Eye Tracking in the Wild: the Good, the Bad and the Ugly

Otto Lappi
University of Helsinki

Modelling human cognition and behaviour in rich naturalistic settings and under conditions of free movement of the head and body is a major goal of visual science. Eye tracking has turned out to be an excellent physiological means to investigate how we visually interact with complex 3D environments, real and virtual. This review begins with a philosophical look at the advantages (the Good) and the disadvantages (the Bad) in approaches with different levels of ecological naturalness (traditional tightly controlled laboratory tasks, low- and high-fidelity simulators, fully naturalistic real-world studies). We then discuss in more technical terms the differences in approach required "in the wild", compared to "received" lab-based methods. We highlight how the unreflecting application of lab-based analysis methods, terminology, and tacit assumptions can lead to poor experimental design or even spurious results (the Ugly). The aim is not to present a "cookbook" of best practices, but to raise awareness of some of the special concerns that naturalistic research brings about. References to helpful literature are provided along the way. The aim is to provide an overview of the landscape from the point of view of a researcher planning serious basic research on the human mind and behaviour.

Keywords: Eye tracking methods, naturalistic studies, simulators, oculomotor events, gaze behaviour, AOI methods, fixation, frames of reference, conceptual issues

Introduction

Modelling human cognition and behaviour in rich naturalistic settings and under conditions of free movement of the head and body – "in the wild" – is a major goal of visual science and experimental brain research. Understanding complex behaviour in information-rich real 3D environments – such as driving, aviation and sports – requires a highly interdisciplinary effort. Developing explicit computational models of the motor patterns – and their underlying neurocognitive basis – requires combining methods from behavioural and brain sciences, engineering, and computer science, in addition to the more traditional experimental psychology approach. The methods and theories have applications in engineering, ergonomics, entertainment, and education.

Here, eye tracking has turned out to be an excellent physiological means to investigate the sensory, motor and cognitive processes involved in our interactions with the real world. Eye movements provide a useful window into the workings of the nervous system, not least because in eye movement studies subjects can be engaged in tasks involving eye–hand coordination (e.g. tool manipulation), social interaction, and even locomotion (either on foot or in a vehicle). Thus, integrative visual function can be observed in a natural ecological context, which is generally not the case with, say, brain imaging methods such as fMRI, or basic neurophysiological methods such as single-cell recording.

This means that eye tracking methods are ideally suited for taking experimental behavioural research outside of the lab and into the real world, while still maintaining high standards of rigorous and precise measurement. This is important, because it has long been acknowledged that excessive focus on confined experimental designs, based on strictly controlled but potentially unnatural or uninformative stimuli and responses, can hamper theory development in psychology and cognitive science (Newell, 1973; Neisser, 1976; Broadbent, 1991).

Relatively inexpensive measuring technologies (physiological sensors, positioning equipment) as well as large localization datasets are available both commercially and in the open source/open data domain. High-fidelity dynamical and rendering simulation models suitable for creating immersive 3D virtual environments are also available, both as open source projects and commercially. However, no "off the shelf" solutions exist for integrating

these data sources into computational models of behaviour – let alone automatic algorithmic solutions for operations relevant to addressing research questions in the behavioural and brain sciences. Innovative research ideas and methodological development are still necessary to take advantage of the opportunities presented by the available technological developments.

With mobile measuring equipment becoming ever more inexpensive and widely available, the past 25 years have seen a proliferation of studies that venture out of the laboratory and "into the wild", to study human visual behaviour in naturalistic settings and outside the restrictions and confines of traditional laboratory experiments. This line of research has led to important insights into the visual strategies humans use in coping with the complexity and ambiguity of real-world tasks (for reviews see Steinman, Kowler & Collewijn, 1990; Regan & Gray, 2000; Land, 2006, 2007; Tatler et al., 2011).

Naturalistic research is necessary to determine which of the many possible visual strategies made possible by the flexibility of the human oculomotor system are actually used in a task, and what roles eye movements serve in these strategies. On the other hand, controlled laboratory experiments can reveal the internal workings of oculomotor mechanisms at a level of physiological detail that is not attainable in a naturalistic setting. But this comes at the cost of restricting the behavioural context to much simplified sensory and motor tasks, and often imposing a rather artificial trial–structure. These approaches therefore complement, rather than compete with, one another.

This review takes a philosophical look at the advantages (*the Good*) and the disadvantages (*the Bad*) in approaches with different levels of ecological naturalness (low- and high-fidelity simulators, fully naturalistic real-world studies). We also look at the methodological pitfalls (*the Ugly*), and how the unreflecting application of lab-based terminology, methods and tacit assumptions may result in poor experimental design or even spurious results.

The paper is written from the point of view of a researcher or a team wanting to implement the available methods to do basic research on the human mind and behaviour. The idea is not to present a "cookbook" of things to do, or even a roadmap of steps to take. Many of the themes are sufficiently complex to warrant a careful review in their own right, and the danger with default solutions or even heuristic rules of thumb is that they become enshrined as "best practices" that may be applied without sufficient consideration and forethought. The paper should be considered more as a tool for building up one's mental checklist of things to consider, in order to make an informed choice when one is weighing one's options on the level of ecological naturalness in the eye tracking setup and experimental design. Is it better to go for maximum control and clarity of analysis, at the expense of ecological naturalness and generalizability? Or should one do a field experiment, so that one can be confident what one observes is more or less what happens in natural conditions in the real world? (But where limitations in analysis methods and experimental control mean that one may not understand what *is* happening as clearly). There is no one correct way to go about this, and in reality a compromise must be struck between maximal control or maximal ecological validity. This review is written in part to raise awareness of some of the special concerns that doing naturalistic research brings about.

We start off by looking at the advantages and the disadvantages of experimental approaches with different levels of ecologicality. Then, some specific concerns about "high-fidelity simulators" (easily presumed to be "more naturalistic" and hence ecologically valid) are raised. Finally, we consider some fundamental issues that crop up when one wants to do research in naturalistic contexts. In particular, differences in the required analysis methods and conceptual approach – compared to the "received" lab–based methods, conceptual terminology, tacit assumptions and analysis methods – are discussed. The issue of defining a "fixation" as a class of gaze behaviour is examined in more detail.

Naturalistic Studies "in the Wild" vs. Laboratory Experiments (The Good and the Bad)

Much of what we know (or think we know) about the involvement of different oculomotor control circuits in complex tasks is based on extrapolating from simple laboratory experiments. These typically isolate a specific oculomotor event (OE) type, and then proceed to model the underlying circuit behaviour. The (implicit) assumption is that these OE circuits act as "modules" selected and activated in naturalistic tasks according to task demands. Many concepts, analysis methods, terminology

and assumptions (explicit or tacit) are borrowed directly from the lab-based tradition of OE classification and analysis – even when the experimental task and stimulus context sometimes go well beyond the original domain of application.

Some theoretical and methodological papers analyse the geometry and linked dynamics of the eye, the head, and the body with a good deal of sophistication, and develop methods for gaze analysis in mobile applications (e.g. Epelboim et al., 1995; Duchowski et al., 2002; Reimer & Sodhi, 2006; Munn, Stefano & Pelz, 2008; Munn & Pelz, 2009; Vidal, Bulling & Gellersen, 2011; Kinsman et al., 2012; Hayhoe et al., 2012; Diaz et al., 2013a; Larsson et al., 2014). Others unfortunately attempt to use the manufacturer-provided event detection algorithms to parse the gaze signal – perhaps a sign of immaturity of the field. When this is done unreflectingly, without careful consideration of the implications that real or simulated locomotor/head movement have on the proper analysis of gaze data (indeed the very definition of what counts as a "fixation" vis-à-vis other classical oculomotor events such as pursuit, VOR, or optokinetic reflex), then results from different studies can become difficult to accumulate.

There are both advantages and disadvantages in naturalistic task settings, compared to restricted laboratory designs. Simulators, depending on the level of visual complexity and physical fidelity, may be closer to one or the other (simulators are discussed in the next section). The individual researcher will need to weigh the importance of each of the advantages and each of the disadvantages – as well as more practical restrictions such as the availability of equipment and analysis methods – relative to the inherent interest in the research questions that could be addressed.

Some of the major advantages (*Good*) and disadvantages (*Bad*) are listed in Table 1[1]. Moving from left to right, realism increases in terms of task organization and stimulus information – but at the cost of reduced experimental control and increasing uncertainty over which stimulus information is actually used by the subject, and how. In the leftmost column, we have the typical eye movement studies in the laboratory (with tasks like reading a text, looking at pictures on a computer screen, performing visual search, or responding to geometrically simple visual targets). In the other columns, we move towards progressively "less domesticated" experimental paradigms, in simulator settings and ultimately fully naturalistic real-world experiments "in the wild".

In a lab experiment, typically the body and the head do not move and the head may be restrained with a chin rest or a bite bar. Oculomotor control in this case reduces to controlling the movement of the eyes in their sockets. The main characteristics of eye movement patterns in these conditions are fairly well established in the eye tracking literature. The canonical OE types identified in laboratory studies are fixations, (micro)saccades, pursuit movements, optokinetic nystagmus, vestibulo-ocular reflex, and vergence and their oculomotor parameters have been exhaustively researched for over 100 years (e.g. eye velocity, event duration, frequency of occurrence with different stimuli or task conditions etc.). Moreover, oculomotor circuit behaviour underlying the canonical eye movement patterns have been modelled in great detail (for reviews, see Ilg, 1997; Miles, 1997; Scudder, Kaneko & Fuchs, 2002; Sparks, 2002; Krauzlis, 2004; Martinez-Conde, Macknik & Hubel, 2004; Munoz, 2004; Angelaki & Hess, 2005; Thier & Ilg, 2005; Engbert, 2006; Collewijn & Kowler, 2008; Barnes, 2008; Martinez-Conde et al., 2009; Rolfs, 2009; Ibbotson & Krekelberg, 2011). What is more, the procedures for identifying them – nowadays increasingly by using dedicated oculomotor event detection and classification algorithms – have been codified to the point where many "off the shelf" solutions exist, bundled with eye trackers or available commercially or as open source projects.

Because the human oculomotor system provides such a large suite of movement patterns that can be quite flexibly integrated into ongoing behaviour, there are very many different possible ways humans might be using controlled gaze stabilization and gaze shifts to accomplish a given task. So we can only know from experiment which ones are actually used. (See for example Ballard & Hayhoe, 1995; Grasso et al., 1996, 1998; Pelz & Canosa, 2001; Hayhoe et al., 2003, 2012; Itkonen, Pekkanen & Lappi, 2015).

The main advantages of highly naturalistic studies are that they can reveal *what* visual cues are used (or at least fixated) in a given task, and how the sampling of visual

[1] *This Table reflects recurring themes the author has encounterd in papers and during review processes. They are probably familiar to most researchers with behavioural science methods training and experience in running experiments.*

information from the 3D scene is arranged in *time*, depending on the imminent sub–goals in each task phase (Regan & Gray, 2000; Hayhoe & Ballard, 2005; Land, 2006; Tatler et al., 2011).

Laboratory settings also have many advantages that are absent in more ecologically realistic paradigms. First, in a laboratory setting, the stimulus can be largely constructed from nothing but known physical parameters, including the ones of theoretical interest to the experimenter. Second, the task can be designed to be simple, at least potentially dependent on the chosen stimulus parameter of interest (the stimulus contains most of the available information relevant to the task). Third, behaviour is easy to express in parametric terms (e.g. reaction time from stimulus presentation). Finally, the task can be explicitly instructed, and the level of task difficulty controlled. These are all Good.

Extrapolating from laboratory experiments and simulator studies into the real world is not always as sound as one might hope, however. It is all too easy to leave the relation between the much simplified task and stimulus set–up in the experiment, and some putative real–world task at the level of intuitive analogy, or just an introductory vignette (which is of course Bad). To draw sound conclusions from laboratory/simulator findings, it is necessary to validate the assumption that the behaviour of interest is quantitatively (or at least qualitatively) similar in the experimental task and in the real world – at the level of dependent variables or specific performance measures.

Field experiments and laboratory/simulator experiments therefore need one another: field data are needed for validating laboratory (and simulator) results, and laboratory (and simulator) data are needed to test alternative mechanistic hypotheses underdetermined by data from fully naturalistic tasks.

The kind of precise control of stimulus parameters and behaviour available in laboratory studies, which is so useful to differentiate between hypotheses, is not possible in the wild. This means that at the moment field experiments can rarely identify oculomotor mechanisms or establish causal dependencies between specific stimulus variables and behaviour with sufficient rigor.

In fact, the modelling aim in most naturalistic studies is actually better characterized as attempting to (1) identify systematic pattern in behaviour (ideally using computational parameterization of the geometry of oculomotor and/or locomotor behaviour, but typically still painstaking manual frame-by-frame annotation), and to (2) identify strategies and/or stimulus parameters that are used to control this behaviour (this typically requires an accurately measured model of the stimulus environment).

In a laboratory experiment, the relevant stimulus parameters are known – because they are chosen and constructed by the experimenter. With richer naturalistic stimuli (including realistic simulators), instruction and task structure increasingly make a difference to the cue value of stimuli. Thus, uncertainty over *what stimuli the subject is actually using* increases.

In the wild, choosing what to represent about the ”stimulus” or the “environment” become the essential methodological challenges. *Parameterizing the behaviour and the stimulus in the first place,* and doing this in a way that facilitates uncovering systematicity in fragments of behaviour under the control of stimulus parameters, is a fundamental aim of modelling complex behaviour in naturalistic environments.

For example, in the context of car driving, it is evident that drivers “look where they are going” or “look at the road”. But this is unilluminating. Most studies of curve negotiation have followed Land and Lee (1994) in parameterizing gaze in terms of tangent point orientation (i.e. gaze direction samples classified by whether they fall within a threshold distance from the tangent point), interpreted to reflect strategies where the driver is “steering by the tangent point” (Land & Lee, 1994; see also Raviv & Herman, 1991; Land, 1998). Now, the generality of this strategy may be contested (for review see Lappi, 2014), and other parameterizations can reveal complementary information (Lappi, Pekkanen & Itkonen, 2013; Itkonen, Pekkanen & Lappi, 2015). The fundamental point is that progress beyond simple visual inspection of gaze overlaid on scene images (“car drivers are looking at the road”) is made by developing and refining the parametric representation of stimulus and gaze. An eye tracker can reveal *where* in the scene gaze is directed, but not what stimulus features or task goals have determined that gaze should be there, at that particular point in time. (We will be returning to this issue, and this example, later).

Table 1.
Rather than one type of research environment being superior to the other across the board, laboratory experiments, low- and high-fidelity simulators and fully naturalistic real-world experiments all offer complementary advantages ("the Good", marked as +) and disadvantages ("the Bad", marked as –).

	Laboratory	Simulator (simple)	"In the Wild"	
			Simulator (high–fidelity)	Naturalistic (real–world)
Stimulus	(–) Simple, sparse (+) Constructed from physical parameters chosen by the experimenter: parameterized *a priori*, varies along the dimensions of theoretical interest ("independent variables") (–) Usually restricted to sedentary settings (–) Information available to subject (visual cues) highly restricted (+) But the cues are known (+) Subject is isolated as much as possible from "confounding" stimuli	(–) Simpler than real world (–) Resolution/field of view limitations (+) Typically more realistic than lab stimuli (+) Constructed from physical parameters chosen by the experimenter: largely reduced to dimensions of primary theoretical interest (–) The subject may not always use the intended cues (only)	(+) Complex, rich (–) Resolution/field of view limitations (+) Constructed to reproduce physical parameters of real world. (–) Limited locomotor dynamics (–) The richer and more complex ("realistic") the stimulus, the more confounds found in natural settings are reproduced (–) The most relevant information and the required fidelity to achieve good behavioural validity is not usually known.	(+) Complex, rich (+) Full field of view of unlimited resolution (+) The stimulus is real physical world (+) Completely natural locomotor dynamics (–) Rarely known with good accuracy (instead of modelling the 3D layout of the scene or workspace, gaze is typically projected onto a scene camera image) (–) Parameterization usually not known *a priori* (–) Information available includes all the "confounds" occurring in natural settings
Task	(–) Given by instruction (–) Rarely naturalistic (require practice) (–) Discrete "events": experimenter–imposed trial structure (+/–) Repetitively performed at the experimenter's discretion	(+) Embedded in ongoing behaviour (continuous dynamic interaction with the simulation) (–) Given partly by task instruction/framing (+) Can be quite naturalistic (some training required) (+) Subtasks may be isolated and repeated at the experimenter's discretion	(+) Embedded in ongoing behaviour (+) Quasi–naturalistic (+/–) Subject to ecological task constraints (optimization strategies or heuristics adapted to real-world) (+) Subtasks may be isolated and repeated at the experimenter's discretion	(+) Embedded in ongoing behaviour (+) Naturalistic (well–learned before experiment) (+/–) Subject to ecological task constraints (optimization strategies or heuristics adapted to real-world) (–) Occurrence of (sub)tasks of interest constrained by real–world events
Behaviour	(–) Restrained movements (+) Critically depends on known stimulus features ("confounding" behaviours prevented) (–) Only simple discrete actions (e.g. button press, eye saccade) (+) Straightforward to express parametrically and epoch (e.g. reaction time from stimulus onset) (+) Eye movement physiology, and the procedures for identifying and reporting eye movement patterns well established in the literature	(–) Fully or partially restrained head movement, sedentary (–) Only simple actions (but continuous, e.g. steering) (–) Limited or minimal (simulated) locomotor kinematics & dynamics (+) Straightforward to express parametrically (but may not present clear epochs) (–) Eye movement physiology and eye tracking methods less well established	(+) Free head movement, simulated and/or real body motion (vection, moving base) (+) Complex multi–joint sequential actions (–) Many degrees of freedom, challenging to measure, model and analyse: requires sophisticated signal analysis (–) Eye movement physiology and eye tracking methods less well established	(+) Free head and body movement & locomotion (+/–) Complex multi–joint sequential actions (–) Many degrees of freedom, challenging to measure, model and analyse: requires sophisticated signal analysis (–) Eye movement physiology and eye tracking methods less well established (+/–) Eye movements cannot be considered in isolaton as oculomotor events: gaze behaviour essentially consists of head and body movement, which need to be modelled in 3D as well

The same applies to modelling the spatiotemporal organization of the behaviour itself: it cannot be trivially parameterized as e.g. response reaction times to discrete, *a priori* determined stimulus events.

Simulator Studies – the Best of Both Worlds?

Simulators are widely used as a tool for operator training in commercial aviation, maritime industries, and the military (air, ground and sea forces). The automotive industry uses simulators in driver evaluation as well as research and development of vehicle dynamics and driver assistance systems (both road car manufacturers and racing teams). In research, simulators are increasingly used as a complementary or even an alternative to doing labour intensive fieldwork.

Compared to field experiments, on the one hand, and traditional laboratory tasks on the other, simulator studies potentially combine the best of both worlds. They offer the unique potential for combining the richness of naturalistic behaviour and ecologically realistic tasks of field research with a relatively noise-free environment, highly repeatable conditions, and experimenter control of stimulus parameters. They also offer possibilities for experimental manipulation that are difficult or impossible to implement physically.

In the real world, the stimulus situation is complex, dynamic, and constantly evolving; it is not always immediately clear how the behaviour itself should be expressed parametrically, or how to determine the relevant stimulus parameters controlling that behaviour. Notably, "the stimulus" is not presented on a rigid trial-by-trial basis, but instead changes dynamically depending on the subjects' motor actions (locomotion and eye movements). These aspects of naturalistic behaviour can be captured and partially brought under experimental control when physical events are simulated in dynamically interactive virtual reality environments with realistic displays and controls. Simulators offer a relatively cost-effective alternative to fully naturalistic, physical setups – with the added benefit that the complex 3D stimulus environment need not be measured and modelled. Instead, the researcher can construct an environment, customize it to the needs of a specific research question, and manipulate it in a way that would not be possible or practical in a physical environment. However, the more complex and rich the simulation environment, the more one is presenting potentially confounding stimuli to the participants, making analysis of the results and validation of the simulator more difficult.

Maximum Realism: Good or Bad?

There is always a danger that impressionistic assessments of "realism" get substituted for experimentally demonstrated validity of a simulator as a research tool. Impressions can be swayed by a few superficial, or task-irrelevant properties (such as how naturalistic the textures look, what the angular extent of the field of view is, or whether kinaesthetic/vestibular feedback is present). For sure, these may be important features for particular applications, but introspection alone cannot establish how important they are for a particular task (or which features are the most important ones to get right), and whether they are reproduced sufficiently accurately (what are the tolerances for "sufficiently accurate" reproduction).

Realism is no substitute for validity, and therefore a high-fidelity simulator is not by default Good, and low-fidelity simulator Bad. Indeed, research on virtual environments has shown that the sense of presence ("being there") is less dependent on whether the display is visually rich and impressively rendered, and quite dependent on features such as frame rate, sound, and response rate in head tracking (the faithful replication of a number of "minimal cues", Slater, 2002). Increasing the complexity of the system may in fact increase the chance of imperfections that can shatter the illusion of presence!

The concept of realism has been analysed and developed in the literature on complex virtual reality. A difference is made between *immersion* and *presence* (Sanchez-Vives & Slater, 2005; Slater & Wilbur, 1997; Slater et al., 2009). Immersion refers to the degree of physical fidelity of sensory stimuli representing the simulated virtual environment (and the isolation of the participant from those stimuli in the real world that would be in conflict with the representation). These include the instantaneously visible display field of view (FOV), and rendering details such as correctness of the geometry, response latencies, resolution, stereoscopy, texture, lighting and frame rate. Presence, on the other hand, refers to the subjectively reported experience of "being there". This is distinct from immersion because it cannot be assessed based on the technical specifications of the system alone, it can only be assessed behaviourally. Immersion is not the only factor affecting presence: also, the motivation and engagement

of the subject, the level of naturalism in the task, and persuasiveness of the instruction given, and the framing of the task can make a big difference.

For using simulators as a *research* tool (i.e. a more controlled surrogate for real-world experimental settings), *external validity* is the most essential measure of realism, however. Like presence, external validity is different from immersion – it cannot be assessed from the technical specifications of the setup. But whereas presence is a holistic concept, referring to behaving, feeling and thinking in the VR/simulation environment "as you would in similar real-world circumstances", validity refers to more specific correspondence between specific performance measures (or physiological measures) of interest. Methodologically, there is also the difference that presence can be assessed by self-report questionnaires (asking about feelings, thoughts, physical sensations and the subjectively judged similarity of behaviour), whereas establishing external validity requires validation experiments that can show the correspondence in the real-world and simulator data (for further discussion of different types of simulator validity see Kemeny & Panerai, 2003).

Ideally, what is needed is that one should be able to demonstrate (convincingly by validation experiments) that:

1. The *relevant* variables ("minimal cues") have been reproduced with high fidelity. These are the ones that make a difference to the measures of theoretical interest, and the ones people have been shown to actually use (external validity).

2. *Spurious* variables that can be used to perform the simulator task (in the restricted simulator environment) in a different way to real-world performance have not been introduced. In other words: the cue value of stimulus variables have not been inadvertently dramatically changed.

3. In abstracting from the real environment and real task constraints, the *task analysis or priority ordering* for the participant have not been inadvertently changed in some essential way.

The more complex the simulator, the more difficult it is to validate these assumptions. High fidelity and immersion perhaps give a simulator "more realistic" face value, but can lead to problems as well. There are more variables to validate, and there may be more variables that are not reproduced with sufficient fidelity to maintain behavioural validity. For example, the physical intensity or timing to the dynamic events may be off. This may detract from the cue value of the variable, compared to the real world (a cue that is important in the real world is not used in the simulator because the information is not accurate enough). This may lead to behavioural strategies different from those used in the ecologically normal situation. Low-fidelity input may have a detrimental effect on overall performance. For example, vestibular stimulation that is subtly out of sync with other simulated events may even worsen the sense of motion, a possible source of disorientation and simulator sickness. In this case, it might actually be better if the cue were not reproduced at all.

In a complex simulation, there are also more variables, in addition to the variable(s) experimental interest that act as confounds and make the analysis of behavioural data more difficult. This actually detracts from one of the attractive properties of the simulator compared to the real world: the researcher being in control of the relevant stimulus variables.

Any simulator, however crude, will resemble real physical environments in some respects, and any simulator, however sophisticated, will likewise differ from the real-world physical stimuli in some respects. For a simulator to be a useful tool for research, the assumption must be made that some behaviour of interest is qualitatively or quantitatively similar in the simulator and in the real world, so that behaviour in the real world can be explained and predicted by behaviour in the simulator.

So, how realistic, and hence how complex "should" a simulator be? One should be wary of the tendency to view maximally realistic high-fidelity immersive simulators that reproduce the phenomenology of "being there" as being the best. While this may be the case for entertainment purposes, for doing research this is not so clear-cut. *The more complex the simulator is, the more difficult it is to validate empirically*. Likewise, the challenges in the analysis of patterns in the data become closer and closer to the difficulties in real-world studies (in particular the problems of parameterizing the complex behaviour and identifying the relevant stimulus parameters).

The richness of the stimuli and the complexity of the task is what differentiates a simulator from sparse stimuli

and simple laboratory tasks, which abstract a very restricted set of stimulus variables and behaviours for detailed study. As one moves from tightly controlled settings "into the wild", the same problems of analysis and interpretation arise – *even if the environment is virtual rather than physical.* There is thus an argument to be made that it is not probably useful to try and reproduce, in a simulator, everything as close as possible to the way it is in the in real world.

The more complex and realistic the simulation, the less one can fall back on established lab–based OE analysis methods, and instead one needs to adopt the methodological and conceptual approach typical of naturalistic studies.

Methodological and Conceptual Issues Specific to Eye Tracking "in the Wild" (The Ugly)

Compared to traditional laboratory eye movement studies, extra layers of complexity in the analysis and classification of eye movements are generated by free head movement and locomotion. This is not entirely due to the difficulty of reliable measurement, but also the more conceptual issue of relativity physical motion to the choice of a frame of reference. When analysing eye-tracking data in the wild, specifying the appropriate coordinate systems and transformations to represent the data is the key to capturing phenomena of interest.

In a sedentary laboratory task with the head fixed, the head, body and laboratory (allocentric) frame of reference are identical[2]. In contrast, when the eye, head, body and the 3D scene can all move relative to one another, complex frame of reference transformations are at the very heart of understanding the pattern of eye movements (Figure 1).

In the head-fixed condition, rotation of the eye in its socket and rotation of gaze in the 3D scene are equivalent. (This is indicated in Figure 1A by the dashed boxes and arrows for eye-in-head and head-in-world coordinate system transformations: they can be ignored when the point of vantage is fixed, e.g. by a bite bar).

However, in head–unrestrained locomotor settings (Figure 1B), changes in the eye-in-head angle (oculomotor events, OE) are no longer equivalent to gaze behaviour (GB, i.e. rotation and translation of the line of sight, the 3D vector from the point of vantage to the point of fixation). This has implications for the calibration of the eye tracking equipment (mapping the eye tracker signal to scene objects), the range of application of traditional oculomotor event detection and classification algorithms, the theoretical interpretation of the eye tracker signal, and the different ways to define "a fixation".

The choice of reference frames also becomes a major consideration for the representation of stimuli and behaviour. Should one think of stimuli as 3D objects in the allocentric scene, or bundles of visual features in the subject's visual field? Does one use a head-centred or body centred visual field? Or should one think of "the stimulus" as the image pattern on the retina (theoretically appealing, but in practice very difficult to measure)?

Likewise, should one think of "eye movement behaviour" in terms of sampling the 3D world with the point of fixation, or in terms of sampling the visual field with the point of regard? Or should one follow the lab-based definition of eye movements as rotation of the eye in the head (equivalent to POR movement in the head-centred visual field, but not in the body–centred or locomotor visual fields)? There is no one right answer to these questions, or even a general "best practice" to fall back on as a default choice. (For detailed discussion of the trigonometry involved in making the choice, see Epelboim et al., 1995; Duchowski et al., 2002; Diaz et al., 2013a).

How to Define "a Fixation" in the Wild? (And Why it Matters)

An eye tracker measures the position and orientation of the eye relative to the head (wearable eye trackers) or relative to elements in the fixed 3D scene (cameras in remote eye trackers). This gives the origin and orientation of the line of sight (gaze vector). Points of regard can be computed if the eye tracker is calibrated to a reference surface fixed to the head (wearable scene camera) or the allocentric frame of reference of the lab (a display).

2 "Frame of reference" is used here to refer to a set of reference directions that is fixed to objects or locations that maintain their spatial arrangement over time. A frame of reference can be used to represent space, i.e. as a basis for a coordinate system for representing space. Specifying a "coordinate system" requires, in addition, a distance metric and a point of origin. Therefore, the head and the laboratory can be said to have identical frames of reference, but different coordinate systems.

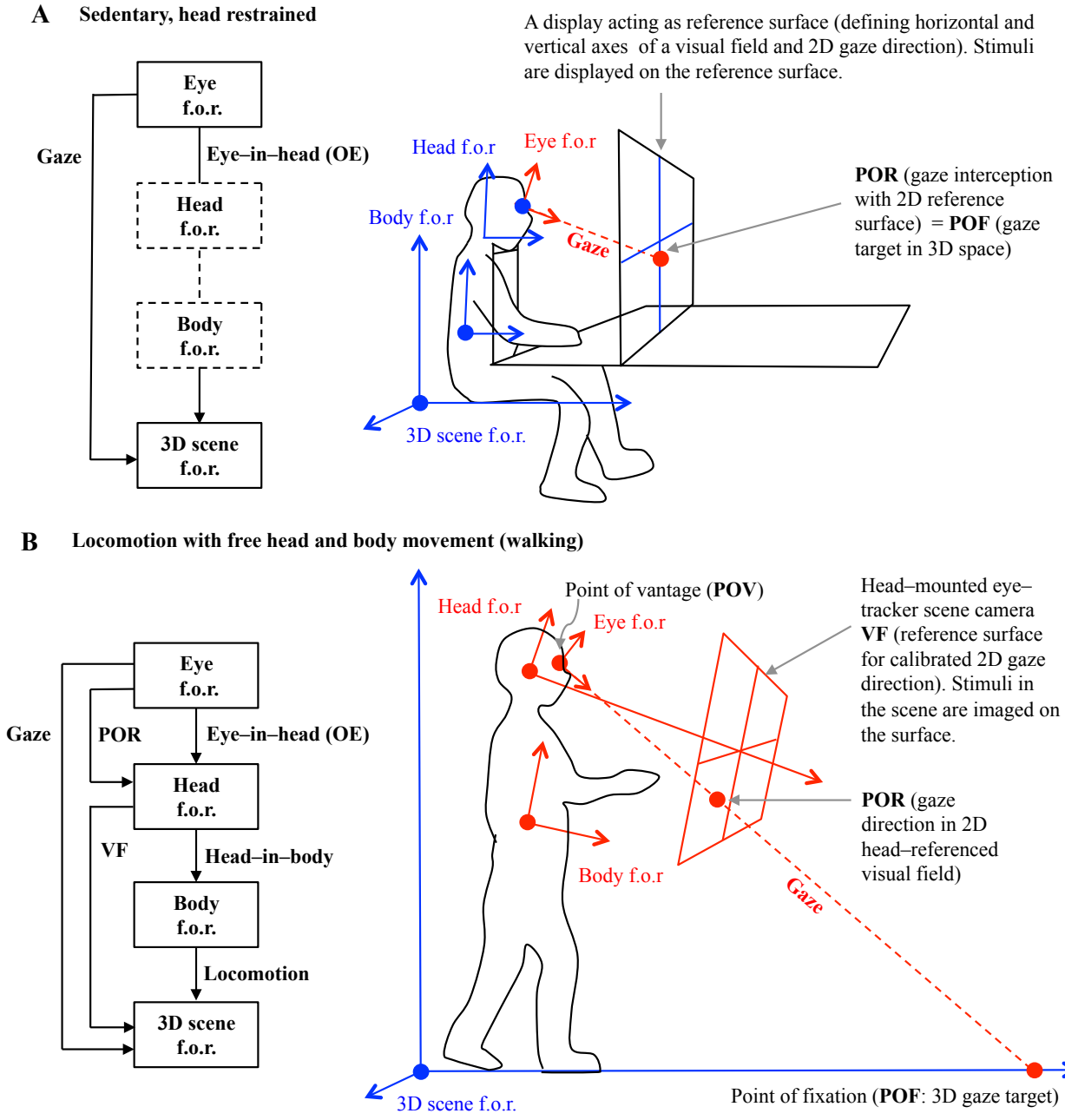

Figure 1. Descriptive terminology used to refer to eye movement patterns in different frames of reference (f.o.r.). The "moving bits" (potentially variable signals) in each case are indicated with red dots. Top: In a sedentary task with head restraint, the head, body and allocentric 3D scene frames of reference are identical. Eye position directly specifies gaze in 3D, and its projection to a a 2D reference calibration surface. Bottom: The decomposition of gaze (eye + head + body) in the 3D scene into point of regard (POR: eye) and visual field (VF: head + body). Naturalistic eye tracking using a head–mounted tracker in free locomotion. While "gaze targets" (and hence AOI's) may be identified in an eye trackers VF, determining gaze and the point of fixation (POF) in 3D requires accurate positioning of the head in the 3D scene f.o.r. (In physical settings this may be done e.g. by triangulating visible landmarks with known 3D locations in the VF image, or by using motion capture – as is required in a VR setting for updating the virtual camera position and orientation).

Typically, when the reference surface is a display screen the stimuli are geometrical patterns displayed *on* the reference surface.

For most tasks typical in eye movement research (laboratory, simulator and naturalistic alike), the most immediately striking feature of the eye tracker output signal is how periods of relative stability ("fixation events") are interspersed with rapid eye movements shifting gaze to a new location ("saccade events"). This fixation/saccade dichotomy is a natural way to set off analysing the signal, and most eye movement research is based on identifying fixations and/or saccades (or other events such as pursuit and vestibulo-ocular responses).

"Fixation behaviour" is the most commonly reported eye movement behaviour in both laboratory and simulator/naturalistic experiments. What is usually reported in laboratory studies are results based on *oculomotor event parameters*, such as fixation durations, total fixation time on target, saccade velocities or latencies, or microsaccade frequencies etc. The motivation for this approach is that fixations are considered to be of interest because they stabilize gaze relative to the stimulus, creating a time window for the acquisition of high–resolution visual information required for higher level perceptual and cognitive processing. The same rationale is usually present, explicitly or implicitly, in naturalistic studies. What is most often reported are "fixation" locations, counts, (cumulative) durations, and gaze position distribution in the scene. For example dwell times within "areas of interest" (AOI's). The eye movements themselves – what the "fixations" are like – is rarely quantitatively described.

But as "fixation" here usually refers to stability of the point of regard (at or near a visual target defining the AOI), or keeping the point of fixation at a physical object or location, it follows that insofar as head rotation or locomotion is present the oculomotor event type is actually a pursuit movement ("tracking fixation"), complemented by compensatory eye movements (optokinetic and vestibulo–ocular slow eye movements). This then implies that a very different physiological state – different oculomotor circuit activity – is involved as far as the theoretical definition of "a fixation" is concerned.

Also pertinent to the present issue is that event detection algorithms for fixation detection from eye–in–head position signal will not work: what is required is *gaze fixation* detection not *oculomotor fixation* detection.

The term "fixation" originally refers to oculomotor fixation, and under this interpretation has a definite physiological meaning: stabilizing the eye in the head. When the observer moves in relation to the environment (and the environment moves in relation to the observer) movement or stability of the eye in relation to the head does not correspond to movement or stability in relation to a visual target. Maintaining a visual target in foveal view may involve the optokinetic reflex and/or smooth pursuit when the target moves in relation to the observer and when the observer moves in relation to the target. In this case, a functionally defined "fixation" – looking at an object stationary with respect to the external world – will require a slow eye movement in the egocentric frame of reference. Gaze fixation as an eye movement class thus may consist of multiple oculomotor events: oculomotor fixation, smooth pursuit, vestibulo-ocular and/or optokinetic reflex.

As an example, consider again the case of a car driver "fixating" a point on his future path (for example a puddle on the road appearing over a crest or from behind a bend in the road). As he approaches the visual target, the horizontal eccentricity and the vertical declination of the target point change continually. Thus, a functional "fixation" that maintains the target on the fovea is actually a pursuit movement in driver centered egocentric frame of reference. Additionally, this pursuit movement corresponds in magnitude and direction to the large-scale optical flow of road texture at and around the location of interest, thus, potentially, recruiting the optokinetic reflex. Finally, VOR will stabilize gaze against perturbations caused by bumps in the road.

Localizing *gaze* in a complex 3D scene with free motion implies that instead of a reference surface stationary relative to both the 3D scene and the subject, the point of vantage and the point of fixation can be represented in a 3D model. (Objects moving in the scene, such as the participant's hand, also should be tracked and the tracking data synchronized with the eye tracker to determine points of fixation on the objects).

Gaze shifts (combined eye–head saccades: gaze shift = eye movement + head movement) and oculomotor saccades are *functionally* similar but, again, the *oculomotor characteristics* differ. The eye–in–head velocity and amplitude no longer fall on the main sequence which is the operational definition of the oculomotor saccade OE class. This is because the movement of the eye is accom-

panied by a synergistic head movement, and the OE characteristics (eye-in-head velocity) depend on the contribution of synergistic head rotation (Collewijn et al., 1992).

Thus, both the definition and identification of a "fixation" (gaze fixation, not oculomotor fixation) and "saccade" (gaze shift, not main sequence OE) need to incorporate compensatory eye movements. At a terminological level, confusion may occur when the same term is used both for stabilizing the eye in the head and for maintaining an object or location as the current target of foveal gaze. When the head and body are fixed to the 3D frame of reference these are the same thing, but when movement is free they are not. And unless this is taken into account in processing the eye tracker output into "fixations", spurious results may be generated. For example fixation duration and counts may be highly unreliable unless compensatory vestibulo–ocular and optokinetic eye movements are properly taken into account (Kinsman et al., 2012), and a tracking fixation can be a pursuit movement – possibly fast enough to be confusable with saccades on gaze velocity alone (Hayhoe et al., 2012).

In complex naturalistic settings, accurately describing eye movement behaviour or "fixation behaviour" is not as straightforward as in a sedentary head–stabilized setup, and cannot ignore the contribution from head rotation on the stability and lability of gaze. Multiple frames of reference and the intricate ways they are interrelated must be considered, and OE and gaze behaviour (3D rotation of the visual axis, or the 2D scanpath of the POR in the visual field) no longer correspond to each another.

Oculomotor Event Identification vs. 3D Gaze Behaviour

Before one can compute global variables that can be tested statistically, and given a psychological interpretation, several processing steps are applied to the raw gaze position signal from the eye tracker (Figure 2). Typically, it is partitioned it into oculomotor events drawn from a small number of different OE "types" (usually the canonical classification separating fixation, saccade, and the slow eye movements, namely pursuit, VOR and OKR). This process is often referred to as event identification.

Traditionally, event identification was done by visual inspection. Today, algorithmic methods are favoured, because they are suitable for analysing large volumes of data, and considered "objective". Nevertheless, expert visual inspection still acts as a kind of practical gold standard, and algorithm output is typically argued for by comparing the results to visual inspection (e.g. Salvucci & Goldberg, 2000, p.71, Nyström & Holmqvist, 2010, p.197; Mould et al., 2012).

It is not trivial how these stages of analysis from raw eye/gaze positions to fixations (and other events) are performed: the choices made can affect the results and theoretical conclusions one can draw (Salvucci & Goldberg, 2000; Shic, Chawarska & Scassellati, 2008; Shic, Scassellati & Shawarska, 2008).

OE identification is performed after signal preprocessing (filtering, rejection of blinks and bad data). It typically consists of sample classification (e.g. finding prospective fixations by a position dispersion threshold criterion), event detection (e.g. determining fixation onset and offset points), event rejection, and merging of detected events (e.g. combining fixations separated by "small" saccades into a fixation with longer duration, and position at the average).

Different algorithms use different eye/gaze–signal properties to detect and classify OE's. These are drawn partly from physiological properties of oculomotor behaviour established in paradigmatic laboratory tasks, partly from rules of thumb in the eye tracking literature. There is no one best set of criteria and classification rules; differences in equipment (such as sampling rates, or signal to noise ratios) and task (such as whether are movements or compensatory eye movements are present) may require different approaches.

Event identification algorithms developed for sedentary applications may use methods that depend on assumptions about the signal, and the behaviour, that are not met in more naturalistic experiments: oculomotor fixation detection is not the same thing as gaze fixation detection. Lab-based analysis methods, terminology and habits of thinking should not therefore be applied in an unreflecting way.

Dispersion based OE identification algorithms identify a sequence of gaze position observations as a fixation if they satisfy a spatial and a temporal constraint. The temporal constraint is minimum fixation duration. A fixation event is detected by comparing the spread of successive gaze position observations against a spatial threshold parameter. Different dispersion measures have been used.

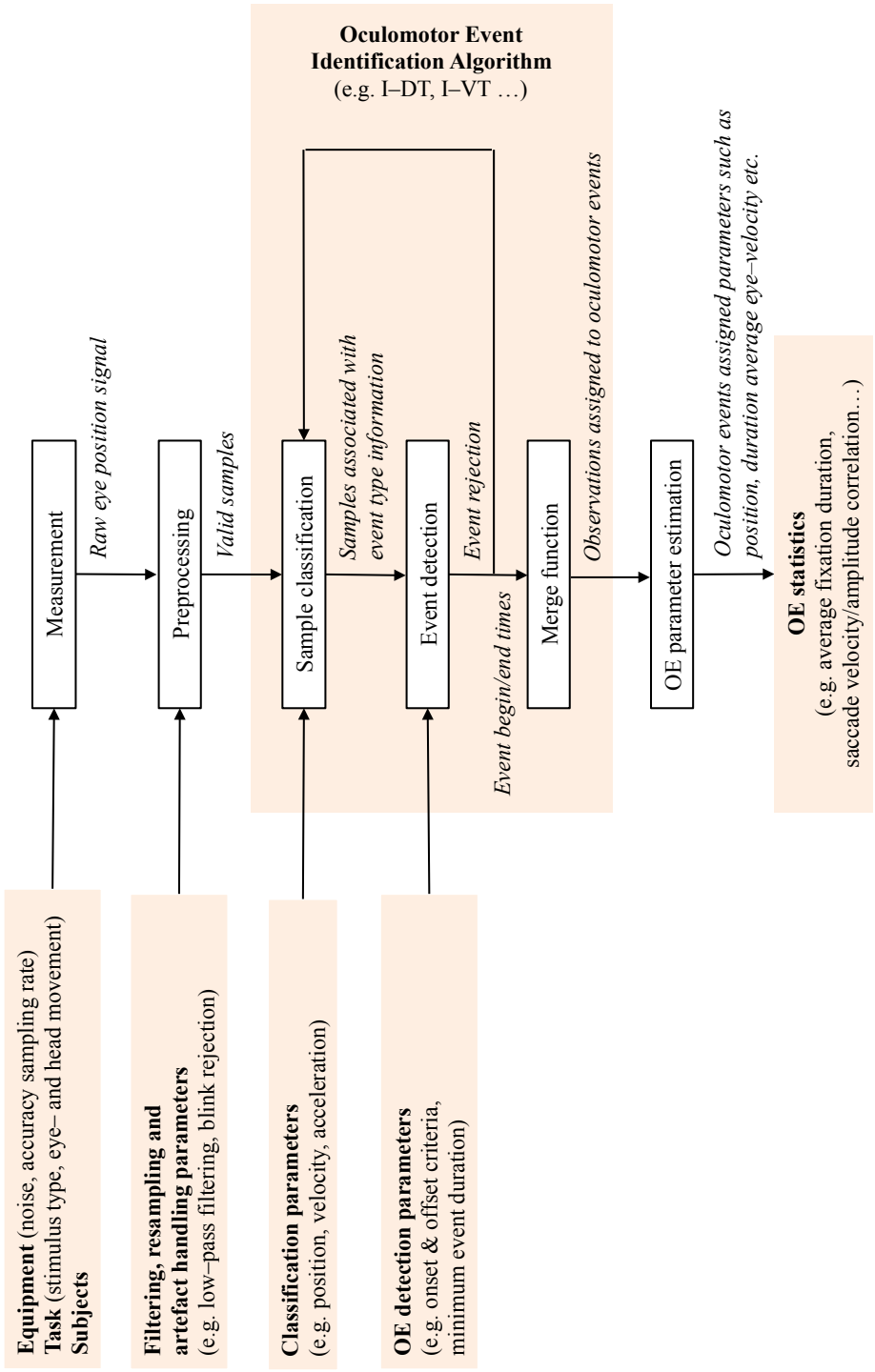

Figure 2. Oculomotor event identification workflow for the most commonly used approaches to partitioning of the eye position signal into discrete oculomotor events (OE). Several processing steps occur before OE statistics such as fixation durations or frequencies, or saccade amplitudes and velocities are computed. How the steps should be taken, and how decisions at different stages are interdependent are generally not very well established in the literature – even for laboratory tasks, let alone more complex simulator and real–world settings. I–DT: dispersion threshold identification. I–VT: velocity threshold identification.

One possible measure is maximum pairwise distance (all data points belonging to a fixation must be within a specified distance, say, 1°, from all other data points included in the same fixation event). This is computationally costly, because the number of comparisons grows exponentially with the number of data points. More efficient is to use maximum centroid distance (all data points within a putative fixation must be within a specified distance from the "moving average position" or centroid of the preceding data points in the fixation). But because the centroid shifts slightly with each new observation, this rule does not ensure that all observations within a fixation remain within any given region of space or that the pointwise distances remain within any given value. Shic, Chawarska & Scassellati (2008) consider this property to make the dispersion measure less "transparent and interpretable". Note, for example, that a smooth pursuit could be classified as a fixation using this dispersion measure, but not the other two measures.

The most often referenced I–DT algorithm (Salvucci & Goldberg, 2000) uses the dimensions of a bounding box, requiring that the sum of the distances of horizontal max and min and vertical max and min values should be below threshold. This is quite efficient, as each new observation need only be compared to the max and min values, and they are the only ones that must be kept in memory. However, the area (size) of the region of *xy* space within which the fixation points are will depend on the shape of the bounding box.

A sequence of data points satisfying the spatial constraint will only be classified as a valid fixation, however if they also satisfy a minimum fixation duration parameter. There is no single value for the minimum physiologically sensible minimum fixation duration. Values between 50 ms (Rayner, 1998) and 100–200 ms (Salvucci & Goldberg, 2000) have been used. (Nyström and Holmqvist, 2010, recommend a low value of 40 ms because they "manually identified several oculomotor fixations in the data, especially during reading, with durations below 50 ms", p.197).

Dispersion thresholding is clearly designed to identify oculomotor fixation, not gaze fixation when the subject's head or body is in motion.

Velocity-based OE identification algorithms set a minimum velocity parameter. The conceptually most straightforward way is to identify time points when the eye is moving faster than threshold: in this case the observation belongs to a saccade, otherwise it belongs to a fixation (provided a minimum fixation duration criterion is satisfied, as above). However, usually the way the velocity signal (derivative of position, difference) is used is to first identify velocity peaks, and then, if the peaks are fast enough to qualify as saccades, search for saccade onset and offset based on velocity profile (e.g. requiring that the saccade velocity profile be symmetrical), and also some reference values for eye velocity or acceleration to determine saccade launch and landing (Smeets & Hooge, 2003; Nyström & Holmqvist, 2010).

Nystrom & Holmqvist (2010) present an adaptive velocity threshold algorithm, where the saccade peak velocity threshold is set algorithmically, based on the data, rather than chosen *a priori* by the user. In the first iteration, an initial peak value is used (somewhere between 100°/s and 300°/s). The mean and SD of valid samples (below threshold) are computed. The second-pass threshold is set at six SD above the mean, and the procedure is iterated until convergence (when successive threshold values do not differ by more than 1°/s).

Unfortunately, although saccades are called "rapid eye movements" (REM), while pursuit and VOR are called "slow eye movements" (SEM), the velocity ranges overlap so that a simple velocity cut–off point cannot be used to define a saccade. Eye velocity during a saccade depends on saccade amplitude in an approximately linear manner (Robinson, 1964; Becker & Fuchs, 1969; Bahill, Clark & Stark, 1975). While large saccades are very rapid indeed (eye velocity saturates at around 500°/s), small saccades can be quite slow (10–100°/s), while "slow" eye movements can be quite fast (pursuit > 100°/s, Lisberger et al., 1981, Hayhoe et al, 2012, VOR > 500°/s, see Sparks, 2002). Thus, only saccades with a large amplitude are reliably detected by using a velocity criterion, and only in the absence of fast target or head movements eliciting pursuit/VOR.

This is a potentially fundamental problem for all algorithms that attempt to classify fixations and saccades based on the tacit assumption that during saccades the eye moves "rapidly", whereas during fixation (including fixational eye movements, FEM) and SEM the eye moves

"slowly"; including more sophisticated methods than thresholding[3].

This assumption is problematic, regardless of the sophistication of the algorithm, because the velocity signal itself simply may not contain sufficient information to positively identify OE's. And with free movement, eye–in–head/POR velocity does not determine fixation stability in 3D at all – head and body movement must be modelled as well.

Eye acceleration is used as the event detection criterion by, for example, Behrens, MacKeben & Schröder–Preikschat (2010). The rationale for using acceleration rather than velocity – even though it is a potentially more problematic signal to estimate, in terms of signal to noise ratio and filtering requirements (Nyström & Holmqvist, 2010) – is the aforementioned overlap in the velocity range of different OE types making the velocity signal inherently ambiguous. The filtered position signal is used to estimate instantaneous acceleration, and the adaptively acceleration threshold for saccade detection is computed from the acceleration signal variability (they use 3.4 sigma of the preceding 200ms, provided no saccade was detected within that window)

Using velocity or acceleration–based algorithms require differentiating the position signal. This can be a problem if the sampling rate is low (as it often is with mobile eye trackers), and because the signal needs to be heavily filtered (especially if the noise level is high), which can affect subsequent OE parameters.

As discussed, in naturalistic studies, it is also paramount that the parameters of movement (or fixation) of the eye must be explicitly based on movement in a specific frame of reference. The right choice of coordinate system is essential both for posing meaningful hypotheses and for being able to compute useful data. Changes in eye position (oculomotor events) only really make sense in the context of known frame of reference transformations.

Handling of the coordinate transformations involved in measuring and modelling eye movements in the wild present additional methodological challenges – happily, they are being addressed increasingly in the technical literature (Duchowski et al., 2002; Reimer & Sodhi, 2006; Munn, Stefano & Pelz, 2008; Munn & Pelz, 2009; Vidal, Bulling & Gellersen, 2011; Kinsman et al., 2012; Hayhoe et al., 2012; Diaz et al., 2013a; Larsson et al., 2014). This kind of methodological development is essential for advancement of the field.

Unfortunately, while many algorithms are available both commercially and open source, no generally accepted standards or guidelines have emerged, even for laboratory studies – less so field and simulator experiments – leaving each research group to tackle the same problems over and over, as they go along. It has been long recognized that the current dearth of scientifically rigorous standards and guidelines in how the identification of OE's from raw gaze position observations is performed – the choice of algorithm and parameters – is hampering progress in eye movement research (Karsh & Breitenbach, 1983; Salvucci & Goldberg, 2000; Nyström & Holmqvist, 2010; Shic, Scasselati & Chawarska, 2008; Shic, Chawarska & Scassellati, 2008; Komogortsev et al., 2009, 2010).

The quality of the signal (noise levels, sampling frequency, missing data), and the experimental setup (viewing a static image under head–stabilized conditions vs. free head movement in a dynamic setting) can interact with the choice signal analysis (filtering, artefact removal) and event detection methods in subtle and unintuitive ways. Different criteria for determining whether or not a fixation or a saccade occurs – or when it begins and ends

3 More advanced eye velocity signal processing methods – still based on the eye velocity signal as the basis for OE identification – use Markov models or Kalman filters. Instead of comparing the data–point pattern to fixed reference values directly, the events (fixation, saccade) are thought instead as (latent) states associated with different velocity distributions, and event detection is the task of estimating the underlying hidden state, given observed data. In I–HMM (Salvucci & Goldberg, 2000) there is a two–state Markov model, where each state is associated with an observation probability distribution (the probability of observing each gaze velocity value, given the state) and transition probability distribution (the probabilities of changing state and remaining in the current state). The two states are intended to represent saccade and fixations states, respectively, with observation probability distributions centered on correspondingly high and low velocity values. The algorithm classifies data points by maximizing the conditional probability of the observation sequence, given the HMM. I–KF (Komogortsev & Khan, 2007; Komogortsev et al., 2009, 2010) models the eye as a dynamical system – with the states position and velocity – going through an observation sequence and updating the state estimate based on the new observation and the previous state estimate, and a noise model. An eye velocity prediction is compared to observed eye velocity, using a $\chi 2$ test to detect saccade onsets.

– make it difficult to compare results across studies. What is more, a poor choice of signal pre–processing or event identification algorithm can create artefacts, or otherwise substantially affect the results. It is therefore worrying whenever i. identification algorithms are chosen by convenience (e.g. using manufacturer provided software packages whether or not the present), ii. algorithms are described only qualitatively and informally, iii. their fit for current purpose is argued only on the basis that some previous study has produced "meaningful" or "reliable" results with a similar method, iii. signal pre-processing performed in a nonchalant way (using rules of thumb or "commonly" used filtering methods). All three worries are regrettably often warranted.

Let us close the discussion on the technical problems (and partial technical solutions) of oculomotor identification by returning to the more fundamental question: is it conceptually or methodologically *essential* to definitely identify the OE "types" or oculomotor "fixations" in naturalistic tasks?

The original motivation for fixation identification, remember, is that fixations are considered to be periods of acquisition of high-resolution visual information because they stabilize gaze relative to the stimulus. This is true for, say, reading, but *not* for complex behaviours involving full-body self motion and target motion. In this case attempting to parse *the eye tracker signal* without taking into consideration the spatial stimulus context of *gaze control* would not work.

For example, Hayhoe et al. (2012) analysed gaze behaviour in squash. They found that about 0.2 seconds *before* the ball bounces from a wall, the point of regard is taken near the bounce location in the visual field, and after the bounce the gaze tracks the ball with a pursuit until shortly before bat contact is made. This is a visual strategy that is *not* based theoretically on OE fixation detection, or OE parameters, but consideration of gaze stability and lability in 3D and relative to specific task events.

Note also that while the *gaze* is stable the head and body are moving. This means the "fixation" is not an OE fixation, and that because the gaze arrives in the bounce area *before* the ball, there is no target feature in the stimulus array at that time (which is precisely why the authors are able make an argument for predictive control of gaze).

Problems with Using AOI Methods "in the Wild"

In practice, gaze behaviour in naturalistic studies is often reduced to AOI data, where observations are lumped into AOI's (that is, regions in the visual field where the point of regard is within some specified threshold of a putative *a priori* identified target). This does not require parsing the eye tracker signal into oculomotor events, but it does require AOI's to be decided upon. Relative gaze frequencies in different AOI's are then computed for psychological interpretation. (These are sometimes referred to as "fixation behaviour" and "fixation durations" but, to be precise, the term *glance behaviour* and glance duration or *AOI dwell time* should be used when the oculomotor/gaze behaviour within the AOI is not known). Inferences about cognitive processes or control strategies are then made based on the background theory that was used to identify the putative gaze target, around which the AOI was formed.

This is a relatively straightforward methodology and – at least with a lot of manual labour – can be performed for moving targets as well (these present dynamical AOIs that do not occupy a constant position in the visual field). The approach is, however, unsatisfactory from a couple of perspectives. First, lumping gaze position observations into AOI's loses information about the detailed pattern of movement. The method is probably used mostly out of convenience. Accurate, reliable and easy to use means to parse naturalistic eye tracker data are not well established in the field. Also, it may represent baggage from lab-based studies: in a sedentary task with very high positional accuracy an AOI method can be a good way to identify fixations (although it will lose information on fixational eye movements). In a mobile task with high noise levels (requiring larger AOI's), it is much less satisfactory.

The size of a useful AOI depends on the positional accuracy of the equipment, and this creates perhaps the biggest problem for the use of the AOI method with naturalistic settings where measurement accuracy is often less. The placement of AOI's is critical for meaningful interpretations. In particular: the relevant targets in the scene must not be so close together that the AOIs would overlap. This becomes a problem when the AOI's must be relatively large because of limited tracking accuracy, and their placement is not under the control of the experimenter.

If there are multiple hypotheses about which targets the subject may be using as cues, the AOI's derived from the different hypotheses must be non-overlapping to be able to differentiate between the hypotheses. If one only cherry-picks one or a few of the possible alternatives as the basis for defining one's AOI's and interpreting one's data, serious confirmation bias will result. Hits in an AOI cannot be interpreted as support for the hypothesis that was used to define the AOI if other potential targets, related to other, competing hypotheses, are also present in the AOI.

An example in the field of driving is the analysis of curve–driving gaze strategies. (See Lappi, 2014, for a more detailed run–down of the argument in this context, and Lappi, Lehtonen, Pekkanen and Itkonen, 2013, where the argument is worked out with on–road gaze data). On–road gaze tracking has showed that in curves drivers' gaze is quite stereotypically drawn towards the inside of the bend, into the near vicinity of the tangent point (Land & Lee, 1994). In other words, a lot of the time the POR can be found in an AOI placed around the tangent point (see replications in Chattington et al., 2007; Kandil, Rotter & Lappe, 2009, 2010).

But how much of this is due to drivers actually looking at the tangent point, and how much is to do with the tangent point just being there, geometrically close to some other point(s) on the road that the driver is steering towards? While most on-road studies (e.g. Underwood et al., 1999; Kandil, Rother & Lappe, 2009, 2010) have chosen to interpret their findings in terms of replicating the "TP orientation" result, some recent studies (Lappi, Pekkanen & Itkonen, 2013; Itkonen, Pekkanen & Lappi, 2015) have cast doubt on the tangent point as "the" point that we look when we steer through a bend. Importantly, these new results are based on analysing optokinetic pursuit parameters – not the standard AOI techniques (which can only produce ambiguous data when the alternative hypotheses predict gaze targets to fall very close to one another in naturalistic settings), or "fixation" counts or durations.

Virtual vs. Physical Settings

A virtual reality/simulator setup does not automatically solve the more fundamental problem of *how* to parameterize gaze and behaviour. But it can be useful in that in order to render the environment to the participant, it already needs to have been geometrically modelled. This is advantageous not only because a higher calibration accuracy can be achieved, but because it facilitates an algorithmic approach to modelling gaze (points of vantage and points of fixation in 3D, rather than relying on point of regard analysis – especially intensive manual frame–by–frame analysis of head camera videos is not practical for large datasets). For an instructive example, compare the aforementioned Hayhoe et al. (2012) study – where the gaze strategy was identified – to the follow-up studies (Diaz, Cooper & Hayhoe, 2013; Diaz et al., 2013b) where the behaviour is investigated in more detail using a VR version of the original task.

This only helps with solving the problem of measuring and modelling the physical environment and localizing the participant accurately (a big step, admittedly!). The problem of parameterizing the stimulus and behaviour for the purpose of answering a particular research question is a conceptual, not only technical problem.

Clever design of the experimental paradigm is still key to achieving meaningful results. As in the real world, good research cannot come from putting people into a high fidelity simulator and recording "what they look at". In a simulator this question is at least relatively well defined in terms of points of fixation/regard having coordinates in known frames of reference (objects and locations must have at least some representation before they can be presented in a simulator in the first place) – this is not often the case in real-world experiments.

Conclusion

In contrast to the typical laboratory eye tracking set-up, most of our everyday behaviour (at least for most people in most cultures) does not occur with the person stationary in one location, seated in front of a static 2D visual display. Instead, humans and animals move about, to observe a 3D scene from one point of vantage to another. Running, driving, dancing, bicycling and sports are of course "locomotor tasks" – but also cooking, infant care, tool use, and many forms of social interaction are inherently non-sedentary. If we do not understand how active movements of the body and the eye are used to update our representation of visual space (and how the resulting changes in point of vantage and gaze direction are integrated in visual perception and action) *then we will not understand vision* (Ballard, 1991; Tatler & Land, 2011).

But the real world is often a messy place. Experiments in the wild can be fraught with complexities and uncertainties that do not present themselves in controlled laboratory conditions. Thus, to understand behaviour in the real world, methodologies and arguments that rest on many kinds of assumptions that apply in the laboratory but which may not be valid in the field may need to be modified.

Simulators can offer advantages over more restricted laboratory tasks in terms of "realism". However, as the level of realism increases, the same problems confronted in the real world start to crop up (with the added problem of undetermined behavioural validity). How realistic, and hence how complex, "should" a simulator be? There may be a tendency to view maximally realistic high-fidelity immersive simulators that reproduce the phenomenology of "being there" as being the best. But while this is certainly the case for entertainment purposes, for doing research this is not so clear-cut. The more complex the simulator is, the more difficult it is to validate, the new challenges come up in the analysis and interpretation of patterns in the data. In particular, the problems of parameterizing the complex behaviour and identifying the relevant stimulus parameters

When research is taken from simple and controlled laboratory task to complex and information rich environments – real or simulated – there is a subtle change in one's philosophical approach to model building. The emphasis shifts from statistical models of relatively trivially identified parameters (such as reaction times) towards finding out the behavioural strategies and stimulus parameters people actually use when engaged in complex naturalistic tasks. As opposed to demonstrating a statistical dependence of one predefined physiological response parameter (such as saccade latency) on a controlled (*a priori* known) stimulus parameter, saccades are instead approached as gaze shifts that move foveal gaze from one target of interest to another, and fixations as periods of relative stability of the gaze in space, rather than the eye in the head. The question of interest then becomes: What role does the gaze (re)alignment with this visual feature serve in the organization of the task? What information is gleaned from this fixation? What is the significance of stabilization/retinal flow patterns that the eye movement pattern generates in the periphery? (See e.g. Regan & Gray 1999; Wann & Land 2000; Wilkie, Wann & Allison, 2008).

These are questions that we are only beginning to address – partly because of technical challenges, partly because of a preoccupation of the field with "fixation behaviour". For example, asking only when (or how often within an average "trial") foveal gaze falls within a pre–defined AOI rather than inquiring in detail in the pattern of eye movements during gaze stabilization and gaze shifts, and how these are coupled to head, hand and body movement.

An eye tracker – when it is calibrated to a reference surface or a world model – can tell us *where* the subject is looking, but does not directly tell us *what* the subject is looking at. For example, in the case of predictive gaze in squash the player appears to be looking at "the wall" or "the floor", but the *real* stimulus parameter that is controlling gaze landing is the predicted future bounce point of the ball. Classifying these gaze-stability events as "looking at the wall" would be uninformative or just plain wrong. (This could happen with the use of manual video frame annotation *without having first analysed the task structure and visual strategy as a whole*). In the driving case the driver is similarly looking at "the road", and the stimulus parameters and cognitive processes involved remain to be discovered.

So, eye tracking in the wild – visually rich naturalistic settings, with unrestrained head movement and dynamically complex locomotion – cannot be pursued simply by putting an eye-tracker on the participant and meticulously recording "where they look".

References

Angelaki, D. E., & Hess, B. J. (2005). Self-motion-induced eye movements: effects on visual acuity and navigation. *Nature Reviews Neuroscience*, *6*(12), 966–976.

Bahill, A. T., Clark, M. R., & Stark, L. (1975). The main sequence, a tool for studying human eye movements. *Mathematical Biosciences, 24*(3*),* 191–204.

Ballard, D.H. (1991). Animate Vision. *Artificial Intelligence,* 48, 57–86.

Ballard, D., Hayhoe, M., & Pelz, J. (1995). Memory representations in natural tasks, *Journal of Cognitive Neuroscience*, 7(1), 66–80.

Barnes, G. R. (2008). Cognitive processes involved in smooth pursuit eye movements. *Brain and Cognition, 68*(3), 309-326.

Becker, W., & Fuchs, A. F. (1969). Further properties of the human saccadic system: eye movements and correction saccades with and without visual fixation points. *Vision Research, 9*(10), 1247–1258.

Behrens, F., MacKeben, M., & Schröder-Preikschat, W. (2010). An improved algorithm for automatic detection of saccades in eye movement data and for calculating saccade parameters. *Behaviour Research Methods, 42*(3)*,* 701–708.

Broadbent, D. E. (1991). A word before leaving. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV* (pp. 863–879). Cambridge, MA: Bradford Books/MIT Press.

Chattington. M., Wilson, M., Ashford, D. & Marple-Horvat, D.E. (2007). Eye–steering coordination in natural driving. *Experimental Brain Research*, *180*: 1–14.

Collewijn, H., Steinman, R. M., Erkelens, C. J., Pizlo, Z., & van der Steen, J. (1992). Effect of freeing the head on eye movement characteristics during three-dimensional shifts of gaze and tracking. In: (eds.) *The head-neck sensory motor system.* Oxford University Press, Oxford, pp. 412–418.

Collewijn, H., & Kowler, E. (2008). The significance of microsaccades for vision and oculomotor control. *Journal of Vision, 8*(14), 1–21.

Diaz, G., Cooper, J., Kit, D., & Hayhoe, M. (2013a). Real-time recording and classification of eye movements in an immersive virtual environment. *Journal of Vision, 13*(12), 1–14.

Diaz, G., Cooper, J., & Hayhoe, M. (2013). Memory and prediction in natural gaze control. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1628), 20130064. DOI: 10.1098/rstb.2013.0064

Diaz, G., Cooper, J., Rothkopf, C., & Hayhoe, M. (2013b). Saccades to future ball location reveal memory-based prediction in a virtual-reality interception task. Journal of vision, 13(1), 1–14.

Duchowski, A., Medlin, E., Cournia, N., Murphy, H., Gramopadhye, A., Nair, S., ... & Melloy, B. (2002). 3-D eye movement analysis. *Behaviour Research Methods, Instruments, & Computers, 34*(4), 573–591. DOI: 10.3758/BF03195486

Engbert, R. (2006). Microsaccades: A microcosm for research on oculomotor control, attention, and visual perception. *Progress in Brain Research, 154*, 177–192.

Grasso, R., Glasauer, S., Takei, Y., & Berthoz, A. (1996). The predictive brain: anticipatory control of head direction for the steering of locomotion. *Neuroreport*, *7*(6), 1170–1174.

Grasso, R., Prévost, P., Ivanenko, Y. P., & Berthoz, A. (1998). Eye-head coordination for the steering of locomotion in humans: an anticipatory synergy. N*euroscience Lett*ers, *253*(2), 115–118.

Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behaviour. *Trends in Cognitive Sciences, 9*(4), 188–194.

Hayhoe, M. M., McKinney, T., Chajka, K., & Pelz, J. B. (2012). Predictive eye movements in natural vision. *Experimental Brain Research, 217*(1), 125–136.

Ibbotson, M., & Krekelberg, B. (2011). Visual perception and saccadic eye movements. *Current Opinion in Neurobiology, 21*(4), 553–558.

Ilg, U. J. (1997). Slow eye movements. *Progress in Neurobiology, 53*(3), 293–329.

Itkonen, T., Pekkanen, J., & Lappi, O. (2015). Driver Gaze Behaviour Is Different in Normal Curve Driving and when Looking at the Tangent Point. *PloS ONE*, *10*(8), e0135505.

Kandil, F., Rotter, A. & Lappe, M. (2009). Driving is smoother and more stable when using the tangent point. *Journal of Vision*, *9*, 1–11.

Kandil F, Rotter A & Lappe M (2010) Car drivers attend to different gaze targets when negotiating closed vs. open bends. *Journal of Vision, 10*, 1–11.

Karsh, R., Breitenbach, F. (1983). Looking at looking: The amorphous fixation measure. In: R. Groner, C. Menz, DF Fisher, RA Monty (Eds.), *Eye Movements and Psychological Functions: International views*, Erlbaum, Hillsdale, NJ

Kemeny, A., & Panerai, F. (2003). Evaluating perception in driving simulation experiments. *Trends in Cognitive Sciences, 7*(1), 31–37.

Kinsman, T., Evans, K., Sweeney, G., Keane, T., & Pelz, J. (2012). Ego-motion compensation improves fixation detection in wearable eye tracking. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 221–224). ACM.

Komogortsev, O. V., Gobert, D. V., Jayarathna, S., Koh, D. H., & Gowda, S. M. (2010). Standardization of automated analyses of oculomotor fixation and saccadic behaviours. *IEEE Transactions on Biomedical Engineering*, *57*(11), 2635-2645. DOI: 10.1109/TBME.2010.2057429

Krauzlis, R. J. (2004). Recasting the smooth pursuit eye movement system. *Journal of Neurophysiology, 91*(2), 591-603.

Land, M.F. & Lee, D.N. (1994). Where we look when we steer. *Nature*, *369*, 742–744.

Land, M.F. (1998). The visual control of steering. In: Harris, L.R. & Jenkin, M. (eds.): *Vision and action*. Cambridge university press: Cambridge. Pp. 163–179

Land, M. F. (2006). Eye movements and the control of actions in everyday life. *Progress in Retinal and Eye Research, 25*(3), 296–324.

Lappi, O. (2014). Future path and tangent point models in the visual control of locomotion in curve driving. *Journal of Vision, 14*(12), 1–22.

Lappi, O., Lehtonen, E., Pekkanen, J., & Itkonen, T. (2013). Beyond the tangent point: gaze targets in naturalistic driving. *Journal of Vision, 13*(13), 1–18.

Lappi, O., Pekkanen, J., & Itkonen, T. H. (2013). Pursuit eye-movements in curve driving differentiate between future path and tangent point models. *PLoS ONE*, *8*(7), e68326.

Larsson, L., Nyström, M., Schwaller, A., Stridh, M., & Holmqvist, K. (2014). Compensation of head movements in mobile eye-tracking data using an inertial measurement unit. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication* (pp. 1161–1167). ACM.

Lisberger, S.G., Evinger, C., Johanson, G. & Fuchs, A.F. (1981). Relationship between eye acceleration and retinal image velocity during foveal smooth pursuit eye movements in man and monkey. *Journal of Neurophysiology, 46*, 229–249.

Martinez-Conde, S., Macknik, S. L., & Hubel, D. H. (2004). The role of fixational eye movements in visual perception. *Nature Reviews Neuroscience, 5*(3), 229–240

Martinez-Conde, S., Macknik, S. L., Troncoso, X. G., & Hubel, D. H. (2009). Microsaccades: a neurophysiological analysis. *Trends in Neurosciences, 32*(9), 463–475.

Martinez-Conde, S., Otero-Millan, J., & Macknik, S. L. (2013). The impact of microsaccades on vision: towards a unified theory of saccadic function. *Nature Reviews Neuroscience, 14*(2), 83–96.

Miles, F. A. (1997). Visual stabilization of the eyes in primates. *Current opinion in Neurobiology, 7*(6), 867–871.

Mould, M. S., Foster, D. H., Amano, K., & Oakley, J. P. (2012). A simple nonparametric method for classifying eye fixations. *Vision Research, 57*, 18–25. DOI:10.1016/j.visres.2011.12.006

Munn, S. M., Stefano, L., & Pelz, J. B. (2008). Fixation-identification in dynamic scenes: Comparing an automated algorithm to manual coding. In *Proceedings of the 5th symposium on Applied perception in graphics and visualization* (pp. 33–42). ACM. DOI: 10.1145/1394281.1394287

Munn, S. M., & Pelz, J. B. (2009). FixTag: An algorithm for identifying and tagging fixations to simplify the analysis of data collected by portable eye trackers. *ACM Transactions on Applied Perception* (*TAP*), *6*(3), 16:1–16:25.

Munoz, D. P. (2002). Commentary: saccadic eye movements: overview of neural circuitry. *Progress in Brain Research, 140*, 89–96.

Neisser, U. (1976). Cognition and reality. San Francisco: W.H.Freeman.

Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In: Chase, W. G. (ed.): *Visual Information Processing: Proceedings of the Eighth Annual Carnegie Symposium on Cognition, Held at the Carnegie-Mellon University, Pittsburgh, Pennsylvania, May 19, 1972*. Academic Press. pp.283–308.

Nyström, M., & Holmqvist, K. (2010). An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data. *Behaviour Research Methods, 42*(1), 188-204. DOI: 10.3758/BRM.42.1.188

Raviv, D. & Herman, M. (1991). A new approach to vision and control for road following. Proceedings of the IEEE workshop on visual motion, 217–225.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372–422.

Regan, D., & Gray, R. (2000). Visually guided collision avoidance and collision achievement. *Trends in Cognitive Sciences, 4*(3), 99–107.

Reimer, B., & Sodhi, M. (2006). Detecting eye movements in dynamic environments. *Behaviour Research Methods, 38*(4), 667–682.

Robinson, D. A. (1964). The mechanics of human saccadic eye movement. *Journal of Physiology, 174(*2), 245–264.

Rolfs, M. (2009). Microsaccades: small steps on a long way. *Vision Research, 49*(20), 2415–2441.

Salvucci, D. D., & Goldberg, J. H. (2000, November). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications* (pp. 71-78). ACM. DOI: 10.1145/355017.355028

Sanchez-Vives, M. V., & Slater, M. (2005). From presence to consciousness through virtual reality. *Nature Reviews Neuroscience*, *6*(4), 332–339.

Scudder, C. A., Kaneko, C. R., & Fuchs, A. F. (2002). The brainstem burst generator for saccadic eye movements. *Experimental Brain Research, 142*(4), 439–462.

Shic, F., Chawarska, K., & Scassellati, B. (2008). The amorphous fixation measure revisited: with applications to autism. In *30th Annual Meeting of the Cognitive Science Society, Washington, DC*. DOI: 10.1145/1344471.1344500

Shic, F., Scassellati, B., & Chawarska, K. (2008). The incomplete fixation measure. In *Proceedings of the 2008 symposium on Eye tracking research & applications* (pp. 111-114). ACM. DOI: 10.1.1.145.8123

Slater, M. (2002). Presence and the sixth sense. Presence: *Teleoperators and Virtual Environments*, 11(4), 435-439.

Slater, M., Khanna, P., Mortensen, J., & Yu, I. (2009). Visual realism enhances realistic response in an immersive virtual environment. *IEEE Computer Graphics and Applications*, 29(3), 76-84.

Slater, M., & Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence: Teleoperators and virtual environments*, *6*(6), 603–616.

Smeets, J. B., & Hooge, I. T. (2003). Nature of variability in saccades. *Journal of Neurophysiology, 90*(1), 12–20. DOI: 10.1152/jn.01075.2002

Sparks, D. L. (2002). The brainstem control of saccadic eye movements. *Nature Reviews Neuroscience, 3*(12), 952–964.

Steinman, R. M., Kowler, E. & Collewijn, H. (1990) New directions for oculomotor research. *Vision Research*, *30*, 1845–1864.

Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision, 11*(5), 1–23.

Tatler, B. W., & Land, M. F. (2011). Vision and the representation of the surroundings in spatial memory. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1564), 596–610.

Thier, P., & Ilg, U. J. (2005). The neural basis of smooth-pursuit eye movements. *Current Opinion in Neurobiology, 15*(6), 645–652.

Underwood G, Chapman P et al. (1999) The visual control of steering and driving: Where do we look when negotiating curves? In: A.G.Gale, I.D.Brown, C.M.Haslegrave and S.P.Taylor (eds) *Vision in Vehicles VII*. Amsterdam, Elsevier.

Vidal, M., Bulling, A., & Gellersen, H. (2011). Analysing EOG signal features for the discrimination of eye movements with wearable devices. In *Proceedings of the 1st international workshop on pervasive eye tracking & mobile eye-based interaction* (pp. 15–20). ACM.

Wann, J., & Land, M. (2000). Steering with or without the flow: is the retrieval of heading necessary?. *Trends in Cognitive Sciences, 4*(8), 319–324.

Wilkie, R. M., Wann, J. P., & Allison, R. S. (2008). Active gaze, visual look-ahead, and locomotor control. *Journal of Experimental Psychology: Human Perception and Performance, 34*(5), 1150–1164.