video signals that interfered with “slippage” by introducing manual verification procedures and “slippage” influence on the gaze position prior to video analysis. Future procedures might consider either appropriate corrections of the recorded signal or automatic opt for eliminating such trials in order to
preserve data quality that can vastly influence the research results (Holmqvist et al., 2012; Holmqvist et al., 2022). Moreover, Tobii eye trackers have already demonstrated the data loss in the altered gaze direction (Niehorster et al., 2020) which is also confirmed in our study and we had to exclude data from 4 participants due to the eye tracker positioning that affected gaze direction in respect to the device.

Although measurement uncertainty type B is relatively small as a result of resolution of 20 ms, our analysis may have hindered the gaze accuracy by other effects such as varying distances already reported by the manufacturer. In the course of writing this article, an empirically based minimal reporting guideline for the eye tracker was published in (Holmqvist et al., 2022). We performed a post-assessment of our method to verify whether our procedure conforms to the noteworthy instructions for eye-tracking studies proposed in (Holmqvist et al., 2022). Subsequently, we stress that overall recording environment comprising setup and geometry, measurement space and monitor size, distance between participant and the eye tracker were all kept as constant as possible, although we could not control for every single possible effect. Our results show that despite all potential influences, the ST parameter was able to discern fit-to-drive group of patients.

4.3 Is Sensing Time a Useful Parameter for Evaluating Fitness to Drive?

Direct comparison of STs could be made only with the so-called SL presented in (D’Addario & Donmez, 2019) and for time to first fixation in (Ciceri et al., 2013) both assessed only in healthy participants. Namely, SLs for pedestrian hazards were 0.34 s (mean), 0.37 s (median), 0.00 s (min), 0.58 s (max), and 0.17 s (SD), which are partly in line with our results of 0.163 s (mean), 0.150 s (median), 0.120 s (min), 0.320 s (max), and 0.470 s (SD) for fit-to-drive patients. SLs in healthy subjects are, to our surprise, more in accordance with STs in conditionally-fit-to-drive patients. However, these comparisons should be taken with a grain of salt. In other words, PR time cannot be explored independently of the road collision situation as even tiny variations may have significant effect on collision management (Ciceri et al., 2013; Jurecki & Stańczyk, 2018). We would stress that the same principle applies for the ST parameter as existing evidence for time to first fixation reveals direct relation to the testing conditions (Ciceri et al., 2013). Mean times to first fixation ranged from 0.06 s to 0.53 s with SDs from 0.14 s to 0.42 s being to some extent in line with our results. The test variations that could cause shorter STs in our study may be related to the fact that two children pedestrians were running across the street behind the bus. Consequently, the test could contribute to the urgency of the risk avoidance and visual detection. This is in
accordance with previous studies where faster moving pedestrian presented a more challenging response scenario by reducing TTC (Chrysler et al., 2015). Although driving simulators present an excellent tool to evaluate driver’s behavior in a real-life complex environment, the presence of the investigators and somewhat unusual conditions may have generated more caution in subjects (Olson & Sivak, 1986) which is in our case especially visible for speed before the collision which was relatively low (Table 2). Compensate mechanism could have taken place in this slow drive as well, as older people tend to respond slowly in general and to counterbalance for reduced cognitive skills by lower speeds on the road (Green, 2000).

In (Cizman Staba et al., 2020) reaction times were considered valid in a range from 0.5 s to 4 s as shorter and longer reaction times were considered “cheat” and “miss”, respectively. Similarly, in (Edwards et al., 2003) the PR times could not be calculated in some cases due to the lack of response or due to the advanced reaction. By following and adapting this approach to ST, we excluded STs that were larger than 500 ms as typical glance is between 500 ms and 2 s as stated in ISO 15007 standard. Despite this rejection, our main conclusion related to fit-to-drive discrimination did not change.

ST is a subdivision of a response or PR time which can be influenced by a variety of factors and diseases. For example, mean PR time to vehicle block was slow after alcohol (2.21 s) and fexofenadine (1.95 s) consumption (Weiler et al., 2000). Also, some mixed findings were reported on driver’s age on PR time, but majority of studies agree that increased age is related to the increased PR time (Broen & Chiang, 1996; Edwards et al., 2003; Lerner, 1993; Olson & Sivak, 1986). Reported PR times for pedestrian sudden appearance in the intersections for the young age (19-23 years) group was 0.97±0.46 s and for the older group (65-83) it was 1.44±0.45 s (Edwards et al., 2003). These PR times are higher than STs presented in this study which is expected. In subjects with Alzheimer’s disease, mean visual reaction time was 511.00±63.20 ms, while healthy controls had 390.00±29.50 ms and age-matched neurological normal controls had 384.00±31.80 ms reaction time (Frittelli et al., 2009). This suggests that fit-to-drive patients could correspond to more healthy-like “normal controls” in (Frittelli et al., 2009) indicating possible differentiation among patient groups by reaction times and reasoning for our findings.

No significant mean effect was observed between conditionally-fit- and unfit-to-drive patients (Table 2). It is stated that patients tend to focus visual attention to a nearer parts of the scene compared to the healthy subjects (Motnikar et al., 2020). Moreover, visuospatial skills are among the best predictors of driving ability (Carr et al., 2019) and ST relation with
the driving fitness is expected. Therefore, the sensing parameter is expectedly smaller in fit-to-drive group as it would be reasonable that they explore both close and distant zones. The absent difference between conditionally-fit- and unfit-to-drive groups could be explained by a possible similar attention deficit that is common in many neurological disorders and causes patients to neglect parts of the scene (Motnikar et al., 2020). Indeed, cognitive distraction leads to delayed visual detection even in healthy sample (D’Addario & Donmez, 2019).

As expected, greater likelihood of simulator crashes was found in previous studies (Motnikar et al., 2020; Thompson et al., 2018) in patient population as well as a poorer driving performance. In our study only one subject had an actual crash. An “almost crash” situation showed expected tendency: fit- and conditionally-fit- had relatively low TTC and much less time to perform the avoidance maneuver, while this was not the case with unfit-to-drive patient. Turns out that TTC influenced management of hazardous situations. Although TTC parameters were in range of those in previous studies (in (Jurecki & Stańczyk, 2018) TTC was in range from 0.6 s to 3 s), statistical tests did not find any significant influence of TTC, neither of IGD, nor speed to the ST in our sample. There are two possible reasons for such finding. Firstly, TTC and speed did not manifest larger discrepancies among three groups of patients (Table 1). Secondly, this may be related to the cognitive processes that are responsible for ST duration. We can conclude similar for IGD influence on ST and assume that ST may be likely affected by cognitive impairment. The only statistically significant multiple interaction is among fitness, IGD, and TTC on ST. TTC to ST in different fitness groups may be related as shorter TTC calls for more urgent/faster action. The logical assumption would be that TTC affects reaction (steering and braking), but we may not exclude possible effect on ST especially in cases when IGD is relatively short.

Our results clearly indicate discernment among fit-to-drive patients and other groups that should be further reproduced and potentially replicated in a larger sample. It would be interesting if ST sensitivity could be used for capturing early driving ability degeneration and risk behavior in drivers as proposed in (Carr et al., 2019) or for assessment of visual attention in other complex environments.

4.4 Limitations of the Study

This study is not without drawbacks. We recognize the following limitations in our single-parameter approach for fitness to drive assessment:
1. The simulation design could incorporate the ability for a researcher to specify time to arrival as a trigger to control the event occurrence as proposed in (Chrysler et al., 2015).

2. Simulator-based drive cannot compensate for some individual preferences such as to the subjects’ own vehicles to avoid adaptation to the experimental conditions (Lerner, 1993). However, our subjects did not have expectations of an exact emergency situation (collision with children), so it may be reasonable to assume that excessive adaptation to the simulator drive did not take place.

3. We did not consider different crashing patterns with different variables, factors, and considerations as proposed in (Chrysler et al., 2015; Ciceri et al., 2013) for calculating and comparing ST in neurological patients. Moreover, we did not test the pedestrian intrusion direction as this could influence PR time duration (Jurecki & Stańczyk, 2018). Further variations for ST calculation may as well include angles at which pedestrians appear, nighttime conditions, etc. (D’Addario & Donmez, 2019). Yet, our single scenario guaranteed that adaptation and expectation did not take place, which could be a drawback for ST evaluation.

4. Our sample included a heterogeneous group of neurological patients (Cizman Staba et al., 2020) that may have caused the absence of the significant differentiation between unfit- and conditionally-fit-to-drive patients. Previous studies went even further by proving the differences in response time tests between stroke patients with left and right lesions (Kaizer et al., 1988), so ST in neurological disorders should be taken with precaution as more thorough clinical research is required.

5. Looking at pedestrians does not necessarily mean seeing them (D’Addario & Donmez, 2019). We assumed that the gaze direction is associated with the perception of hazard onset, but this may not be the case. Although expected, the difference between looking and seeing could not entirely explain the best ST of 100 ms in an unfit-to-drive patient as this patient had IGD of 60 pixels (37 other patients had smaller IGD). Despite the fact that lower limit of a saccade duration is less than 120 ms (ISO 15007 standard), we decided to keep this sample throughout analysis.

6. The scene design should take into account possible emotional distress especially in patients prone to such reactions. Subject with actual crash in our study did not report emotional reaction to the crash incidence, but our model did not visually display the physics of a true crash once it happened as in (Chrysler et al., 2015).
5 Conclusions

With simple technology such as eye tracker video and open-source fast YOLO object detector, we were able to exploit a subcomponent of perception–response time termed sensing time in a controlled environment enabled by a driving simulator for driving performance assessment in neurological patients. The sensing parameter alone proved significantly efficient in fit-to-drive identification. For discerning among conditionally-fit and unfit-to-drive patients, a more comprehensive approach with additional neurophysiological and driving simulator markers is needed.

Acknowledgements

J.S. kindly acknowledges University Rehabilitation Institute Soča employees and the Nervtech team. Authors gratefully appreciate the support from Nenad B. Popović, PhD from University of Belgrade – School of Electrical Engineering for his valuable assistance in design of illustrations and for provided feedback for the initial manuscript structure. Also, both Authors thank Nebojša Jovanović, MSc from University of Belgrade - School of Electrical Engineering for his kind contribution to earlier stages of the project, especially for his work on developing Python code to capture sensing time parameter.

Declarations

Funding

N.M. acknowledges amiable support from the Grant No. 451-03-68/2022-14/200103 funded by the Ministry of Education, Science, and Technological Development, Republic of Serbia. This research was financially supported by Slovenian Research Agency within the research program ICT4QoL - Information and Communications Technologies for Quality of Life, grant number P2-0246, and the research project Neurophysiological and Cognitive Profiling of Driving Skills, grant number L2-8178.

Conflict of Interest

The Authors have no relevant interests to disclose.
Ethics Approval

The results of this study have been previously published and the current research was performed retrospectively. The study was approved by the Medical Ethics Committee at the University Rehabilitation Institute Soča in Slovenia.

Consent to Participate

All participants signed Informed Consents in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

Consent for Publication

All data published as part of this article are anonymized and no personal data are shared. Therefore, consent for publication was not obtained.

Open Practices Statement: Data and Code Availability

Python source code and sample eye tracker video is available on GitHub under GNU General Public License v3.0 and released on Zenodo with DOI (Miljković & Sodnik, 2022a), while .csv table with relevant parameters and R code for statistical analysis is available on Zenodo repository under Creative Commons Attribution 4.0 International (Miljković & Sodnik, 2022b).

References

Al Amin, I. H., & Arby, F. H. (2022). Implementation of YOLO-v5 for a real time social distancing detection. Journal of Applied Informatics and Computing, 6(1), 01-06. https://doi.org/10.30871/jaic.v6i1.3484

BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, OIML. (1993). Guide to the Expression of Uncertainty in Measurement. British Standards Institution, London.

Broen, N. L., & Chiang, D. P. (1996, October). Braking response times for 100 drivers in the avoidance of an unexpected obstacle as measured in a driving simulator. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 40, No. 18, pp. 900-904). Sage CA: Los Angeles, CA: SAGE Publications. https://doi.org/10.1177%2F154193129604001807

Carr, D. B., Stowe, J. D., & Morris, J. C. (2019). Driving in the elderly in health and disease. Handbook of Clinical Neurology, 167, 563-573. https://doi.org/10.1016/B978-0-12-804766-8.00031-5

Chrysler, S. T., Ahmad, O., & Schwarz, C. W. (2015). Creating pedestrian crash scenarios in a driving simulator environment. Traffic Injury Prevention, 16(sup1), S12-S17. https://doi.org/10.1080/15389588.2015.1015001

Ciceri, M., Ruscio, D., Confalonieri, F., Vangi, D., & Virga, A. (2013, October). Hazard detection in driving context and braking reaction time. In Proceedings of the 22nd EVU Annual Congress. Florence (pp. 17-19).
Cizman Staba, U., Klun, T., Stojmenova, K., Jakus, G., & Sodnik, J. (2020). Consistency of neuropsychological and driving simulator assessment after neurological impairment. *Applied Neuropsychology: Adult*, 1-10. https://doi.org/10.1080/23279095.2020.1815747

D'Addario, P., & Donmez, B. (2019). The effect of cognitive distraction on perception-response time to unexpected abrupt and gradually onset roadway hazards. *Accident Analysis & Prevention*, 127, 177-185. https://doi.org/10.1016/j.aap.2019.03.003

Edwards, C. J., Creaser, J. I., Caird, J. K., Lamsdale, A. M., & Chisholm, S. L. (2003, July). Older and younger driver performance at complex intersections: Implications for using perception-response time and driving simulation. In *Proceedings of the Second International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design* (pp. 33-38). Iowa City, IA: Public Policy Center.

Farooq, U., Ahmed, S., Madabush, J., Kolli, V., Esang, M., Kotapati, V. P., ... & Stanciu, C. N. (2018). Survey of physician attitudes towards psychogenic nonepileptic seizures and driving. *Epilepsy & Behavior*, 83, 147-150. https://doi.org/10.1016/j.yebeh.2018.03.031

Fisher, D. L., Pradhan, A. K., Pollatsek, A., & Knodler Jr, M. A. (2007). Empirical evaluation of hazard anticipation behaviors in the field and on driving simulator using eye tracker. *Transportation Research Record, 2018*(1), 80-86. https://doi.org/10.3141/2018-11

Frittelli, C., Borghetti, D., Judice, G., Bonanni, E., Maestri, M., Tognoni, G., ... & Judice, A. (2009). Effects of Alzheimer's disease and mild cognitive impairment on driving ability: A controlled clinical study by simulated driving test. *International Journal of Geriatric Psychiatry: A Journal of the Psychiatry of Late Life and Allied Sciences*, 24(3), 232-238. https://doi.org/10.1002/gps.2095

Green, M. (2000). "How long does it take to stop?" Methodological analysis of driver perception-brake times. *Transportation Human Factors*, 2(3), 195-216. https://doi.org/10.1207/STHF0203_1

Holmqvist, K., Nyström, M., & Mulvey, F. (2012, March). Eye tracker data quality: What it is and how to measure it. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 45-52). https://doi.org/10.1145/2168556.2168563

Holmqvist, K., Örbom, S. L., Hooge, I. T., Niehorster, D. C., Alexander, R. G., Andersson, R., ... & Hessels, R. S. (2022). Eye tracking: Empirical foundations for a minimal reporting guideline. *Behavior Research Methods*. https://doi.org/10.3758/s13428-021-01762-8

Jocher, G., Nishimura, K., Mineeva, T., & Vilarinho, R. (2020). yolov5. Code repository https://github.com/ultralytics/yolov5.

Jovanović, N. (2021) Analysis of eye-tracking movements during driving simulation with application in neurology [Master thesis, University of Belgrade – School of Electrical Engineering].

Jurecki, R. S., & Stańczyk, T. L. (2018, April). Analyzing driver response times for pedestrian intrusions in crash-imminent situations. In *2018 XI International Science-Technical Conference Automotive Safety* (pp. 1-7). IEEE. https://doi.org/10.1109/AUTOSAFE.2018.8373339

Kaizer, F., Korner-Bitensky, N., Mayo, N., Becker, R., & Coopersmith, H. (1988). Response time of stroke patients to a visual stimulus. *Stroke*, 19(3), 335-339. https://doi.org/10.1161/01.STR.19.3.335

Korner-Bitensky, N., Sofer, S., Kaizer, F., Gelinas, I., & Talbot, L. (1994). Assessing ability to drive following an acute neurological event: Are we on the right road?. *Canadian Journal of Occupational Therapy*, 61(3), 141-148. https://doi.org/10.1177%2F000841749406100303
Lan, W., Dang, J., Wang, Y., & Wang, S. (2018, August). Pedestrian detection based on YOLO network model. In *2018 IEEE International Conference on Mechatronics and Automation (ICMA)* (pp. 1547-1551). IEEE. https://doi.org/10.1109/ICMA.2018.8484698

Lerner, N. D. (1993, October). Brake perception-reaction times of older and younger drivers. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 37, No. 2, pp. 206-210). Sage CA: Los Angeles, CA: SAGE Publications. https://doi.org/10.1177/2F154193129303700211

Miljković, N., & Sodnik, J. (2022). NadicaSm/Sensing-Time-Calculation-from-the-Eye-Tracker-Videos (v1.0.0). [Software code], Zenodo. https://doi.org/10.5281/zenodo.6560419

Miljković, N., & Sodnik, J. (2022). Parameters for Statistical Evaluation of Sensing Time Effectiveness for Assessment of Fitness to Drive [Data set]. Zenodo. https://doi.org/10.5281/zenodo.6560246

Motnikar, L., Stojmenova, K., Štaba, U. Č., Klun, T., Robida, K. R., & Sodnik, J. (2020). Exploring driving characteristics of fit-and unfit-to-drive neurological patients: A driving simulator study. *Traffic Injury Prevention, 21*(6), 359-364. https://doi.org/10.1080/15389588.2020.1764547

Niehorster, D. C., Santini, T., Hessels, R. S., Hooge, I. T., Kasneci, E., & Nyström, M. (2020). The impact of slippage on the data quality of head-worn eye trackers. *Behavior Research Methods, 52*(3), 1140-1160. https://doi.org/10.3758/s13428-019-01307-0

Oliphant, T. E. (2006). *A guide to NumPy* (Vol. 1, p. 85). USA: Trelgol Publishing.

Olson, P. L., & Sivak, M. (1986). Perception-response time to unexpected roadway hazards. *Human Factors, 28*(1), 91-96. https://doi.org/10.1177/001872088602800110

Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*. https://doi.org/10.48550/arXiv.1804.02767

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 779-788). https://doi.org/10.1109/CVPR.2016.91

Schanke, A. K., & Sundet, K. (2000). Comprehensive driving assessment: Neuropsychological testing and on-road evaluation of brain injured patients. *Scandinavian Journal of Psychology, 41*(2), 113-121. https://doi.org/10.1111/1467-9450.00179

Schultheis, M. T., Garay, E., & DeLuca, J. (2001). The influence of cognitive impairment on driving performance in multiple sclerosis. *Neurology, 56*(8), 1089-1094. https://doi.org/10.1212/WNL.56.8.1089

Team, R. C. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.r-project.org

Thompson, T., Poulter, D., Miles, C., Solmi, M., Veronese, N., Carvalho, A. F., ... & Uc, E. Y. (2018). Driving impairment and crash risk in Parkinson disease: A systematic review and meta-analysis. *Neurology, 91*(10), e906-e916. https://doi.org/10.1212/WNL.0000000000006132

Valiati, G. R., & Menotti, D. (2019, June). Detecting pedestrians with YOLOv3 and semantic segmentation infusion. In *2019 International Conference on Systems, Signals and Image Processing (IWSSIP)* (pp. 95-100). IEEE. https://doi.org/10.1109/IWSSIP.2019.8787210

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., ... & Van Mulbregt, P. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods, 17*(3), 261-272. https://doi.org/10.1038/s41592-019-0686-2
Weiler, J. M., Bloomfield, J. R., Woodworth, G. G., Grant, A. R., Layton, T. A., Brown, T. L., ... & Watson, G. S. (2000). Effects of fexofenadine, diphenhydramine, and alcohol on driving performance: A randomized, placebo-controlled trial in the Iowa driving simulator. *Annals of Internal Medicine, 132*(5), 354-363. https://doi.org/10.7326/0003-4819-132-5-200003070-00004

Wickham, H., Francois, R., Henry, L., & Müller, K. (2015). *Dplyr: A Grammar of Data Manipulation*. R *Found. Stat. Comput., Vienna*. R package version 0.4. 3. https://CRAN.R-project.org/package=dplyr. https://CRAN.R-project.org/package=dplyr

Yi, Z., Yongliang, S., & Jun, Z. (2019). An improved tiny-yolov3 pedestrian detection algorithm. *Optik, 183*, 17-23. https://doi.org/10.1016/j.ijleo.2019.02.038