Published: 02 June 2025



ECSEE 2025: European Conference on Software Engineering Education
*June 2 - 4, 2025*
*Seeon, Germany*

RESEARCH-ARTICLE

# A Unified Holistic Model of Visual Perception for Code Reviews

**FLORIAN HAUSER**, Regensburg University of Applied Sciences, Regensburg, Bayern, Germany

**TIMUR EZER**, Regensburg University of Applied Sciences, Regensburg, Bayern, Germany

**LISA GRABINGER**, Regensburg University of Applied Sciences, Regensburg, Bayern, Germany

**JÜRGEN HORST MOTTOK**, Regensburg University of Applied Sciences, Regensburg, Bayern, Germany

**HANS GRUBER**, University of Turku, Turku, Southwest Finland, Finland



ECSEE '25: Proceedings of the 6th European Conference on Software Engineering Education (June 2025)
https://doi.org/10.1145/3723010.3723017
ISBN: 9798400712821

# A Unified Holistic Model of Visual Perception for Code Reviews

Florian Hauser
Laboratory for Safe and Secure Systems
OTH Regensburg
Regensburg, Germany
florian.hauser@oth-regensburg.de

Timur Ezer
Laboratory for Safe and Secure Systems
OTH Regensburg
Regensburg, Germany
timur.ezer@oth-regensburg.de

Lisa Grabinger
Laboratory for Safe and Secure Systems
OTH Regensburg
Regensburg, Germany
lisa.grabinger@oth-regensburg.de

Jürgen Mottok
Laboratory for Safe and Secure Systems
OTH Regensburg
Regensburg, Germany
juergen.mottok@oth-regensburg.de

Hans Gruber
University of Regensburg
Regensburg, Germany
University of Turku
Turku, Finland
Hans.Gruber@ur.de

## Abstract

Despite being a well-structured domain, software engineering lacks standardized definitions, metrics, and theories for analyzing eye movements in domain-specific tasks. This gap can be addressed by adapting models from other fields, such as radiology and psychology. In particular, the *holistic models of image perception* provide a suitable framework for software engineering applications. This paper introduces a *unified model of visual perception* focused on eye movements during code reviews. It is based on prior research, findings from other studies, and cross-domain theories. Empirical studies on C and C++ code reviews confirm a phase-based process, where experts switch between global scanning and focal viewing. In addition, significant differences in the fixation rate, fixation duration, number of saccades, and AOI-specific metrics highlight the role of expertise in visual processing. The proposed model offers a structured framework for eye-tracking analysis in software engineering, defining relevant metrics and supporting future refinements across various software engineering tasks.

## CCS Concepts

**• Social and professional topics → Software engineering education**; **• Human-centered computing → Empirical studies in visualization**.

## Keywords

Software engineering education, visual perception, eye tracking, expertise, code reviews

**ACM Reference Format:**
Florian Hauser, Timur Ezer, Lisa Grabinger, Jürgen Mottok, and Hans Gruber. 2025. A Unified Holistic Model of Visual Perception for Code Reviews. In *ECSEE 2025: European Conference on Software Engineering Education (ECSEE 2025), June 02–04, 2025, Seeon, Germany*. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3723010.3723017



*ECSEE 2025, Seeon, Germany*

ACM ISBN 979-8-4007-1282-1/25/06
https://doi.org/10.1145/3723010.3723017

## 1 Introduction

The *holistic models of image perception* have been successfully applied for decades in fields such as medicine, psychology, and other domains to analyze the visual strategies of experts, novices, or general users in specific scenarios [13, 27, 38, 44]. Although the term *"image"* suggests their application is limited to visual or graphical stimuli, these models have demonstrated a much broader applicability [38, 44]. For instance, they have been utilized in software engineering to analyze eye movements during code reviews or to gain insights into the examination of UML diagrams [17–19].

In this context, it becomes evident that the *holistic models of image perception* can further contribute to the standardization of eye-tracking studies in software engineering [17–19, 35, 41, 42]. While software engineering in general is a well-defined domain [6, 29, 45], it still lacks consistent terminology, metrics, and methodologies in the context of eye tracking research [35, 42].

This paper aims to address the current challenges in software engineering when it comes to conducting and analyzing eye tracking studies (see section 2). In section 3 it will provide an overview of various *holistic models of image perception*, additional theories and illustrate how they can be applied within the domain of software engineering. Subsequently, key findings from previous research and own studies will be extracted in section 4, leading to the proposal of a *unified model of visual perception*, focused on code reviews. The proposed model will be discussed in detail (see section 5), including its limitations (see section 6), potential future modifications and enhancements (see section 7).

## 2 Challenges in conducting and analyzing eye tracking studies in software engineering

From a research perspective, software engineering in general can be described as a well-defined domain [22, 45]. This is historically rooted, as software engineering emerged as a solution to the *software crisis*, with the primary goal to professionalize the

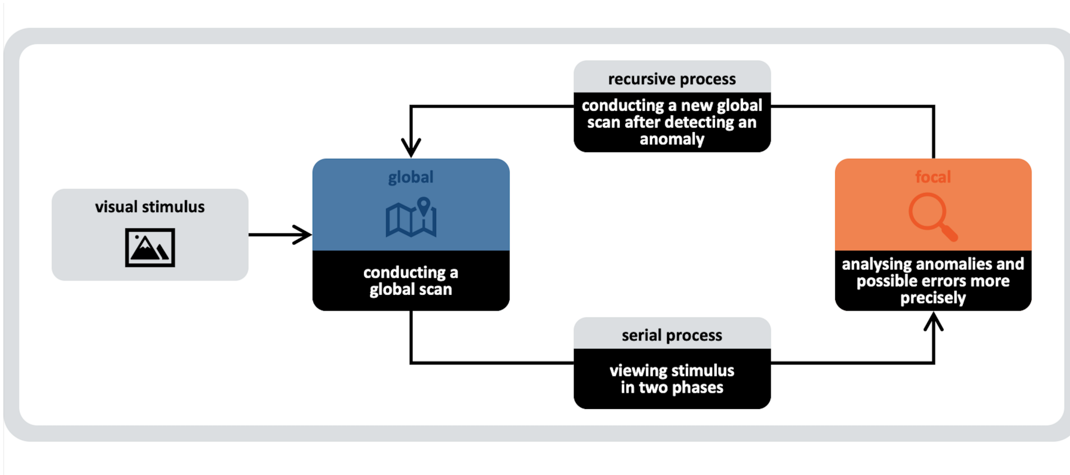


**Figure 1: Global-focal search model [31], taken from Hauser et al. [18]**

development processes and to establish commonly accepted standards [6, 29]. The intention behind this is summarized by Naur and Randell: *"The phrase software engineering was deliberately chosen as being provocative, in implying the need for software manufacture to be based on the types of theoretical foundations and practical disciplines, that are traditional in the established branches of engineering"* [29, p.13] Despite these favorable foundations, it becomes evident that software engineering currently faces a gap when it comes to eye tracking research [35, 41, 42].

This gap is extensively discussed in two systematic literature reviews by Sharafi et al. [42] and Obaidellah et al. [35]. According to them, a key challenge in applying eye-tracking to software engineering is the lack of standardization. Currently, there are no consistent terms, definitions, metrics, or methods in this field. Sharafi et al. [42] highlight this issue in their meta-study, advocating for the adaptation of established eye-tracking standards from other domains [7, 23, 38, 44]. They further emphasize the need for standardized guidelines [35, 41, 42].

Even a comparison of experimental software from providers like Tobii, SMI, and SR Research shows inconsistencies [25, 36, 37, 39], suggesting that this is a broader issue in eye-tracking research. It remains to be seen how the software engineering domain will address this challenge moving forward.

While initial attempts, such as those by Bednarik and Tukiainen [2], have been made, these are difficult to generalize. A more substantial contribution was made by Sharafi et al. in 2020 with their practical guide on conducting eye-tracking studies in software engineering, which provides a concise overview of relevant metrics [41, pp.3173-3179]. Despite this progress, the lack of standardization remains a critical issue [35, 42].

## 3 Theoretical Background

The following part of the paper will provide an in-depth description of the *holistic models of image perception* (see 3.1). This includes an examination of the global-focal search model (see 3.1.1), the two-stage detection model (see 3.1.2, and the *holistic mode vs. search-to-find* (see 3.1.3). Following this, the three models are compared to each other and similarities are presented (see section 3.2). Finally, section 3.3 will provide a detailed overview of additional theories to the *holistic models of image perception.*

### 3.1 The holistic models of image perception

The *holistic models of image perception* originate from radiology and psychology, and are typically used to examine and interpret differences between experts and novices in the viewing of visual stimuli [13, 27, 38, 44]. While they are most commonly applied to images (e.g. MRI and X-Ray scans, chess positions, or everyday situations), they are not limited to these contexts [13, 27, 28, 31–34, 38, 44, 49]. A detailed comparison of these models, including an in-depth discussion of their unique features and relevant metrics, can be found in a literature review by Sheridan and Reingold [44]. In the following sections, the most important models for code reviews are described.

*3.1.1 The global-focal search model.* The global-focal search model (depicted in Figure 1) was originally developed by Nodine and Kundel back in 1987 and has been studied and refined over the years [32, 34]. At its core, the model is phase-based and can be used to explain the approach of experts when analyzing a visual stimulus [31, 44]. In the global phase, experts perform an initial scan of the stimulus to gain an overview and absorb the information it contains. During this process, they compare the observed information with their prior knowledge, which includes prototypical representations of normal cases and abnormalities. These cognitive constructs are referred to a schema within the theory [31]. This comparison helps to identify potential abnormalities, distractors, or deviations early on [31, 33, 44]. The focal phase involves a detailed examination and analysis of the abnormalities identified during the initial scan. This phase is particularly evident in the observer's eye movements. Experts tend to focus their fovea (the region of the eye responsible for detailed vision) on the relevant areas, signaling increased attention to these regions [31, 33, 44]. Nodine and Mello-Thoms [32, p.869] describe the global-focal search model as a linear sequence but emphasize that the process can be iterative when analyzing

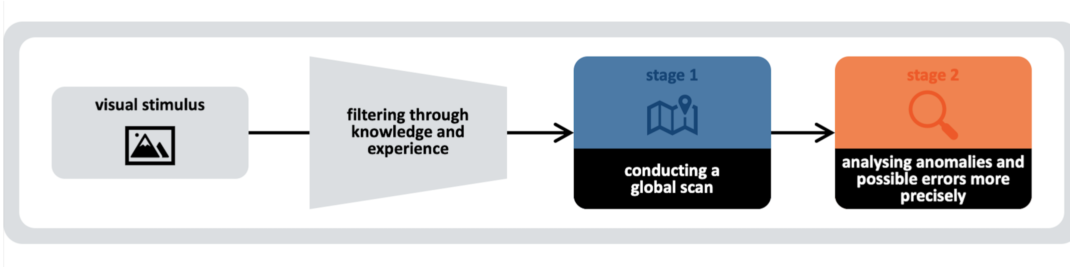


**Figure 2: Two-stage detection model [49], taken from Hauser et al. [18]**

a stimulus: *"Attention shifts back to the medical image for a new global impression flagging another perturbed region, focal analysis searches it, a new object may be regognized, and recursive testing for abnormalities continues until the observer is satisfied that enough evidence has accumulated to make a diagnostic decision."* [32, p.869]

*3.1.2 The two-stage detection model.* The two-stage detection model by Swensson [49] shares several similarities with the global-focal search model [31, 32, 34], particularly in emphasizing a holistic approach to image analysis. Both models agree that experts can quickly assess large regions of a visual stimulus by utilizing their parafoveal and peripherals vision to gain an overview. Detected abnormalities are then examined in detail using foveal vison [31, 32, 34, 49]. Like the global-focal search model [31], the two-stage detection model is based on the assumption that the analysis of a visual stimulus involves two processing stages. However, unlike the former, Swensson'smodel does not consider these processes to be recursive. According to Swensson [49], the first phase involves automatic filtering mechanisms (that are a product of a persons level of expertise) to identify and select abnormalities. In the subsequent phase, these selected abnormalities receive the observer's full attention for detailed analysis. During this analysis, the abnormalities are compared to prior knowledge. Based on this comparison, cognitive evaluation processes are conducted to determine whether, and with what level of confidence, the abnormality meets the criteria under investigation [49]. A schematic representation of the two-stage detection model can be found in Figure 2.

*3.1.3 Holistic mode vs. search to find.* The third relevant model in the context of this paper is called *holistic mode vs. search to find* [28]. It is depicted in Figure 3. Using the *holistic mode*, experts conduct a rapid yet thorough scan of the visual stimulus, enabling them to identify errors and anomalies. Areas of interest identified by using this method are subjected to more detailed examination by using the *search-to-find* approach. The authors [28] state that the two modes in this model can operate simultaneously. Global information can still be processed even when the viewer is already using *search-to-find* to analyze anomalies in-depth. The model assumes that the ability to employ the *holistic mode* is correlated with the expertise of the viewers. Novices, lacking certain knowledge and abilities, typically have to use the slower *search-to-find* approach [28, 44].

## 3.2 Similarities and metrics of the *holistic models of image perception*

The three previously introduced models share several similarities. All of them divide the viewing process into distinct phases, with certain goals: one phase serves to get an overview, while the other one focuses on a closer examination of discovered anomalies or errors.

Another common feature is their application in the study of visual expertise (see Figure 4). These models are capable of identifying and describing the strategies employed by advanced practitioners and novices [13, 17–19, 27, 38, 44]. Sheridan and Reingold have explored this topic extensively through a systematic literature review [44]. Their analysis examined various studies that applied *holistic models of image perception* across different domains and investigated which eye tracking metrics were used as dependent variables. The authors extracted the key findings of each study and presented them in tabular form, showing how these metrics change with increasing expertise [44, p.5]. A summary of their insights can be found in Table 1.

## 3.3 Additional and complementary theories to the *holistic models of image perception*

As previously outlined, the *holistic models of image perception* provide a valuable framework for analyzing and interpreting eye movements across various domains [13, 28, 31, 38, 44, 49]. However, these models exhibit certain gaps and require complementary theories to account for aspects such as underlying cognitive processes and performance differences between experts and novices.

A similar challenge was addressed in the doctoral thesis of Ellen Kok, who applied these models to study visual expertise in radiology [27]. Kok extended the holistic models of image perception by incorporating two additional theories: the *information reduction hypothesis* [15, 16] and the *theory of long-term working memory* [10, 26]. These theories help to further explain information processing and the influence of experience. Building on Kok's work [27], prior research [17] has also integrated the *cognitive load theory* [46–48]. This theory similarly aims to refine the understanding of cognitive processing and the role of expertise. The underlying theories of visual expertise for this paper are depicted in Figure 4 and will be described in the following.

*3.3.1 Information-reduction hypothesis.* The *information-reduction hypothesis*, introduced by Haider and Frensch [15, 16], suggests that experts optimize information intake by focusing on relevant

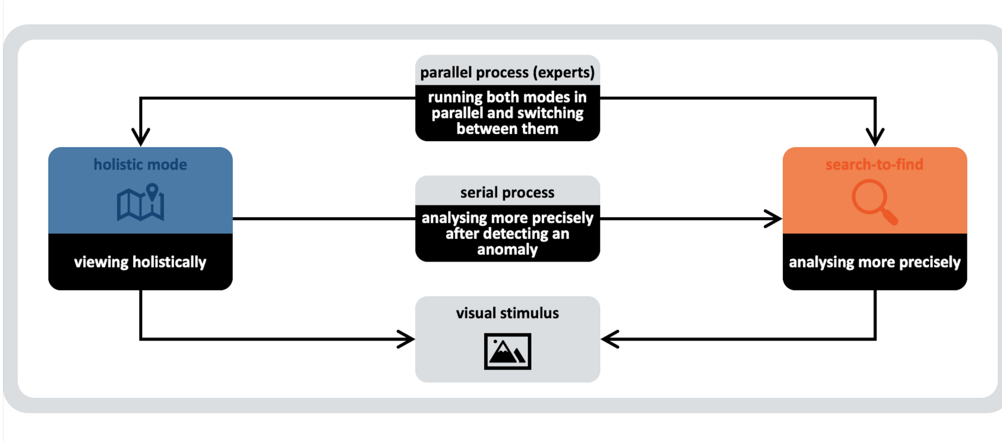


**Figure 3: Holistic mode vs. search to find [28], taken from Hauser et al. [18]**

details while ignoring irrelevant ones. Novices, in contrast, tend to be distracted by salient but unimportant features.

Empirical evidence supports this theory across various domains, particularly in medicine. For example, Kok [27] found that radiology novices fixated on irrelevant air pockets in the stomach, whereas experts ignored them. Similarly, Rubin et al. [40] showed that lung specialists scanned only 26% of lung tissue but detected 75% of malignant nodules.

Beyond medicine, research by Antes and Kristjanson [1] and Charness et al. [5] demonstrated that experienced artists and chess players concentrate their fixations on task-relevant areas more than novices.

Gegenfurtner et al. [13] summarized the theory's core assumption: experts make fewer but more targeted fixations on relevant areas while reducing fixations on redundant ones. This highlights the importance of eye tracking metrics related to attention distribution. Due to its broad applicability and domain-independent nature, the *information-reduction hypothesis* can also be applied to software engineering and code reviews.

*3.3.2 Theory of long-term working memory.* The *theory of long-term working memory*, introduced by Ericsson and Kintsch [10], explains how expertise transforms memory structures. It suggests that as expertise increases, individuals develop a more structured long-term memory, allowing them to quickly process, store, and retrieve information for future use. This enables efficient application of past experiences to new cases, effectively overcoming the short-term memory limitations [10, 13, 26, 27].

Gegenfurtner et al.[13] propose applying this theory to visual expertise research, particularly through fixation-based and AOI-based eye-tracking metrics. They argue that if experts encode and retrieve information faster than novices, this should be reflected in shorter fixation durations.

In software engineering, the *theory of long-term working memory* is relevant to code reviews. Due to their experience, experts possess extensive knowledge of coding errors, guidelines, and critical areas, allowing them to identify issues more efficiently than novices [3, 17–21, 24].

*3.3.3 Cognitive load theory.* The *cognitive load theory* focuses on the cognitive processing capacity. The theory categorizes cognitive load into three types [46–48]:

**Table 1: Sheridan and Reingolds list of relevant metrics for the *holistic models of image perception* (based on [44, p.5])**

| Eye tracking metric | Influence of expertise |
|---|---|
| *Total viewing times* | With increasing expertise, less time is required to process visual stimuli. |
| *Number of saccades/fixations* | Fewer fixations and saccades occur as expertise increases. |
| *Saccade length* | With increasing expertise, longer saccades are occurring. |
| *Scanpath length* | The length of scan paths decreases with increasing expertise. |
| *Time to first fixation on anomaly* | The time until an anomaly is first fixated decreases with increasing expertise. |
| *Proportion of time fixating relevant regions* | With increasing expertise, relevant areas are fixated for a proportionally longer time. |
| *Dwell times* | With increasing expertise, general dwell time decreases, but increases for anomalies. |
| *Fixation times* | Fixations become shorter with increasing expertise. At the same time, it is also observed that the fixation rate (number of fixations per second) increases with expertise. |

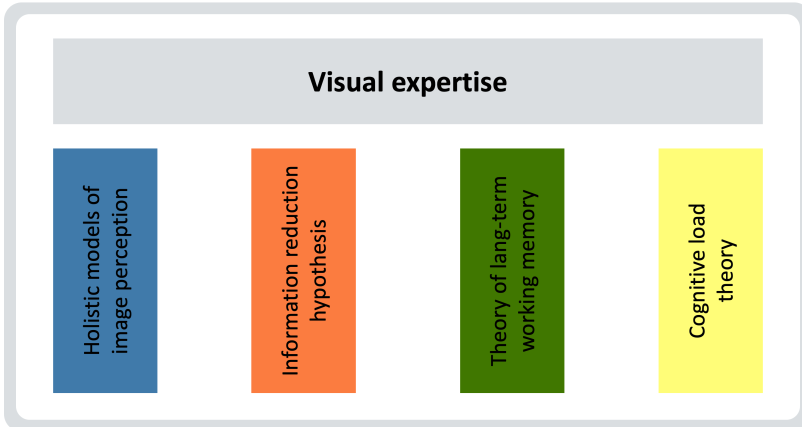


**Figure 4: Model of visual expertise, taken from Hauser [17, p.27]**

- *Extraneous cognitive load:* Caused by external distractions, such as a noisy work environment, which can hinder performance. Instructional design aims to minimize this load to free up cognitive resources for the actual task.
- *Intrinsic cognitive load:* Determined by the complexity of the task itself, which cannot be reduced but can be managed through expertise development, allowing experts to apply heuristics and automatism more efficiently [8, 12, 46–48].
- *Germane cognitive load:* Represents the cognitive effort invested in solving the task. The goal is to maximize this load while minimizing extraneous cognitive load for optimal performance.

In the context of code reviews and visual expertise, the *cognitive load theory* provides insights into stimulus complexity. The interaction between *intrinsic* and *germane cognitive load* can highlight differences between experts and novices, offering a deeper understanding of how domain-specific experience influences review performance.

## 4 Methods

The following section 4 explains the methodology used to develop the *unified model of visual perception* for the application in code reviews. In a first step (see section 4.1), related studies from software engineering are analyzed and it is examined to which extent the domain already provides potential solutions. Next, it outlines which components of the models are suitable for the analysis and interpretation of eye movements during a code review and how they complement the existing concepts (see section 4.2). Finally, findings from former studies that applied these models in the context of code reviews are incorporated, highlighting the identification and verification of relevant metrics (see section 4.3).

### 4.1 Existing concepts in software engineering

In software engineering, a systematic application of *holistic models of image perception* or of similar approaches has yet to be established. The only exception so far is the prior work to this publication (which will be discussed at section 4.3). However, some surprisingly similar concepts can be found in the domain.

The work of Uwano et al. [50] investigates how programmers conduct code reviews. They observe that participants initially scan the source code to gain an overview, reviewing approximately 73% of the code during the first 30% of the review time. They refer to this behavior as a *scan pattern*. In 2012, the study by Uwano et al. [50] was replicated by Sharif et al. [43], who reported rather similar findings. They also confirmed that during the initial phase of a code review, participants aim to develop an overview of the code under examination.

Expanding beyond the initial phase, the work of Begel and Vrzakova [3] provides insights into the review process as a whole. Their study revels that participants switch between an initial scan of the code and more detailed analysis later in the process. They also show that this switching between global scanning and focused examination is not limited to the early stages but occurs repeatedly throughout the entire review process.

In a slightly different yet relevant direction for this paper, the work of Busjahn et al. [4] provides important insights. Their research focuses on the reading behavior of source code, demonstrating that it differs significantly from reading natural text. Source code is read in a far less linear fashion, characterized by a distinct non-sequential and fragmented pattern of attention. This jumping behavior can not be observed to the same extent in the reading of natural texts.

### 4.2 Application of the holistic models of image perception

A closer comparison of the findings presented in the previous section 4.1 with the *holistic models of image perception*, reveal significant similarities [3, 4, 28, 31, 43, 49, 50].

In software engineering, it is well understood that reviews can be separated into different phases, each associated with characteristic eye movement patterns [3, 17–19, 43, 50]. The transitions between global and focal viewing described by Begel and Vrzakova

[3] closely align with the global-focal search and two-stage detection models [31, 49]. Additionally, there is consensus within the domain that reading source code bears little to no resemblance to reading natural text. This fundamental difference results in unique eye movement patterns specific to code comprehension [4, 17–19, 21, 30, 35, 42].

### 4.3 Results from own studies

In prior studies to this paper, the *holistic models of image perception* are applied to analyze eye movements during code reviews in C and C++ [17–21]. Despite being separate studies, both experiments are using a rather similar empirical design and serve as a key component for the *unified model of visual perception* described in section 5. In both studies, experts (with at least five years of programming experience) and novices (with basic knowledge of the relevant programming languages) are compared in their code review performance. Eye movements are recorded as participants asses both correct and faulty code samples, providing evaluations and error descriptions. The C study presents data from a sample of 24 participants (18 novices, 6 experts) [19, 20], while the C++ study includes 34 participants (18 novices, 16 experts) [18, 21].

In both studies, the eye-tracking metrics presented by Sheridan and Reingold in their literature review [44, p.5] are used as dependent variables. Unlike other studies, these experiments are making use of a phase-based approach, which allows a temporal analysis of eye-tracking metrics and their behavior during the different phases of the code reviews. Inspired by Uwano et al. [50], the reviews are divided into three equal phases based on the individual completion time (Phase 1: 0-33%; Phase 2: 34-66%; Phase 3: 67-100%). The changes in the relevant metrics between the phases are analyzed for both the C and C++ studies using a series of Friedman Rank Sum Tests to evaluate their statistical significance. The results and a detailed description for each metric are provided in Table 2 and 3.

The metrics presented in Table 2 and 3 indicate that eye-tracking measures fluctuate throughout the code review rather than remaining constant. While not all results reach statistical significance, the observed trends align with the patterns described by Sheridan and Reingold [44]. Both the C and C++ studie [18, 19] suggest that the relevant eye-tracking metrics follow the principles of the *holistic models of image perception* [28, 31, 49].

The combined metric changes support a phase-based approach [17–19]: In line with the global-focal search model [31–34] and two-stage detection model [49], the review process begins with a global scan of the source code, during which errors and anomalies are identified. This is followed by a focal phase, where these issues are examined in more detail. Especially the behavior of the *fixation rate* in both studies supports this assumption. It starts with a relatively high number of fixations per seconds and decreases during the experiment, which can be seen as an indicator for gloabal and focal viewing [31, 44, 49]

Raw data from both studies further indicate that these phases occur recursively, repeating multiple times throughout the review. While this recursion is already mentioned in the global-focal search model [31], it appears more frequent in code reviews [3, 18, 19, 30], creating challenges in data analysis. Each phase transition introduces noise, which impacts individual phase metrics and weakens statistical significance.

## 5 The unified model of visual perception

Based on the findings from the C and C++ studies [17–21], expertise research [8, 9, 11–14], and in alignment with the results and theories from other researchers [3, 4, 10, 15, 16, 26, 27, 43, 46–48, 50], a *unified model of visual perception* for code reviews is created and described in the following section 5.

The *unified model* for interpreting eye movements during code reviews is largely based on the global-focal search model [31] while incorporating participants' prior knowledge and expertise, as emphasized by Swensson [49]. The model, depicted in Figure 5, combines the global-focal search model [31] with the two-stage detection model [49] and complements them with additional insights, such as the information-reduction hypothesis [15, 16] and the theory of long-term working memory [10, 26].

The *unified holistic model* emphasizes domain-specific prior knowledge, enabling experts to identify vulnerable sections of code more effectively and direct their attention in a strategic way [15, 16]. Prior knowledge also supports global and focal viewing during the review process by facilitating the classification of code, using experiences from former cases, and by supporting the validation of the results [10, 26]. In addition, experts can use their skills and knowledge to provide detailed error descriptions that are valuable for other reviewers and programmers [17–21, 30].

In summary, the *unified holistic model* is based on the *holistic models of image perception*, expertise research and additional theories. While currently being focused on code reviews, it provides a foundation for broader applications, such as UML in software engineering. Future research will validate and refine the model in other domains.

## 6 Limitations

Since the *unified model of visual perception* [17] is based on a transfer of different theories from other domains, it still faces limitations in its current version when applied to software engineering (especially in the context of code reviews). It should be considered that image perception and reading comprehension rely on fundamentally different cognitive processes, even within the context of natural language. Analyzing an X-ray scan differs significantly from reading, particularly when it comes to understanding source code [4]. The latter requires lexical and linguistic processing mechanisms that cannot be directly generalized from image perception [3, 4, 17–19]. This means that cognitive processes during a code review cannot be fully explained by using existing models from image processing [17, 44]. Furthermore, while the current approach shows promise, it does not completely resolve all the identified challenges. Given the wide variety of artifacts in software engineering [35, 42], the model requires topic-specific metrics and further refinements. These aspects were not sufficiently addressed in the foundational work of the *unified model* and will require further investigation in future studies. [3, 4, 17–19, 35, 38, 41, 42, 44]

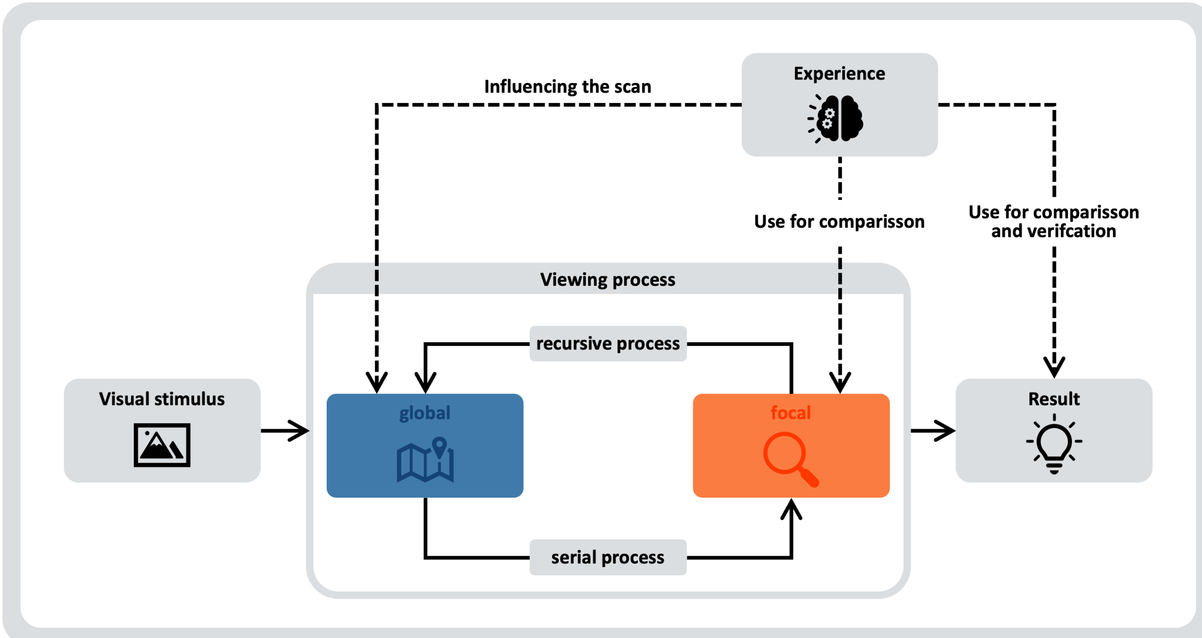


**Figure 5: Unified model of visual perception, taken from Hauser [17, p.227]**

While the metrics proposed by Sheridan and Reingold [44, p.5] already provide valuable insights into relevant eye movements during code reviews, further adjustments and extensions are necessary in this area (and in software engineering in general, when it comes to eye tracking) [17, 35, 41, 42]. Additionally, it should be considered that many eye-tracking studies in software engineering rely on artificial examples, such as short and highly simplified code snippets or unrealistic scenarios [3, 4, 17, 35, 41–43, 50]. At the same time, sample sizes are often small (approximately N = 6–30), raising questions about the generalizability of the findings [35, 42, 50]. Moreover, the proposed metrics and their behavior during domain-specific tasks need to be explored in greater detail [17]. The goal should be to establish a robust basis for interpreting eye movement data within the domain of software engineering. The literature review by Sheridan and Reingold [44], which achieved a similar outcome for medical imaging, could serve as a benchmark for this effort. Consequently, future research should further refine both the methodological approach and the theoretical understanding of cognitive processes involved in code reviews.

## 7 Conclusion

The *unified model of visual perception* presented in this paper is an updated and extended version of the *holistic models of image perception* [27, 28, 31–34, 38, 44]. Its applicability is not limited to images as visual stimuli, and the model places a stronger focus on the role of experience in the viewing process [17]. The results of the C and C++ study, which made use of the *holistic models of image perception* and (at a later stage) the *unified model of visual perception* for the analysis of eye movement data, confirm its suitability for code reviews and indicate its potential for broader application in the domain of software engineering [17–21].

In summary, the *unified model of visual perception* addresses the gap in software engineering, as mentioned in section 2, regarding the design, analysis, and interpretation of eye-tracking studies [35, 42]. It provides an empirical foundation and corresponding metrics that can be used to plan and conduct such eye-tracking studies in software engineering [17]. In its current state, it represents an initial step that can be adapted and modified to fit different requirements in this domain. Being the result of a transfer from other domains, the *unified model of visual perception* still faces limitations (see section 6) in its current version when applied to software engineering, particularly in the context of code reviews. Future research should address these challenges and contribute to further refine and expand the model, ensuring its effectiveness across different software engineering tasks [17]. The model will benefit from additional studies that provide more eye movement data, allowing future adaptations and improvements.

## Acknowledgments

The authors wish to acknowledge the use of ChatGPT and DeepL in the writing of this paper. This tools were used for translations and to assist with improving the language in the paper. The paper remains an accurate representation of the authors' underlying work and novel intellectual contributions. We thank the funding project FH-Invest (FKZ: 13FH101IN6) run by Prof. Dr. Jürgen Mottok for providing equipment for the eye tracking laboratory and Prof. Dr. Christian Wolff from the University of Regensburg for arranging the laboratory areas. The present paper is based on former results from the EVELIN project (FKZ: 01PL12022F, project sponsor: DLR) which

**Table 2: Results of the Friedman Rank Sum Tests and Medians for Each Eye Tracking in the C study [19, p.6]**

| Eye tracking metric | Group | Differences between phases | | | Median per phase | | |
|---|---|---|---|---|---|---|---|
| | | Phase 1 vs. 2 | Phase 1 vs. 3 | Phase 2 vs. 3 | Phase 1 | Phase 2 | Phase 3 |
| **Number of fixations** | **Complete** | No | No | No | 491.000 | 455.000 | 474.000 |
| | **Experts** | No | No | No | 457.380 | 448.250 | 468.380 |
| | **Novices** | No | No | No | 534.130 | 528.130 | 544.000 |
| | Complete: $\chi^2(2)$=2.835, p=.242; Experts: $\chi^2(2)$=2.250, p=.325; Novices: $\chi^2(2)$=1.322, p=.516 | | | | | | |
| **Total fixation duration in [ms]** | **Complete** | Yes | Yes | No | 172566.100 | 168975.100 | 165037.800 |
| | **Experts** | No | Yes | No | 140756.300 | 130923.800 | 130940.800 |
| | **Novices** | Yes | Yes | No | 173846.600 | 169994.400 | 183559.800 |
| | Complete: $\chi^2(2)$=23.304, p=.000; Experts: $\chi^2(2)$=10.750, p=.005; Novices: $\chi^2(2)$=12.933, p=.002 | | | | | | |
| **Average fixation duration in [ms]** | **Complete** | Yes | Yes | No | 371.130 | 349.080 | 358.100 |
| | **Experts** | No | Yes | No | 355.530 | 361.770 | 340.890 |
| | **Novices** | Yes | Yes | No | 385.510 | 349.080 | 372.140 |
| | Complete: $\chi^2(2)$=9.652, p=.008; Experts: $\chi^2(2)$=7.750, p=.210; Novices: $\chi^2(2)$=8.133, p=.017 | | | | | | |
| **Fixation rate** | **Complete** | Yes | Yes | Yes | 2.450 | 1.160 | .790 |
| | **Experts** | No | Yes | No | 2.500 | 1.120 | .810 |
| | **Novices** | Yes | Yes | Yes | 2.440 | 1.190 | .750 |
| | Complete: $\chi^2(2)$=46.000, p=.000; Experts: $\chi^2(2)$=16.000, p=.000; Novices: $\chi^2(2)$=30.000, p=.000 | | | | | | |
| **Number of saccades** | **Complete** | No | No | No | 414.000 | 448.000 | 493.000 |
| | **Experts** | No | No | No | 379.000 | 383.500 | 352.500 |
| | **Novices** | No | No | No | 414.000 | 470.000 | 505.000 |
| | Complete: $\chi^2(2)$=.783, p=.676; Experts: $\chi^2(2)$=.750, p=.687; Novices: $\chi^2(2)$=.400, p=.819 | | | | | | |
| **Number of visits (on errors)** | **Complete** | Yes | Yes | No | 17.000 | 32.000 | 44.000 |
| | **Experts** | Yes | Yes | No | 14.000 | 28.000 | 37.000 |
| | **Novices** | Yes | Yes | No | 17.000 | 34.000 | 44.000 |
| | Complete: $\chi^2(2)$=27.714, p=.000; Experts: $\chi^2(2)$=10.903, p=.004; Novices: $\chi^2(2)$=16.933, p=.000 | | | | | | |
| **Dwell time (on errors) in [ms]** | **Complete** | Yes | Yes | No | 7931.690 | 11179.460 | 13758.470 |
| | **Experts** | No | No | No | 5551.750 | 9623.600 | 10786.920 |
| | **Novices** | No | Yes | No | 9875.860 | 11690.970 | 14277.190 |
| | Complete: $\chi^2(2)$=15.217, p=.000; Experts: $\chi^2(2)$=6.750, p=.034; Novices: $\chi^2(2)$=8.933, p=.011 | | | | | | |

was supported by the Ministry of Education and Research (BMBF) of the Federal Republic of Germany. The collection and analysis was done in the context of the HASKI project (FKZ: 16DHBKI035), also sponsored by the German Federal Ministry of Education and Research (BMBF).

## References

[1] James R. Antes and Arlinda F. Kristjanson. 1991. Discriminating Artists from Nonartists by Their Eye-Fixation Patterns. *Perceptual and Motor Skills* 73 (12 1991), 893–894. Issue 3. https://doi.org/10.2466/pms.1991.73.3.893

[2] Roman Bednarik and Markku Tukiainen. 2006. An eye-tracking methodology for characterizing program comprehension processes. In *Proceedings of the ACM Symposium on Eye tracking research and applications (ETRA)* (San Diego). ACM, 125–132. https://doi.org/10.1145/1117309.1117356

[3] Andrew Begel and Hana Vrzakova. 2018. Eye movements in code review. In *Proceedings of the Workshop on Eye Movements in Programming (EMIP)* (Warsaw). ACM, 1–5. https://doi.org/10.1145/3216723.3216727

[4] Teresa Busjahn, Roman Bednarik, Andrew Begel, Martha Crosby, James H. Paterson, Carsten Schulte, Bonita Sharif, and Sascha Tamm. 2015. Eye Movements in Code Reading: Relaxing the Linear Order. In *Proceedings of the 23rd IEEE International Conference on Program Comprehension (ICPC)* (Florence). IEEE, 255–265. https://doi.org/10.1109/ICPC.2015.36

[5] Neil Charness, Michael Tuffiash, Ralf Krampe, Eyal Reingold, and Ekaterina Vasyukova. 2005. The role of deliberate practice in chess expertise. *Applied Cognitive Psychology* 19 (2005), 151–165. Issue 2. https://doi.org/10.1002/acp.1106

[6] Edsger W. Dijkstra. 1972. The humble programmer. *Commun. ACM* 15 (10 1972), 859–866. Issue 10. https://doi.org/10.1145/355604.361591

[7] Andrew T. Duchowski. 2017. *Eye Tracking Methodology* (3rd ed.). Springer. https://doi.org/10.1007/978-3-319-57883-5

[8] Karl Anders Ericsson. 2016. Expertise and individual differences: The search for the structure and acquisition of experts' superior performance. *Wiley Interdisciplinary Reviews: Cognitive Science* 8 (2016), 1–6. Issue 1. https://doi.org/10.1002/wcs.1382

[9] Karl Anders Ericsson. 2018. *The Differential Influence of Experience, Practice, and Deliberate Practice on the Development of Superior Individual Performance of Experts.* Cambridge University Press, 745–769. https://doi.org/10.1017/9781316480748.038

[10] Karl Anders Ericsson and Walter Kintsch. 1995. Long-term working memory. *Psychological Review* 102 (6 1995), 211–245. Issue 2. https://doi.org/10.1037/0033-295X.102.2.211

[11] Karl Anders Ericsson, Ralf Thomas Krampe, and Clemens Tesch-Römer. 1993. The role of deliberate practice in the acquisition of expert performance. *Psychological Review* 100 (1993), 363–406. Issue 3. https://graphics8.nytimes.com/images/blogs/freakonomics/pdf/DeliberatePractice(PsychologicalReview).pdf

[12] Karl Anders Ericsson and Tyler J. Towne. 2010. Expertise. *Wiley Interdisciplinary Reviews: Cognitive Science* 1 (2010), 404–416. Issue 3. https://doi.org/10.1002/wcs.47

[13] Andreas Gegenfurtner, Erno Lehtinen, and Roger Säljö. 2011. Expertise Differences in the Comprehension of Visualizations: A Meta-Analysis of Eye-Tracking Research in Professional Domains. *Educational Psychology Review* 23 (2011), 523–552. Issue 4. https://doi.org/10.1007/s10648-011-9174-7

[14] Hans Gruber. 2007. Bedingungen von Expertise. *Begabt sein in Deutschland* (2007), 93–112. Issue 25. http://books.google.com/books?hl=en&lr=&id=CpmZTK3CGwMC&oi=fnd&pg=PA93&dq=Bedingungen+von+Expertise&ots=HLWbMqJXHa&sig=KruTvuA6iIaFs9uKbSF_hXiRfGM

[15] Hilde Haider and Peter A. Frensch. 1996. The role of information reduction in skill acquisition. *Cognitive Psychology* 30 (1996), 304–337. Issue 3. https://doi.org/10.1006/cogp.1996.0009

[16] Hilde Haider and Peter A. Frensch. 1999. Information reduction during skill acquisition: The influence of task instruction. *Journal of Experimental Psychology: Applied* 5 (1999), 129–151. Issue 2. https://doi.org/10.1037/1076-898X.5.2.129

**Table 3: Results of the Friedman Rank Sum Tests and Medians for Each Eye Tracking in the C++ study [18, p.6]**

| Eye tracking metric | Group | Differences between phases | | | Median per phase | | |
|---|---|---|---|---|---|---|---|
| | | Phase 1 vs. 2 | Phase 1 vs. 3 | Phase 2 vs. 3 | Phase 1 | Phase 2 | Phase 3 |
| Number of fixations | Complete | No | No | Yes | 191.000 | 198.500 | 165.500 |
| | Experts | No | No | No | 205.500 | 226.000 | 206.000 |
| | Novices | No | Yes | Yes | 178.000 | 170.500 | 115.000 |
| | Complete: $\chi^2(2)$=16.133, p=.000; Experts: $\chi^2(2)$=6.000, p=.049; Novices: $\chi^2(2)$=13.775, p=.001 | | | | | | |
| Total fixation duration in [ms] | Complete | No | Yes | Yes | 109195.000 | 108274.000 | 85681.000 |
| | Experts | No | Yes | Yes | 121227.500 | 126577.000 | 117492.500 |
| | Novices | No | Yes | Yes | 95871.500 | 93729.500 | 80500.000 |
| | Complete: $\chi^2(2)$=36.059, p=.000; Experts: $\chi^2(2)$=19.500, p=.000; Novices: $\chi^2(2)$=16.778, p=.000 | | | | | | |
| Average fixation duration in [ms] | Complete | No | Yes | No | 614.157 | 545.704 | 521.099 |
| | Experts | No | No | No | 584.987 | 486.226 | 501.678 |
| | Novices | No | No | No | 666.987 | 569.195 | 588.188 |
| | Complete: $\chi^2(2)$=9.235, p=.001; Experts: $\chi^2(2)$=3.875, p=.144; Novices: $\chi^2(2)$=5.444, p=.066 | | | | | | |
| Fixation rate | Complete | Yes | Yes | Yes | 1.589 | .748 | .316 |
| | Experts | Yes | Yes | Yes | 1.580 | .805 | .388 |
| | Novices | Yes | Yes | Yes | 1.615 | .611 | .247 |
| | Complete: $\chi^2(2)$=68.000, p=.000; Experts: $\chi^2(2)$=32.000, p=.000; Novices: $\chi^2(2)$=26.000, p=.000 | | | | | | |
| Number of saccades | Complete | No | Yes | Yes | 170.500 | 276.000 | 1341.500 |
| | Experts | No | Yes | Yes | 164.000 | 211.500 | 911.000 |
| | Novices | No | Yes | Yes | 206.500 | 371.000 | 1491.000 |
| | Complete: $\chi^2(2)$=32.548, p=.000; Experts: $\chi^2(2)$=13.500, p=.001; Novices: $\chi^2(2)$=20.310, p=.000 | | | | | | |
| Number of visits on errors | Complete | Yes | No | Yes | 4.000 | 11.500 | 7.000 |
| | Experts | Yes | No | No | 3.500 | 16.000 | 7.500 |
| | Novices | No | No | No | 4.000 | 8.000 | 5.500 |
| | Complete: $\chi^2(2)$=14.970, p=.001; Experts: $\chi^2(2)$=15.129, p=.000; Novices: $\chi^2(2)$=2.771, p=.250 | | | | | | |
| Dwell time on errors in [ms] | Complete | Yes | No | No | 5121.500 | 6657.500 | 3702.500 |
| | Experts | Yes | No | No | 2997.500 | 10182.000 | 3960.000 |
| | Novices | No | No | No | 6737.000 | 4887.000 | 3438.500 |
| | Complete: $\chi^2(2)$=7.471, p=.024; Experts: $\chi^2(2)$=7.125, p=.028; Novices: $\chi^2(2)$=4.778, p=.092 | | | | | | |


[17] Florian Hauser. 2024. *Visuelle Expertise bei Code Reviews.* Ph. D. Dissertation. University of Regensburg. https://doi.org/10.5283/epub.59753

[18] Florian Hauser, Lisa Grabinger, Timur Ezer, Jürgen Horst Mottok, and Hans Gruber. 2024. Analyzing and Interpreting Eye Movements in C++: Using Holistic Models of Image Perception. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA)* (Glasgow). ACM, 1–7. https://doi.org/10.1145/3649902.3655093

[19] Florian Hauser, Lisa Grabinger, Jürgen Mottok, and Hans Gruber. 2023. Visual expertise in code reviews: Using holistic models of image pervception to analyze and interpret eye movements. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (Tübingen). ACM, 1–7. https://doi.org/10.1145/3588015.3589189

[20] Florian Hauser, Rebecca Reuter, Ivonne Hutzler, Jürgen Mottok, and Hans Gruber. 2018. Eye movements in software engineering - What differs the expert from the novice?. In *Proceedings of the International Conference of Education, Research and Innovation (ICERI)* (Seville). IATED, 632–642. https://doi.org/10.21125/iceri.2018.1129

[21] Florian Hauser, Stefan Schreistetter, Rebecca Reuter, Jürgen Mottok, Hans Gruber, Kenneth Holmqvist, and Nick Schorr. 2020. Code Reviews in C++. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA)* (Stuttgart). ACM, 1–5. https://doi.org/10.1145/3379156.3391980

[22] Dirk W. Hoffmann. 2013. *Software-Qualität.* Springer.

[23] Kenneth Holmqvist, Marcus Nyström, Richard Andersson, Richard Dewhurst, Halszka Jarodzka, and Joost Van De Weijer. 2011. *Eye tracking. A comprehensive guide to methods and measures.* Oxford University Press.

[24] Alexander Homann, Lisa Grabinger, Florian Hauser, and Jürgen Mottok. 2023. An eye tracking study on MISRA C coding guidelines. In *Proceedings of the 5th European Conference on Software Engineering Education* (Seeon). ACM, 130–137. https://doi.org/10.1145/3593663.3593671

[25] SensoMotoric Instruments. 2017. BeGaze Manual.

[26] Walter Kintsch, Vimla L. Patel, and K. Anders Ericsson. 1999. The role of long-term working memory in text comprehension. *Psychologia* 42 (1999), 186–198. Issue 4.

[27] Ellen M. Kok. 2016. *Developing visual expertise. From Shades of Grey to Diagnostic Reasoning in Radiology.* Ph. D. Dissertation. University of Maastricht. Issue 2016.

[28] Harold L. Kundel, Calvin F. Nodine, Emily F. Conant, and Susan P. Weinstein. 2007. Holistic Component of Image Perception in Mammogram Interpretation: Gaze-tracking Study. *Radiology* 242 (2 2007), 396–402. Issue 2. https://doi.org/10.1148/radiol.2422051997

[29] Peter Naur and Brian Randell. 1968. Software Engineering. In *Report on a conference sponsored by the NATO Scince Committee* (Garmisch). NATO Scientific Affairs Division.

[30] Markus Nivala, Florian Hauser, Jurgen Mottok, and Hans Gruber. 2016. Developing visual expertise in software engineering: An eye tracking study. In *Proceedings of the Global Engineering Education Conference (EDUCON)* (Abu Dhabi). IEEE, 613–620. https://doi.org/10.1109/EDUCON.2016.7474614

[31] C.F. Nodine and H.L. Kundel. 1987. *The Cognitive Side of Visual Search in Radiology.* Elsevier, 573–582. https://doi.org/10.1016/B978-0-444-70113-8.50081-3

[32] Calvin Nodine and Claudia Mello-Thoms. 2000. *The Nature of Expertise in Radiology.* SPIE, 859–894. https://doi.org/10.1117/3.832716.ch19

[33] Calvin F. Nodine, Harold L. Kundel, Sherri C. Lauver, and Lawrence C. Toto. 1996. Nature of expertise in searching mammograms for breast masses. *Academic Radiology* 3 (12 1996), 1000–1006. Issue 12. https://doi.org/10.1016/S1076-6332(96)80032-8

[34] Calvin F. Nodine and Claudia Mello-Thoms. 2010. *The role of expertise in radiologic image interpretation.* Cambridge University Press, 139–156.

[35] Unaizah Obaidellah, Mohammed Al Haek, and Peter C.-H. Cheng. 2018. A Survey on the Usage of Eye-Tracking in Computer Programming. *Comput. Surveys* 51 (1 2018), 1–58. Issue 1. https://doi.org/10.1145/3145904

[36] Tobii Pro. 2020. Tobii Spectrum Pro. https://www.tobiipro.com/de/produkte/tobii-pro-spectrum/

[37] Tobii Pro. 2021. Tobii Pro Lab Description. https://www.tobiipro.com/siteassets/tobii-pro/products/software/tobii-pro-lab/Tobii_Pro_Lab_Product_Description.pdf/?v=1.181

[38] Eyal M. Reingold and Heather Sheridan. 2012. Eye movements and visual expertise in chess and medicine. *The Oxford Handbook of Eye Movements* (2012),

528–550. https://doi.org/10.1093/oxfordhb/9780199539789.013.0029
[39] SR Research. 2022. Data Viewer. https://www.sr-research.com/data-viewer/
[40] Geoffrey D. Rubin, Justus E. Roos, Martin Tall, Brian Harrawood, Sukantadev Bag, Donald L. Ly, Danielle M. Seaman, Lynne M. Hurwitz, Sandy Napel, and Kingshuk Roy Choudhury. 2015. Characterizing search, recognition, and decision in the detection of lung nodules on CT scans: Elucidation with eye tracking. *Radiology* 274 (2015), 276–286. Issue 1. https://doi.org/10.1148/radiol.14132918
[41] Zohreh Sharafi, Bonita Sharif, Yann Gaël Guéhéneuc, Andrew Begel, Roman Bednarik, and Martha Crosby. 2020. A practical guide on conducting eye tracking studies in software engineering. *Empirical Software Engineering* 25 (2020), 3128–3174. Issue 5. https://doi.org/10.1007/s10664-020-09829-4
[42] Zohreh. Sharafi, Zéphyrin Soh, and Yann-Gaël Guéhéneuc. 2015. A systematic literature review on the usage of eye-tracking in software engineering. *Information and Software Technology* 67 (2015), 79–107. Issue 7. https://doi.org/10.1016/j.infsof.2015.06.008
[43] Bonita Sharif, Michael Falcone, and Jonathan I Maletic. 2012. An eye-tracking study on the role of scan time in finding source code defects. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA)* (New York). ACM Press, 381. https://doi.org/10.1145/2168556.2168642
[44] Heather Sheridan and Eyal M. Reingold. 2017. The holistic processing account of visual expertise in medical image perception: A review. *Frontiers in Psychology* 8 (2017), 1–11. https://doi.org/10.3389/fpsyg.2017.01620
[45] Ian Sommerville. 2018. *Software Engineering* (10 ed.). Pearson.
[46] John Sweller. 1988. Cognitive Load During Problem Solving : Effects on Learning. *Cognitive Science* 285 (1988), 257–285. Issue 12.
[47] John Sweller. 1994. Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction* 4 (1 1994), 295–312. Issue 4. https://doi.org/10.1016/0959-4752(94)90003-5
[48] John Sweller. 2010. Element Interactivity and Intrinsic , Extraneous , and Germane Cognitive Load. *Educ Psychol Rev* 22 (2010), 123–138. https://doi.org/10.1007/s10648-010-9128-5
[49] Richard G. Swensson. 1980. A two-stage detection model applied to skilled visual search by radiologists. *Perception and Psychophysics* 27 (1980), 11–16. Issue 1. https://doi.org/10.3758/BF03199899
[50] Hidetake Uwano, Masahide Nakamura, Akito Monden, and Ken ichi Matsumoto. 2006. Analyzing Individual Performance of Source Code Review Using Reviewers' Eye Movement. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA)* (San Diego). ACM, 133–140. https://doi.org/10.1145/1117309.1117357