What Makes Driving Difficult?
Perceived Effort and Eye Measures Follow Visible Semantic Complexity Factors

Christopher D. D. Cabrall[a] , Riender Happee[a], Joost C.F. de Winter[b]

[a]*Cognitive Robotics - Intelligent Vehicles,* [b]*Biomechanical Engineering,*
Delft University of Technology, Mekelweg 2, 2628 CD Delft, the Netherlands
C.D.D.Cabrall@tudelft.nl, R.Happee@tudelft.nl, J.C.F.deWinter@tudelft.nl

Abstract

A majority (95%) of crashes can be attributed to humans, with the highest cause category (41%) involving errors of recognition (i.e., inattention, distraction, inadequate surveillance). Driving safety research often claims that as much as 90% of the information that drivers use is visual. However, these claims have been hampered by a lack of numerical measurement systems. Presently, we develop an ordinal visual driving scene complexity measurement based on human judgments and eye behavior. Mimicking the rebuilding of situation awareness in take-over conditions we presented 60 randomly ordered video clips (3 s duration), varying complexity factors of traffic density, road curvature, and miscellaneous visual features. Eyes of 15 participants were recorded while viewing the clips, and participants provided ratings between 0 and 100 to the question of "how much effort for you to take control and drive within that segment?". Effort ratings showed a monotonic increase with the number of complexity factors present. A statistically significant increase was also found for saccade amplitude, whereas a statistically significant decrease was found for fixation duration. Pupil size also showed a significant increase but only between 2 complexity levels and at a relatively less convincing strength. In conclusion, the present complexity factor coding scheme apparently corresponds to subjective effort. Further consideration should be given to relating eye tracking measures to visual driving scene components and task demands. In real-time driving systems, both human occupant(s) and computerized processes may observe the same scene at the same time, and matching the machine quantification of the situation to intuitive human judgments is expected to aid in the adherence to advisories and acceptance of automated aids.

Keywords

Intelligent Vehicles; Transition of Control; Workload; Highly Automated Driving; Driver State Monitoring

1. Introduction

Background Motivation. According to historic reports from the National Highway Traffic Safety Administration (NHTSA) in the USA [1], a substantial drop can be seen in fatal motor vehicle crash data from about a decade ago (Figure 1). However, in February 2017, the National Safety Council public service organization in the USA released relatively disturbing record-setting trends in driving deaths: an increase of +6% in the last year and +14% increase between 2014 and 2016; and with an exceedance of 40,000 for the first time in the last decade, this altogether represents "the most dramatic two-year escalation in 53 years" [2].
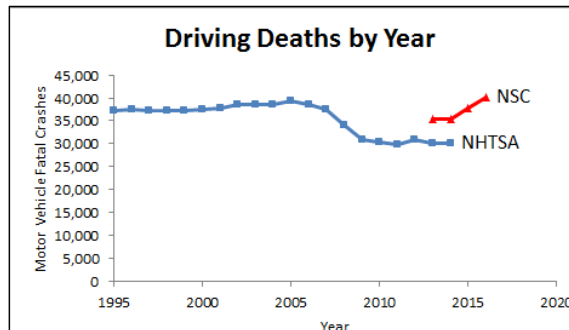
Figure 1: Number of fatal motor vehicle crashes, from the NHTSA (1995 to 2014) and from NSC (2013 to 2016). Note: data may not be directly comparable in absolute terms because NHTSA counts fatalities within 30 days of the crash while NSC allows for up to a full year.

Common accounts for the drop in the number of fatalities (Figure 1) are the wider fleet adoption of advanced driver assistance systems (ADAS) (e.g., electronic stability control, lane departure warnings, collision warnings, adaptive cruise control, backup cameras, etc.). Indeed, upon its creation in 2004, the Wikipedia page for

1

"Advanced driver assistance systems" only listed about a handful of examples while it now stands at around 30 [3]. Furthermore, in a 2017 driver safety public opinion poll conducted by the NSC, 42% of drivers reported having such safety features in their vehicles but with an alarming 37% of these drivers claiming agreement with shutting off the safety technology because it is confusing, irritating, or giving too many false positives [4] while an overall 5% of respondents indicated that they felt such new safety features should be avoided in next car purchases.

In the face of a plateauing (or even worse, reversing) trend of reducing traffic fatalities, the proliferation of new driving technology alone may not be enough to reach visionary national traffic safety ambitions looking to eliminate deaths (i.e., "Road to Zero" in USA, "Vision Zero" in Sweden, etc.). With new technological gains, there are complicating challenges such as behavioral adaptation [5] and human automation interaction issues remaining to be addressed whenever human eyes and brains are both involved [6]. For example, a majority (95%) of determined critical causes of crashes have implicated the driver, with the highest cause category (41%) being not physical/mechanical actuation issues but cognitive/perceptual errors of recognition (i.e., inattention, internal/external distractions, inadequate surveillance) [7]. Correspondingly, driving safety research literature seeks to understand how humans process visual information, which has been claimed to account for 90% of the all information employed by drivers. However, these claims have been hampered by a lack of numerical measurement systems [8], let alone ones that might be accepted by the human public in all contexts, cf. [9].

In summary, a question worth examining is what things can be seen in driving scenes by humans that connote difficulty for driving. Having such knowledge on how much perceived subjective effort it takes to drive within certain driving situations may aid the design of driver warning, assistance, and transitions with autonomous control systems by nature of reducing false alarms and issues of automation surprise, thereby improving acceptance, use, and effectiveness as well as ensuring driver readiness. Thus, the aim of the current study was to develop a numerical driving scene complexity measurement informed from human judgments. Furthermore, because these judgments would be formed only from visual differences presented to our participants (i.e., driving video clips), their correspondent eye behavior in evaluating presented driving scenes was also analyzed.

Eyes, Effort, and Driving Scenes – Getting Situated. Paired with the ubiquity of mobile communication and Internet connected devices, camera technology has reached a recent inflection point in catalytic response to increasing demands for commodity convenience (small, cheap) without compromising on quality (resolution, capability). In both hardware and software terms then, the utilization of visual information channels in driving safety domains is a prevalent development area ripe for potential value considering the assumption that in addition to whatever cameras a manufacturer builds into their vehicle, consumers bring at least a few of their own already (i.e., cell phone, dash cams, etc.). Such trends might be expected to continue even more so in the future as both professional/personal off-the-shelf and after-market camera-based solutions/services continue to advance (e.g., tele-conference/communications, live streaming, computer vision, augmented reality, etc.). Indeed, a wide area of applied research in ADAS is progressing with cameras pointed both in the direction of the driver and/or aspects of the road scene [10].

In regards to the information contained in the eyes of the driver, a recent review has been conducted to summarize literature on the relation of eye measurement parameters and driver mental workload [11]. While non-consensus results were found regarding some parameters (i.e., blink rate), other measures derived from blinks (i.e., duration, latency), fixations (i.e., duration, gaze variability) and pupils (i.e., dilation, ICA) were shown to have directional promise as estimates of mental effort (see Table 1).

Table 1: Relation of Eye-Related Physiological Measures and Drivers' Mental Workload

Measure		Mental workload
Blinks	Rate	+ / −
	Duration	−
	Latency	+
	PERCLOS	+
Fixations	Duration	+
	Gaze variability	−
Pupillometry	Pupil dilation	+
	ICA	+

Note. In the literature, PERCLOS has mainly been associated with sleepiness and performance decrements rather than workload per se [11].

In particular, the controversy regarding blink rate was explainable due to situational aspects (i.e., visual demands) and the theoretical differentiation of these situational aspects to cognitive demands [11]. For example, high densities of driving scene visual demands such as traffic lights, road signs, road markings, billboard/advertisements, etc. may reduce blink rate as the observer prolongs looking exposures to take in and process the large amount of visual information. On the other hand, cognitive demand does not necessarily follow from an increase in visual demand, as drivers may not in fact be tasked to take it in all information but instead prioritize along subsets of the most subjectively relevant/meaningful aspects. Although showing more directional consistency evidenced across the reviewed literature in relation to driver mental workload (Table 1), the remaining eye measures may also deserve circumspection would such situational aspects be more finely scrutinized (i.e., consideration and comparison of visual versus mental driving demands).

Situational aspects from ecological systemic [12] [13] and hierarchical perspectives [14], [15] have a long standing presence in driving safety theory. Attempting a systematic concept to describe driving activity as early as 1938, Gibson and Crooks [12] introduced the notion of fields of safe travel protruding in front of vehicles that are impinged by obstacles to be avoided (e.g., other traffic and road curvature). Michon conveyed [14] and further elaborated [15] on dealing with dangers in driving by employing three performance levels associated with specific risk evaluation stages. Herein, these are enumerated in reverse order from higher to lower levels but described per Michon's own terminology: 3) strategic level with issues of overall trip planning/goals and risk acceptance; 2) tactical/maneuvering level with management of risk probabilities through negotiations like speeding up, slowing down, turning, and overtaking, and 1) operational/control level with threat coping aims to avoid acute, perceived danger wherein the basic skills of steering and braking come directly into play. Lastly, a cross-continental group of around a dozen subject matter experts from both the US and the EU recently convened and published a "conceptual framework and taxonomy for understanding and categorizing driver inattention" [13]. While drawing on both of the aforementioned theoretical accounts of safety margins and functional levels, the authors summarized their dynamic perception-action cycle framework by characterizing dangers/difficulty in driving (i.e., inattentive driving) as a mismatch between exogenous contexts containing specific task demands/predictability versus endogenous resource allocation.

Between these different approaches, the importance and clarification of situations is commonly communicated through use of various metaphors.. Gibson & Crooks availed to probing tongues, valence fields, and affordances [12]. Michon described three structured levels with progressive temporal urgency, and differentiated true dangers from perceived threats, calling upon the terminology of 'risks' from human information processing research and 'stress/threat-coping' from psychophysiological and bio-medical research [14], [15]. Engström et al. [13] used adaptive attention and resource demand/capacity models, and repeated the classic conception of a beach-traversing ant: "Viewed as a geometric figure, the ant's path is irregular, complex, hard to describe. But its complexity is really a complexity in the surface of the beach, not a complexity in the ant" [16]. Thus, within the situated theory approaches to driving safety introduced above, there are many explanatory diagrams, metaphors, and models which underscore an importance of treating a context beyond that of just a purely driver-centric approach.

Eyes, Effort, and Driving Scenes – Getting Specific. When it comes to detailed and specific accounts of driving task demands, examples from the theoretical literature may on the one hand appear considerably vague ("road surface conditions, road infrastructure layout, visibility and the behaviour of other road users", "normal conditions (e.g., daylight, dry road surface, sparse traffic, wide lane)" – [13]) or anecdotal ("when a driver sees two cars approaching on a two-lane road and the rear car swings out to pass with the intention of cutting in before the various paths meet" – [12]). Additionally, driving vigilance task demand operationalization from the empirical literature has been found at times overly contrived or under-specified [17]. Consequentially, a recently validated approach is found in [18] by employing the Internet to tap the eyes/brains of 200 different human crowdworkers from 46 different countries in one and a half days with 12,819 different examples of driving scenes. Therein a 36-item content categorization scheme for describing driving scenes was found to yield both accurate and reliable responses even though the crowdworkers viewed those scenes through limited resolution video clips that lasted for only a few seconds.

However, a pressing question remains. How might the apparent plethora of possible driving situations be resolved towards a standardized complexity assessment? The current study aimed to address this question by assuming a focus on readily visible features (i.e., for robust application with common low-level camera solutions) and by categorizing items in line with those proposed in the above situated approaches and theoretical frameworks.

Within the US-EU working group taxonomy [13], the authors defined the first two lower levels of Michon [14], [15], with "operational" being dependent on "safe margins to the road edges and other road users" and

"tactical" as including aspects of decision, for example, "whether to yield at an intersection". Furthermore, they [13] defined "activities critical for safe driving" to include "all levels that are required to maintain acceptable safety margins" with both operational level examples such as "maintaining headway, keeping in the lane, visually scanning an intersection for oncoming vehicles" and tactical level examples of "deciding whether to yield and interpreting safety-related traffic signs". Such a division between lower level immediacy and higher level consideration is also found in common accounts of cognitive processing, that is, System 1 for fast intuitive reactions and instantaneous decision, and System 2 for more deliberate thinking involved in focused reasoning, assessment, and analysis [19]. Thus, in the present study, we prioritized more immediate control aspects of course conflict and collision conflicts as separate unique complexity category factors with more narrowed/directed emphasis while we combined a broader category of less immediate behavioral/infrastructural aspects. In other words, in our complexity factoring grouping scheme, like Michon [14], [15], we prioritize fast control aspects (safety margins, more immediate/acute, threat coping) such as presence of other cars and presence of road curvature for two different first-order complexity factors (e.g., "probable danger") likely to elevate driving difficulty/effort. Additionally, we consider another diluted third complexity factor by grouping a wider category of aspects regarding more interpretation/decision (signage, road layout, etc.) to manage probabilities/potentials but not usually as immediate/urgent in kind.

In summary, ADAS challenges of acceptance and trust are becoming recognizable both in recent layman surveys [6] and expert observations [9]. An alignment of human and machine judgements should reasonably be expected to facilitate driver-automation interaction. Thus, to contribute knowledge towards such alignment and to take advantage of a burgeoning resource of in-vehicle cameras and computational power, the present study aimed to assess correspondent ordinal relations between visual complexity in videos, self-reported supposed driving control effort and eye measures. Such a motivation follows in line with a concise call-to-arms as put forth by [20]:

> "ADAS will become an increasingly powerful force in directing the attention of the drivers **to what the ADAS estimates** to be critical elements of the roadway environment. **To the extent that the ADAS correctly identifies** critical elements, it will mitigate distractions but could also distract drivers if it directs attention away from these critical elements...distraction reflects **a mismatch** between the attention demanded by the road environment and the attention devoted to it" (p. 35) [emphasis added]

We hypothesized that immediate driving control aspects of road curvature and other vehicular presence have enough effort priority to constitute each their own factors of scene complexity compared to a wider more general category of miscellaneous infrastructure/behavioral activity. We expected that scenes with lower levels of these complexity factors would receive lower effort judgments by human observers compared to higher levels, with higher perceived effort. Additionally from the general reviewed literature of eyes and driving workload [11], it would be presumed for eye data to follow in line with previous results such that with higher effort ratings we might expect to see increased pupil sizes, increased fixation durations, and decreased gaze variability (herein captured as saccade amplitude).

2. Methods

We implemented an empirical approach to investigate the effects of various visible driving scene characteristics on human participant perceived effort ratings and correspondent eye behavior.

Participants and Apparatus. Written informed consent was obtained from all participants and the research was approved by the Human Research Ethics Committee of the Delft University of Technology under the title 'Driving video ratings' (16 December 2015). The experiment was completed by 15 participants (six female, nine male) aged between 18 and 36 (M = 26.60, SD = 4.26) with an average driving experience of around seven years since obtaining the driver's license (M = 7.20, SD = 4.20).

The experiment apparatus consisted of an isolating partition, a stimulus display monitor, eye tracker camera with integrated IR source and dedicated head/chin rest mount, as well as a gaming steering wheel (Figure 2). The display of the monitor measured 24 inches diagonally with a resolution of 1920 x 1080 pixels and was positioned about 95 cm in front of the participant and about 35 cm behind the eye tracking camera/IR source. The boundaries of the stimulus display area subtended within 32/25 degrees of horizontal/vertical viewing angle per the setup guidelines of the SR Research Eyelink 1000 Plus eye tracker which, after individual participant calibration/validation was used to record eye behavior data. The eye event parser was set according to the default psychophysical configuration recommended by the manual for research containing smooth pursuit movements and containing measurements of saccadic amplitude: saccade velocity threshold of 22 degs/sec; saccade acceleration threshold of 3800 degs/sec^2, and saccade motion onset delay threshold of 0 degrees.

The gaming steering wheel was a Logitech G27 but was not connected to anything and along with the isolating partition was used to facilitate driving video stimulus immersion. Participants also made use of a standard USB desktop mouse to input effort ratings on the stimulus display monitor.
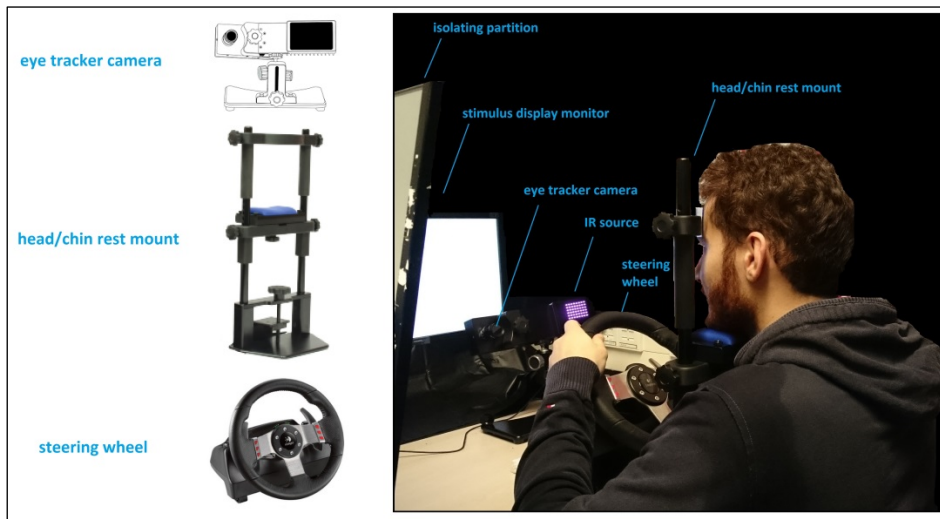

Figure 2: Diagram of experiment apparatus components and arrangement

Procedure. Participants were encouraged to sit up straight and the height of the head/chin rest mount, as needed for precise eye tacking, was adjusted to each participant to reduce potential neck/shoulder strain. Participants kept their heads stationary within the mount throughout the experiment except for voluntary rest breaks made available to them around every five minutes across about 15 minutes of driving video viewing and rating trials. Each trial began with an online drift correction dot in the center of the screen to which participants needed to fixate and click the mouse at the same time to begin. A three second long driving video clip was then played during which participants were tasked to move their hands to the wheel while imagining that they were taking over control (i.e., from automated driving) and that they must drive within that scene.

Examples of several driving video clips from each complexity level are shown in Figure 4. These stimuli consisted of a randomly ordered set of 60 clips across a multi-level grouping extended from a semantic content categorization scheme and annotated driving video clips developed in [18]. Three binary grouping complexity factors (presence of other traffic, non-straight road geometry, miscellaneous infrastructure/vehicle behavior) were used to construct four balanced levels based on 0, 1, 2, or all 3 complexity factors being present, for a subtotal of 15 video clips per level. The duration of 3 seconds was validated as adequate for human observers to reliably relate driving scene contents [18] which include dynamic aspects not possible to see in still images.

After the clip finished and disappeared, an effort rating scale ("How much effort for you to take control and drive within that segment?") was presented on the upper half of the screen and participants moved a vertical mouse cursor to click on the scale to input their answer from between "Very Low" to "Very High". Cursor click horizontal positions were divided by the pixel length of the scale and rounded to a single point resolution from 0 to 100. The presented horizontal effort scale contained 21 equally spaced demarcations from left to right following from those described within the seminal NASA TLX (Task Load Index) subscales [21] and widely adopted across driver workload assessment research (see Figure 3).
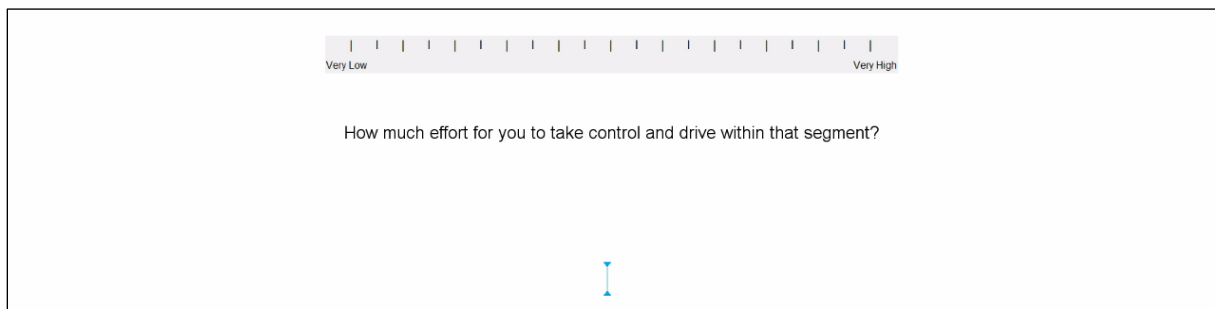
Figure 3: Driving effort response scale and cursor used to position on top of scale

Level 3

yes road curvature,
yes other traffic,
yes misc. feature (signage) — 1, 1, 1

yes road curvature,
yes other traffic,
yes misc. feature (road marking, traffic light, signage) — 1, 1, 1

yes road curvature,
yes other traffic,
yes misc. feature (turning on/off road, road marking, signage) — 1, 1, 1

Level 2

yes road curvature,
yes other traffic,
no misc. feature — 1, 1, 0

yes road curvature,
no other traffic,
yes misc. feature (signage) — 1, 0, 1

no road curvature,
yes other traffic,
yes misc. feature (signage) — 0, 1, 1

Level 1

yes road curvature,
no other traffic,
no misc. feature — 1, 0, 0

no road curvature,
yes other traffic,
no misc. feature — 0, 1, 0

no road curvature,
no other traffic,
yes misc. feature (signage) — 0, 0, 1

Level 0

no road curvature,
no other traffic,
no misc. feature — 0, 0, 0

no road curvature,
no other traffic,
no misc. feature — 0, 0, 0

no road curvature,
no other traffic,
no misc. feature — 0, 0, 0

Figure 4: A dozen examples out of the full set of 60 to illustrate the grouping levels of complexity factors. The first factor pertained to road curvature, the second factor to other traffic, and the third factor to any combination of miscellaneous infrastructure/behavior, that is, road markings (excluding lane boundaries), mutli-lane roads, signs/signals facing the driver, traffic lights, lane changes (self/other), and turning on/off road (self/other) [18].

3. Analysis and Results

Dependent measure data were grouped by complexity Level (0, 1, 2, or 3) and shown in a boxplot (Figures 5, 6, 8, 9) to visualize medians, first and third quartile information with outliers presented beyond 1.5x the interquartile range. Next in order to test hypothesized differences, multiple repeated measures single factor Analysis of Variance (ANOVAs) were performed each time with complexity as the independent variable and with effort ratings and eye behavior data (i.e., averaged as the mean across 3 seconds of measurement: pupil size, fixation duration, saccade amplitude) as the dependent variables. Rank transformations were first applied to dependent measures prior to each ANOVA [22] for the benefit of being more robust to effects of outliers. Thus, the repeated-measures ANOVA was conducted on a matrix (15 participants * 4 levels) consisting of mean ranks across 15 videos per level. Post hoc comparisons with Bonferroni correction (i.e., alpha = 0.05/6) were performed for each of the six possible inter-complexity level comparisons to ascertain where that difference evidenced.

Effort Ratings. An apparent increase in average effort rating responses was found across low to high complexity level driving videos (Figure 5). A one-way repeated measures ANOVA revealed a significant effect of complexity level category on effort ratings ($F(3,42) = 51.7$, $p < 0.001$). Post hoc comparisons indicated that mean ranks of effort ratings were statistically significantly different between each of the six combinations of Levels.
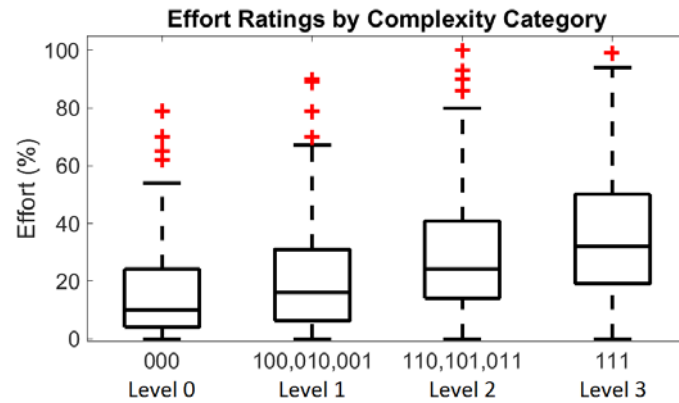
Figure 5: Responses to the question "How much effort for you to take control and drive within that segment?" on a scale from "very low" to "very high" (see Fig. 3) presented immediately following a driving video clip grouped by driving video complexity level.

Pupil Size. No apparent increase was found in average pupil size across the low to high complexity level driving videos (Figure 6). A one-way repeated measures ANOVA, however, did reveal a significant effect of complexity level category on effort ratings ($F(3,42) = 6.44$, $p = 0.001$). Post hoc comparisons showed Level 1 to be statistically significantly different from Level 0.
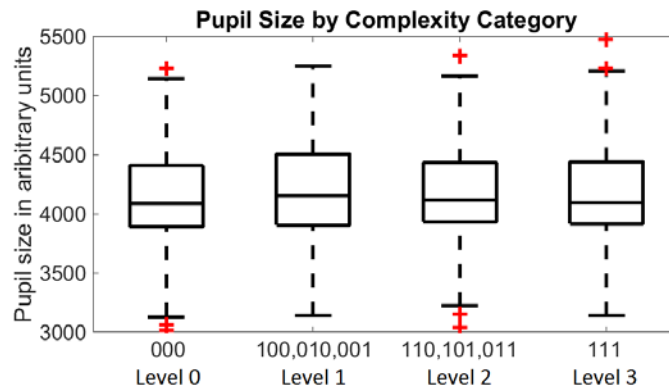
Figure 6: Eye tracker measured pupil size in arbitrary/relative units grouped by driving video complexity level.

To further examine the lack of apparent differences in pupil size between the various complexity levels, a heat map visualization was generated for the dependent variable of pupil size by averaging and sorting across all participants and all video clips in an orthogonal tabularization (Figure 7). Differences between driving videos did not seem to evidence obvious differences in average pupil size overall (row range of 291; max. of 4298 to min.

of 4007) nor when looking relatively within individuals separately where ratios of standard deviations divided by means of a participant's pupil size across all videos averaged around 0.036 indicating not a lot of variance per participant on this measure.
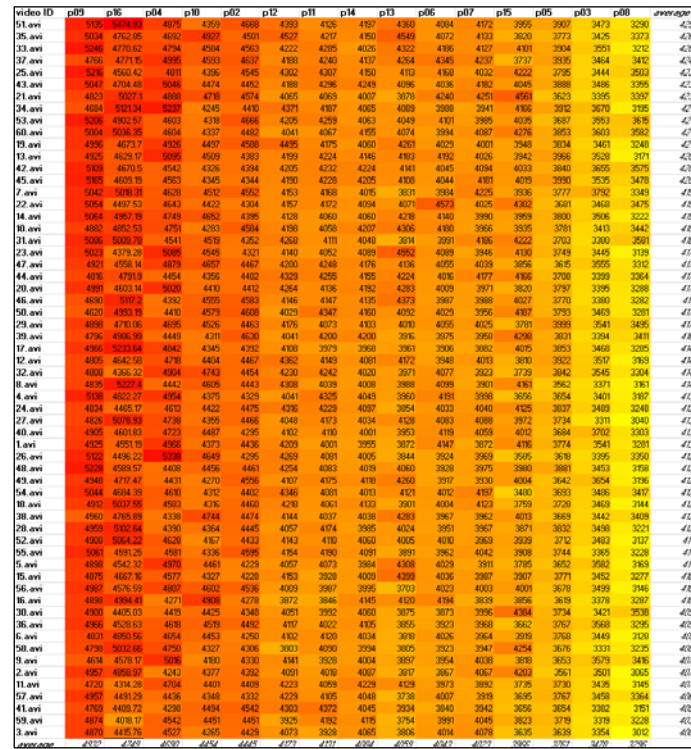

Figure 7: Heat map visualization of pupil size data in a combination sorted arrangement for all participants (i.e., left-to-right largest-to-smallest average pupil size) crossed with all driving video clips (i.e., top-to-bottom largest-to-smallest average pupil size).

Fixation Duration. An apparent decrease in average fixation durations was found across the low to high complexity level driving videos (Figure 8). A one-way within subjects ANOVA revealed a significant effect of complexity level category on fixation duration ($F(3,42) = 18.2$, $p < 0.001$). Post hoc comparisons showed significant decreases between Level 0 versus Level 3, Level 1 versus Level 2, and Level 2 versus Level 3.
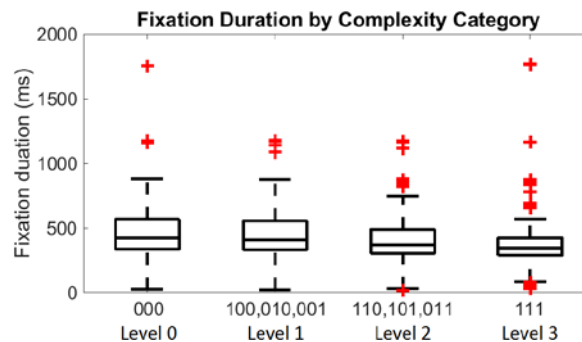

Figure 8: Eye tracker measured fixation durations grouped by driving video complexity level.

Saccade Amplitude. An apparent increase in saccade amplitude was found across low to high complexity level driving videos (Figure 9). A one-way within subjects ANOVA revealed a significant effect of complexity level category on saccade amplitude ($F(3,42) = 68.0$, $p < 0.001$). Post hoc comparisons indicated that mean ranks of saccadic amplitude for the Level 3 complexity category were significantly higher than all of the other complexity categories, as well as Level 2 versus Level 0.
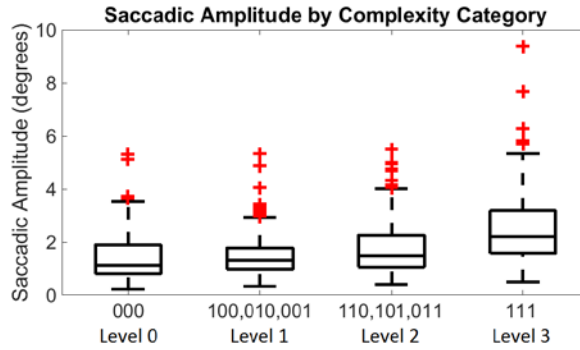
Figure 9: Eye tracker measured saccade amplitude grouped by driving video complexity level.

4. Discussion

Our utilized complexity level groupings of visible driving scene components in terms of specific factors of road straightness/curvature, presence/absence of other vehicles, and a broader factor of behavioral/infrastructure aspects appear to contain a meaningful distinction as evidenced by subjective rating responses to the question "How much effort for you to take control and drive within that segment?" Figure 5 presents a compelling picture of monotonically increasing medians along with the complexity factor level groupings with statistically significant differences found between all six possible level comparisons.

We did not find an increase in pupil size convincingly corresponding with the increase in complexity level and effort ratings that might be suggested from [11]. Potentially confounding illumination effects were not presently controlled for between the different driving scenes. Additionally, pupil size was measured here only as a single average over a full 3 second duration, in contrast to the various resolutions and time spans found within the analyses of the reviewed pupillometry studies [11]. However, a distinction in approach and motivation most likely best accounts for our differences in results. The previous pupillometry studies tended towards use of extraneous activities ("secondary cognitive tasks were used to induce mental workload during driving and recorded pupil dilation" [11]), while our question pertained to aspects of the driving scene itself alone that might connote more or less participant perceived effort. Similarly, we did not find a presumed increase in fixation duration with increased driving scene complexity, but instead evidenced a statistically significant reverse effect. While prolonged fixations are often indicative of increased difficulty in information interpretation (and just as with our pupil size results), we did not evidence increased cognitive effort with the increasing driving scene complexity but instead something else underlying the increased effort ratings. Likewise, instead of a presumed decrease in gaze variability (cf. "tunnel vision") effect of increasing complexity, we found significant increases in saccadic amplitude.

Simply put, our complexity factors contain driving-relevant visual attention demands (e.g., more to look at and assess) on a more immediate sensation/perception level irrespective of subsequent higher cognitive/mental demands. Importantly, participants subjectively experienced (i.e., felt and expressed) such differences in their effort response ratings and this has additionally been captured in the eye data (albeit in different directions than traditional driving mental workload assessment approaches and results may suggest and would hypothesize). Lastly, per our complexity factor grouping scheme, and in line with established driving theoretical accounts, we found such demands on drivers to correspond with more weight on visual aspects of the ego-vehicle oriented lower level immediate control than merely an increase in visual information per se.

There are important considerations in common across our measures that should be taken into account. Presently (due to difficulties in valid/reliable manual human annotation), a potentially confounding effect of ego-vehicle speed was not yet controlled/characterized across the videos, and we recommend such an aspect as an interesting mediating or independent factor to investigate in future studies. Furthermore, our participants were rapidly deposited into assessing a driving scene from which they were just prior completely removed from and not a part of. Such a truncated exposure of course does not reflect present day naturally contiguous driving experiences. For these reasons, we believe our results may show particular value more towards vigilance monitoring around transitions of control for example in highly automated driving domains wherein drivers are allowed to remove themselves from driving (allowing for reading, viewing movies, sleeping, etc.) until points where they re-uptake control. In regards to driver state monitoring in lower levels of automated driving (i.e., assistive and manual control), drivers are assumed to be more continuously engaged in the driving and on account of aspects like memory and expectation decreasing uncertainty, their perceived effort may not be as high as visual driving demands might expect, and assessments of their eye measures should might thus be appropriately adjusted.

5. Conclusions

In conclusion, the present complexity factor coding scheme refers to semantic items, in the sense that humans can evidently both explicitly label such items in reliable ways [18] and from the present study ascribe meaning to such items, specifically here in terms of perceived driving effort. Furthermore, such meaning can be extracted from eye tracking data on short time scales relevant to real-time applications and especially here shown when the driving scene is being assessed anew for difficulty from being most immediately previously detached. Further consideration should be given to exacting relations between eye tracking and driving workload/demands, because in real-time automated/assistive driving systems both human occupant(s) and computerized processes may observe the same scene at the same time. Matching the machine quantification of the situation to intuitive human judgments is expected to aid in the adherence to advisories and acceptance of automated aids. For example, in most cases if a driver chooses to pick up their phone while driving, they probably will also have made some visual assessment of the difficulty/ease of the driving scene first. If an intelligent vehicle had the same scene effort evaluation knowledge as humans, its judgements might be better qualified and consequently better accepted/trusted and ultimately of greater application utility and benefit.

References

1. NHTSA (2017), *Fatality Analysis Reporting System (FARS) Encyclopedia*. National Highway Traffic Safety Administration. Retrieved online Feb. 27, 2017 from https://www.fars.nhtsa.dot.gov/Main/index.aspx.
2. NSC (2017), *2016 Motor vehicle deaths estimated to be highest in nine years*. National Safety Council, Fatality Estimates. Retrieved online Mar. 20, 2017 from http://www.nsc.org/learn/NSC-Initiatives/Pages/Fatality-Estimates.aspx
3. Wikipedia (2017), *Advanced driver assistance systems*. Retrieved online Feb. 27, 2017 from https://en.wikipedia.org/wiki/Advanced_driver_assistance_systems.
4. NSC (2017), *Driver Safety Public Opinion Poll*, February, 2017. National Safety Council. Retrieved online Feb. 27, 2017 from www.nsc.org/NewsDocuments/2017/Driver-Safety-Poll.pdf.
5. Martens, M.H., and G.D. Jenssen (2012), *Behavioral adaptation and acceptance*. In: Eskandarian, A. (Ed.), Handbook of Intelligent Vehicles. 2013: Springer-Verslag, London, p. 117-138
6. Baxter, G., J. Rooksby, Y. Wang, and A. Khajeh-Hosseini (2012), *The ironies of automation: Still going strong at 30?* Proceedings of the 30th European Conference on Cognitive Ergonomics, ACM, Aug 2012. Edinburgh, United Kingdom
7. NHTSA (2008), *National motor vehicle crash causation survey*. Report to Congress. Report No. DOT HS 811059. Washington, D.C.: National Highway Traffic Safety Administration.
8. Sivak, M. (1996), *The information that drivers use: Is it indeed 90% visual?* Perception, 1996. **25**(9): p. 1081-1089.
9. Sierhuis, M. (2016), *Socially acceptable urban self-driving*. Presented at Future of Driving Symposium, TU Delft, Nov 2016. Delft, the Netherlands. Retrieved online Feb. 28, 2017 from https://collegerama.tudelft.nl/Mediasite/Play/5e14f4baa6ec448ea8a8007256ff040e1d? playfrom=0&player=f2d8bc5380f242c6a84e3dd837fa8b270a.
10. Ohn-Bar, E., & Trivedi, M.M. (2016). Looking at humans in the age of self-driving and highly automated vehicles. *IEEE Transactions on Intelligent Vehicles*, 1, p. 90-104.
11. Marquart, G., C.D.D. Cabrall, and J.C.F. de Winter (2015). *Review of eye-related measures of drivers' mental workload*. Proceedings of the Applied Human Factors and Ergonomics Conference, July 2015. Las Vegas, USA. Procedia Manufacturing, 3, p. 2854-2861.
12. Gibson, J.J., and L.E. Crooks (1938). *A theoretical field-analysis of automobile-driving*. The American Journal of Psychology, 1938. **51**(3): p. 453-471.
13. Engström, J., C.A. Monk, R.J. Hanowski, W.J. Horrey, J.D. Lee, D.V. McGehe, M. Regan, A. Stevens, E. Traube, M. Tuukkanen, T. Victor, and C.Y. Yang (2013), *A conceptual framework and taxonomy for understanding and categorizing driver inattention*. Brussels, Belgium: European Commission. Retrieved online Feb. 28, 2017 from https://ec.europa.eu/digital-single-market/news/us-eu-inattention-taxonomy-report
14. Michon, J.A. (1979). *Dealing with danger*. Technical Report no. VK 79-01, Traffic Research Centre of the University of Groningen, The Netherlands. Retrieved online Feb. 28, 2017 from http://www.jamichon.nl/jam_writings/1979_dealing_with_danger.pdf
15. Michon, J.A. (1985). *A critical view of driver behavior models: What do we know, what should we do?* In L. Evans and R.C. Schwing (Eds.) Human behavior and traffic safety. 1985: Plenum Press, New York, p. 458-520. Retrieved online Feb. 28, 2017 from http://jamichon.nl/jam_writings/1985_critical_view.pdf
16. Simon, H. (1996). *The Sciences of the Artificial*. 3rd ed. 1996: The MIT Press, Cambridge, MA.
17. Cabrall, C.D.D., R. Happee, and J.C.F. de Winter (2016). *From Mackworth's clock to the open road: A literature review on driver vigilance task operationalization*. Transportation Research part F: Traffic Psychology and Behavior, 2016. **40**: p. 169-189.
18. Cabrall, C.D.D., Z. Lu, M. Kyriakidis, L. Manca, C. Dijksterhuis, R. Happee, and J.C.F. de Winter (2016), *Validity and reliability of naturalistic driving scene categorization judgments from crowdsourcing*. Proceedings of the Road Safety on 5 Continents (RS5C) conference, May, 2016. Rio de Janeiro, Brazil.
19. Kahneman, D. (2011), *Thinking fast and slow*. 2011: Farrar, Straus, and Giroux, New York.
20. Lee, J.D., K.L. Young, and M.A. Regan (2008), *Defining Driver Distraction*. In M.A. Regan, J.D. Lee, and K.L. Young (Eds.) Driver Distraction: Theory, Effects, and Mitigation. 2008: CRC Press, Boca Raton, FL, USA, p. 31-40.
21. Hart, S., and L. Staveland (1988), *Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research*. Advances in Psychology, 1988. **52**: p. 139-183.
22. Conover, W.J., and Iman, R.L. (1981). Rank transformations as a bridge between parametric and nonparametric statistics. *The American Statistician*, **35**(3):, p.124-129.