



Psychophysical studies on Monte Carlo rendering noise visual perception

Vasiliki Myrodia

Université de Lille

École Doctorale Sciences de l'Homme et de la Société (SHS)

Équipe Action Vision et Apprentissage (AVA)

Laboratoire SCALab UMR 9193, CNRS

Date de soutenance: 15 Décembre 2021

Thèse de doctorat en Psychologie, psychologie clinique, psychologie sociale

Directeur de thèse Pr. Laurent Madelain, CNRS, Université de Lille

Membres du jury

Dr. Pascal Mamassian, CNRS, École Normale Supérieure, Rapporteur

Pr. Sebastien Mielle, University of Wollongong | UOW, Rapporteur

Dr. Anna Montagnini, CNRS, Institut de Neurosciences de la Timone, Examinatrice

Pr. Christophe Renaud, Université Cote d'Opale, Président du jury

Directeur de thèse

Pr. Laurent Madelain

CNRS, Université de Lille

Membres du jury

Dr. Pascal Mamassian

CNRS, École Normale Supérieure, Rapporteur

Pr. Sebastien Miellet

University of Wollongong | UOW, Rapporteur

Dr. Anna Montagnini

CNRS, Institut de Neurosciences de la Timone, Examinatrice

Pr. Christophe Renaud

Université Cote d'Opale, Examineur

L'université n'entend donner aucune approbation ni improbation aux opinions émises dans les thèses. Ces opinions doivent être considérées comme propres à leurs auteurs.

Abstract

Computer-generated images are now commonly used in printed or electronic media. The physically-based rendering using the Monte Carlo method to produce these images induces the presence of visual noise which decreases when the computation time increases. Our research aims at better understanding the human perception of this noise to optimize the computation time without detectable loss of image quality. However, investigating noise perception creates some methodological challenges. The conventional paradigms used in visual search and scene viewing tasks are not well suited to measure noise perception because the definition of noise is an unfamiliar concept to naive participants. In our first study, we varied the noise level of a part of the scene using the adaptive method Quest+. The perceptual threshold at 50% was obtained from the estimated psychometric function. In a second task, observers were asked to detect a quality difference using only their peripheral vision (chapter 5). Our results revealed that participants are using primarily their most central vision to detect a degradation in image quality. Ecological studies in image quality research are needed to understand noise perception under real-world conditions. We implemented an online study (chapter 6) and collected data in both conditions (laboratory, online). The comparison of the results showed that there was no significant difference between the thresholds measured in the different conditions (chapter 7). Finally, we investigated the effects of scenes and textures on perceptual threshold and fixation paths (chapter 8). These findings revealed that the non-textured and brightest areas are the most fixated and the most used to detect the presence of noise. In order to predict human fixations we proposed a new approach by computing a saliency map of the difference of two images with different

noise levels. This map is a better prediction than the saliency map calculated on a single image for the noise detection task. Overall, our results, grounded on human visual perception, may contribute to improving realistic physically-based rendering methods.

Keywords: eye movements, visual noise, physically-based rendering, Quest+

Résumé

Les images photoréalistes générées par des algorithmes de rendu physique utilisant la méthode de Monte Carlo induisent la présence de bruit visuel qui diminue lorsque le temps de calcul augmente. Nos travaux ont pour but de mieux comprendre la perception humaine de ce bruit afin d'optimiser le temps de calcul sans perte détectable de qualité des images. Le concept de bruit dans des images est une notion mal connue des participants naïfs. Cela pose certains défis méthodologiques car il nous a fallu adapter les paradigmes conventionnellement utilisées dans les tâches de recherche visuelle. Au cours de notre première étude, nous avons fait varier le niveau de bruit d'une partie de la scène en utilisant la méthode adaptative Quest+. Le seuil perceptif à 50% a été obtenu à partir de l'estimation de la fonction psychométrique. Dans une seconde tâche, les observateurs devaient détecter une différence de qualité en utilisant uniquement leur vision périphérique (chapitre 5). Les résultats de cette étude ont révélé que les participants utilisent principalement leur vision centrale pour détecter une dégradation de la qualité de l'image. Les études écologiques dans la recherche de la qualité de l'image sont nécessaires pour permettre de comprendre la perception dans des conditions réelles. Nous avons mis en place une étude en ligne (chapitre 6) et nous avons collecté des données dans les deux conditions (laboratoire, en ligne). La comparaison des résultats a montré qu'il n'y a pas de différence significative entre les seuils mesurés dans ces différentes conditions (chapitre 7). Enfin, nous nous sommes intéressé aux effets des scènes et des textures sur le seuil perceptif et les fixations (chapitre 8). Ces analyses nous ont permis de remarquer que les zones non texturées et les plus claires sont les plus fixées et les plus utilisées pour déterminer la présence de bruit. Afin de prédire les fixations humaines nous avons

proposé une nouvelle approche en calculant une carte de saillance sur la différence de deux images ayant des niveaux de bruit différents. Cette carte est une meilleure prédiction que la carte de saillance calculée sur une seule image pour la tâche de détection du bruit. L'ensemble de nos résultats, s'appuyant sur la perception visuelle humaine, peuvent contribuer à améliorer les méthodes de rendu physique réalistes.

Mots clés : mouvements oculaires, bruit visuel, rendu physique réaliste, Quest+

Acknowledgments

My warmest thanks to my PhD supervisor Laurent Madelain, Professor at the University of Lille. Thank you for the opportunity to participate in this project, for the bright ideas, and for your endless patience and optimism. I could not imagine a better supervisor.

I would like to thank the members of my PhD committee, Pascal Mamassian, Sebastien Miellet, Anna Montagnini, and Christophe Renaud, for evaluating my work.

Many thanks to Samuel Delepoulle and Jerome Buisine from the ULCO. It was great to work with you. Thanks to Raphaelle Radenne for her participation in this project and her support and Laurent Ott for his technical support and his kind comments. Also, I would like to thank Emmanuelle Fournier and Sabine Pierzchala for their help.

I would like to thank the members of Scalab. I am grateful for your kind attention, your affection, your support, and all the pleasant moments.

I would like to thank my friends in Athens and France for always being there for me. Ιδιαίτερα ένα μεγάλο ευχαριστώ στον Alexis, στη Βάσω και στη Σοφία. Τέλος, θα ήθελα να ευχαριστήσω τους γονείς μου και την αδερφή μου για την υποστήριξη, την φροντίδα και την αγάπη τους όλα αυτά τα χρόνια.

Prologue

The production of photorealistic computer graphics requires the use of lighting simulation methods, characterised by the presence of visual noise. This noise gradually disappears as the methods used converge, but at the cost of very high computation times. This is the challenging part for the production of 3D image sequences, which require the computation of several distinct images. This thesis is part of the ANR project *Prise-3D* (ANR-17-CE38-0009). The main aim of the project *Prise-3D* is to gain a better understanding of the perceptual mechanisms underlying the perception of a quality image to determine criteria for automatically stopping calculations, which can be used in the context of high-quality audiovisual productions, but also to guide the rendering of interactive algorithms.

Contents

Abstract	v
Résumé	vii
Acknowledgments	ix
Prologue	xi
Contents	xiii
Acronyms	xvii
I Introduction	1
1 Photorealistic computer graphics	3
1.1 Computer-generated images (CGI)	3
1.2 Physically-based rendering	7
1.3 Path tracing	9
1.4 Production and applications	11
2 Visual attention and saliency maps	15
2.1 Visual attention theory	15
2.1.1 Spatial attention	16
2.1.2 Feature-based attention (FBA)	17
2.1.3 Object-based attention	19

2.2	Saliency maps	20
2.3	Image quality metrics	25
3	Psychophysical research	29
3.1	Conventional and adaptive psychophysics	29
3.2	Online Behavioral Experiments	32
3.3	Pros and cons of online research	35
4	Saccades & fixations	37
4.1	Eye movements in reading	40
4.2	Eye movements in scene viewing	42
4.3	Eye movements and visual search	44
4.4	Eye movements and image features	47
4.5	Context and objectives of the thesis	49
II	Studies	53
5	Effects of eccentricity on assessing the quality of computer-generated images using Quest+ algorithm	55
5.1	Abstract	55
5.2	Introduction	56
5.3	Methods	60
5.4	Stimuli	61
5.5	Participants	63
5.6	Task 1 : Perception threshold experiment	63
5.6.1	Procedure	63
5.7	Task 2: Peripheral vision experiment	64
5.7.1	Procedure	64
5.8	Data processing	65

5.9	Results	66
5.9.1	Task 1 : Behavioural performance	66
5.9.2	Task 2	68
5.10	Discussion	71
5.11	Supplementary experiment	75
6	A general Python based framework for online behavioural experiments	81
6.1	Abstract	81
6.2	Introduction	81
6.3	Technical challenges	85
6.3.1	Python	85
6.3.2	Django	87
6.3.3	Developed web platform	88
6.4	Case study: online deployment of Quest+	90
6.4.1	Quest+	90
6.4.2	Quest+ in a 2 alternative forced choice online experiment	91
6.4.3	Results	94
6.4.4	Server / Client exchanges during the experiment	96
6.5	Conclusion	97
7	Comparison of threshold measurements in laboratory and online studies using a Quest+ algorithm	99
7.1	Abstract	99
7.2	Introduction	100
7.3	Method	102
7.3.1	Apparatus	102
7.3.2	Stimuli	103
7.3.3	Procedure	104

7.4	Study 1	105
7.4.1	Participants	105
7.5	Study 2	107
7.5.1	Participants	107
7.6	Data processing	108
7.7	Results	109
7.7.1	Study 1	109
7.7.2	Study 2	113
7.8	Discussion	116
8	Measured and predicted visual fixations in a Monte Carlo rendering noise detection task	121
8.1	Abstract	121
8.2	Introduction	122
8.3	Previous works	123
8.4	Methods	125
8.5	Results	128
8.6	Discussion	132
III	Conclusions	133
9	General Discussion	135
9.1	Main findings	135
9.2	Perspectives	139
	Bibliography	141

Acronyms

2AFC	two-alternative forced choice
CGI	computer-generated images
FBA	feature-based attention
FIT	feature integration theory
GBVS	Graph-Based Visual Saliency
HVS	human visual system
IQ metrics	Image quality metrics
MSE	mean square error
NSPP	number of samples per pixel
PBR	Physically-based rendering
PSNR	peak signal-to-noise ratio
RI	reference image
SSIM	structural similarity index measure

List of Figures

1.1	Evolution of noise during increase of computing time and NSPP . . .	6
1.2	The room is rendered from the perspective of a camera	9
2.1	General architecture of a saliency map based on Itti & Koch model.	22
2.2	Graph-Based Visual Saliency (GBVS) algorithm flowchart	24
3.1	Graphical illustration of the Quest+ procedure	32
3.2	A general example of an online study with the method of Quest+ .	34
4.1	Schematic of a sequence of four fixations and three saccades	46
4.2	Schematic illustration of the studies presented in this thesis	52
5.1	Changes in visual noise depending on the NSPP	58
5.2	Evolution of thresholds across trials for the 3 scenes	67
5.3	Frequency of trials in which observers reported the same quality across the image with respect to the size of the mask for the 3 scenes	69
5.4	Pattern of fixations for all trials for scene 1, 2 and 3 and fixation map for the foveal ($< 1.09^\circ$) and for the peripheral ($> 5.82^\circ$) masks	70
5.5	Schematic of the stimuli presented in the supplementary study	75
5.6	Fixation maps for the two scenes under different conditions	77
5.7	Fixation maps for the two scenes with data from all trials and saliency maps for both scenes calculated by the method of Itti & Koch	78
5.8	We defined the areas more often and less often fixated. We covered the information from all over the scene and we kept only these areas visible.	79

5.9	The PSNR between each image and the RI is lower for the more viewed areas than for the less viewed areas in both scenes	80
6.1	Overview of the experiment protocol using the Django framework as a server	89
6.2	Timeline of the experiment	94
6.3	Summary of data collected in a single participant	95
6.4	Description of the main server / client communications during the experimental procedure	97
7.1	Example of a scene at different stages of computation. The visual noise decreases as the computing time increases.	102
7.2	Timeline of the experiments and examples of pictures	106
7.3	Results of the calculated thresholds in the laboratory and online experiments for 3 scenes	111
7.4	Evolution of threshold and entropy for both conditions	112
7.5	Convergence's trials for the 3 scenes and for both conditions	113
7.6	Results of estimated thresholds by Quest+ in the laboratory and online experiments for 5 scenes.	114
7.7	Results of the convergence trials in both conditions for 5 scenes . .	115
7.8	Raw data of thresholds from the five sessions and for the five scenes in both conditions	116
8.1	Effect of the number of light path estimation on visual noise	123
8.2	Same bathroom scene with some texture differences in order to change the location of the generated noise	126
8.3	Five different inside and outside scenes	127
8.4	The four images used in the experiment	130
8.5	The five images used in the experiment	131

9.1	Stopping criterion applied during rendering process using noise saliency map model	140
-----	--	-----

List of Tables

4.1	Average fixation durations and average saccade lengths for 4 activities, silent reading, oral reading, scene perception, and visual search	39
6.1	Description of Quest+ adaptive psychometric method	92
7.1	Correlation matrix for 5 scenes on online condition	117
7.2	Correlation matrix for 5 sessions on online condition	117

Part I

Introduction

1 Photorealistic computer graphics

1.1 Computer-generated images (CGI)

Image synthesis or rendering is a combination of imaging methods that aim at generate digital images from the description of a 3D scene. The designer of the synthetic image has full control of all the parameters and can create a final image that corresponds perfectly to his needs. Rendering relies on computer science, physics, mathematics, signal processing, and of course visual perception. It is widely used in many fields with many applications such as computer animation, video games, decoration, architecture, advertisement, and so on.

The first computer-generated images (CGI) appeared in the 1960s (Appel, 1968). But, either in terms of modelling or in terms of quality, these images were far from being a realistic reproduction of an environment. In the 1980s, the rapid development of computer hardware marked the beginning of realistic image synthesis techniques. The evolution of computer science and computer graphics has led to better performances and new requirements in rendering.

The process of producing CGI may be separated in two main steps, the design of the scene and the illumination (rendering). The design of the scenes consists in the modelling of the content, the geometric coordinates, the light sources, the colours, the materials, and generally all the features that permit a realistic representation of a scene. Rendering consists in simulating the illumination and the interactions between light paths and the different objects and materials such as its reflections and absorptions.

There are two main types of illumination, local and global illumination. The first techniques for rendering were based on local lighting. These methods calculate the luminosity of a volume by taking into consideration only local criteria and not the interactions between objects. The technique of global illumination, on the other hand, takes into account the interactions between all the objects and surfaces that are present in a scene. The aim is to simulate the different light interactions in a realistic environment, resulting in calculation often quite complex and time-consuming. Global illumination is based on the laws of geometric optics and the physics of propagation of light. Computing global illumination means considering illumination due to light scattering at least twice before reaching the camera. This means that this method takes into account not only the light that comes directly from the light source but also light rays from reflections by other surfaces in the scene. The simulation of global illumination requires to precisely model the type and position of the light sources, the geometry and position of the objects, and the colour and properties of the materials. The scene will then be used for modelling light interactions such as reflections, transmissions, and absorptions. These interactions include all real effects of light phenomena and their propagation in a scene. The simulation of lighting therefore relies on a precise understanding of the physics of reflection and optics. In this thesis, we will focus on scenes generated using a global illumination method.

In 1968, Kajiya implemented the first mathematical modelling related to the global lighting of a scene in the form of rendering equation (Kajiya, 1986). This equation has become the core of the global illumination calculation unifying all the rendering algorithms. Kajiya proposed to solve the rendering equation by the Monte Carlo method (see 1.3). The Monte Carlo method is mainly used to calculate, by a probabilistic method, a numerical estimate of multidimensional integrals that are difficult to solve analytically. The iterative and random process offers an accurate and precise approximation of the real value of the analytical solution.

The overall idea is to stochastically track each ray of light intersecting a point in the scene and recursively evaluate the luminance at that particular point as the average of the luminance at each point with an impact in the scene. However, the recursive and iterative aspect of these methods considerably increases the computation time. For each ray-traced, the stochastic process is repeated until a light source is reached. This process often requires a large number of iterations. It becomes even more difficult for complex scenes and small point sources with close to zero probability of intersection. Importantly, the random part of this process creates a form of visual noise in the final image. This noise is not easily described or related to any known type of noise. The solution to improve or even eliminate this noise is to increase the number of paths traced per pixel until the rendering algorithm converges to a good quality image. However, this increase in the number of paths affects the computation time that increases in proportion.

The Monte Carlo recursive sampling process does not have a criterion for automatically stopping the calculations when it is considered that all the visual artefacts are eliminated. Until now, one solution is to maximize the number of samples for each pixel (i.e. the number of samples per pixel (NSPP) will be used throughout the manuscript) to obtain the best possible result. Unfortunately, this solution is costly as increasing the NSPP mechanically increases the computing time.

An example is presented in Figure 1.1 which shows how the quality of an image improves as the NSPP and the computation time increase. The limit at which the image has reached a satisfactory perceptual quality or, on the contrary, presents visible errors that require further processing is unknown beforehand. The definition of this limit could optimize the computation time so that an observer would be unable to perceive the presence of noise in the image. To this end, the human perception of this visual noise in CGIs remains to be better understood.

There are many works that have focused on improving the CGI computation time. Among these methods, the path tracing with next step estimation (Shirley

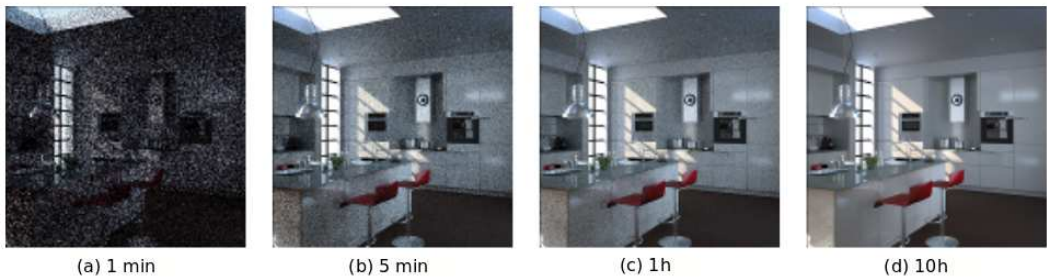


Figure 1.1: Evolution of noise during increase of computing time and NSPP. (a) The result of an image generated after 1 min of calculation. It has a low quality with high visual noise, i.e. random colour values for the pixels in different regions of the image. (b-c) The images are generated after 5 minutes and 1 hour respectively. There is a strong reduction in visual noise but the quality is still low for some regions of the image. (d) This image is generated after 10 hours of computing time. The visual noise is not detectable by an observer and the details are computed precisely to make it look like a photo of an actual scene.

et al., 1996) and the bidirectional path tracing (Lafortune & Willems, 1993) have been widely tested. All these methods we will describe later in more details.

Despite these progress, the problem of improving significantly the computation time remains unsolved. Exploring the human perception of this Monte Carlo noise to improve the rendering methods appears to be a necessary first step. The quality of final image is strongly related to human observers' evaluation and so the perception should be integrated into the production of CGI.

The objective of our research is to accurately measure and model the noise sensitivity of an observer. During our studies, we have used sets of images computed by photorealistic rendering engines, with different noise levels. Our findings could contribute to the improvement of the computation time of the stochastic algorithms used to produce CGI without a detectable loss of visual quality.

1.2 Physically-based rendering

The objective of computer graphics nowadays is to be as realistic as possible. Physically-based rendering (PBR) has become widely used for this reason. The evolution of computer science has led to more computationally demanding approaches to rendering and PBR has become a viable solution. PBR proposes precise modelling of light scattering by using the principles of physics to model the interaction of light and matter.

Although a physically based approach seems to be the most obvious solution for rendering, the first time this approach was presented was in the 1980s. The first was Whitted who introduced the idea of ray tracing for global lighting effects (Whitted, 1980). His approach was the first attempt to simulate precisely the distribution of light in scenes and the rendered images were significantly improved.

After this pioneer contribution, Cook and Torrance managed to render metal surfaces which was a great challenge for researchers at the time (Cook & Torrance, 1981). A generalization of Whitted's approach was developed by Cook et al. (Cook et al., 1984). They showed that ray tracing could generate important lighting effects, for example, blur and reflection from glossy surfaces.

In 1986, Kajiya introduced the application of the Monte Carlo method in the field of rendering to solve the rendering question (Kajiya, 1986). This approach proposed a derivative of ray-tracing algorithms called path tracing in which the evaluation of the luminance is performed according to the Monte Carlo theory. This process is known as stochastic global illumination and it is a method in which stochastic paths are generated from the camera's point of view toward the 3D scene. The integration of the rendering equation by the Monte Carlo method has significantly contributed to the development of complete and accurate solutions of the most complex light simulations such as colour reflections. Since Kajiya's model was presented with the path tracing algorithm, various stochastic methods for global illumination have been

developed. The main goal of these methods is the improvement of the computing time.

One of the most important improvements that have been proposed was the model of Lafortune and Willems that built paths from both the camera and the light sources and tried to connect the different path nodes (Lafortune & Willems, 1993). Their goal was to improve each pixel of the scene and consequently reduce the visual noise faster. A crucial step for PBR was Veach's work on new algorithms like bidirectional path tracing and sampling that improved the final images (Veach, 1997). In 1998, Szirmay-Kalos proposed a model in which the lighting paths are chosen randomly so there could be high frequency colour variations through the image (Szirmay-kalos, 1998). The Monte Carlo approach ensures that this process will converge to the final value of the pixels. The number of samples to reach the visually perfect image is not known in advance and for two different scenes the number of samples needed to converge to the final image will not be the same.

Instead of distributing the samples randomly, sampling strategies (also called adaptive sampling or adaptive rendering) that distribute the sample more intelligently have been proposed. A new sampling method adapted to the image, based on local regression theory, has been developed by Moon et al. and has given more interesting results than previous approaches (Moon et al., 2014).

PBR makes extensive use of ray tracing. This work in combination with the evolution of processing power has had a very important effect on rendering complex scenes with physical approaches that are now widely used. The path-guiding techniques aim at targeting in a more adaptive way the areas that have a high contribution to the scene (Vorba et al., 2019). These areas of interest require more calculations which can be done by using learning algorithms and extracted data or other scene information. The question that still remains is how to stop the computation in each area of the image when it is considered visually noiseless. In order to reduce the computation time but without a loss in quality.

1.3 Path tracing

Path tracing is an extension of the ray-tracing technique (Appel, 1968). Path tracing is the first unbiased Monte Carlo light transport algorithm used in graphics (Kajiya, 1986). The rendering equation was:

$$L_0(x, \omega_0) = L_e(x, \omega_0) + \int_{\Omega} L_i(x, \omega_i) \cdot f_r(x, \omega_i \rightarrow \omega_0) \cdot \cos \theta_i \, d\omega_i \quad (1.1)$$

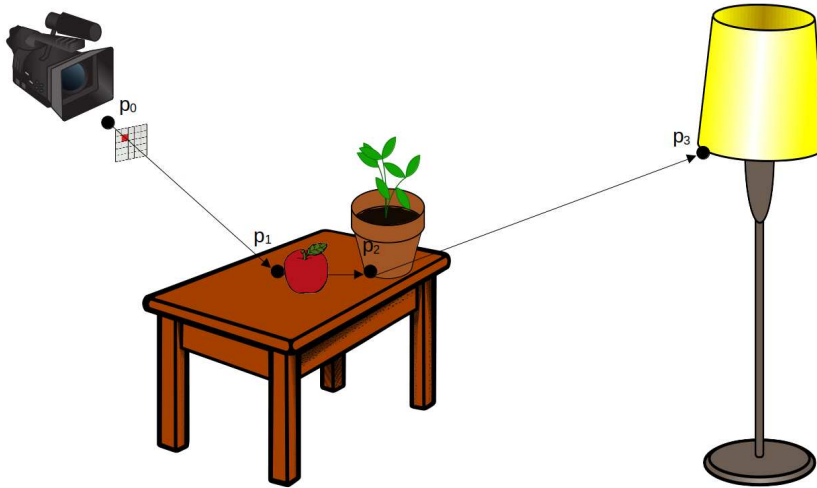


Figure 1.2: The room is rendered from the perspective of a camera. We render a 5x5 pixel image from the camera, that is the white grid. A ray is traced from the camera, p_0 , through the pixel to find out. The ray hits a point p_1 and after an intersection with a series of positions in the scene, it arrives on the light, p_3 . The color of the pixel is calculated by the rendering equation.

The path-tracing process consists of tracing rays from the camera towards an object (figure 1.2). At each intersection of a ray with a surface, a direction of reflection or refraction is drawn randomly. The luminance and colour of the first point (pixel) of impact in the scene are recursively evaluated. For each pixel, according to the Monte Carlo method, the final luminance is the average of the contributions of all the paths used. The recursive calculation stops when the ray reaches a light source. However, the intersection of a ray with one of the light sources has a very low

probability because the latter constitutes only a small percentage of the scene. This induces considerable time spent to try to reach a source relying on random paths. The recursive calculation can be stopped by using the Russian roulette. Russian roulette was introduced to graphics by Arvo and Kirk in 1990 (Arvo & Kirk, 1990). It is an unbiased Monte Carlo technique that allows stopping in a probabilistic way a recursive stochastic algorithm (Spanier et al., 1970). The Russian roulette is an unbiased solution but it increases the variance which results in the visual noise of a scene.

The calculation error according to the Monte Carlo method is of the order of $\frac{1}{\sqrt{N}}$ where N represents the NSPP. The computation error, in this case, corresponds to the noise of a scene. This means that for an image with good quality, it is necessary to increase the NSPP even if that means that simultaneously the computation time increases. In general, thousands of samples per pixel that require many hours or even days of computation may be necessary for a high-quality final image. These two parameters, noise and computation time, are two important variables to take into consideration during a generation of a scene.

Path tracing is the most implemented algorithm in the field of production scenes. It can efficiently handle sampling many lights, occlusion, indirect and direct illumination, and caustic lighting. Many approaches were developed over the years for all these sampling problems. Global illumination is a group of algorithms that aim to simulate precisely the lighting effects in a virtual environment.

Among the most common global illumination algorithms is the Monte Carlo method that we have mentioned before. The Monte Carlo method allows the integration of stochastic methods to the global illumination calculation. This step improves the simulation of all light transfers. In our studies, we use global illumination as it is the most popular method of light simulation in a 3D scene. Although it is longer to generate a scene, global illumination results are often more realistic than other methods like local illumination. The progressive exploration of the space ensures the visual convergence of the algorithms towards the final image. At each iteration, we obtained the whole scene. The intermediate stages of computation contain the scenes with incomplete information, which appears as visual noise heterogeneously distributed over the image surface.

1.4 Production and applications

The evolution of computers in the 1980s changed a lot of things in the field of computer graphics. Computer graphics are used more and more for animation and film production. There are early examples of using visual effects in films like *Voyager 2 Flyby of Saturn* (1981) and *Star Trek II: The Wrath of Khan* (1982). The limits of using widely computer graphics were that the computing power was not sufficient for complex reflection models and global lighting effects.

On the other hand, researchers believed that PBR was not the best solution for visual effects as it was more important to achieve a desired artistic effect even if it was not close to physical accuracy. This idea changed in the late 1990s and early 2000s when Blue Sky Studios introduced a physically based pipeline (Ohmer, 1997). The short film that they made, named *Bunny* in 1998, used an early version of Monte

Carlo global illumination. It had a remarkable difference in quality and it contributed to the development of the ray-tracing system.

In 2001 the first version of Arnold renderer by Marcos Fajardo was proposed. It was the first attempt to generate scenes with complex geometry, textures, and global illumination. It was also faster than any other rendering algorithm known at the time. This renderer was the first to be used by Sony Pictures Imageworks on the movie *Monster House*.

In the mid-2000s, Pixar presented his version of rendering named *RenderMan*. It started with a combination of ray-tracing and rasterization and in 2015 evolved into a physically-based ray tracer. PBR is now widely used for producing computer-generated imagery for a wide range of applications. As we mentioned before, these images are used for films visual effects, video games, and other applications in many fields as architecture, design, and advertisement.

Well-known production companies as Warner Bros and Metro-Goldwyn-Mayer use PBR for more realistic visual effects. Films such as *Gravity* (2013) and *The Hobbit: The Battle of the Five Armies* (2014) were rendered using a physically-based rendering model. From 2004, IKEA decided to evolve from the use of traditional photography for the IKEA catalogue to CGI. The reason was that it was very expensive, time-consuming, and inefficient to build each set for different cultures and different rooms. Now, up to 75% of their pictures are CGI. This could be the textures, the furniture, the walls, the light. The procedure for the generation of these scenes is separated into two steps. The first step is the creation of a highly detailed, physically accurate 3D model for every product. The second step is the creation of photo-realistic digital textures for every material of every product. The substantial difference in quality is the use of ray-tracing technology which allows creating a realistic scene by controlling light. Other big companies like Nobilia and Miele use 3D imagery in their marketing and communications. The benefit is the ability to control every detail and make changes easily and with limited financial impact.

To automate the process of detecting visual convergence, we need to define the probability of looking at an area/object in a scene. In this thesis, our first step to investigate this question is the attention theories and their direct link to saliency maps. The second step concerns some of the most common studies and paradigms related to eye movements on different tasks.

2.1 Visual attention theory

Visual attention is a vast research topic that has attracted the interest in many scientific domains such as psychology, neurophysiology, perception, computational neuroscience, and psychophysics. Visual attention theories intend to explain how visual attributes are processed and how some are selected over others. Attention is considered a selective process that allows us to optimize our performance in visual tasks. Selective attention enables us to collect certain information and guides our behaviour according to our needs as it is not possible to process all the visual information available simultaneously. Space plays an important role in visual selection but it is not the only reference frame. Objects can overlap in space while viewing a natural scene. It is clear that attentional selection is based on features and objects as well.

There are three main categories of visual attention (for review: (Carrasco, 2011)):

- spatial attention: when attention is deployed to locations. For example, if we are looking for a friend to come, it is possible to deploy attention to the door of the room.
- feature-based attention (FBA): when attention is deployed to specific aspects of a scene such as colours and orientations. For example, if we are looking for a red shirt, it is possible to selectively deploy attention to red objects.
- object-based attention: when attention is controlled by objects' structure. For

example, if we are looking for our keys, it is possible to deploy attention to objects of a particular size and structure.

2.1.1 Spatial attention

Spatial attention refers to the orientation of attention to a specific region and helps monitor a specific location (Eriksen & Hoffman, 1972). Spatial attention is categorised into overt attention and covert attention. Overt attention occurs when the gaze is attracted toward a location of attention and overt attention consist in attending to an area in the periphery without moving the gaze to this area. Spatial attention is also often equated with the spotlight attention (Eriksen & Hoffman, 1972; Posner, 1980), a zoom lens (Müller et al., 2003; A. Treisman & Schmidt, 1982) and the mind's eye (Jonides, 1983).

Overt attention concerns everyday activities, for example when we move our eyes or our heads in the direction of an object or a location. The mechanisms of overt attention select locations to direct the high-resolution processing available in the fovea. Overt visual attention is also related to perception and object recognition. A recent study showed that there is a causal influence of overt attention on the object perception (Kietzmann et al., 2011).

Covert attention is very often used in everyday life while driving or performing a competitive sport. Covert attention was first described by Helmholtz (Helmholtz, 1925). When he was looking inside a wooden box through two holes, Helmholtz concentrated on a particular region of his visual field (without moving his eyes in that direction). He found that while illuminating a scene in a viewing box he only had an impression of objects in the region to which he had been paying attention, thus showing that attention could develop independently of eye position.

In 1890, William James described two different modes of attention, one is passive and involuntary while the other is active and voluntary (James, 1891). Later Posner

developed his cueing task and also proposed two kinds of covert visual attention. The first is called exogenous and it refers to bottom-up processing stimuli. It is an involuntary system that reacts to external stimulation that appeared suddenly. The second mode is endogenous and it refers to a top-down system. It is a voluntary system to monitor information at a given location.

Endogenous attention is also known as sustained attention and exogenous attention is also known as transient attention. The terms exogenous/transient and endogenous/sustained are often used as synonyms. These two types of attention have a difference in temporal dynamics. We need approximately 300ms to deploy endogenous attention. Many studies have shown that observers are capable of sustaining the voluntary deployment of attention to a certain position as long as it is necessary for a task. However, the exogenous deployment of attention is transient: it peaks by 100 ms, rises, and decays quickly (Cheal et al., 1991; Hein et al., 2006; Ling & Carrasco, 2006). Endogenous attention influences the performance with cue validity (Giordano et al., 2009; Mangun & Hillyard, 1990; Sperling & Melchner, 1978). Exogenous attention influences performance even when cues are not informative (Barbot et al., 2011; Herrmann et al., 2010).

2.1.2 Feature-based attention (FBA)

Feature-based attention refers to the enhanced sensitivity to a feature value. Attention is selectively deployed to visual features such as colours, orientations, directions independently of their locations (Boynton, 2009; Maunsell & Treue, 2006). FBA is particularly useful in visual search when for example an observer is searching for a target that has known features.

A popular theory in visual search has been proposed by Treisman and Gelade (A. M. Treisman & Gelade, 1980). This theory, named feature integration theory (FIT) is widely used in modelling the visual attention mechanism. Treisman et al.

investigated the processing mechanisms of a visual stimulus to determine whether the properties of the stimulus are processed in parallel or not. They measured the reaction time required to find a target visually different among a set of distractors. They defined two types of search, the feature search, and the conjunctive search. The first occurs when the target is distinguished from the distractors by a visual feature. It creates a pop-out effect which only occurs if there is a single target that differs from its surrounding while all distractors (or the rest of the scene) are homogeneous (Wolfe, 1994). In this case the reaction (search) time is constant regardless of the number of distractors. This search is associated with bottom-up mechanisms.

Conjunctive search occurs when the target and distractors share several visual features. In this case, the reaction time increases with the number of distractors. This search can be described as sequential and requires a high degree of voluntary attention. It is associated with top-down mechanisms (A. M. Treisman & Gelade, 1980; Wolfe et al., 2010).

Another model that follows the principles of FIT is the guided search model made by Wolfe et al. (Wolfe et al., 1989). This study showed that in a conjunctive search, stimuli are divided into distractors and potential targets in a parallel process. In another study, Wolfe proposes to inventory the perceptual attributes capable of guiding the orientation of attention. A number of visual attributes such as colour, motion, orientation, and size were found to guide attention (Wolfe & Horowitz, 2004).

The FBA is particularly useful in visual search and it has been used widely to discover the ability to detect a target with specific features among distractors. Many studies showed that a selection bias could be created when the target features were the same from trial to trial. This means that the FBA affects the guidance during visual searching (Carrasco et al., 1998; Wolfe & Horowitz, 2004). FBA is independent of object appearance and so it is different from object-based attention (Xiao et al., 2014).

2.1.3 Object-based attention

The object-based view suggests that attention is directed to objects based on the gestalt principles. Gestalt principles describe how humans combine similar elements, recognize patterns and simplify complex images when perceiving objects. In 1984, Duncan published a seminal work for the field of object-based attention (Duncan, 1984). He explored the limits of attention by measuring the number of objects that could be detected simultaneously. The aim was to present a stimulus and observers had to judge one or more features of the objects. He found that observers were able to judge two features of the same object without any loss of accuracy while their performance decreased when they judged two features of different objects simultaneously.

Object-based attention affects the efficiency of visual search. Search efficiency improves with increasing similarity among the distractors and decreasing similarity between the distractors and the target (Wolfe & Bennett, 1997). Sohn et al. showed that attention to a particular feature of an object enhances the processing of other features of that object (Sohn et al., 2004) even if these other features were irrelevant for the task (Lu & Itti, 2005).

Object-based attention has also been studied in a variety of other ways and methods such as how it affects eye movements. Studies revealed that the processing of the targets of saccades starts even before the eyes begin to move. Observers had an improvement in visual discrimination tasks, starting 100ms before a saccade (Castet et al., 2006; Montagnini & Castet, 2007). This improved performance supports an effect of presaccadic attention.

2.2 Saliency maps

Most vertebrates use visual attention as a way to select a part of the information to identify and interact with their environment. This process scans the environment, guides the eye, and helps to focus on the most salient information and ignore the others. The goal of understanding this behaviour and also of using it for computational applications, is a priority for researchers. A considerable effort has been devoted to understanding the mechanism of visual attention. The computation of visual salience would allow the detection of perceptually important regions in a scene. Koch & Ullman were the first to introduce the saliency maps and the idea of predicting a regularity of visual fixation patterns (Koch & Ullman, 1985). A saliency map is a 2D map that assigns a degree of perceptual importance to each pixel in the image. The aim is to define the most salient area of a scene and to predict the possible target for eye movements.

For each eye movement in the direction of a scene or an object, our visual attention is fixated on particular regions which are distinct from their surroundings for example a red balloon in a blue sky. These areas are dependent on the content of the scene and independent of the observer's behaviour. The classical saliency models proposed in the literature are biologically inspired and based on feature selection (Itti et al., 1998). This allows to replace the geometric attributes used for saliency computation with perceptual attributes, and it has been showed that these perceptual models succeed in modelling correctly the observer's eye movement (Kim et al., 2010).

Several complex processes are involved in visual attention, and more precisely, visual attention that concerns the visual saliency of a region. Visual salience can be defined as the perceptual information that allows certain objects or regions of a scene to stand out from their surroundings and thus be easily noticeable. These elements would attract attention, and thus the fixations of observers, more than others. As

we have previously discussed there are two attentional processes in human vision that influence the focus of visual attention on a particular region of the scene:

- Bottom-up processes, also called stimulus-driven, are exogenous mechanisms that depend on the inherent properties of the visual stimulus such as contrast, texture, and shape. The observer does not have any goal or intention to move his eyes to these regions of the scene. The salience in this context is invariant and depends solely on the visual attributes of the specific region.
- Top-down processes, also called goal-driven, are endogenous mechanisms that depend on the observer's intention and what he is looking for in the scene. The eye movements are influenced by the task given to the observer, the semantics of the stimulus, but also by his own experience.

Several authors showed that visual selection is predominantly bottom-up (Itti & Koch, 2000; Nothdurft, 2002; Theeuwes, 1991, 1992, 1994) and others that goal-driven mechanisms play a role but after attention is captured by salient elements (Theeuwes, 1992, 1994). Studies on these attentional processes showed that bottom-up mechanisms are faster and precede top-down mechanisms (D. Parkhurst et al., 2002; Tatler et al., 2005; Wolfe et al., 2010).

Itti et al. have proposed an implementation of the saliency map based on the bottom-up aspect (Itti, 2005a; Niebur et al., 2002). The bottom-up model of visual attention considers the change in the sensitivity of the human visual system (HVS) as a function of visual parameters like intensity, colours, and orientation. To describe it shortly, subsampling and low-pass filtering are performed on the input image to decompose it into 9 spatial scales. From these 9 scales, three visual features are detected. These types of local structure, intensity, colours (Red, Green, Blue, Yellow), and orientations (0° , 45° , 90° , 135°) are computed in parallel in feature maps. A total of 42 feature maps are produced. 12 maps are produced for colour, 6 for intensity and 24 for orientations. Each feature is then combined at different scales according to the center-surround differences and normalized in a conspicuity map. Finally, the

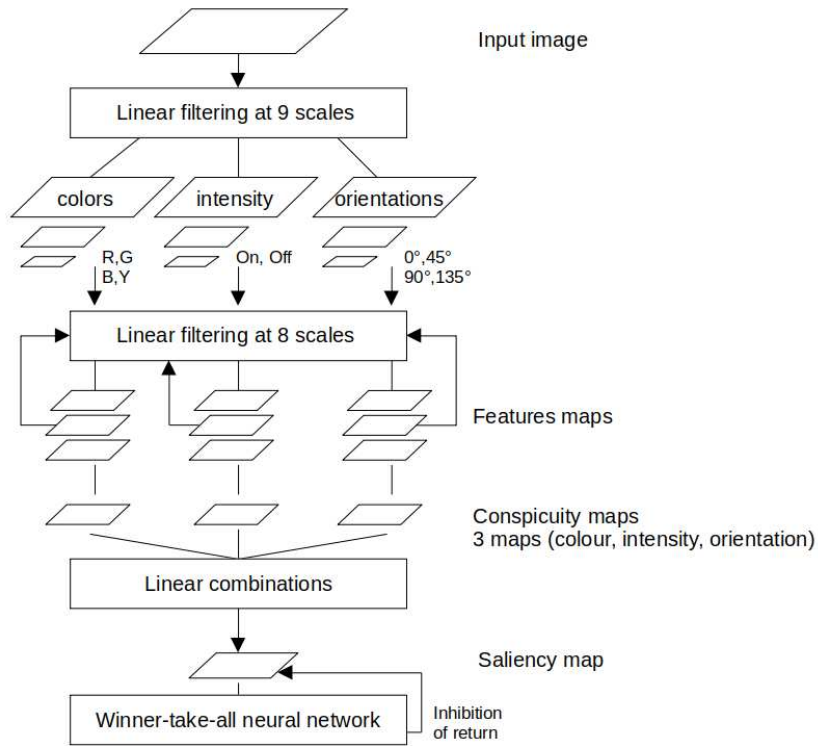


Figure 2.1: General architecture of a saliency map based on Itti & Koch model.

three conspicuity maps of intensity, colour, and orientation, are combined in one saliency map. A diagram summarising the architecture of this model is shown in Figure 2.1.

Currently, many versions of these maps are proposed in the literature. Researchers tested different parameters to detect the set of parameters that affect more the visual attention of an observer. Bruce and Tsotsos define bottom-up saliency based on maximum sampling (Bruce & Tsotsos, 2005). Another interesting idea is the effort to formally and mathematically define surprise. Itti et al. defined the sense of surprise by computing a saliency map in a classical way for each of the features and then using another function to highlight local variation (Itti & Baldi, 2009).

One of the most interesting model is the Graph-Based Visual Saliency (GBVS) model. This model also uses the steps that we have described before but exploits

the computational power, topographical structure, and parallel nature of graph algorithms to achieve natural and efficient saliency computations (Harel et al., 2007). The comparison of results of the GBVS model with the classical model of Itti & Koch showed that GBVS predicts fixations on 749 variations of 108 images with a success 98% versus 84% (Harel et al., 2007; Lin & Lin, 2014). The GBVS model works by extracting image features and linear filtering. The result is a unique feature map for each channel, intensity, colour, and orientation. The next step is to create the activation maps, for this, the value of each location i corresponding to a weight that is calculated according to all other positions. This weight depends on the distance and the difference between i and j . For example, a location point somehow unusual in its neighbourhood will correspond to high values of activation.

The next part is the normalisation process of the activation maps. If the dissimilarity is not concentrated on each activation map, then the resulting map may be less informative. Even if this step sounds trivial, it is a necessary part of the method.

The classical models of saliency map had great results in many applications such as object detection (Li & Itti, 2011; Rosin, 2009; Rutishauser et al., 2004; Seo & Milanfar, 2009), image quality assessment (Ma & Zhang, 2008; Ninassi et al., 2007) and action detection (Seo & Milanfar, 2009). However, how accurately a saliency model can predict human fixations is still a question (Nuthmann et al., 2017; Wahid et al., 2019).

Among many implementations for the saliency maps, we have decided to use the GBVS model to compare with our fixation maps. We decided to use this method as it shows remarkable consistency with human subjects and also, compared to established models, it has better performance. But, it is noteworthy that the notion of visual noise lacks from the existing saliency models. The ability of a human observer to detect or persist in the presence of noise is impressive, while the accuracy of a saliency map deteriorates with the presence of visual noise.

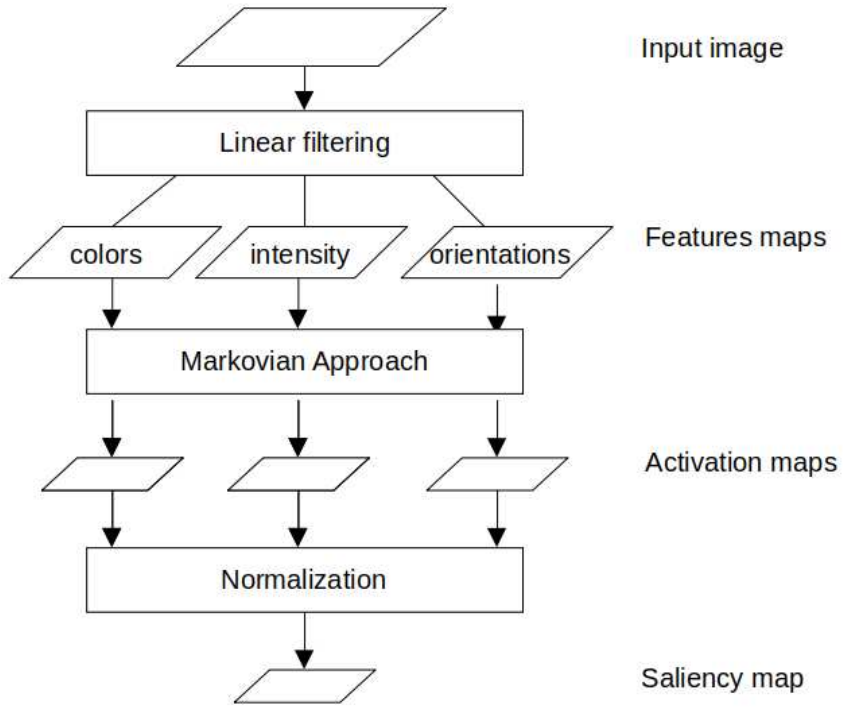


Figure 2.2: Graph-Based Visual Saliency (GBVS) algorithm flowchart.

Despite the wide variety of saliency models, all of them need a clean and distortion free input to predict the salient regions. However, when the task is to detect visual noise, the existing saliency models failed to consider the property of noise. This may happen because of the input of these models. If the input is a noisy map instead of a clean image, it is possible to have better results. In our case, as a noise map, we consider the deviation of two images with different levels of noise. The results of this experiment are presented in study 8.

2.3 Image quality metrics

For evaluating image quality there are two methods, one subjective and the other objective (Avcibas et al., 2002). The first requires a number of observers to look at the images and evaluate them according to their opinion. This method is expensive and time-consuming so it is not very popular. The latter needs an automatic algorithm to assess image quality without human evaluation.

Image quality metrics (IQ metrics) are designed to measure the degradation of an image in order to improve the quality of the result. Measuring the quality of an image is a complicated process. It is even more difficult to compare the results of this process with human data as human opinion is affected by many individual parameters. Throughout the years, a great variety of image quality measurements based on different techniques such as pixel-difference, correlation and human visual system (HSV) have been proposed.

Most IQ metrics calculate the difference between two images, the distorted or noisy image and the original. The IQ metrics are categorized in full-reference, reduced-reference, and no-reference. The full-reference requires a reference image, reduced-reference requires a reference image that exists partially for a number of features and could be used for comparison and finally, no-reference is a blind quality assessment.

The most common and widely used methods are mean square error (MSE), peak signal-to-noise ratio (PSNR), and structural similarity index measure (SSIM). The first approach, MSE, is a mathematically convenient metric that is based on the intensity of distortions. It is computed by the equation 2.1:

$$MSE = \frac{1}{NM} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} e(m, n)^2 \quad (2.1)$$

where, $e(m,n)$ is the error between the pixel at position m,n of the original image

and the distorted image. The PSNR is also a mathematical measurement based on the pixel difference between two images. It is calculated by the following equation :

$$PSNR = 10 \cdot \log\left(\frac{Max^2}{MSE}\right) \quad (2.2)$$

The PSNR value approaches infinity when the MSE value is equal to zero. This shows that a higher PSNR value indicates a lower error and so a better image quality. Finally, SSIM is a measurement based on the structural similarity between two images (Wang et al., 2004). A high-quality image is one that has a structural content such as object boundaries most similar to the original image. SSIM describes any distortion by three principal factors, luminance, contrast, and structural comparisons. It is defined by the equation 2.3.

$$SSIM(f, g) = l(f, g)c(f, g)s(f, g) \quad (2.3)$$

where,

$$\left\{ \begin{array}{l} l(f, g) = \frac{2\mu_f\mu_g+C_1}{\mu_f^2+\mu_g^2+C_1} \\ c(f, g) = \frac{2\sigma_f\sigma_g+C_2}{\sigma_f^2+\sigma_g^2+C_2} \\ s(f, g) = \frac{2\sigma_f\sigma_g+C_3}{\sigma_f\sigma_g+C_3} \end{array} \right.$$

According to the literature, MSE and PSNR can be poor predictors of subjective ratings as they do not take into consideration the human visual system. Studies showed the score obtained by the IQ metrics are not always compared to the subjective evaluation of observers (Girod, 2005; Pappas et al., 1996). SSIM is considered to be correlated with the quality perception of the human visual system.

An interesting study compared the mathematical relationship between PSNR and SSIM (Horé & Ziou, 2010). They found that the PSNR is more sensitive to additive Gaussian noise while the opposite is noticed for jpeg compression. Both

measurements are equally sensitive to Gaussian blur and the values of the PSNR can be predicted from the SSIM and vice-versa.

During our studies, the images that are used are generated in such a way that there is not a reference image that can be considered as a noise-free image (see 1.1). All these methods for evaluating image quality calculate the difference between two images the noisy and the reference. This means that these methods could not be used directly in our case but can give us an index to compare human data with the values calculated by the algorithms.

3

Psychophysical research

3.1 Conventional and adaptive psychophysics

Psychophysical research aims at evaluating the transformation of a physical stimulus by the perceptual system and to explain how this transformation works for different contexts and stimuli. This transformation function can not be measured directly but must be estimated through the subjective responses of the participants. The traditional methods of psychophysical experiments do this by collecting the participants' answers over several iterations for the same stimulus. This method requires a large number of trials for each stimulus and so long experimental time. But these are not the only constraints of the conventional methods. Another challenge is when there are non-linear interactions between stimulus features, for example when the probabilities of responses have a complex relationship to stimulus parameters. Also, another limit of these methods is that it is necessary to perform different experiments to estimate different features of the psychometric function. According to the classical model of discrimination, Weber's law, sub-threshold, and supra-threshold are modelled separately (Georgeson & Sullivan, 1975; Guan & Banks, 2016).

Classical psychophysics often aims to measure three parameters: the detection threshold (DT), the just-noticeable-difference (JND), and the point of subjective equality (PSE) (Macmillan & Creelman, 2005). DT is the limitation of sensitivity of an observer to detect correctly a stimulus. The JND shows that the usual measure of slope is half the distance of the stimulus between the 25th and 75th percentile. The

PSE is when the stimulus intensity is judged equal to 50% of the time. Conventional methods are the method of constant stimuli, the method of limits, and the method of adjustment. In a constant stimuli procedure a stimulus from a fixed set of stimuli is presented repeatedly to the participants. The stimulus value giving a detection response 50% of the time is considered as the threshold. For a conventional method of constant stimuli, the number of stimuli grows exponentially according to the number of dimensions and the number of points per dimension, leading to experiments that need 2500 trials per subject (Guan & Banks, 2016).

In the method of limits, the stimuli are presented in either ascending or descending order to determine the smallest detectable amount. Finally, the method of adjustment differs from the method of limits because here the participant does not need to respond to a stimulus but adjusts continuously the stimulus until he can or cannot perceive it. As mentioned before, these methods are also time-consuming. It is necessary to repeat the same exact stimulus in order to properly estimate the response probability. Also, in the method of adjustment is not possible for a participant to adjust the threshold in more than one or two dimensions. So, if more than two dimensions are probed, the experiment should be separated into more sessions. The limitations of these methods and the advantages are vastly searched and discussed (Klein, 2001) but none of these methods are capable of evaluating stimuli with more than two dimensions.

Adaptive procedures have been developed to address some of these challenges. The general idea of the adaptive methods is to define a threshold with high accuracy but in fewer trials (Leek, 2001). They attempt to estimate a threshold by systematically varying the parameters of the stimulus and finding the model that fits best the participant's answers. The main adaptive methods are PEST (Taylor & Creelman, 1967) and Quest (Watson & Pelli, 1983).

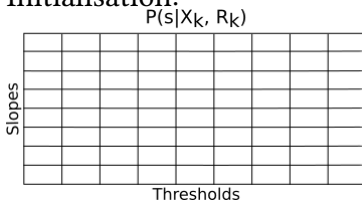
PEST (Parameter Estimation by Sequential Testing) is an adaptive procedure for rapid and efficient psychophysical experiments. It is designed to place trials

at the most efficient location along the range of stimuli to increase measurement precision with the minimum number of trials required to estimate a threshold. In the beginning, it is necessary to select an initial level and step size. After each presentation at a selected level, a statistical test is performed which estimates if it is higher or lower than the targeted performed level. According to the outcome of this test, the next level will be chosen, and the new level could be higher or lower. The final threshold is the final value in the experiment, which means that it does not take into consideration the performance of previous trials. The limitation of this method is that it assumes that the stimulus varies along a single dimension.

On the other hand, Quest uses all the information available from the previous trials to choose the next stimulus. Quest supports multidimensional parametric models in the new version named Quest+ (Watson, 2017). Quest+ operates using Bayes' Theorem. The primary role of Quest+ is to compute the posterior probability of each parameter value, stimuli x , and observed responses r . Each update of the posterior requires to compute the likelihood of every possible response given every possible parameters combinations and stimulus values combinations. Also, Quest+ has two useful functions. First, it can propose the most suitable next stimulus. Quest+ computes the stimulus that minimizes the expected Shannon entropy. The stimulus that has the minimum value can be considered as the most informative at this moment and it should be presented to the participant. Second, Quest+ proposes an automatic stop of testing based on the entropy. It is possible to define a stopping criterion for the entropy so that when it is reached the testing stops. A low value of Shannon entropy means that the highest probability of distribution is found to this value. In Figure 3.1 a graphical illustration of the procedure is presented.

In our experiments, in both laboratory and online versions, we have used the Quest+ method to collect our data and estimate the perceptual thresholds or the limits of the visual central and peripheral vision to detect noise in a scene. In 2021, a new non-linear generalization of traditional psychophysics theory has been

Initialisation:



$$p(s|X_k, R_k) = p(s) \prod_{i=1}^k p(r_i|x_i, s) \quad (3.1)$$

$$p(r_i|x_i, s) = \begin{cases} \Psi(x_i, s) & \text{si } r_i = 1 \\ 1 - \Psi(x_i, s) & \text{si } r_i = 0 \end{cases} \quad (3.2)$$

Update:

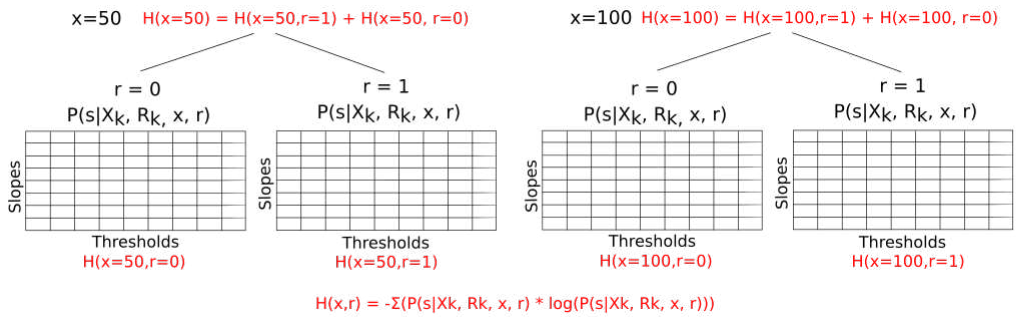


Figure 3.1: Graphical illustration of the Quest+ procedure. The vector of stimulus x , the parameters s of the psychometric function Ψ and the responses r . In the initialisation the probabilities of all the parameters are uniform. After k trials, the posterior probability is calculated by equation 3.1. After trial k , we can compute the probability of a response r for any stimulus x . Also, the function that describes the entropy H is calculated for each stimulus x and response r . The value x that minimizes the entropy is the next stimulus to be presented.

proposed. The model proposed an open-source package, named AEPsych, that is very recent and available for testing (Owen et al., 2021). The idea of this model is that it does not need a parametric assumption about the interaction between intensity and the other dimensions.

3.2 Online Behavioral Experiments

Since Internet has entered the way in which people communicate and search information, scientific disciplines have integrated the Web as a tool for collecting data. Social and psychological studies have started to include online experiments

and to have access to a much wider population, even to populations that were not possible to reach until this moment. Online experiments have become very popular recently because of the Covid-19 pandemic that has added many restrictions to human interactions and to research with human participants. Many researchers have redirected or replaced their lab-based experiments with an online version of the studies. Online research has grown exponentially in the last 20 years and so the interest in tools and platforms that offer an easy and fast way to do this is developing continuously.

The main three steps for conducting an online study are building, hosting, and recruiting participants. Building concerns the programming of the experiment. There are many options in this step to choose from programming languages to experimental builders but here we will present only a selection of the most popular techniques. The most widely used programming language for web development is JavaScript. Although, there are new libraries proposed for web development in other languages such as python. PsychoPy is the easiest way to do a transition from a laboratory to an online study but it is limited to less complex experiments (J. Peirce et al., 2019). It offers a transpilation for the python script to JS script and then an easy way to publish the study on their server. There is also the option of OpenSesame which is a graphical experiment builder that supports python scripting (Mathôt et al., 2012). A new builder for reliable online experiments and with precise time recording for the tasks is Gorilla (Anwyl-Irvine et al., 2020).

The next step is where to host an online study so as to be available for online distribution. Most of the laboratories have their own server where researchers could host and maintain their studies. But there are other options in case that this solution does not exist. A great number of providers for building an online study, propose hosting services. PsychoPy that we have mentioned before, offers an easy hosting service named Pavlovia. A study can be very easily be exported by PsychoPy Builder directly to their server and then be published directly.

Last but not least is the recruitment of the participants. The dominant advantage of online studies is that they give access to a wide range of participants. This means that in a couple of hours, it is possible to collect data from hundreds of participants. In this step, one may publish a study to the University site, social media, and mailing lists. But for increasing the number of participants, there is the option of recruiting platforms such as Amazon Mechanical Turk (MTurk) (Crump et al., 2013; Mason & Suri, 2012; Paolacci et al., 2010) and Prolific (Palan & Schitter, 2018). Both of them maintain an active pool of participants with great diversity and they offer payment handling services. An important difference between them is that Prolific requires an ethical reward for the participants with a minimum possible compensation per hour. The last point to take into consideration before starting to create an online study is to check the compatibility from one step to another.

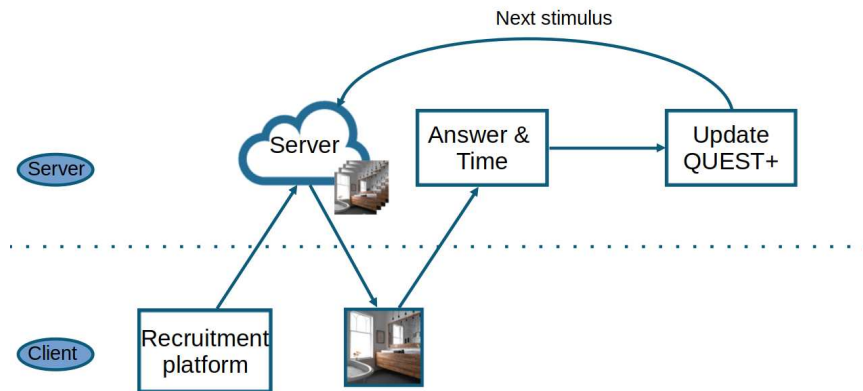


Figure 3.2: A general example of an online study with the method of QUEST+. The two sides show the client and the server. The stimuli are presented to the participants while their answers, and reaction time are saved to the server directly. QUEST+ algorithm updates to the server's side and proposes the most suitable stimulus for the next trial.

3.3 Pros and cons of online research

Internet-based studies have become very popular and so it is important to understand the advantages and disadvantages of online data collection. First, psychological research online provides access to a wide range of populations, diverse, unique and this means a substantial reduction of necessary time to collect data (Birnbaum, 2004; U. D. Reips, 2002b). One may consider that online studies may save money (U. D. Reips, 2002b) but it is not always the case. For example, researchers who have access to free participant pools or to students' groups that participate for course credits, do not need to spend money for collecting data. The crucial point is that results from online studies are compared to laboratories results and they are consistent with findings from traditional methods (Gosling et al., 2004; Weigold et al., 2013). Second, online experiments may have higher ecological validity than laboratory conditions (Dandurand et al., 2008). We can take as an example our own study: the evaluation of image quality is more realistic in a real-life environment than in a sterile laboratory.

There are also disadvantages to running an online study rather than a laboratory study. First, there are issues with high percentages of dropouts and repeated participation (Birnbaum, 2004; U. D. Reips, 2002b). Over the years, researchers have proposed specific solutions for avoiding these problems (Kraut et al., 2004; U. D. Reips, 2002a, 2002b). Some researchers showed that a bonus compensation could motivate the participants (Biemer et al., 2018; Bosnjak & Tuten, 2003). Repeated participation could only be reduced by tracking the Internet protocol (IP) address. This is not accepted by the ethical committees. When the recruitment happens via a platform the number of repeated participation decreases as the recruitment platforms are responsible for the unique identity of their participants.

Another limitation is that participants are often less attentive or motivated (Chandler et al., 2014). For example, participants might understand the instructions

wrongly or might not remember the objective of the task. This could happen if they participated in the study in a noisy environment with many distractors. The proposed solution is the insertion of attention check questions (ACQs). Studies showed that ACQs can reduce low-quality data (Aust et al., 2013; Oppenheimer et al., 2009). For example, an ACQ tends to be an easy calculation based on the basic arithmetic operations or a simple question of the instructions to verify if it is clear.

Pilot testing may also improve the quality of the collected data. Online studies need to be pilot tested because participants have the instructions given in text only while in the laboratory the instructions are often given both in text and verbally. After this initial test, the study could be fixed if it is required and then it can be pilot testing again for the new version or publish on the experiment server.

4

Saccades & fixations

Eye movements are widely used to study cognitive processes during everyday tasks. Our oculomotor system has two important functions, the eyes can move to a new fixation point and they can remain fairly still and new information is acquired from the environment between the movements. These two components are named saccades and fixations.

Saccades are rapid movements that are made to bring objects of interest onto the fovea which is a central, highly sensitive part of the retina. The fovea is a small region, about 1° of visual angle in size, that gives the best acuity to the eyes. Acuity decreases rapidly with eccentricity. For example, while reading, the eyes move to bring a specific part of the retinal image of the text on the fovea. The eyes normally generate on average 2 to 3 saccades each second (Land, 2012) and each saccade lasts at most 40-50ms in humans (Becker, 1989).

During saccades, visual resolution decreases dramatically (Volkman et al., 1968). The eyes are moving so quickly during a saccade that it is not possible to acquire any new information from a stable visual stimulus would be perceived. Moreover, a phenomenon termed saccadic suppression actively removes the visual stimulation from awareness during the saccade (Matin, 1974) so that we are practically blind while our eye move. Saccades correspond to 11% of our lifetime and according to the long-held assumption that our eyes are temporarily blind during saccades, this means that 11% of our life we are blind. However, a very recent study revealed that our eyes may have access to some information from the environment during saccades (Schweitzer & Rolf, 2021).

Saccades are often characterized by specific parameters such as amplitude, duration, peak velocity, and saccadic reaction time (latency). Amplitude corresponds to the size of the saccade and is often measured in degrees of visual angle; it is used to estimate the accuracy of the movement and is often transformed in a gain which is the ratio of the amplitude to the target distance. A gain inferior to 1 indicate hypometric movements (too small), while a gain greater than 1 indicate hypermetric movements (too large). The duration of the saccade is intrinsically linked to the amplitude. It corresponds to the amount of time during which the eyes move. The duration of saccades depends on the amplitude. For example, a 2° saccade takes about 30ms, while a 5° saccade takes around 40 – 50ms (Abrams et al., 1989). This relationship is named the main sequence. The duration of the saccade increases as a function of the amplitude; for amplitudes up to 15 – 20° (Bahill et al., 1975) and this increase is strongly non-linear for amplitudes larger than 20° (Baloh et al., 1975). Peak velocity is the maximum velocity of the saccade measured in degrees/s. Latency is the interval between target presentation and when the eyes start to move in a saccade and it is on average 175 – 200ms (Becker & Jürgens, 1979). Saccade latency is influenced by instruction (Reddi & Carpenter, 2000), the urgency to make a decision (Montagnini & Chelazzi, 2005), environmental contingencies (Vullings & Madelain, 2018), context (Vullings & Madelain, 2019) and prior experience (Wong et al., 2017).

Fixations are the pauses over informative regions of interest before and after the saccades. Since vision is mostly suppressed during a saccade (Thiele et al., 2002), only fixations are used to acquire any new information (Matin, 1974). The retina processes visual information during this period. During an actual fixation, the eyes are not absolutely still. There are three main types of fixational eye movements: microsaccades, smooth drifts, and tremors. Small involuntary saccades that occur during fixation are called microsaccades. The slow drifts are small and tend to prevent fading of vision. High-frequency low-amplitude tremor is present in foveate

Table 4.1: Average fixation durations and average saccade lengths for 4 activities, silent reading, oral reading, scene perception, and visual search

	Fixation Duration (ms)	Saccade Length	
		Degrees(°)	Letters
Silent reading	225 – 250	2	7 – 9
Oral reading	275 – 325	1.5	6 – 7
Scene perception	260 – 330	4 – 5	
Visual search	180 – 275	3	

and nonfoveate species, but it is uncertain whether ocular tremor aids vision in humans.

It is reasonable to think that the saccadic eye movements in different tasks are controlled by the same mechanisms, and that the same principles regarding eye movements should apply to all tasks. However, it is not so easy to generalize eye movement behaviour. For example, fixation durations and saccade lengths in reading differ from those measured in scene perception and visual search (Rayner et al., 2007). There are many studies of eye movements to investigate cognitive processes during different activities as reading, scene perception, face recognition, object recognition, and visual search.

Table 4.1 shows the range of average fixation durations associated with 4 common activities (Keith Rayner, 2009). As we can see in table 4.1 fixations are longer in oral reading than silent reading because the eyes move faster than the reader can enunciate a word and so an observer will stay longer on a word to follow his voice. In scene perception, fixations are often longer than in reading. The scan path length in scene perception and visual search is usually larger than in reading. This is explained because more information can be obtained on each fixation while reading. Lastly, the large range of fixations duration in visual search depends on the stimulus. For example, we need more time to find information when a complex scene is presented such as an apartment with many objects than a simple scene with a single object. Also when a scene is complex, saccades are often shorter.

In this thesis, we are interested in the eye movements during a visual search task and more precisely during attempting to detect the visual noise of CGIs. Eye movements are widely used to investigate cognitive processes during different tasks. In the next sections, we will briefly present some eye movements studies for three principal topics, reading, scene viewing and, visual search.

4.1 Eye movements in reading

In this thesis, we did not study the task of reading. Although, we will describe shortly the main axes of the literature on this subject because it follows a common strategy with visual search and scene viewing. Reading remains a vast field of scientific exploration. The task is well-defined and precisely studied. Also, many paradigms are inspired by the research in reading and are applied in other topics.

A very good example of a sequence of saccades and fixations to extract information occurs when reading a book. The recording of eye movements during reading shows that the role of saccades is to find the words that are decoded during fixations while the retinal image of the world is located on the fovea. Readers normally move their gaze along a line of text in a sequence of fixations separated by saccades. The average fixation duration in reading is 225 – 250ms and the average saccade length is 7 – 9 letters for the Latin alphabetic writing system. For reading studies the usual metric of distance is letters but in order to compare the results with other tasks we need to transform it to degrees of visual angle as in Table 4.1.

Readers decide when and where to move their gaze independently (Rayner & McConkie, 1976; Rayner & Pollatsek, 1981). They have an individual preferred viewing location (PVL) which tends to be between the middle of a word and the

beginning of that word such that their saccades tend to fall short from the world center (Rayner, 1979). Fixation duration of a word is affected by its predictability (Rayner, 1998). Predictability is the transitional probability between two words (Binder et al., 1999; Rayner, 1998). Readers have prior knowledge of the semantic relations with the previous words and they can predict the next word and decide if it is necessary to fix it to understand the meaning of a sentence (Ashby, 2006; Folk & Morris, 2003; Juhasz & Rayner, 2003, 2006).

Viewers can allocate attention independent of eye position and in case of reading, readers skip words (Posner et al., 1980). Skipping words depends on the word length and the context. Short words are more probable to be skipped than long words (Brysbaert et al., 2012; McConkie et al., 1989; Rayner, 1998; Vergilino-Perez et al., 2004). Words that possess semantic content and contribute to the meaning of the sentence are fixated about 85% of the time (Gautier et al., 2000). If a word is skipped during reading, this does not mean that it is not processed. Skipped words are processed by the parafoveal vision which means that there is a decrease in acuity and processing velocity (Rayner & Morrison, 1981).

Many paradigms are developed to control how much information an observer can process during a fixation. To this end, researchers are using a gaze-contingent moving-window paradigm: by controlling the size of the window one, controls the information available. In the case of reading, the idea is to determine how large the window must be for the reader to read normally or vice versa. In the classic moving-window paradigm, the text is displayed normally inside the window, but outside the window the letters are replaced either with other letters or X or homogenous masking patterns (McConkie & Rayner, 1975, 1976; Rayner & Bertera, 1979). Miellet et al. have created a modification of the classic moving-window paradigm, named parafoveal magnification (Miellet et al., 2009). With this modification, they managed to keep the size of the letters around the fixation location normal and increased the size of letters in eccentric vision to compensate for the decrease of acuity. The

results showed that central vision is necessary for reading (Fine & Rubin, 1999a, 1999b).

The studies in the field of eye movements during reading revealed three important aspects (i) that a reader do not fixate every word to understand the meaning, (ii) the prior knowledge of a language helps a reader choose where to look and what to look for and finally, (iii) that readers need their most central vision for reading. As a particular visual search task readers are looking for certain targets through information. In the next section, some characteristics of eye movements during scene viewing are discussed.

4.2 Eye movements in scene viewing

The distinction between scene viewing tasks and visual search task is not always clear. Our experiments are not traditional studies of either of the two categories. In our studies, participants had to look at a scene and then search for visual noise to evaluate the quality of the image. It is necessary to present the studies that have been held in both fields to understand better the choices that we have done for our studies.

The study of scene viewing goes back to the '30s with the pioneer contribution of Buswell on how people look at pictures (Buswell, 1935). This study has had a profound influence on the other seminal work in the field of eye movements and vision, the study of Yarbus (Yarbus, 1965). Yarbus revealed that when different observers looked at the same painting, the pattern of eye movements was similar but not identical. Importantly, Yarbus observed that, depending on the task and the information that people must obtain, the fixations patterns will be different. During scene perception, fixation durations are longer compared to reading, and the average

saccade length is usually larger (see Table 4.1). The fixation duration depends on both the features of the scene and the purpose of the task.

Many studies had examined the importance of saliency (Foulsham & Underwood, 2008; Henderson et al., 2007) and the scene context (Henderson, 2007) during free viewing. The salient features attract attention and guide eye movements during a visual search (D. Parkhurst et al., 2002). Bottom-up properties of the scene define the locations that are capable of attracting focal visual attention. As we have seen, saliency is defined as a set of features in a scene such as contrast, colour, intensity, brightness, etc.

Mannan et al. investigated the relationship between image features and fixation locations. They found that edges and intensity contrast attract fixations when viewing complex scenes (Mannan et al., 1996, 1997). Scene context also guides the search behaviour to likely target locations. The consistency of a scene affects the eye movements of an observer. Observers rapidly move their eyes to an emotional (Harris et al., 2014) or unusual (Bonitz & Gordon, 2008; Underwood et al., 2008) objects of a scene. In scene viewing tasks, observers often have prior knowledge allowing predicting where to look for information. Scene-consistency means that for example a car is unlikely to appear in the sky. When the target is presented on the expected area, search time is on average 19% faster. Observers focus their fixations in the most informative areas of the scene (Neider & Zelinsky, 2006).

As the observers explore a scene, they move their gaze several times per second and they focus on different areas of a scene. This raises the question of the respective limits of peripheral and central vision in scene viewing. As it is mentioned in the previous section, the best way to control the visual field that an observer is using during a task is by using gaze-contingent paradigms. The technique of the moving window/mask paradigm to determine the visual sensitivity is used widely in this field as well (Van Diepen et al., 1999; Van Diepen & Wampers, 1998). The Window paradigm is used in various tasks such as object identification (Henderson et al.,

1997), scene recognition (Saida & Ikeda, 1979) and scene perception task (M. L. Võ & Henderson, 2011).

Peripheral vision influences the gist of a scene as the scene gist is acquired from information across the entire image and not only in the foveal region (Boyce & Pollatsek, 1992; Boyce et al., 1989). However, degrading peripheral information during the initial fixations did not have an effect on scene exploration (Van Diepen & Wampers, 1998). Parafoveal and peripheral vision is useful for object recognition. The role of each part of vision depends on the task difficulty, the image size, the contrast, and the eccentricity. In difficult tasks like detection of a change between two scenes with only one critical object, central vision is mostly used as viewers succeed only if their fixations are within 2° of the critical object-scene (Henderson et al., 2003). While in easier tasks like when detecting whether a scene contains an animal, viewers had a good performance using only peripheral vision with fixations within 50° of the critical object (Thorpe et al., 2001).

The study of eye movement patterns during scene viewing contributes to the understanding of how information in the visual environment is acquired. Scene exploration attempts to analyse objects in context with respect to the structure of the scene, semantic relationships between objects and the features of these objects. Eye movement behaviour during scene viewing reveals useful pattern of information that is acquired during the fixations. In the next section, we will present some important studies on visual search.

4.3 Eye movements and visual search

The study of eye movements in visual search concerns all the tasks in which viewers search through stimuli to find specific targets. Table 4.1 shows that fixation durations in search tasks are highly variable. Saccade length is slightly larger than

that in reading and shorter than that in scene perception. This variability depends on how difficult is to detect the target, the aim of the search task, and the viewer's strategy (Van Zoest et al., 2004). Search for targets in visual search is closer to scene viewing than reading. Reading is a well-specified task while scene viewing depends on the scene. The number of fixations and fixation duration increase when the search task is complicated or the scene is dense with many objects and distractors (Vlaskamp & Hooge, 2006).

Theories of visual search are generally viewed as being either stimulus-driven or goal-driven. Stimulus-driven or bottom-up selection is a fast and often compulsory mechanism. Stimulus-driven selection happens when visual selection depends on the stimulus properties. For example, a red object among a group of black objects will automatically capture attention.

Goal-driven or top-down selection is a slower mechanism that takes place when observers can select the stimuli according to a certain goal of the task. For example, the aim of a task (Land & Lee, 1994; Yarbus, 1965) or internal models (Henderson & Hollingworth, 1998) can affect eye movements. Goal-driven control can influence visual search only if an object has been selected in a stimulus-driven selection. Fast eye movements are completely stimulus-driven, while slower eye movements are goal-driven (Van Zoest et al., 2004).

Like in the previous research fields, the moving-window/mask paradigm is often used during the visual search to determine the perceptual spatial constraints. The findings are close to the findings of scene viewing. The increase of the mask size affects the increase in the search time, the number of fixations, and the duration of the fixations (Bertera & Rayner, 2000). Also, when searching for an object in real-world scenes, scene context may contribute to directing attention toward regions containing potential search targets (M. L. H. Vö, 2021; M. L. H. Vö et al., 2019). Peripheral vision seems to explore a scene while central vision seems to analyse regions of interest with high resolution (Nuthmann, 2014). A very recent study

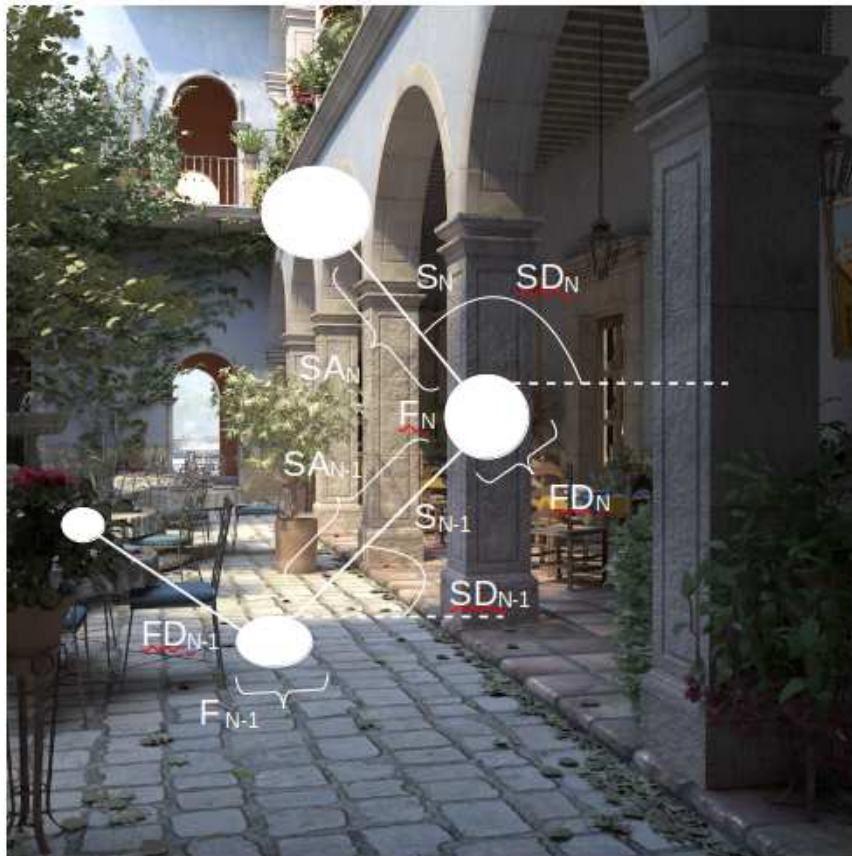


Figure 4.1: Schematic of a sequence of four fixations (circles) and three saccades (lines) between these fixations during a visual search. The principal characteristics of these fixations and saccades are fixation duration (FD), saccade amplitude (SA), and saccade direction (SD) for each of the current (N) fixation or saccade.

shows that the role of central vision in identifying objects in natural scenes seems to be minor, while the role of peripheral vision in visual search may have been underestimated (David et al., 2021).

The literature for visual search while seeing noisy stimuli is still limited. One study investigated the perception via a gaze-contingent window and degradation of information, for example, blur, outside this window (Loschky et al., 2005). They found that the parafoveal degradation is not detectable by viewers as they did not notice that the scene outside the window was blurred. Rajashekar et al. found that

the location of fixations was not random during a presentation of a noisy scene but depended on the features of the scene. They showed that during a visual search task, global image features are used by viewers and that viewers were looking for a component feature of a target rather than for the target itself (Rajashekar et al., 2002). Viewers are influenced by many visual cues like in scene perception tasks such as colour (Williams, 1967), size, and shape (Findlay, 1997; Murray et al., 2003) to detect their target and program their saccades during a visual search. This study showed how certain features of a scene can affect eye movements. In the next section, we explore the impact of image features on eye movements and define image features.

4.4 Eye movements and image features

The interest of researchers to investigate image features is related to the high interest in creating a computational model based on intelligent image processing. Privitera and Stark created an image-processing algorithm that mimics eye movements in detecting regions of interest. They compared their predictions with actual human fixations and found that half of the predictions were accurate (Privitera & Stark, 2000). Features may be specific structures in the image such as points, edges and objects. Among all these features there are some that are related to quality. Pedersen et al. proposed six parameters of quality attributes for the evaluation of image quality: colour, lightness, contrast, sharpness, artefacts such as noise, and physical parameters such as glossy surfaces (Pedersen et al., 2009).

A technique to analyse the relationship between features and fixations is to investigate the correlation of fixations and simple image features such as contrast and pixel intensity. It has been found that observers fixate high-contrast regions and that the intensities of neighbouring pixels at the fixations were less correlated than in image regions selected randomly (Reinagel & Zador, 1999).

Scan paths in free viewing are sometimes viewed as depending on the local contrast or salient features (Itti & Koch, 2000). Indeed, a number of studies showed that salience models can provide a more effective account for a free viewing task than for a search task (Underwood & Foulsham, 2006; Underwood et al., 2006). When no specific target, objective, time, or other constraint is specified to an observer, image features may play a predominant role in guiding attention, and eye movements, toward a potential but general point of interest (Itti, 2005b). The most usual strategy for evaluating the influence of visual features in eye movements has been to find differences between visual features at the fixated locations and at selected control locations. These studies showed robust differences between the two locations and the greatest differences occurred for luminance contrast and edges (Mannan et al., 1997; D. Parkhurst et al., 2002; Tatler et al., 2005; Tatler et al., 2006). Tatler examined the link between a set of visual features (brightness, chromaticity, contrast, edges, etc.) in scenes and the distribution of fixations while freely viewing scenes. He found little or no correlation between these features and fixation distributions.

The studies presented above used different features such as contrast, pixel-intensity, brightness, etc. to investigate how they affect eye movements. All these parameters are crucial for the evaluation of image quality. Another feature of images that is important for image quality is the noise. Some images could have a high level of noise but observers could not perceive it. Noise could be more easily detectable to a specific area of an image. A study which targetted the influence of image content on the perception of noise showed that image content influenced the perception of noise. More precisely, it was more difficult to estimate the quality of an image with many details than with fewer details. Also, they showed that participants did not perceive noise in the same way as image quality metrics (Juric et al., 2016).

4.5 Context and objectives of the thesis

The goal of realistic CGI is to produce complex images that are not distinctive from real scenes. However, the computation of realistic CGI is an expensive process. Even with the latest technology, realistic rendering using global illumination takes time. The lack of a reference image that can be considered as a noise-free image leads to only one solution to stop the algorithm which is that the evaluation of the final image quality by a human observer.

The idea of taking into consideration the properties of the HVS to improve the existed methods of rendering is therefore quite promising for the development of the field. Some properties of the image may not be perceived and some modifications or improvements may be ignored. So, by taking advantage of the limits of the HVS we can hope to optimize the rendering algorithms. More precisely, the integration of human perceptual data could make it possible to carry out only the necessary calculations in each area of the image based on the actual visibility of the image quality. The aim is an interactive photorealistic rendering, by distributing the calculations according to the user's current task, in order to ensure a quality rendering on the focal areas, and a lower quality rendering in the peripheral areas.

The present work concerns the exploration of human perception when they interact with CGI to provide knowledge that may be used to define an automatic stop of the calculations when the current image has perceptually converged. The long term objective is to automate the process of detecting the visual convergence thresholds of stochastic lighting simulation algorithms and to rank image areas and/or objects in the scene that should be rendered first. This process would be more efficient than the present methods.

The experimental work presented in this thesis is divided in three parts.

In the first study, we focused on the effects of eccentricity when judging the quality of CGI. In particular, we investigated the role of central and peripheral

vision when assessing the quality of an image. The question was addressed in two ways. First, a task was designed to define the individual perceptual threshold of our observers by using the adaptive method of the Quest+ algorithm. Second, we used a gaze-contingent paradigm to display a high-quality image through a Gaussian transparency mask at the gaze position so that the lower quality image was visible only in the periphery. The mask diameter was adjusted on each trial using the Quest+ algorithm. This permitted to create fixation maps for each scene and define the areas more often fixated.

We mentioned in the introduction the importance of online studies for psychophysics. Ecological studies in the field of CGI are necessary. In the real-life, we often look at CGI in conditions that can not be compared to the controlled conditions of a laboratory. We created online studies by using the same adaptive method of Quest+ that we have used in the laboratory condition. We present the technique that we used to create this study by using only Python and the python framework named Django. Then, we conducted online studies for measuring the perceptual threshold for the same scenes used in the laboratory and we compared the results between the two conditions. We estimated the learning effects on judging the quality of a CGI over a short period of time. For this experiment, we used indoor and outdoor scenes with different features. We separated the experiment into five sessions, one session for each scene, and the same participant evaluated the quality of each scene. We measured the perceptual threshold of each scene in both a laboratory and online condition.

In a fourth study, we tested the effects of the scenes and the features on judging the quality of an image. Viewers were separated in two groups. The first group evaluated five different indoor and outdoor scenes to study the effects of the scene and the second group participated in the same study but with four variations of the same indoor scene. We manipulated some of the textures of the image according to the measured probability of looking at an area in a scene without changing the

global image structure. For both studies, we recorded eye movements and then we constructed fixation maps to quantify the visual exploration of the scenes. These maps were then compared to noise maps and GBVS maps computed over the visual noise to assess whether the fixation of visible noisy parts of a scene might be predicted. However, although the existing saliency maps predict the more salient areas of an image, they do not take into consideration the visual noise and so the integration of this parameter to a saliency model has many restrictions. This research project could be viewed as a first step to better characterize the perception of visual noise in CGIs.

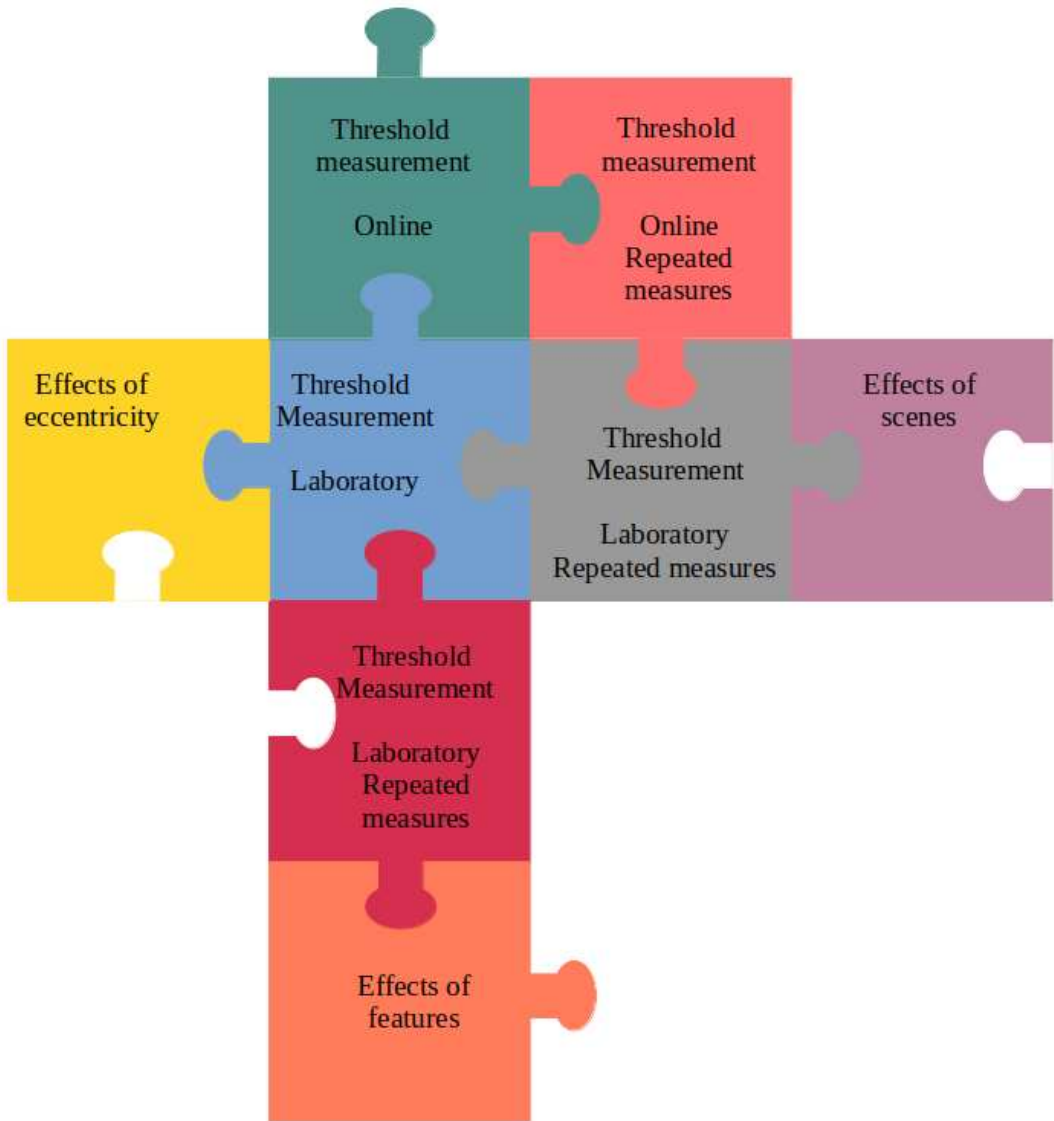


Figure 4.2: Schematic illustration of the studies presented in this thesis.

Part II

Studies

5.1 Abstract

The production of a very high-quality computer-generated image (CGI), termed photorealistic, needs lighting simulations that require significant computing power. The CGIs algorithms induce visual noise, which varies inversely with computation time. In this study, we investigated the effects of eccentricity on the CGI quality perception. To find the role of peripheral vision when assessing the quality of a CGI, we conducted a 2-task experiment using sets of images at different stages of computation (i.e. with various noise levels) from three different CGIs. In the first task, we quantified the 50% perceptual threshold for each participant. We used a QUEST+ Bayesian adaptive algorithm to select the quality of the next stimulus. In the second task, we investigated the role of peripheral vision when assessing the image quality. Participants observed pictures composed of a high-quality image named reference image (RI) and the image that corresponds to their threshold estimated in the first task. We used a gaze-contingent paradigm to display the RI through a Gaussian transparency mask at the gaze position so that the lower quality image was visible only in the periphery. The mask diameter was adjusted on each trial using the QUEST+. Results indicate that a mask of about 100 pixels (3.64°) impaired the observer's ability to report a quality difference. This study revealed that only a small central position of the image needs to be computed in high quality, while for more eccentric parts of the image one can use lower quality, allowing faster image rendering.

5.2 Introduction

Photorealistic CGI are now commonly used in printed or electronic media. For instance, about 75% of the pictures in the IKEA catalogue are not actual photographs taken by a camera but images rendered using 3D softwares. Realistic image computation usually rely on the global illumination method (Kajiya, 1986) in which the physical interactions of light between the objects, lights and cameras lying within a modelled 3D scene are simulated. The formal rendering equation cannot be analytically solved and Monte Carlo approaches are conventionally used to estimate the value of the pixels of the final image. Sampling is performed through the computation of random light paths between the camera and the light sources in the 3D scene. A ray of light is sent from the camera location through a pixel and is randomly reflected by the surface of the first object it encounters. A large number of light paths are created for a given pixel and, due to the law of large number and the Monte Carlo process, the average of the samples for each pixel converges to the solution. Importantly, the rate of convergence is inversely proportional to the NSPP (Shirley et al., 1996).

This has important practical consequences. First, the visual aspect of the final image directly depends on the NSPP. Second, for a given scene, the NSPP required to obtain a realistic final image – i.e. to achieve convergence - is unknown. One solution is to systematically have a very large NSPP but this comes at a cost as the computation time is directly related to the number of samples used to obtain an image and convergence often requires several hours or even days of computation to obtain a high-quality image. For example, the production of a single image for a film can represent several hours of computation. This duration must be multiplied by the number of images composing the motion picture and by the number of points of view for 3D movies. Particularly, the rendering time of Disney's feature film "The New Heroes" exceeds 1 million hours of CPU (central processing unit) time

(and a storage capacity of 5 petabytes) (Stuart, 2014). Deciding when a satisfying visual convergence is reached for an image is therefore an important challenge as computing too many samples unnecessarily increases production costs while stopping computation too early strongly impacts the image perceived quality (see Fig. 5.1).

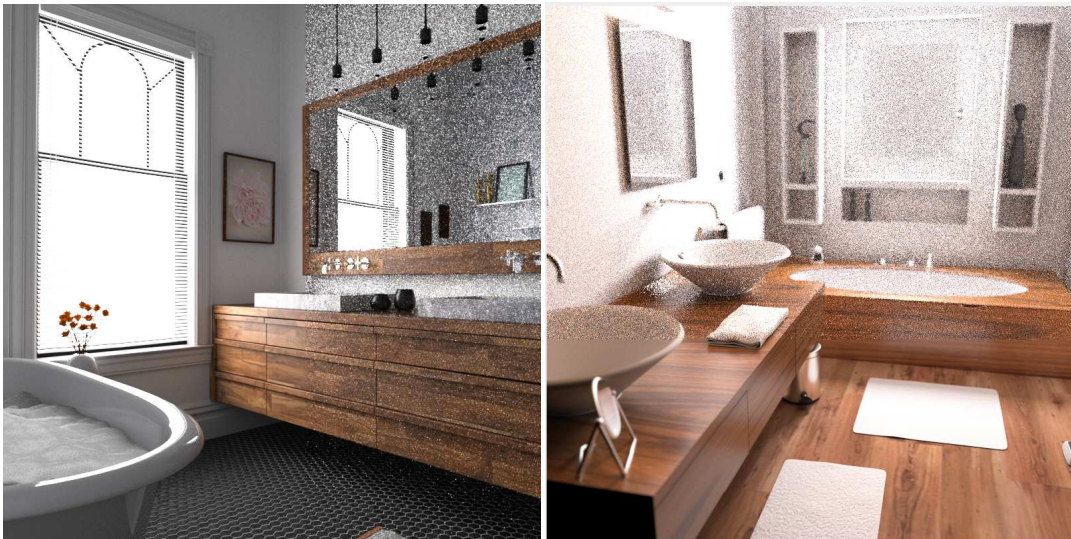
Image quality is often considered as depending on several parameters, such as colour, lightness, contrast, sharpness, artifacts (noise, contouring, banding) and physical parameters (e.g. gloss) (Pedersen et al., 2009). In the case of CGIs noise is a crucial parameter for perceived quality because Monte Carlo methods suffer from



(a)

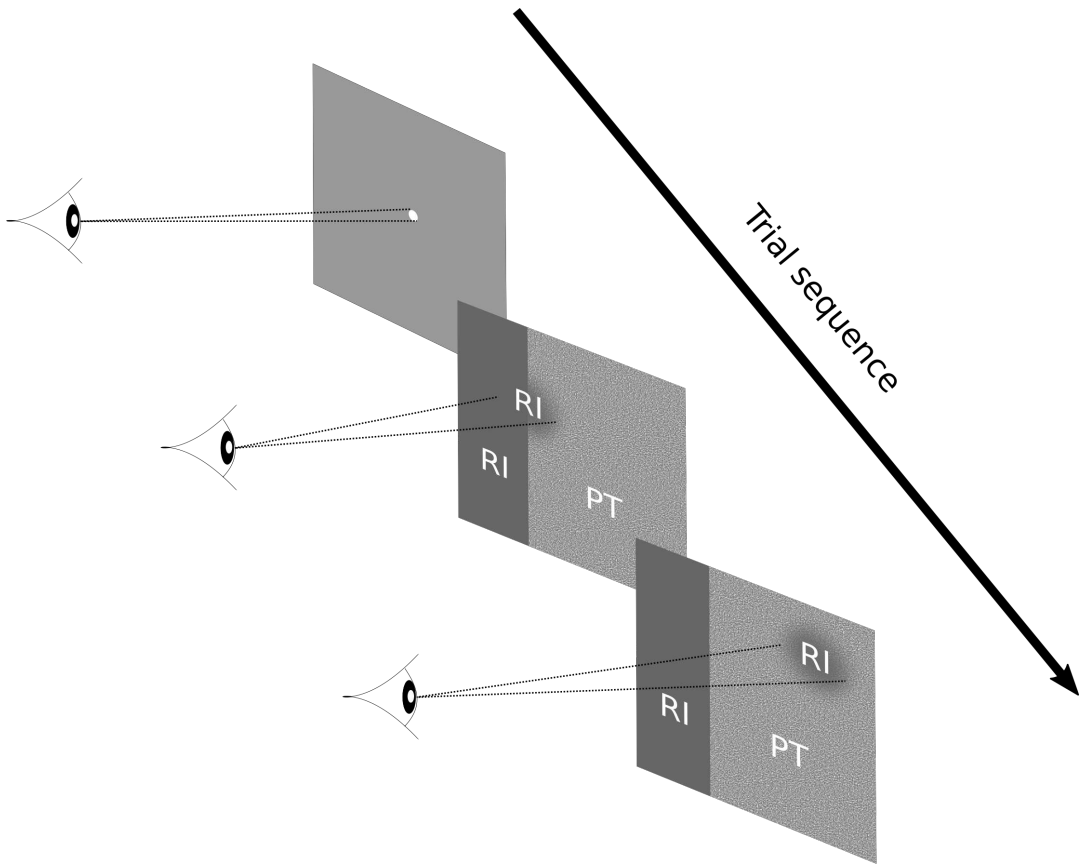
(b)

(c)



(d)

(e)



(f)

Figure 5.1: Changes in visual noise depending on the NSPP. (a) The picture generated with 20 NSPP has a low quality with high visual noise, i.e. random colour values for the pixels in different regions of the image. (b) The same scene now generated with 200 NSPP. There is a strong reduction in visual noise but the quality is still low for some regions of the image. (c) The same scene generated with 10000 NSPP. The visual noise is not detectable by an observer and the details are computed precisely. The image looks like an actual photo of a kitchen. (d-e) Examples of pictures created by combining two images with different NSPP. (d) Picture composed of two images, the left part originating from the reference image (i.e. largest NSPP), the right part originating from a low NSPP image. The visual noise is clearly visible in the right part of the picture. (e) Same as (d) for a different scene. The picture is now composed of two horizontal images, the low NSPP one in the upper part of the picture while the bottom part originates from the reference image. (f) Time course of a trial with a central mask (gaze contingent task). A central fixation point is presented for 700 – 1000 ms before displaying the picture. A part of the image is again composed of the reference image (RI) and the other part is composed by the image at perceptual threshold as estimated in task 1. A gaussian mask displaying the RI is always presented centrally. The diameter of the central mask is adjusted on each trial using the Quest+ algorithm. The trial ends with the observer's answer.

variance in the estimated pixel values, which appears as visual noise (Zwicker et al., 2015). Before full convergence a fraction of the pixels would have random colour values, an effect that progressively disappears when increasing the NSPP (see Fig. 5.1). Because this Monte Carlo noise is intrinsically stochastic and arises from an unknown random distribution function it differs from well-known noise models and is quite challenging to characterize. Moreover, the visibility of this artefactual noise depends on the specific image characteristics (e.g. the luminance, textures, colours). Subsequently, the extent to which observers can detect this noise, or, conversely, the actual improvement in CGI quality in relation to the increase in NSPP, is still unknown. Even though it is well established that noise reduction is achieved when increasing the number of light path samples, we lack a reliable method to define a criterion for stopping the calculations (Buisine et al., 2021). In practice, human observers usually subjectively decide when visual convergence is reached and image quality is satisfactory.

Using the optimal NSPP, that is finding an automatic stop criterion, for noise-free image rendering is a complex matter in part because it requires to take in consideration the properties of human visual perception. Indeed, all Monte Carlo methods result in noisy images but, from a practical point of view, what really matters is the extent to which the noise is visible to the observer. Meeting the challenge at hand therefore requires a better description of the psychophysics of Monte Carlo noise in CGIs. One research strategy is to probe whether the properties of human visual perception established for natural visual scenes extend to CGI noise perception. In the present research we focus on the respective role of central and peripheral vision as, so far, no characterization of the contribution of visual information available from each part of the visual field to the perception of noise in CGI has been provided.

Visual perception across the retina is not uniform, and there is a strong decrease in acuity with increased retinal eccentricity (Loschky et al., 2005). In scene viewing,

peripheral vision is effective to explore the scene and decide where to look next, while central vision is used to analyse the regions of interest with high resolution (Nuthmann, 2014). Peripheral visual information greatly contribute to vision, both in scene perception (Boyce & Pollatsek, 1992) and object recognition (Thorpe et al., 2001). Loss of central vision has a minor impact to targets identification in a 3D environment while peripheral vision is reliably used to identify targets (David et al., 2021). Disentangling the role of foveal versus peripheral vision has been often studied using gaze-contingent methods such as the stimulus is absent (Saida & Ikeda, 1979), degraded (Van Diepen & Wampers, 1998) or is selectively removed (David et al., 2021; Miellet et al., 2010; Van Diepen et al., 1999). Here, we developed a gaze-contingent paradigm using a hybrid stimulus inspired by the “iHybrid” technique (Miellet et al., 2011), displaying a central mask presenting a high-quality CGI over the noisy image so that noise would be visible only in periphery. The extent of the central mask was adjusted based on the observer’s ability to report the presence of noise, allowing us to estimate the spatial limits on noise perception.

5.3 Methods

To investigate the role of central versus peripheral vision for noise perception we first estimated the individual perceptual thresholds using a Quest+ algorithm. These thresholds were then used in a second, gaze-contingent, task in which the threshold-quality version of the image was presented peripherally while a high-quality portion of the image was shown centrally. The diameter of this central portion was dynamically adjusted during the course of the experiment, again using a Quest+ algorithm.

5.4 Stimuli

The scene stimuli (768x800) were 3 coloured, realistic computer-generated images of indoor scenes created with physically-based rendering. Physically-based techniques attempt to replicate the natural process of picture acquisition by simulating the interaction of light and matter within a 3D scene (Pharr et al., 2016). Based on a physical model of the scene a MonteCarlo process is implemented to estimate the value of each pixel of the image and the progressive exploration of the scene space ensures a visual convergence of the algorithms towards the final image. At each iteration random light paths between the camera and the light sources are computed. An image generated at 10 000 NSPP is thus an image for which the lighting for each pixel of the image grid has been simulated through 10 000 iterations. Along the course of the image generation iterative process we saved the computed images after fixed numbers of iterations in order to create a set of images of the same scene with increasing visual quality and decreasing visual noise. For a single scene, we therefore had a first image – FI (which had the lowest quality), a number of intermediate images depending on the fixed step-FS (that is the number of iteration between two saved images) and a reference image – RI (the final high quality image). The NSPP for the reference image was subjectively set by a human observer who stopped the computation when there was no detectable reduction in the visual noise. Importantly, the final NSPP depended on the physical characteristics of the scene.

For the present experiment we chose to use three indoor scenes of rooms rather than isolated objects to provide observers with a wide range of surfaces, materials and areas of interest when performing the detection task. The scenes differed in terms of colours, luminance, surface scattering, indirect light transport and ray propagation. Because the visual quality strongly depends on the scene's physical characteristics, the NSPP of an image should not be directly compared to the NSPP of another image. The first set, which rendered a bathroom, was composed of

Chapter 5 Effects of eccentricity on assessing the quality of computer-generated images using Quest+ algorithm

400 stimulus images (FI =50 NSPP, FS=50 NSPP, RI=20000 NSPP). The second set, picturing a different bathroom scene, comprised 200 stimulus images (FI =10 NSPP, FS=10 NSPP, RI=2000 NSPP). The third set, a kitchen, comprised 500 stimulus images (FI=20 NSPP, FS=20 NSPP, RI=10000 NSPP). Estimating the perceptual threshold of the quality of a CGI poses some specific difficulties. Indeed, the understanding of the term quality is quite subjective and might cover very different properties of an image for different observers. Similarly, the notion of visual noise is not typically used in the day-to-day assessment of image quality and people tend to have a subjective sense of what visual noise looks like. Moreover, it is important to recognise that some noise is always present in physically-based rendered CGIs, even when a very high NSPP is used. This situation made it difficult to use a simple detection task in which participant would simply report whether some noise is present or absent or whether the image quality is high or low because, as we found out in some pilot versions of the present experiments, observers interpreted the instructions in very different ways. To limit the impact of the instructions we chose to ask observers whether the displayed picture was composed of two images, with different NSPP, or was a single image, in a two-alternative forced choice (2AFC) task. On each trial, the RI was randomly cut either horizontally or vertically. Images were cut from side to side, at least 100 pixels from an edge. The missing part of each image was then replaced by the corresponding portion taken from one of the stimulus images (i.e. with a lower NSPP). Therefore, participants always saw a composite picture in which a part of the picture had the highest possible NSPP (RI) and another part of the image had a lower NSPP. at least 100 pixels from an edge. The missing part of each image was then replaced by the corresponding portion taken from one of the stimulus images (i.e. with a lower NSPP). Therefore, participants always saw a composite picture in which a part of the picture had the highest possible NSPP (RI) and another part of the image had a lower NSPP. We reasoned that reporting seeing a single image implies that the difference in noise is below the perceptual threshold while reporting seeing a composite image implies that the visual noise

present in the stimulus image is above the perceptual threshold. By manipulating the NSPP in the stimulus image across trials one should therefore be able to estimate the individual perceptual threshold.

5.5 Participants

42 adults (mean age = 24.5 years) with normal or corrected-to-normal vision participated in this experiment for course credit or compensation. They were naïves with respect to the objective of the study. Each participant experienced only one of the three scenes. The protocol of the study conformed to the Declaration of Helsinki and was approved by the Ethical Committee of the University of Lille (Agreement n°2019-392-S78). All participants gave informed written consent.

5.6 Task 1 : Perception threshold experiment

5.6.1 Procedure

Stimuli were displayed on a video monitor (Iiyama HM204DT, 100 Hz, 22 in.) using the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997) for MATLAB. Participants were seated on an adjustable stool in a darkened quiet room, facing the centre of the computer screen at a viewing distance of 60cm. To minimize changes in visual distance the participant's head movements were restrained with a chin and forehead rest, so that the eyes in the primary gaze position were directed toward the centre of the screen.

Participants were instructed to report whether they saw a single image, or whether they thought the picture was composed of two images with different amount of

noise in a 2AFC task. They reported their choice by pressing the left or right buttons using a joystick.

Each trial began with a fixation cross displayed in the centre of the screen for 200 ms. The composite picture was then displayed until the observer pressed one of the two buttons on the joystick. The trial was terminated at the button press. The stimulus image used in the next trial was then chosen according to the Quest+ algorithm. Quest+ is based on Bayes' Theorem (Watson, 2017) and allows to estimate the perceptual thresholds for each participant with fewer trials than other adaptive procedures. Quest+ requires computing the posterior probability of each parameter value being true, given vectors of trial-by-trial stimulus values and observed responses. In our case, we used the logistic function defined by a threshold (in NSPP units) and a slope. The algorithm estimates the model that best fits the participant's responses in the 2AFC task. This process was repeated on each trial, until the convergence to the 50% perceptual threshold of the observer (Macmillan & Creelman, 2005). Quest+ suggests when to stop testing based on the entropy. We first conducted a pilot experiment ($N = 3$) using the method of constant stimuli to estimate a Quest+ stopping criterion based on a realistic entropy value.

5.7 Task 2: Peripheral vision experiment

5.7.1 Procedure

The goal of the second task was to estimate the extent of peripheral vision used to detect the presence of a noisy part in the composite images. Participants were instructed to perform the same 2AFC as in the first task. The main difference was that the composite image was now always composed of a portion of the reference image and the other portion was taken from the image with the NSPP corresponding

to the individual 50% perceptual threshold measured in the first task. We used a gaze contingent procedure to limit the vision of the stimulus image to peripheral vision (Fig. 5.1). To this end, a gaussian mask containing the reference image was displayed at the gaze location. The gaussian mask had smooth edges and was not detected by the participants. The diameter of the gaussian mask was adjusted on each trial, using the Quest+ algorithm. To minimize eye movement measurement errors, the participant's head movements were restrained with a chin and forehead rest, so that the eyes in the primary gaze position were directed toward the centre of the screen. Viewing was binocular, but only the right eye position was recorded and digitized in both the vertical and horizontal axes. Eye movements were measured continuously with an infrared video-based eye tracking system (EyeLink, SR Research), sampled at 2,000 Hz. Data were transferred, stored, and analysed via programs written in MATLAB running on an Ubuntu Linux computer. Online processing of the eye-movement signal allowed to change the position of the gaussian mask within 2 frames (i.e. 20ms) following a change in the gaze position.

The second task started with a standard nine-point eye-tracker calibration. The fixation point presented in the middle of the screen at the beginning of each trial was used to estimate possible drifts in the gaze position measurement at the beginning of each trial for 700 – 1000ms. If the measured distance between the gaze position and the fixation point was larger than 30 pixels (about 1 deg of visual angle) for more than 10 trials in a row, the nine-point calibration procedure was used to re-calibrate the eye tracker.

5.8 Data processing

To compare the perceptual thresholds across the 3 scenes measured in the first task we normalized the NSPP of each image sets to a fixed dynamic range (between 0 and 1). We computed the median value and the first and third quartiles of the threshold

distributions. We used the PSNR ratio to estimate the noise level of the images. For each observer we computed the PSNR for the three images corresponding respectively to their 25%, 50% and 75% perceptual thresholds. These PSNR were then compared to the one obtained for the reference image (Al-najjar & Soong, 2012; Avcibas et al., 2002; Corsini et al., 1996).

Fixations and saccades were extracted from the raw eye movement data collected in the second task using the Eyelink software. We extracted the average Z-score values of fixations durations for each image zone. These values were then plotted back into the scene according to their coordinates. We smoothed the fixation duration map by convoluting it with the following two-dimensional Gaussian Kernel function:

$$\text{Kernel } N(O, \sigma^2 I)$$

where I is the input data matrix and $\sigma/2 = 1^\circ$ visual angle (Caldara & Miellat, 2011; Lao et al., 2017).

For the eye-movement analysis, we measured for each trial the fixation durations, the scan-path length (i.e. the cumulated amplitude of the saccades) and the total number of fixations. The behavioural performance was measured by the frequencies of the participants reports in the 2AFC task with respect to the size of the mask and the decision time.

5.9 Results

5.9.1 Task 1 : Behavioural performance

The perceptual 50% threshold is the NSPP at which participants were at chance when detecting whether the image was a composite one. Figure 5.2 plots the evolution of the individual thresholds estimated by the Quest+ algorithm during the course of the sessions. We estimated the threshold convergence using a backward-

moving difference 30 trials window. On average the threshold stabilized at the 168th trial. The Quest+ stopping criterion was estimated based on the minimum Shannon entropy [$SC_{scene1} = 5.46$; $SC_{scene2} = 4.59$; $SC_{scene3} = 6.5$] measured in a pilot experiment.

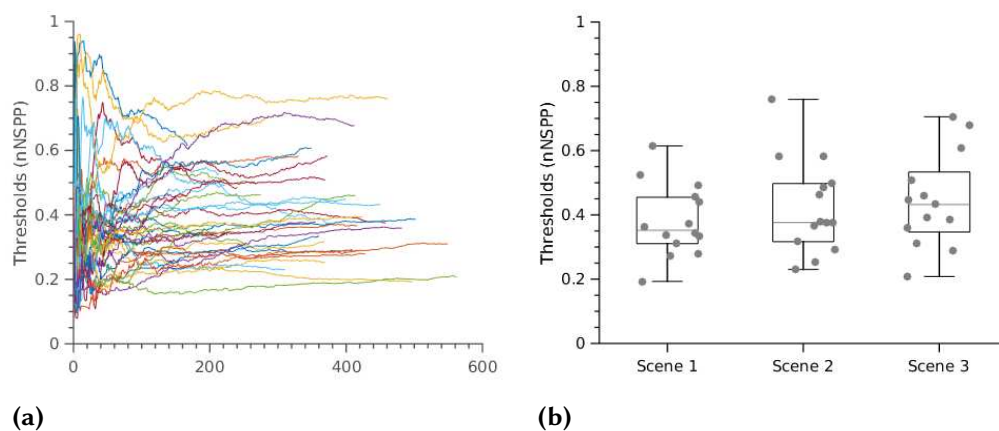


Figure 5.2: (a) Evolution of thresholds across trials for the 3 scenes. The NSPP was normalized between 0 and 1. Each color is a participant. (b) Median values, 1st and 3d quartiles for calculated thresholds for the 3 scenes. The median PT for scene 1 is $Q_2 = 0.351$ ($Q_1 = 0.311$; $Q_3 = 0.454$), for scene 2 is $Q_2 = 0.377$ ($Q_1 = 0.310$; $Q_3 = 0.486$) and for scene 3 is $Q_2 = 0.413$ ($Q_1 = 0.338$; $Q_3 = 0.509$).

The median decision time (i.e. the temporal interval between the image onset and the button press) was similar across the scenes : 2.38 secs. for scene 1 ($Q_1 = 1.76$; $Q_3 = 3.55$), 2.71 secs. for scene 2 ($Q_1 = 1.84$; $Q_3 = 4.47$) and 2.71 secs. for scene 3 ($Q_1 = 1.92$; $Q_3 = 4.06$). A Pearson's correlation test between the individual perceptual threshold and the individual median decision time failed to reach significance ($p > 0.05$) suggesting that there was no relationship between the time spent on a specific scene and the final PT. For each stimulus we computed the difference between the individual threshold and the NSPP as a difficulty index for each trial. We reasoned that a smaller difference would induce a more difficult choice. We found a negative correlation between the difficulty index of each stimulus and the decision time

($R = -0.093, p < 0.001$). Stimuli with NSPPs larger than the individual threshold yielded slower decision time ($F(1, 41) = 46.890, p < 0.001, \eta^2 = 0.534$).

We computed the PSNR, an image quality metric based on the pixel difference between two images, between the RI image and all the other images of the set.

$$PSNR = 10 \cdot \log\left(\frac{Max^2}{MSE}\right) \quad (5.1)$$

Max : maximum possible pixel value of the image

MSE : Mean squared error

We then compared the PSNR of the images with a NSPP that corresponded to the 1st (Q_1) and 3d (Q_3) quartiles of the individual threshold distribution. We found that the PSNR ranged [0.423 0.449], [0.394 0.441] and [0.429 0.458] for the first, second and third scene, respectively.

5.9.2 Task 2

In the gaze-contingent task, we estimated the convergence of the Quest+ algorithm from the 36th trial using a backward-moving difference 30 trials window. The final masks diameter obtained with the Quest+ algorithm had a 3.64° median ($Q_1 = 1.82^\circ, Q_3 = 5.82^\circ$).

Figure 5.3 shows the frequencies of observers reports according to the masks' diameters. For the three scenes we found a strong correlation between the mask diameters and the responses frequencies ($R = 0.94; p < 0.001$).

To further evaluate the extent of peripheral vision necessary to perform the task we chose to compare two extreme categories of masks diameters: perifoveal masks ($< 1.09^\circ$) versus peripheral masks ($> 5.82^\circ$). The extent of the peripheral mask was similar to what has been previously used to prevent central vision (e.g. (David et al., 2021)).

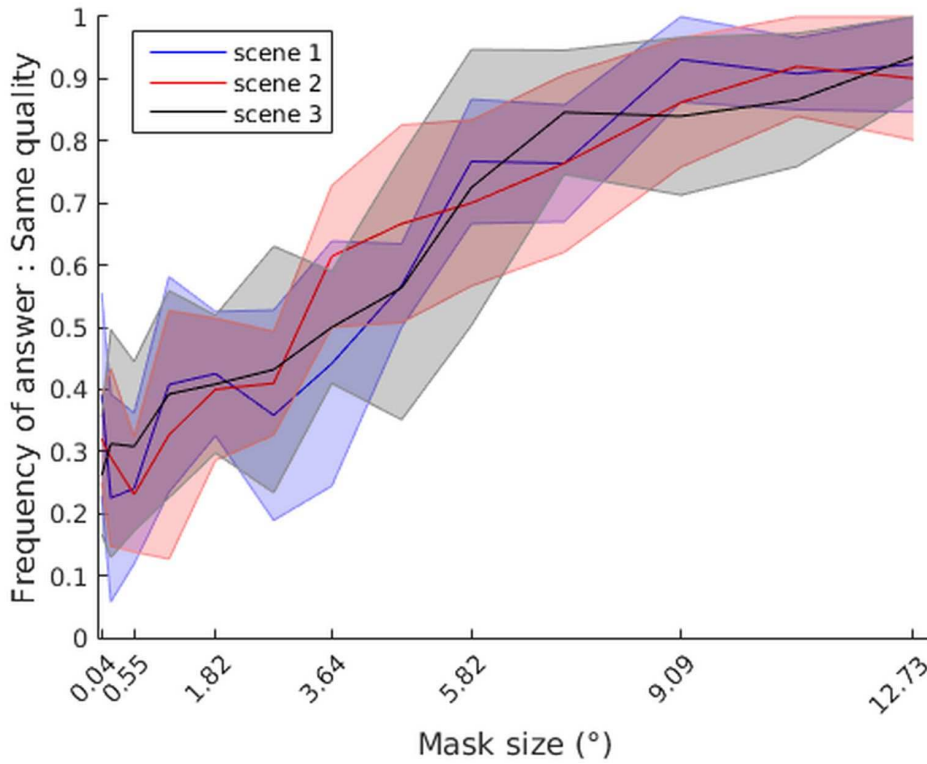


Figure 5.3: Frequency of trials in which observers reported the same quality across the image with respect to the size of the mask (expressed in degrees of visual angle) for the 3 scenes.

Figure 5.4 shows the average fixations durations maps for each of the 3 scenes. In order to determine the magnitude of the fixation duration for each zone of the scene, we extracted the average of the Z-scored fixation durations across all trials. We then decomposed the fixation map of each scene into two separate maps, with either perifoveal or peripheral masks. Comparing the two fixation maps revealed that the fixated areas were independent of the mask's diameter as the euclidian distance between the two fixation maps were lower than the null hypothesis 1% CIs (estimated with 100000 permutation) for each of the three scenes.

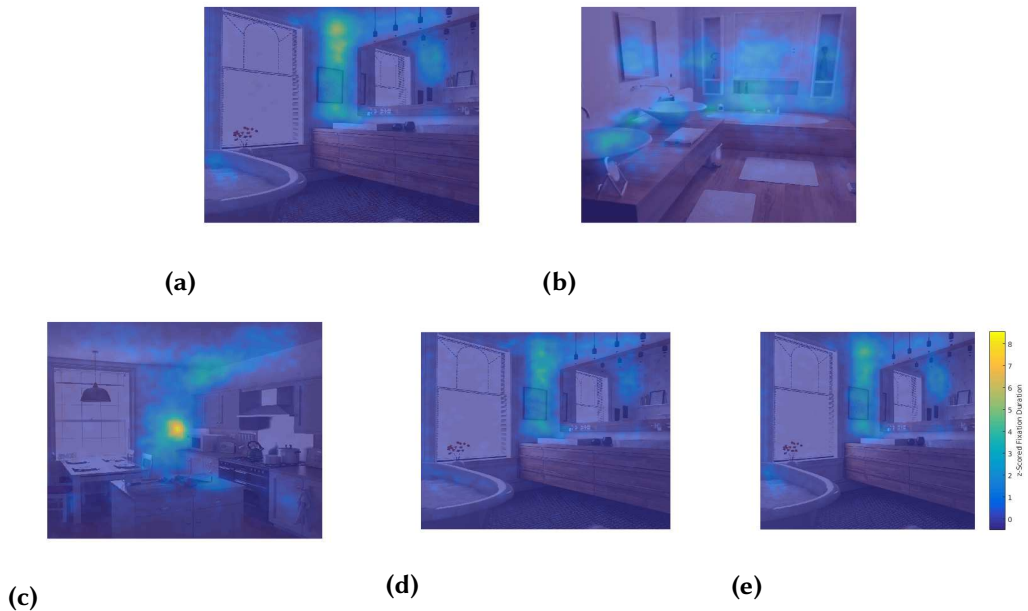


Figure 5.4: (a-b-c) Pattern of fixations for all trials for scene 1, 2 and 3. Fixations maps were created by collapsing z-scored fixation durations across all trials and all masks' sizes. (d) Fixation map of scene 1st scene for the foveal ($< 1.09^\circ$) masks. (e) same as (d) for the peripheral ($> 5.82^\circ$) masks.

Unsurprisingly, we found a strong positive correlation between the average number of fixations and the fixation durations ($R = 0.975, p < 0.001$), the fixation durations and the scan-path lengths ($R = 0.872, p < 0.001$) and the average number of fixations and the scan-path lengths ($R = 0.929, p < 0.001$). Moreover, a repeated measure ANOVA revealed interaction effects between the masks' category (i.e. foveal

or peripheral) and the average number of fixations ($F(1, 41) = 38.372, p < 0.001, \eta^2 = 0.483$), the average fixation duration ($F(1, 41) = 35.446, p < 0.001, \eta^2 = 0.464$) and the length of scan-path ($F(1, 41) = 64.171, p < 0.001, \eta^2 = 0.610$).

Repeated measures ANOVA conducted on the median decision time revealed a significant main effect of the masks' category ($F(1, 41) = 37.995, p < 0.001, \eta^2 = 0.481$). The reaction time was longer while using the peripheral vision. We found a significant effect of behavioural responses on decision time. The analysis revealed that observers had lower reactions when they answered that they did not detect a difference in quality ($F(1, 41) = 87.780, p < 0.001, \eta^2 = 0.682$).

5.10 Discussion

In the present research, we used the adaptive Bayesian psychometric method Quest+ to estimate the spatial extent of the information used by the participants when assessing the presence of visual noise in CGIs. We first estimated the observers individual threshold of noise perception by manipulating the NSPP in the images. This threshold was then used to investigate the relative contribution of central and peripheral vision when searching for the presence of visual noise. Our gaze contingent paradigm simultaneously provides central visual information from the reference image and peripheral visual information from lower NSPP image. Our data show that blocking a limited part of the central vision (3.64°) is sufficient to greatly impair the observers ability in detecting the presence of noise.

These results are in strike to contrast with observations obtained with gaze contingent scotoma paradigms when searching for target objects in natural scenes. Indeed, central masks of up to 6 degrees did not impair object finding, revealing the ability to capture the scene gist and identify objects peripherally (David et al., 2021; Nuthmann, 2014). Moreover, studies with patients affected by central

visual pathologies show that they may consistently use peripheral vision to perform everyday tasks (Boucart et al., 2013; Thibaut et al., 2018; Tran et al., 2010). This discrepancy may be attributed to the difference in the nature of the tasks at hand: searching for visual Monte Carlo noise in a CGI must rely on central vision as the stimulus is, by definition, at the pixel level. On the contrary, visual search of objects may rely on object grammatical constraints and “anchorness” (Draschkow & Vö, 2017), in addition to low level visual information. These differences highlight the necessity to specifically study the perception on Monte Carlo noise in rendered images as it may prove impossible to simply generalize conclusions from established results in other types of search tasks : much as looking for a simple shape among distractors may differ from searching an object in a real-world scene (Eckstein, 2011; Wolfe et al., 2011) deciding whether a CGI image is noisy may rely on specific strategies.

Importantly, our gaze-contingent mask also significantly affected the way observers explored the scenes. It is found that smaller windows led to longer fixations (Loschky & McConkie, 2002). Interestingly we observed the opposite effect as increasing the mask diameter, which mechanically reduces the amount of visual noise in the image which otherwise had a fixed level of noise, resulted in increased fixation duration as well as in the number of fixations and length of scan-paths. Noise detection performance was comparable for the three scenes and eye movements were affected by the increase in mask diameter in a similar way. These changes in eye movements should not be attributed to an artefactual effect due to the mere presence of the mask as all observers reported that the mask was not visible. Instead we postulate that it is because observers were searching for the presence of noise that eye movements changed when the amount of visual noise decreased.

In both tasks we found a significant effect of the task difficulty on the time to reach a decision. In the first experiment, the decision time depended on the level of noise. Observers needed more time to answer for noise levels above their threshold

than for noise levels below their threshold. Importantly, the individual perceptual thresholds were not correlated with the time spent observing during the experiment, indicating that the effect of the difficulty of the task was relative to the observers sensitivity. In the gaze contingent experiment the difficulty depended on the mask diameters as it was positively correlated with decision time.

Observers increased their detection at threshold from 50% at the end of the first part of the study to 30% with the smallest mask (0.04° , or 1 pixel) in the gaze contingent experiment. It is noteworthy that this mask was presented to the observers in only 5 trials on average, because the Quest+ procedure does not promote extreme values. This small number of data implies that one should consider this effect with caution. This threshold improvement may be explained by a learning effect but also could be attributed to well described effects of stimulus range and frequency on mean category scales (Macmillan & Creelman, 2005). In the first part of the study, the observers were exposed to a wide range of stimuli while in the second part we presented only the stimulus at threshold.

As Figure 5.2 shows, we estimated a large range of perceptual thresholds across observers, comprised between 0.1 and 0.9. We found a correlation between the estimated perceptual threshold and the corresponding values of PSNR. We computed that, for the lower and upper quartiles of thresholds, the corresponding PSNR values had very small differences (0.031, 0.048 and 0.029 for the first, second and third scenes, respectively). This indicates that half of the observers had a threshold close to the same level of noise when considering the PSNR rather than the NSPP. The present data demonstrate that even if observers use their central and peripheral vision they will not be able to detect a difference above their perceptual threshold. This could permit to stop the computation of a CGI earlier (i.e. with less NSPP), as the quality of the image can not be perceptually improved. These results could be used to design a new method for establishing a perceptual stopping criterion when generating CGIs (Buisine et al., 2021).

Based on Yarbus's (1967) well-known study, one might expect that the pattern of fixations should depend on the aim the task. Indeed, when exploring an image observers make a series of saccades toward a limited number of local regions that are informative depending on the task demands and these regions are often inspected earlier during the course of the exploration and attract more fixations than other parts of a scene. However, fixations are not only influenced by the aim of the task but also by visual saliency and by global scene properties (Henderson, 2007). It is well established that image features strongly influence human eye movements in natural viewing tasks (D. J. Parkhurst & Niebur, 2003; Reinagel & Zador, 1999): eye movements are attracted to image areas with high contrast and low interpixel correlation (Rajashekar et al., 2002). In our study observers fixated the same areas within a scene regardless of the mask's diameter. However, when searching for the presence of visual noise the regions of interest for each scene appear to vary with the actual features of the scene. We noticed that the areas without a texture were fixated more often, possibly because detecting some noise at the pixel level in low frequency areas is easier than in areas with high visual frequencies due to textured patterns. This has been previously reported and it is established that the human visual system is less sensitive to errors in areas of a CGI that comprise high visual frequencies (Bolin & Meyer, 1998; Myszkowski, 1998; Ramasubramanian et al., 1999). In summary, we have shown that assessing the presence of Monte Carlo visual noise in a computer-generated image strongly rely on central vision. Our studies could have practical use to define an automatic stopping criterion for photo-realistic rendering (Takouachet et al., 2017) and contribute to improving real-time rendering by achieving the optimal level of photorealism at the minimum rendering cost. Moreover, our data show that there are specific zones, with a higher level of visible noise, depending on the features of the scene, that attract more fixations. A realistic description of the characteristics defining these perceptually relevant regions grounded on the properties of human vision may serve adaptive frameworks

for image computation relying on specifically targetting noisy parts of a scene (Vorba et al., 2019).

5.11 Supplementary experiment

One may wonder whether participants fixated the same areas of the scenes due to the presence of the mask or because of the specific features of those areas. To disentangle between these two possibilities we sought to study the scan-paths of scene exploration when performing the 2AFC task without a Gaussian mask. To this end we repeated the same study, separated into two tasks with a new group of observers ($N = 24$). The first task aimed at estimating the perceptual threshold of each observer. For the second task, we constrained the random parameters to create the stimulus configuration : images were cropped horizontally or vertically, either 264 pixels from the left or top or 528 pixels from the right or bottom. The reference image (RI) was cropped and the missing portion was replaced with an image from one of three categories based on the amount of noise; high, at the perceptual threshold, and low (figure 5.5).

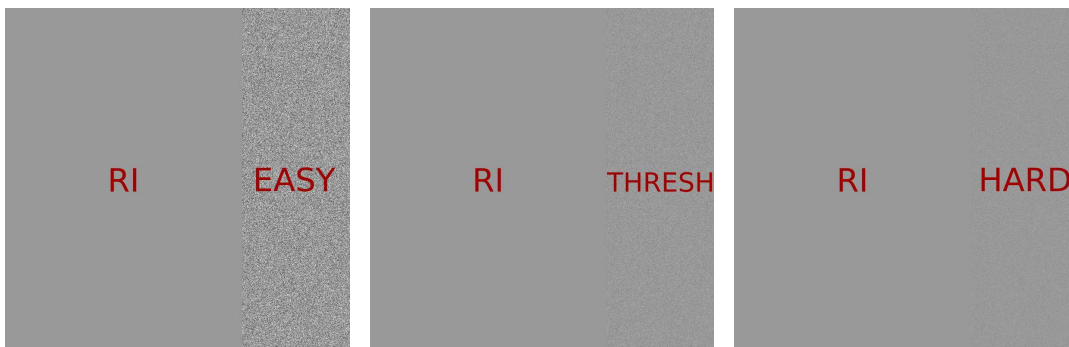


Figure 5.5: Schematic of the stimuli presented in this study. The last generated image of each set (i.e. with less visual noise) was used as the RI and compared to the other images in the set. Each comparison image was cut randomly and the missing part was replaced with the corresponding part of an image belonging to the following three categories: easy (25% perceptual threshold), at perceptual threshold and hard (75% perceptual threshold).

In the second task, we recorded eye movements and used these data to create fixation maps by collapsing z-scored fixation duration across all study conditions. We separated the data according to the two extreme difficulty levels of the presented stimuli (easy, hard) and the responses given (different quality, same quality). Figure 5.6 shows the maps for the four conditions for both scenes. We noticed a strong repetition of the fixation pattern regardless of the condition. The regions of interest were not affected by the difficulty of the presented configuration or the ability of participants to detect a difference in quality.

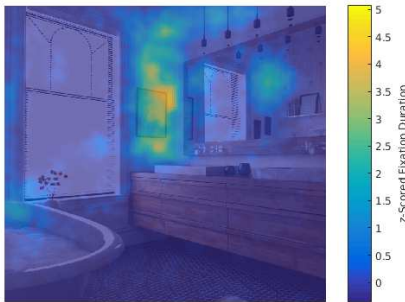
To further answer the question of what attracts their fixations the most, we compared the fixation maps with saliency maps. We used the saliency map as an approximation to estimate where attention might be deployed in the visual scene. Saliency maps were created using the method of Itti et al. shown in figure 2.1. The



(a) Fixations for answer : different quality



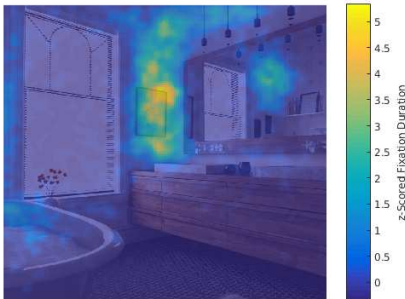
(b) Fixations for answer : different quality



(c) Fixations for answer : same quality



(d) Fixations for answer : same quality



(e) Fixations for a stimulus that corresponds to the category of easy stimuli



(f) Fixations for a stimulus that corresponds to the category of easy stimuli



(g) Fixations for a stimulus that corresponds to the category of hard stimuli



(h) Fixations for a stimulus that corresponds to the category of hard stimuli

Figure 5.6: Fixation maps for the two scenes under different conditions. a-b: maps created only by data from trials for which participants responded that they detected a difference in quality, c-d: maps created by data for which participants responded that they did not detect a difference in quality, e-f: maps created by data for which stimuli came from the easy category, i.e. 25% of threshold, g-h: maps created by data for which stimuli came from the difficult category, i.e. 75% of threshold.

Chapter 5 Effects of eccentricity on assessing the quality of computer-generated images using Quest+ algorithm

saliency maps for both scenes were computed based on three main features, colour, intensity and orientation. The visual noise is not considered in these maps.

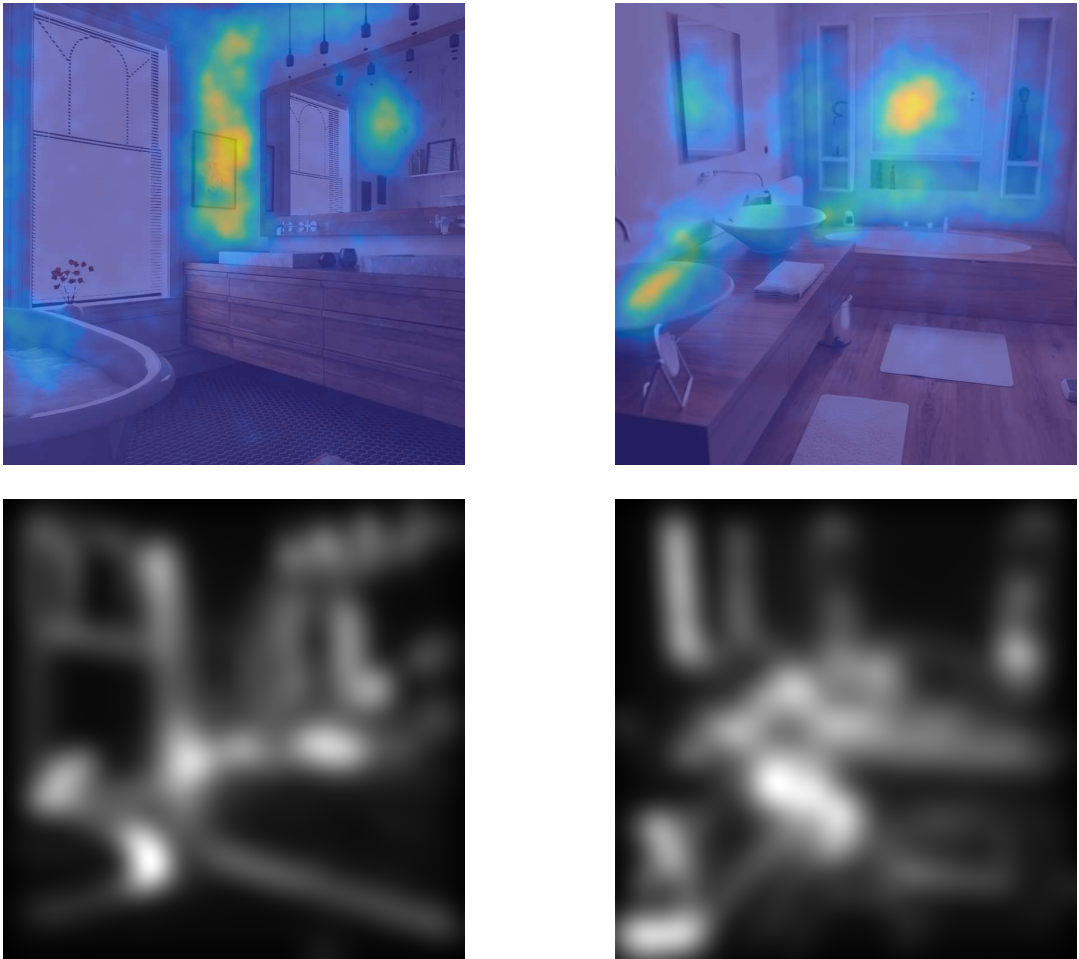


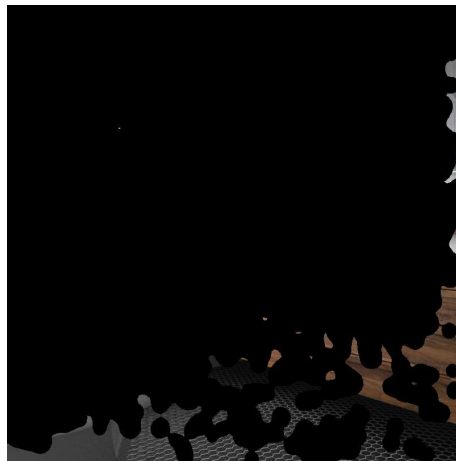
Figure 5.7: Fixation maps for the two scenes with data from all the trials and the saliency maps of these two scenes calculated by the Itti and Koch method (Itti & Koch, 2000).

Figure 5.7 shows the average fixation maps for all trials and the saliency maps for both scenes. Comparison of the fixation and saliency maps shows that the most frequently fixated areas are not the most salient. This is not surprising because saliency is computed on three features of a scene without taking into consideration the visual noise. In our case, the less salient areas attracted more fixations, which

was strongly related to the task objective. Participants focused on finding visual noise in scenes and ignored the salient areas.



(a) Areas more often fixated for scene 1



(b) Areas less often fixated for scene 1



(c) Areas more often fixated for scene 2



(d) Areas less often fixated for scene 2

Figure 5.8: We defined the areas more often and less often fixated. We covered the information from all over the scene and we kept only these areas visible.

The next step in our analysis was to define and isolate the more or less frequently fixated areas for comparison between them. We considered the more frequently fixated areas as those areas where the fixations were greater than 30% of the total fixations. Inversely, the less frequently fixated areas were defined as the areas where fixations were less than 30% of the total number of fixations for that scene. We then

generated two new sets of images using only this information. We covered the entire scene except for these areas. Figure 5.8 shows examples of these new image sets.

We compared the new image sets with the initial image by calculating the PSNR (eq. 5.1). PSNR is used as a measure of image quality and is applied to the new images and the initial image (figure 5.9). The results show longer fixation durations in information-rich areas : more frequently viewed areas have higher noise levels than less frequently viewed areas. The subjective assessment of image quality was therefore validated by this additional analysis of the noise level in each area of a scene using an IQ metrics. This supplementary study was useful in validating our hypothesis that participants focus on the noisiest areas of a scene. Although we used only two scenes for these studies, these results provide an important hint on the perception of visual noise when different features are present.

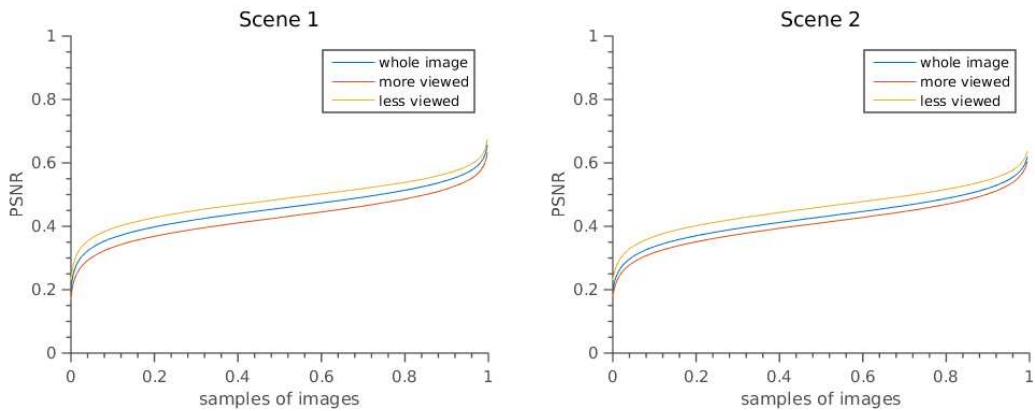


Figure 5.9: The PSNR between each image and the reference image (RI) is lower for the more viewed areas than for the less viewed areas in both scenes ($pvalue < 0.01$). The entire image lies between the other two images sets.

6.1 Abstract

Online studies are increasingly popular among behavioral researchers, allowing to collect data from larger and more diverse samples. However, it is not always trivial to create computationally demanding experiments and ensure the quality of the data. Here we present a method for conducting online studies and exemplify it using the Bayesian adaptive QUEST+ algorithm. Our method is mostly implemented in Python as only the participant's interactions are processed in JavaScript. With the use of the Python web framework named Django, it is possible to build solutions that involve any Python libraries. To illustrate our method, we present a case study and the results we collected online. The example is a simple discrimination task in which we use the QUEST+ method to estimate a perception threshold. This method can be adapted and used for many types of online experimental studies using Python as the main programming language.

6.2 Introduction

Online behavioral experiments have gained popularity in the past decade (Gosling & Mason, 2015) and the COVID-19 pandemic undoubtedly amplified this phenomenon. Online experiments have intrinsic advantages over conventional lab-based behavioral experiments. Specifically, the ease and scale of participant recruitment

for online experiments have profound impacts on our research practices. Indeed, it is now possible to quickly collect data with sample sizes that would be nearly impossible to reach in lab-based settings. In the context of the “replication crisis” in psychology (Pashler & Wagenmakers, 2012), increasing the statistical power to detect effects of limited magnitude is both desirable and necessary (Maxwell et al., 2015). When exploring new questions, the ability to rapidly evaluate the feasibility of a protocol and to quickly estimate the effects of different tasks or experimental conditions also explains the growing use of online studies. Moreover, specific and diverse populations become more easily accessible with online recruitment. This is especially useful when one needs to move away from the typical population of young undergraduate college students to conduct clinical or cross-cultural studies, to probe the particularities of some expertise, or to collect data in hard to reach participants. It has also been argued that research run online improves ecological validity by permitting to study behaviour as it occurs in the real world (Gosling & Mason, 2015). Another great benefit of online experiments concerns the possibility to practice truly reproducible and open science as the exact same experimental protocol may be replicated by any research group - provided the original computer code is shared within the scientific community. There are also obvious limitations when compared to laboratory settings such as data quality or number of dropouts (Birnbaum, 2004; U. D. Reips, 2002b). It is however apparent that online research has become a major tool for experimental researchers.

Moving from lab-based to online research is not always as easy as it might seem at first glance (Grootswagers, 2020). From a practical point of view, running behavioral experiments online require the researcher to i) program an experiment, ii) deploy it on a server and iii) recruit participants. Fortunately, many tools have been recently developed to facilitate implementing the necessary components of online experimentation (see (Sauter et al., 2020) for a recent survey) and there are now commercial solutions integrating all the necessary steps. Importantly, the

programming language used in the large majority of available solutions is JavaScript (JS). JavaScript is mainly a client-oriented language in which scripts are loaded in a web-browser application at each page refresh, allowing the user to manage the page content in a dynamic way. On the other hand, the languages used in the psychology laboratories to program experiments are often C++, MATLAB or Python. This creates a situation in which implementing similar experiments online and in the laboratory is not trivial. One possibility is to use a JS-based solution to program online experiments while continuing to use Python or MATLAB in the laboratory. Although this might be practical in some circumstances it would often be helpful to have the ability to use the same language online and in the lab. Indeed, many researchers have developed specialized libraries that would not be easy to transpile in JavaScript. In the same vein, many codes and dedicated toolboxes have been shared within the scientific community over the years that one may use to ease the implementation of complex experiments in the laboratory. Moreover, it is often the case that companies developing experimental devices used in behavioural experiments, i.e. eye trackers, movements digitizer, EEG recording systems etc., ease the development of computerized experiments by providing software and functions that may easily be integrated within existing programs in Python. The ability to use stripped down versions of these experiments online helps transitioning from laboratory settings and considerably reduce the time and cost associated with implementing similar experimental procedures online and in the laboratory. Another benefit of using a single language online and in the laboratory is the ability to use the exact same code, allowing one to directly compare the data and assess the quality of the measurements across those different settings, possibly even using intra-subjects design.

Importantly, a JavaScript program runs locally on the participant's computer and commercial solutions for JS online experiments such as Pavlovia or Gorilla (Anwyl-Irvine et al., 2020) offer to host the experiments, manage the data storage

and handle participants. On the other hand, non-JS based programs must run on a specific server. Moving away from the JS-based solutions therefore implies that one must setup and maintain a server, a requirement that may be intimidating at first. Fortunately, many labs or academic institution already own and maintain servers that may be used to host online studies, allowing full control and better flexibility at a relatively low cost as a handful of free open source packages are now available allowing to easily setup a server to run experiments online. However, there are no packages readily available dedicated to host and easily deploy non JavaScript online behavioural studies on a server.

In this paper we propose a solution consisting in a generic Python-based framework that may be adapted to many online psychophysics experiments. Indeed, many behavioural experiments share the same generic structure consisting in repeating discrete trials in which the participant is presented with a single or several stimuli. The participant's task is to report whether a stimulus is present or to report which of the stimuli present some specific characteristics. Typically, what varies across experiments is the computation occurring between trials to set the parameters defining what will happen in the next trial. For instance stimuli might be chosen independently from the outcome of previous trials, either randomly, or from a list, within a pre-defined set. In adaptive procedures the parameters of the next trials depend on what has been presented in previous trials and the answers that have been collected. It may also be the case that a visual stimulus needs to be redrawn before each trial depending on the experiment's progress. In other procedures, stimuli parameters might depend on the responses of other participants. From a programming point of view it appears that once the main architecture of a trial is set, i.e. presenting a stimulus and collecting a response, what matters is the specialized computation carried between trials in order to choose the next trial parameters or to construct the next stimulus.

Here we provide a set of functions that may be used to deploy any trial-based

procedure online using the Python Django web framework. An obvious advantage is that any Python package may easily be integrated for advanced computations between trials allowing a wide range of online experiments. Importantly, the Python code used in the laboratory to define the trials parameters might be used to deploy online experiments as well. In addition, our set of functions manages the necessary identification of the participant, the proper data management on the server side, and the communication with the local computer. Interacting with the participant is done locally using a set of JavaScript functions. To exemplify our Python framework we tested an online implementation of the psychophysical Quest+ procedure (Watson, 2017; Watson & Pelli, 1983).

6.3 Technical challenges

In this section we detail the technical choices for implementing Python experiments online.

6.3.1 Python

Python is an easy and general-purpose programming language that became widely used over the years. Python has many advantages over other languages that justify its popularity. Python is a free, open-source, platform independent, interpreted language with a simple syntax that is easy to learn and use in the scientific context (Bassi, 2007). According to PYPL (Popularity of Programming Language), an index that measures how often language tutorials are searched on Google, Python is the first language (30.6%). The popularity of Python is such that there is now quite a rich support available for advanced programming as well as many introductory tutorials available. Because Python is widely used in many academic, industry data

science and engineering communities, a great variety of dedicated scientific and data handling libraries have been developed over time such as NumPy (Oliphant, 2006) and SciPy (Oliphant et al., 2007). For behavioural experiments, specialized open-source applications have been developed to ease the programming of classical procedures and provide useful Python functions such as OpenSesame (Mathôt et al., 2012), Experyment (Krause & Lindemann, 2014) and Psychopy (Brooks, 2019; J. W. Peirce, 2007). Among them Psychopy is a popular library for conducting behavioural experiments (Brooks, 2019; J. W. Peirce, 2007). It provides a useful set of stimuli, methods, and advanced procedures. It is noteworthy that it also allows researchers to create simple python-based experiments using a dedicated builder and automatically transpile them in JavaScript to be hosted on commercial servers such as Pavlovia. However, due to important differences between Python and JavaScript, the automatic transpilation is not fully functional yet, in particular for computationally demanding programs. Recently a python package to program adaptive design optimization experimental procedures, termed ADOpy, has been introduced (Yang et al., 2021). ADOpy is an open-source package which simplifies the implementation of demanding psychophysical model-based procedures. It relies on Bayesian optimization to maximize the information obtained on the next experimental trial. The ability to use these specialized tools when programming experiments in the laboratory explains why Python is now widely used to conduct experiments in the laboratory.

Using Python for online experiments has also some important advantages over a pure JavaScript program. Because JS runs entirely on the participants computer a forced refresh of the page or of the browser's internal storage systems can lead to a loss of client information and therefore of the state of the experiment. It is of course difficult to foresee all the possible actions on the client side in order to ensure seamless data collection over a large sample. Moreover, a JavaScript code can be easily modified as it is downloaded by the client, rendering it vulnerable to malicious manipulations of data. Furthermore, developing complex and mathematically

demanding experimental procedure in JavaScript requires non-trivial technical skills. This might explain why there are many open-source programs to develop complex adaptive procedures such as Quest+ (Watson, 2017; Watson & Pelli, 1983) or other adaptive procedures (Yang et al., 2021) available for Python but not for JavaScript yet. The solution we propose is based on Python and limits the use of JavaScript to the local interaction with the user.

6.3.2 Django

To deploy Python-based online experiments one needs a web framework that enables the easy development of secure websites using Python. We chose a well established high-level Python web framework named Django which includes code from the Python standard library (Django Software Foundation, 2013). Django is a stable and well documented method and many private companies such as Instagram, Pinterest, Mozilla, etc. have been using it for many years. It is open-source Licensed and includes a copy of the Python license for compliance with Python's term. Django is free and open source. Importantly, Django offers an administration interface which may be used to create secure access pages to store the participants data. Another important point is that our proposed code offers a Docker (Anderson, 2015) integration of the Django application. Docker is a free and open-source software using virtualization that has the advantage of simplifying the server installation. A server hosting the application using Docker provides an easy online access, and a set of tools readily available to ease the development and maintenance of web applications. Docker also allows the user to partition the libraries useful for a specific application, without conflict versions, and to run in the background while allowing distant interactions.

6.3.3 Developed web platform

The web application we propose is essentially a generic online experiment integration module in which the majority of the code is in Python. To ease the development of online experiments, we wanted to propose a modular approach where only some part of the code needs to be adapted when developing a new experiment. Figure 6.1 describes the exchange of data involved when running an online study with our solution. The Django server manages the interactions with the client. A unique identifier is used to keep track of the client identity during the experiment. The Django server simply sends the trial parameters to the client. On the client side, a JavaScript code process the keystrokes once the stimulus has been loaded and displayed, allowing to collect both the participant's response and response time (i.e. the time elapsed between the stimulus onset and the keystroke). On the server side, using Python allows the researcher to integrate any pre-existing Python package and source code easily. Importantly, the whole code of the study is stored on the Django server, allowing both persistence and regular backup of the experiment so that no information regarding the client's state may be lost in case of a communication problem. The proposed set of functions is freely available online on Github : <https://github.com/prise-3d/WebExperiment-QuestPlus>

The example application incorporates the Quest+ method to adjust the stimuli parameters throughout the experiment. The use of Quest+ is provided here as an example for a better understanding of the tool and to allow other users to modify the relevant parts as necessary to implement experiments according to their own needs. A detailed documentation describing the functions usage and the data transactions between JavaScript and Django is available as well. The Quest+ adaptive procedure is often used in the laboratory using Python implementations. However, only one implementation of Quest+ for online studies exists in JS (<https://github.com/kurokida/jsQUEST>). Here we used Quest+ to better describe the

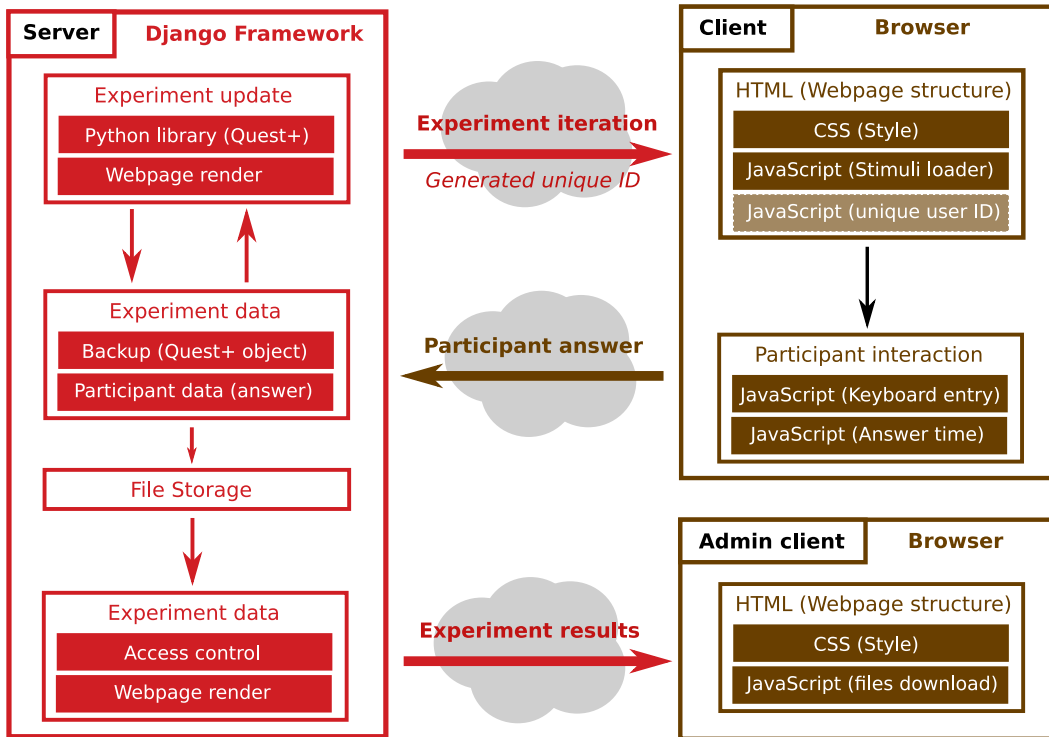


Figure 6.1: Overview of the experiment protocol using the Django framework as a server. The figure describes the implementation of a Quest+ procedure to exemplify the flow of data during an experiment. The Django server initializes the experiment and the Quest+ instance associated with the specific participant. The server then send the content of the web page to the client computer. The JavaScript code loads and display the stimulus within the participant’s web-browser. The participant will then have to respond to the stimulus once by a keystroke. Once a keypress is detected the response and the response time are sent to the server. The server saves the incoming data and updates and stores the Quest+ instance to the server’s file system. The participant is then presented with new web page content that depend on the updated Quest+ instance. The experiment ends once a stopping criterion is reached, either from the Q+ procedure or after a user defined number of iterations. In addition, a unique identifier generated by the server is stored in the client’s browser. This identifier links the server and the client which enable to store the data in relation to this unique identifier. Importantly, the experiment may be stopped and restarted using this unique identifier. On the Django server, the administrator can access the collected data.

tool and to allow other users to modify the relevant parts as necessary to implement any Python based experimental procedures according to their own needs. This study

was tested across several browsers (Google Chrome, Firefox, Safari) and operating systems (Microsoft Windows, Apple macOS, Linux).

6.4 Case study: online deployment of Quest+

To probe our solution and exemplify the usage of our program we implemented an online study using the Quest+ procedure. Many Python implementations of the Quest+ method developed are freely available. In this paper we adapted and used an implementation available here : <https://github.com/mmagnuski/questplus>. This implementation was tested and compared to laboratory results of the same study in Matlab. We found a significant similarity between the two conditions (Myrodia et al., 2021).

6.4.1 Quest+

Quest+ is a generalization of Quest, an adaptive psychometric method that provides an estimate of several parameters of a psychophysical model. Quest+ adjusts the parameters of a psychometric function relying on Bayes' Theorem to find the model that best fits the observer's responses (Watson, 2017; Watson & Pelli, 1983). One of the benefit of the Quest procedure is that it provides a set of individual parameters with fewer trials than other conventional methods such as the constant stimuli method or staircase methods. In principle, it accommodates arbitrary numbers of stimulus dimensions, psychometric function parameters, and trial outcomes. It is therefore suited for a great variety of experiments. The primary role of Quest+ is to estimate the posterior probability of each parameter value, given vectors of trial-by-trial stimulus (x) and observed responses (r). We briefly describe this method using the logistic psychometric functions (Ψ) with two parameters, $s = (Threshold, Slope)$,

as an example. The stimulus values x that minimizes the computed Shannon entropy (H) defines the stimulus that will be presented on the next trial (see table 6.1). Quest+ may therefore also be set to advise the experimenter on when to stop the experiment, with a stopping criterion based on the Shannon entropy. This iterative procedure is repeated until the algorithm converges toward a value that is considered as stopping criterion. The experimenter may also set a maximum number of trials to terminate the experimental session before the stopping criterion is reached. Once the stopping criterion is reached, or the maximal number of trial occurred, Quest+ provides the values of the model parameters that fit best the experimental data. During an experiment, in order to motivate the participant and increase the speed of convergence, the experimenter may, from time to time, decide to randomly pick a stimulus among the weaker cases instead of choosing the stimulus that minimizes the entropy for the next trial.

6.4.2 Quest+ in a 2 alternative forced choice online experiment

We ran a simple discrimination task to estimate the perceptual threshold for differences in computer generated image quality using the Quest+ procedure. The experimental procedure received approval from the Ethical Committee in behavioral sciences of the University of Lille (Agreement n°2019-392-S78) and conformed to the standards set by the Declaration of Helsinki. Realistic image computation simulates the physical interactions of light between the objects, lights and cameras lying within a modelled 3D scene, a process known as global illumination (Kajiya, 1986). Numerous light paths are build per pixel and the mean of the samples for each pixel converges depending on the NSPP (Shirley et al., 1996). Importantly, the perceived image quality increases when the NSPP increases as some visual noise is perceivable if the NSPP is too low. In the present experiment we generated a set of images with a family of NSPP. Each stimulus presented during the experiment was either the

Table 6.1: Description of Quest+ adaptive psychometric method

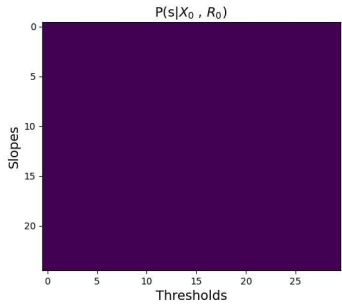
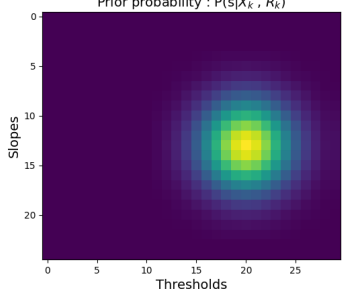
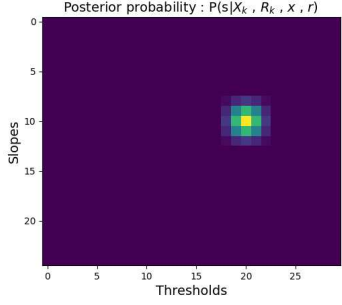
Method illustration	Description
	<p>Initialization This method considers that the initial prior probabilities of each parameter has a non-zero uniform value if no prior data is available.</p>
	<p>Trial k :</p> <p>Posterior probability For a trial k the posterior probabilities of the parameters is the product of the prior probabilities and the likelihood of the collected data.</p> $p(s X_k, R_k) = P(s) \prod_{i=1}^k p(r_i x_i, s)$ $p(r_i x_i, s) = \begin{cases} \Psi(x_i, s) & \text{si } r_i = 1 \\ 1 - \Psi(x_i, s) & \text{si } r_i = 0 \end{cases}$
	<p>On each trial, the posterior probabilities is a combination of the prior probabilities and the likelihood of every possible combination between all stimuli and both alternatives.</p> <p>Entropy Thus, after k trials, the entropy from the response r at stimulus x can be computed using the following equation</p> $H(x, r) = - \sum (P(s X_k, R_k, x, r) \cdot \log(P(s X_k, R_k, x, r)))$ <p>The stimulus that minimizes the entropy will be presented to the next trial x_{k+1}. This procedure is repeated until the entropy reaches a stopping criterion.</p>

image computed with the highest NSPP (the reference image, RI) or a composite picture in which a random part of the picture had the highest NSPP and another part of the image had a lower NSPP (the stimulus image, SI). A participant was asked to report whether he thinks the quality is the same across the whole picture or whether he notices a difference in quality in a 2AFC task. By experimentally manipulating the NSPP of the stimulus image one may estimate the individual psychometric function describing the relation between number of samples per pixels and the image perception. A Python implementation of the Quest+ method (Watson, 2017; Watson & Pelli, 1983) was used to choose the NSPP of the stimulus image for the next trial. To this end, the threshold and slope of a logistic function minimizing the entropy were computed after each trial.

Figure 6.2 displays the timeline of the experiment. The study starts with a set of written instructions. We also ask for the participant's consent. Then we implemented six training trials in which a feedback, a black line displayed to reveal where the two images would meet if there were a difference in quality, signalled the expected answer. Participants are asked to look at the image for as long as they want before pressing either the left or right arrow key to report whether they perceive a difference in image quality or not. No feedback was given for the remaining experimental trials. We chose a stopping criterion, i.e. a minimal value of entropy, based on other similar pilot experiments. For this particular experiment the minimum number of trials was set to 200 and the maximum to 250. As it is often recommended (Sauter et al., 2020) we added a checkbox for attention control during the experiment. At the end of the task, participants are asked to report the setting under which the study was performed (in the dark or not, etc.). We also had an instructional manipulation check (IMC) for estimating how well the participant understood the instructions (Oppenheimer et al., 2009). These additional steps were included because online participants may be more distracted along the course of the session than when physically present in the laboratory (Chandler et al., 2014) (Jun et al., 2017).

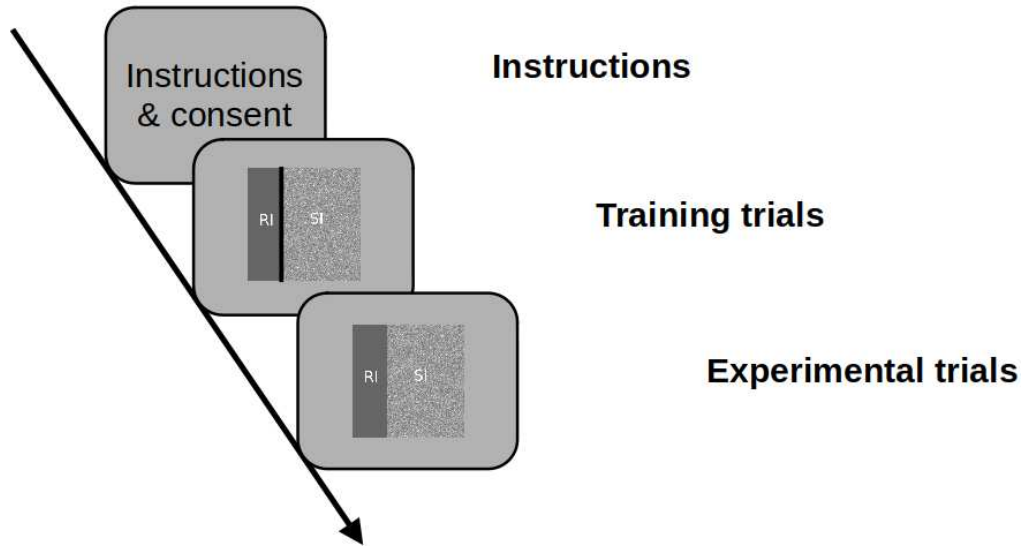
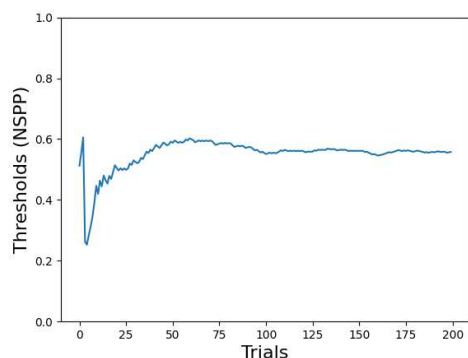


Figure 6.2: Timeline of the experiment. Participants first see a series of instructions. Then they experienced six training trials in which a feedback signals whether the answer was correct or not. In the experimental trials no feedback was delivered and the quality difference between the reference image (RI) and the stimulus image (SI) was adjusted by the Quest+ algorithm.

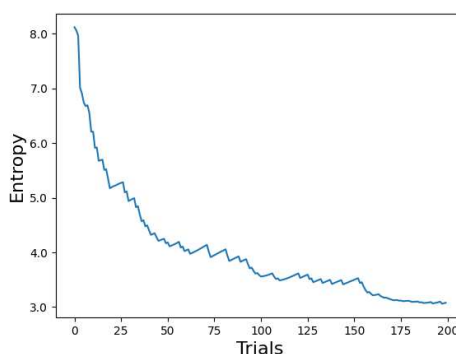
6.4.3 Results

After each trial we recorded the threshold and slope computed by Quest+, the entropy, the participant’s response and the decision time. The decision time was recorded on the client’s side and then sent and saved on the server’s side. Figure 6.3 (a) plots the evolution of the perceptual threshold normalized by the maximum NSPP with respect to the trial number. We estimated a convergence of the threshold after the 100th trial and the threshold remained stable for the remaining trials. This indicates that the Quest+ predicted threshold converged toward the actual perceptual threshold within the first 100 trials. Figure 6.3 (b) shows the corresponding evolution of the entropy. As can be seen, the entropy steadily decreased until the stopping criterion was reached at the 55th trial. Because we imposed a minimum number of trials the experiment continued until the 200th trials. The entropy decrease reflects

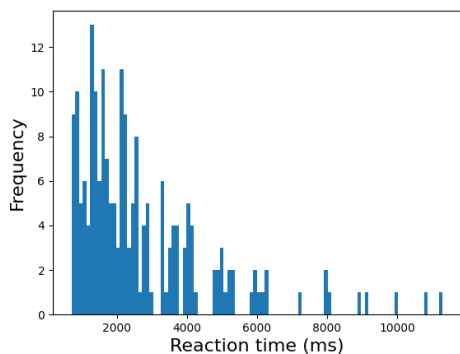
the minimization of the disorder of the model across trials and the convergence to the final values for the two parameters of our psychometric function. The final value for each of these parameters, the threshold and slope of the logistic function, were computed by Quest+. We used these values to plot the sigmoid function (figure 6.3 (d)). Figure 6.3 (c), shows the decision time distribution. The median decision time was 2138 ms (median absolute deviation = 877 ms) for this participant.



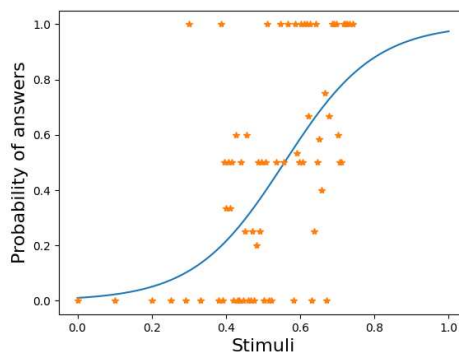
(a) Evolution of thresholds (normalized) across the experimental trials.



(b) Evolution of entropy across the experimental trials.



(c) Decision time distribution (ms).



(d) Probability of reporting seeing a uniform image as a function of the normalized NSPP. Actual data (dots) superimposed to the psychometric function with the threshold and slope estimated by Quest+ (solid line).

Figure 6.3: Summary of data collected in a single participant.

The observed pattern of change of the threshold and the diminution of entropy

is consistent with previously observed effects. Indeed, we have conducted similar studies in the laboratory and compared the results to online data across conditions (Myrodia et al., 2021). We found a significant similarity in the online and laboratory data.

6.4.4 Server / Client exchanges during the experiment

Figure 6.4 shows the HTTP exchanges between the server and the client during the experimental procedure. When a participant first clicks on our study's link, the internet browser sends a request to the Django server to open a new session. A unique identification (ID) number is then generated that will be saved on the client's side before the beginning of the study. The client identifier is stored locally at the level of the browser's local cache to be retained even after the browser or computer has been closed. At the beginning of the study, a Quest+ object is created and saved on the server's side and associated with the user's ID. This object is necessary for the whole study as it includes all the information used to update the algorithm along the course of the experimental session. During a trial the stimulus is displayed until the participant press a key. The choice and decision time are sent and saved on the server. The stimulus parameter and response are used to update the Quest+ object and choose the stimulus that will be presented next based on the computed entropy. This procedure is repeated until the algorithm convergences or until a specific number of trials is reached. The decision time is extracted and recorded on the client's side as the accuracy and precision of JavaScript in measuring reaction time has been proved reliable (de Leeuw & Motz, 2016; Semmelmann & Weigelt, 2017).

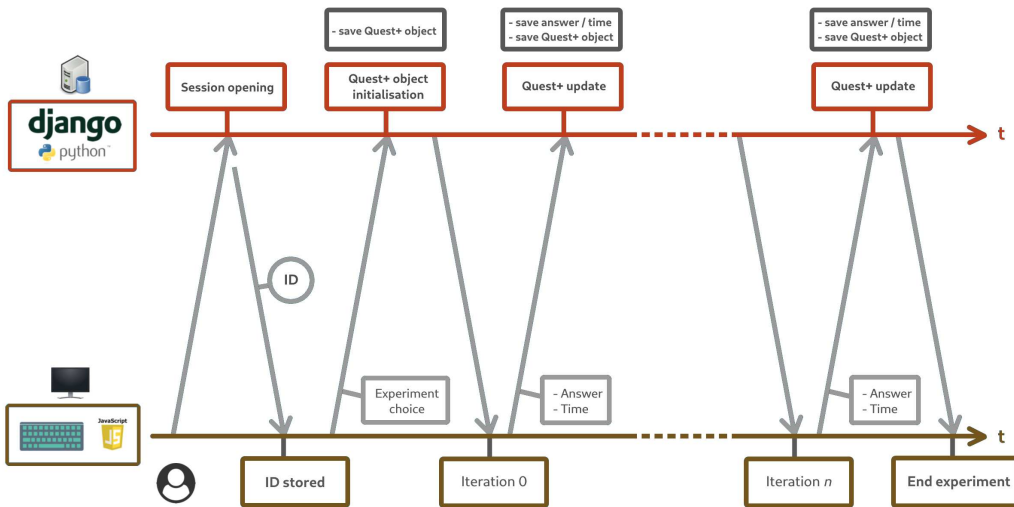


Figure 6.4: Description of the main server / client communications during the experimental procedure.

6.5 Conclusion

In the present article we have presented a general method to design and conduct online behavioral studies using programs developed in Python. We illustrated our solution with an online perceptual threshold measurement experiment using the Quest+ adaptive method. In our framework, most of the computation is done with Python using the Django web framework. The use of JavaScript is limited to display the stimuli and handle the participant interactions within the local browser. Although our solution might be more demanding at first than other pure JavaScript framework already available, it is a suitable option for many researchers willing to implement computationally demanding experimental procedures online. The main benefit is that it is possible to use any Python libraries or pre-existing functions as the Django server handles a version of Python identical to the one typically used on a local computer. Beyond obvious practical advantages, this opens new possibilities for behavioural online studies, expanding the range of procedures one may use

while allowing faster implementation and a greater similarity between lab-based and online methods.

7

Comparison of threshold measurements in laboratory and online studies using a Quest+ algorithm

7.1 Abstract

Online experiments have become popular and it is useful to test how data collected online compare to data measured in the laboratory. Here we compared perceptual thresholds of the perceived quality of CGI in a large-sample ($N = 174$) measured online, and a smaller-sample ($N = 71$) obtained in a laboratory-controlled study. Stimuli were three sets of CGIs picturing different scenes of the interior of an apartment. We found no significant difference between the two conditions ($p = 0.511$). We conducted a second study with repeated measurements in laboratory and online condition. The objective of this study was to define a pattern of thresholds and generalize the findings of a previous study across different scenes with various features. Observers participated in an experiment separated in 5 sessions with 5 indoor and outdoor scenes. We measured perceptual thresholds of the perceived quality of CGI in an online experiment ($N = 60$), and in a laboratory-controlled study ($N = 11$). For the online condition, the Kruskal-Wallis test showed significantly different thresholds between the five scenes ($p < 0.05$). The same significant difference was observed in the laboratory condition ($p < 0.05$).

7.2 Introduction

Many applications such as video games, films, architecture or design heavily rely on realistic computer-generated images (CGI). The main method to compute these images, the physically-based rendering (PBR), was developed by simulating the interactions of light with the textures and materials present in a scene (Pharr et al., 2016). To this end a Monte-Carlo approach based on the global illumination method (Kajiya, 1986) in which stochastic paths of luminance are generated from the camera's point of view toward the 3D scene has been proposed. Importantly, the quality of a CGI strongly depends on the number of paths or samples per pixel (NSPP) because the Monte-Carlo methods induce some visual noise that is inversely correlated to the NSPP : if the NSPP is too low, the algorithm did not converge and some random luminance colour values are attributed to a fraction of the pixels of the image. The appearance of this noise is illustrated Figure 7.1 which shows images at different stages of computation. There are currently no mathematical solution to predict the actual NSPP required to achieve convergence in a given image. This has important practical consequences as stopping the computation too early reduces the computing time but results in low quality images, while stopping the generation too late unnecessarily increase production costs.

Our research focuses on the perception of visual noise when evaluating CGI quality in an attempt to better describe the conditions under which the Monte-Carlo noise appears to be visible to human observers. Importantly, CGIs are used in everyday life and we are interested in exploring the perception of Monte-Carlo noise in ecological settings. We reasoned that conducting online experiments may therefore provide us with a more realistic description of the phenomenon at hand than conventional laboratory settings. Indeed, Monte-Carlo noise perception may be susceptible to the influence of many parameters that are typically neutralized in

laboratory studies such as the nature and configuration of the display, the viewing distance or the effects of ambient light during measurement.

Internet-based testing has several advantages including they are quick, less expensive, a researcher can collect data from a large sample or a representative population (Woods et al., 2015) and online studies have become very popular in the last decades (Gosling & Mason, 2015). There are many examples of successful online studies based on crowdsourcing such for visual search (de Leeuw & Motz, 2016) or reaction time (Sasaki et al., 2017). However, only a limited number of studies have been carried on on sensory perception yet, e.g. on size perception (Brady & Alvarez, 2011) or colour perception (Ware et al., 2019). Until now, there were some questions regarding the quality data collected online. The findings demonstrate that data from online studies can be equivalent from laboratory data (Germine et al., 2012; Gosling et al., 2004). The literature provides evidence that short and simple tasks can be run reliably online, for instance in a problem-solving task (Dandurand et al., 2008). Psychophysical measurements are demanding as it is necessary to implement rigorous control during the whole study. Findings from online behavioural studies are rarely compared to laboratory findings where the condition are controlled but some examples exist: Sasaki and Yamada conducted a contrast threshold study in both conditions and found a significant equivalence in the contrasts between the results of the laboratory and online repetition conditions, even though the thresholds were significantly higher in the online non-repetition condition compared to the laboratory one (Sasaki & Yamada, 2019).

Here we examined the suitability of online experiments on Monte Carlo noise perception task and we first directly compared the estimated thresholds of visual noise perception in an online condition to data collected in the laboratory. We used different scenes with various features to explore the perception of participants in a large variety of scenes. The aim of the second study was to explore the intra-subject effects of the scene on thresholds' estimation in both conditions (online, laboratory).

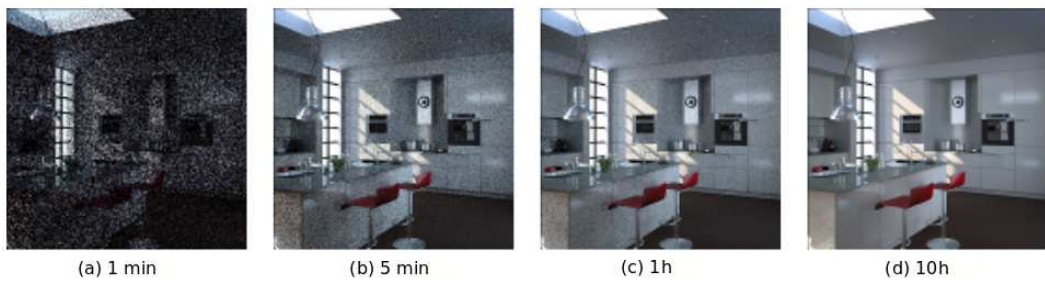


Figure 7.1: Example of a scene at different stages of computation. The visual noise decreases as the computing time increases.

The study was separated into 5 sessions with 5 different scenes presented in a random order.

7.3 Method

This study focused on measuring the visual perceptual threshold of noise in computer-generated images, measured in NSPP, for different scenes. We estimated these perceptual threshold online as well as in laboratory experiments.

7.3.1 Apparatus

In the laboratory condition, stimuli were displayed using Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997) for Matlab R2015 on a video monitor (Iiyama HM204DT, 100 Hz, 22 in.). The stimuli (768x800 pixels) were presented on a screen at a 768×1024 resolution. The observer's visual field was fixed using a chin-and-head rest at a viewing distance of 60 cm from the display. Participants were seated on an adjustable stool in a darkened and quiet room.

In the online condition, the experiment was conducted on python by using Django (Django Software Foundation, 2013), a python web framework. We used Django to

create an online study exclusively in python. The viewing distance and the lightness was not controlled and varied across the participants.

7.3.2 Stimuli

The scene stimuli (768x800) were 5 coloured, realistic computer-generated, indoor and outdoor scenes that were created using physically based rendering with various geometries, materials and lighting effects. Our image database consisted of images from 5 different scenes with a different per-pixel sampling for each scene. The generation of each scene was stopped when the visual noise was not perceived by a human observer. For each scene we obtained a set of images during the rendering process by saving the images at different stages of computation. For scene 1 we recorded a set of 400 images, ranging from 50 samples to 20 000 samples (i.e. with a 50-samples increase). Scene 2 ranges from 10 samples to 2 000 samples with a 10-samples increase (200 images). Finally, scenes 3, 4 and 5 range from 20 samples to 10 000 samples with a 20-samples increase (500 images). The images with the highest NSPP (i.e. the final image of a set) were considered as RI.

In both of the following studies, each RI was randomly cut either horizontally or vertically and the missing part of each image was replaced by the corresponding part of an image with lower NSPP (i.e. with more Monte-Carlo noise) to create a composite image. The location and NSPP of the noisier image was manipulated across trials. Images always were cut from side to side. To avoid large inequalities between the size of the two parts, the cut location was constrain to leave at least 100 pixels from an edge.

7.3.3 Procedure

Figure 7.2 (a) shows the timeline of the procedure in the laboratory condition. After the instructions and the examples, the participants initiated each trial by a fixation point in the centre of the screen for 200ms. The fixation point disappeared the instant the reconstructed configuration of a scene appeared. The scene remained on the display until the observers answered by using a joystick. Participants were asked to report by choosing “Left” or “Right” whether they see a single image with the same quality everywhere, or whether they think it is composed of two images with different qualities in a 2-Alternative Forced Choice (2AFC) task. Following the answer, the quality of the presented reconstructed scene is then adjusted according to the Quest+ algorithm (Watson, 2017).

Quest+ is based on the notion of entropy may help deciding when to stop testing. Prior to the actual experiments, we conducted pilot testing ($n=3$) to find the stopping criterion of the algorithm Quest+ by using the method of constant stimuli in the laboratory. We also defined the initialisation of thresholds and slopes with possible values. The Quest+ algorithm estimates the whole psychometric function but we decided to use the 50% perceptual threshold as the stimulus judged as “same quality” 50% of the time, i.e. the PSE or point of subjective equality (Macmillan & Creelman, 2005). In case that QUEST+ did not converge or converge too fast, we forced a minimum number of possible trials (100) and a maximum number of possible trials (600).

In the online condition, the procedures were identical to that of the laboratory conditions except for a couple of adjustments that we have conducted to ensure the data quality. The participants accessed the experiment either by the mailing list of the university or by the page of crowdsourcing. As it is presented in figure 7.2 (b), we first gave the instructions of the task. We then presented the participants with some examples from the easiest case to the more difficult. The participants could spend as

much time as they needed on each example and when they answered, we provided a feedback indicating the correct answer. If the example was a configuration of two images with different qualities, we presented a bold, black line indicating the limit between the two images. The last step before the beginning of the task was the consent form. After completing the experiment, we provided a 14-digit number or in the case of Prolific, they were redirected automatically to the validation page. With both of these methods, we could verify that the participant completed the task. In the online condition, we decided to reduce by 50% the size of each dataset. We did this by multiplying by two the step of each scene, so as to reduce the necessary time for the convergence of the algorithm Quest+ without a quality loss. For the stopping criterion, we used the entropy. In this case, we calculated the entropy's difference across the 10 last trials, if this difference was inferior to a limit, we considered that the algorithm converged. The minimum possible number of trials was 200 and the maximum 250. The average duration of the study was 20 minutes. The time limit (87 minutes) was calculated by Prolific and it depends on the estimation of the average duration for this study.

Another issue for online study is to deal with the boredom and attention liability in the participants (Chandler et al., 2014). To avoid low-quality data we added a checkbox for attention control during the experiment and at the end of the task, we asked some attention questions and questions regarding the task.

7.4 Study 1

7.4.1 Participants

For the laboratory study we had 71 Participants. (M age = 23, 82% women) with normal or corrected-to-normal vision who participated for course credit or

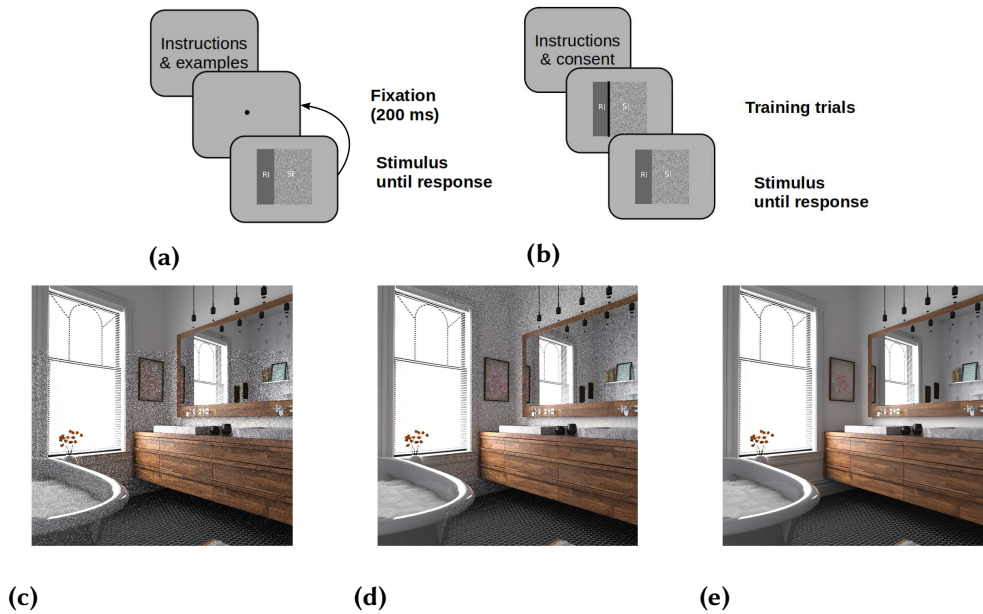


Figure 7.2: (a) Timeline of the experiment in the laboratory condition. For enhanced visibility, we presented the stimulus with the maximal difference in quality between the two subparts in this figure. (b) Timeline of the procedure in the online condition. Before the beginning of the task, we showed some examples to the participants to explain the task. For these examples, we used the same scene as in the main study and provided feedback to the participant with a bold black line in the position of the cut (if there was one). Examples of pictures that were created by combining two images with different qualities. (c) the picture is composed of two images with quite different qualities, the upper part originating from the reference image (i.e. with a large NSPP), the bottom part originating from a low NSPP image (50 NSPP). The visual noise is visible on the bottom part of the picture. (d) in the same scene, note that the picture is now composed of two vertical images, the noisy one being in the right part of the picture while the left part originates from the reference image. In this example, the noisy part consists of an image with 550 NSPP. (e) This image has the same quality everywhere, i.e. is the reference image (RI).

compensation. Among these participants, 33 were recruited specifically for this task and the remaining 36 participants took part in another study as well (these data are both reported here). Two participants were considered as outliers and their data were not analysed. For the online version, we recruited 174 participants (M age = 28.1, 55% women). Thirty four participants were considered as outliers and their data were not analysed. In this first study, we used 3 coloured, indoor scene images.

Each participant performed a single session with only one of the scenes. They were not aware of the objective of this study. The protocol of the study was approved by the Ethical Committee of the University of Lille (Agreement n°2019-392-S78). All participants gave informed written consent before the laboratory study. To obtain consent in the online version we gave analytical instructions before the beginning of the study, and asked participants to take part in the experiments only if they agreed to the instructions. We clarified that they do not have to participate in this study or may choose to leave the study at any time. We recruited our participants either from the University of Lille for course credit or from an active pool of potential participants named Prolific. Prolific provides a recruitment service and payment handling services anonymous and secured (Palan & Schitter, 2018). We also had the opportunity to make our study available only for participants with normal or corrected-to-normal vision.

7.5 Study 2

7.5.1 Participants

Participants were 11 for the laboratory study (M age = 20, 63% women) with normal or corrected-to-normal vision participated in this task for compensation. For the online version, we recruited 60 participants and we had 11 outliers (M age = 25.3, 34.7% women). We used the 5 coloured, indoor and outdoor scenes. Each participant saw all 5 scenes in separate sessions, in random order. They were not aware of the objective of this study. The study aimed at investigating the variability within-subjects across the five sessions of this study. We also aimed at estimating the perception thresholds of Monte Carlo with a variety of scenes with different features to generalize the results of the first study. The method that we used was the

same as in the first study except that measurements were repeated for each scene in 5 different sessions.

The protocol of the study was approved by the Ethical Committee of the University of Lille (Agreement n°2019-392-S78). All participants gave informed written consent before every session in the laboratory condition. For the online condition, participants had access to a new scene 24h after they had finished the previous scene. We gave analytical instructions before the beginning of the study and informed our participants that a bonus payment for their commitment was planned at the end of the 5 sessions.

We specified that it was possible to leave the study at any time. We recruited our participants only from the active pool of Prolific. Our study was available only for participants with normal or corrected-to-normal vision.

7.6 Data processing

Data analysis was performed in Matlab for the laboratory condition and in Python for the online condition. The first step was to detect and exclude the outliers for both conditions. As outliers, we considered participants who completed the study exceptionally fast (3 standard deviations below the mean) or exceptionally slow. For the online condition, we added crucial questions at the end of the experiment. Participant who failed 2 or more of these questions were excluded for further analysis. Finally, to remove participants that answered independently from the displayed stimuli, we used the maximum threshold calculated in the laboratory as an inclusion criterion. We considered this value as the upper limit of a possible threshold. All participants with a perceptual threshold higher than this value after the first 50 trials were considered as outliers. For the first study the maximum threshold in the laboratory was 0.94 and for the second study was 0.93.

Perceptual threshold data were recorded during the experiment and used to define the final threshold for each scene. Data must be normalized in some way because the scenes had a different NSPP during their generation. Each scene was normalized according to the NSPP of the reference image. The normalization allows us to compare the relative thresholds across different scenes and between the two conditions.

Performance was quantified by the stability of threshold evolution and entropy convergence. As our purpose was to examine whether the perceptual thresholds were significantly different for the two conditions, we conducted a permutation test and a Mann-Whitney equivalence test between the 2 conditions.

The median decision time for the online condition was not a precise variable to compare between the two conditions because of the variable connection delays to the server. We measured the velocity of convergence by estimating the convergence trial based on the time course of the threshold. For the second study, we did a pairwise comparison for all the measurements. We compared the thresholds and the convergence trial between the different scenes and sessions. We tested the correlation of thresholds among the five scenes and sessions. We conducted a pairwise comparison for the variable of the decision time only in the laboratory.

7.7 Results

7.7.1 Study 1

In the laboratory condition two participants were excluded because their decision time were exceptionally slow (Mean reaction time = 3.07 ± 1.72 and outliers mean reaction time = 9.83s) so that data from 69 observers were analysed. In the online experiment 34 participants were excluded, 17 because their threshold exceeded the

maximum threshold calculated in the laboratory (0.94 NSPP), 9 participants were excluded because they failed at least two attention check questions (ACQs) and 8 participants failed in both criteria. Data from 135 observers were analysed.

We measured the perceptual threshold of the participants for one of the three different scenes using the Quest+ method. The estimated thresholds in the laboratory and online experiments are shown Figure 7.3. First, we compared the results between the two conditions using a permutation test. Non-significance difference was assessed with permutation test, in which the data were shuffled for 1 000 iterations ($p = 0.511$).

To test for the similarities across conditions we used the Mann-Whitney equivalence test (Wellek, 2010). The corresponding tolerances of the ε'_v are $\varepsilon'_1 = \varepsilon'_2 = 0.1$ and the significance level $\alpha = 0.05$. The Mann-Whitney test yields the Mann-Whitney statistic $W+ = 0.5$, the estimated standard deviation $\sigma[W+] = 0.04$ and the critical bound $CMW(0.05, 0.1, 0.1) = 0.79$. Mann-Whitney test for equivalence leads to the rejection of the null hypothesis of non-equivalence, revealing that the equivalence between the thresholds in the laboratory and online conditions was significant.

To further analyse the observers perception, we plot in Figure 7.4 the evolution of the thresholds and entropy as estimated by the Quest+ algorithm for both conditions and 3 scenes. In both conditions there is a large between subjects variability in their final thresholds but not in the values of entropy. The entropy systematically converged to the limit that we had set or close to this limit. We noticed that the entropy continued to decrease until the end of the experiment in the cases in which the minimum number of trials was not reached. This might be explained by a change in strategy of answering during the experiment. The stopping criterion is calculated based on the minimum Shannon entropy. For laboratory conditions, we performed pilot tests prior to the actual experiments to define the entropy convergence limit [$SC_{scene1} = 5.46$; $SC_{scene2} = 4.59$; $SC_{scene3} = 6.50$]. For the online condition, we

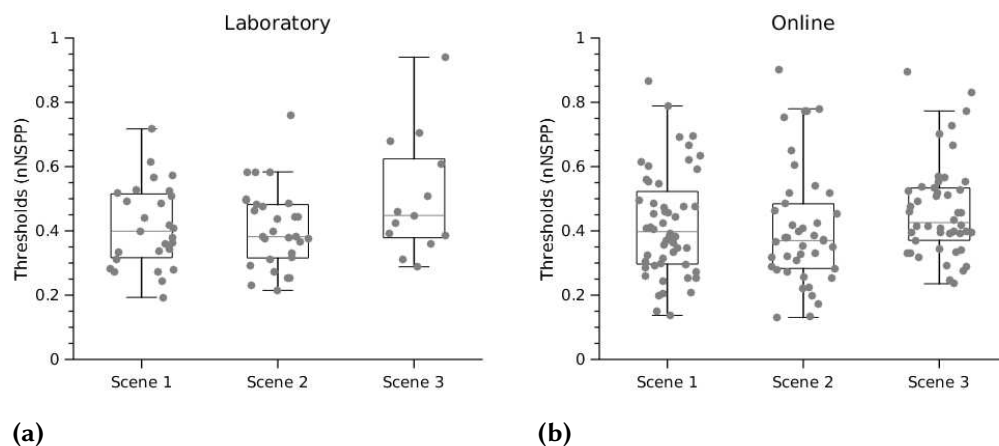


Figure 7.3: Results of the calculated thresholds (a) in the laboratory and (b) online experiments for 3 scenes. The NSPP was normalized between 0 and 1.

calculated the entropy's difference of the 10 last trials, if this difference was less than 0.05 we considered that the algorithm is converged.

The evolution of perceptual threshold for the 3 scenes revealed that the Quest+ algorithm estimated the threshold rapidly and that the threshold remained mostly stable until the end of the experiment. Stability is measured by finding the convergence trial of the thresholds' evolution. We calculated the mean evolution velocity of each subject. We considered as the limit of stability the mean plus 2 times the velocity standard deviation. For each scene and each participant, we compared the evolution's velocity with the stability limit. When the limit was reached, we noted the trial as the convergence trial.

Figure 7.5 plots the convergence trials for both conditions. A permutation test first assessed the non-significance difference of convergence trials between the online and the laboratory conditions. The convergence trials were shuffled for 1 000 iterations ($p > 0.05$). Then a Mann-Whitney equivalence test was conducted did not reveal a significant equivalence with values $W+ = 0.73$, $\sigma[W+] = 0.03$, $CMW(0.04, 0.1, 0.1) = 1.26$. To probe whether this result was due to the different sizes of the image set. As we explained before, in the online condition, we divided

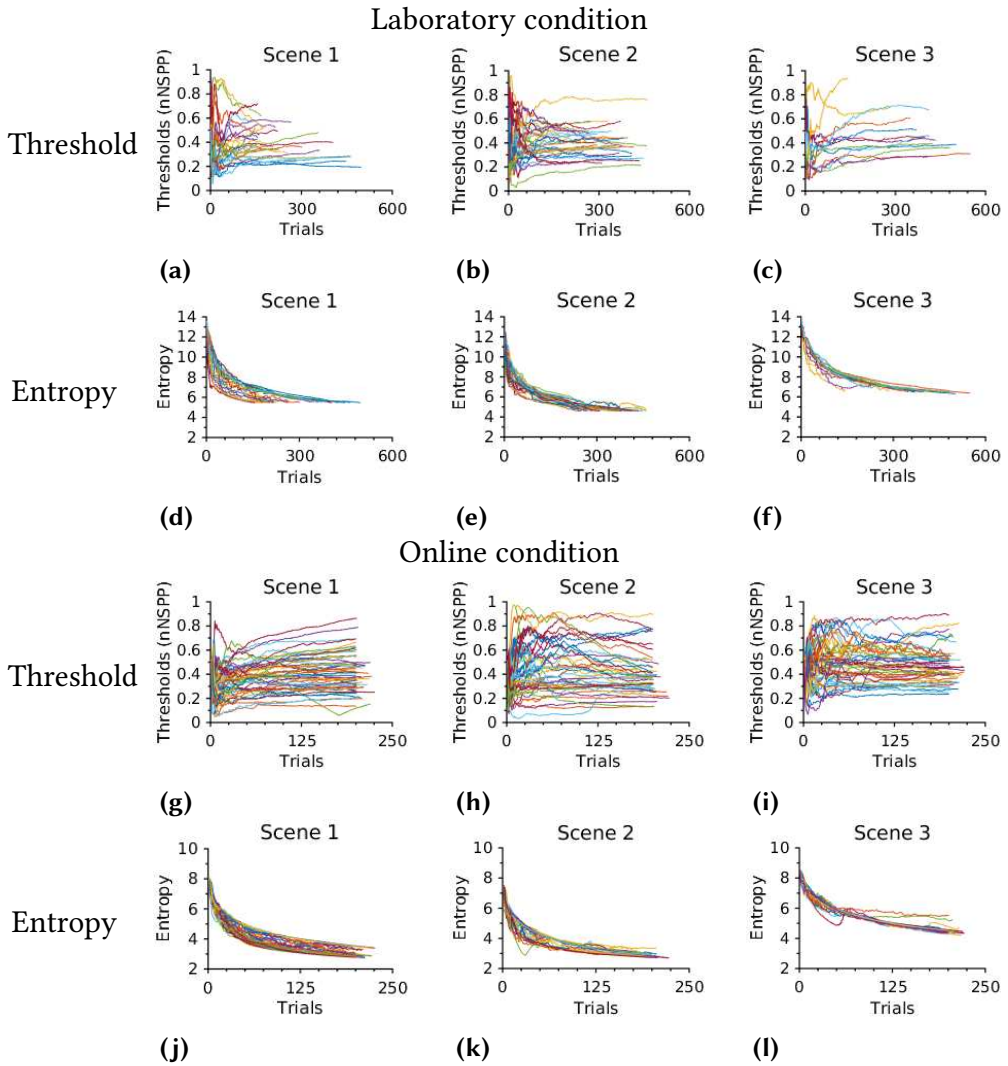


Figure 7.4: For the laboratory condition : (a-c) Evolution of the threshold for the 3 scenes along the course of the sessions. The NSPP was normalized between 0 and 1. Each colour is a participant. (d-f) Corresponding entropy. The entropy decreases until it reaches the convergence limit. For the online condition : (g-i) Evolution of the threshold for the 3 scenes along the course of the sessions. The sizes of the image sets were divided by 2. (j-l): Corresponding entropy. In this condition, the limit of convergence was calculated by considering the difference across 10 consecutive trials.

by two the size of the image sets to decrease the duration of the study. For the first scene, we had 200 images, for the second scene we had 100 images and for the third scene, we had 250 images. With the same parameters used in each actual conditions we simulated data from 120 participants per condition. We found that the velocity of convergence is independent of the size of the image set, as for the simulated laboratory participants we found that the convergence trial was 75 and for the online was 72.

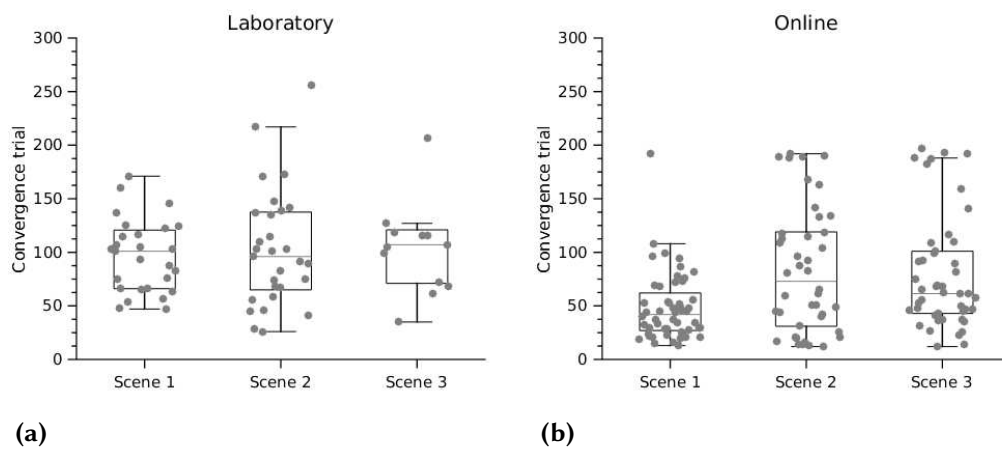


Figure 7.5: Convergence’s trials for the 3 scenes and for both conditions. In the online study the median convergence trial is earlier than in the laboratory.

7.7.2 Study 2

The objective of the second study was to define a pattern of thresholds and generalize the findings of the first study across different scenes. For this study, we collected data from 11 observers in the laboratory condition and 49 in the online version of the study. The exclusion criteria were the same as in study 1. We excluded 11 participants from the online study, 2 because they failed ACQs, 8 had thresholds exceeding the upper limit of threshold calculated in the laboratory (0.93 NSPP) and

1 answered exceptionally slow during scene 3 (Mean reaction time = 4.60 ± 2.17 and outlier mean reaction time = 19.31s).

This study is separated into 5 sessions : in each session, a different scene was presented to the participant. Each subject participated in all 5 sessions and the order that the scenes were presented to each participant was randomized. Figure 7.6 plots the calculated thresholds for the 5 scenes in the laboratory and online conditions. To define a pattern of thresholds we used a multiple comparison test.

For the online condition, a Kruskal-Wallis test revealed a significant difference ($p < 0.01$) between the calculated thresholds of the 5 scenes. A pair-wise multiple comparison test with Tukey's procedure showed that the mean ranks of thresholds for scenes 1, 2 and 3 are significantly higher compared to scene 4 ($p_{1-4} = 0.006$, $p_{2-4} = 0.003$, $p_{3-4} < 0.01$). Thresholds for scene 3 were significantly higher compared with scene 5 ($p_{3-5} = 0.004$). For the laboratory condition, the Kruskal-Wallis test showed a significant difference between the scenes ($p < 0.01$). The multiple comparison tests showed that the mean ranks of scenes 1, 2 and 3 are significantly higher compared with scene 4 ($p_{1-4} = 0.002$, $p_{2-4} = 0.002$, $p_{3-4} < 0.01$).

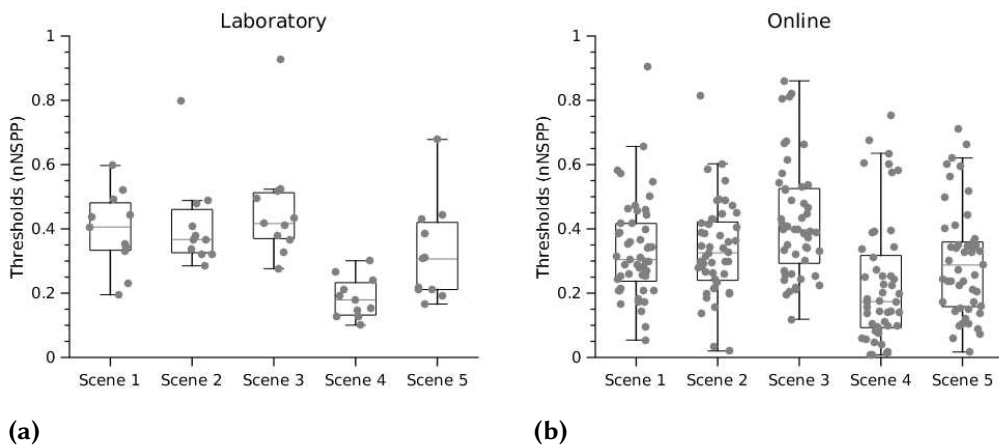


Figure 7.6: Results of estimated thresholds by Quest+ in the laboratory and online experiments for 5 scenes. In both conditions, each participant saw all 5 scenes.

We then compared the thresholds obtained in the 5 sessions. We supposed that

there could be a learning effect between the first and the last session. However, for both conditions (online or laboratory), we did not find a significant difference between the calculated thresholds of the 5 sessions ($p_{online} = 0.250, p_{lab} = 0.272$).

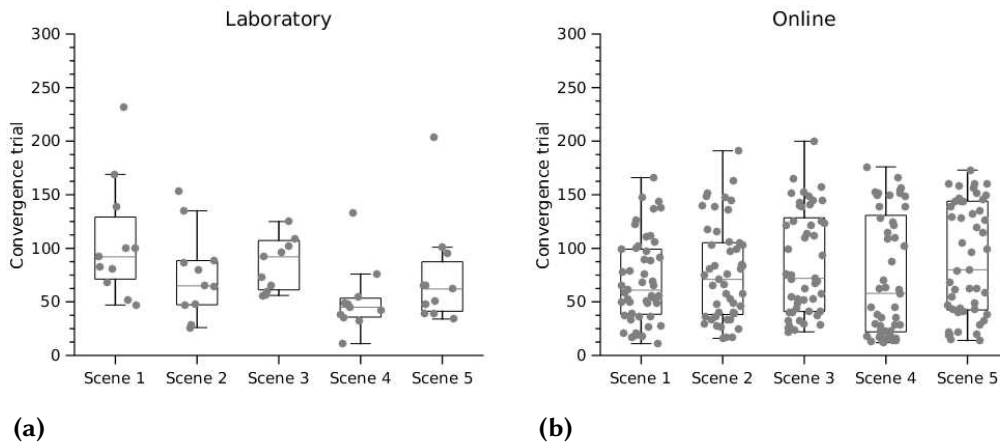


Figure 7.7: Results of the convergence trials in both conditions for 5 scenes.

For the convergence trial, we did not find a significant difference between the 5 scenes in the online condition ($p_{online} = 0.214$). However, for the laboratory condition, the Kruskal-Wallis test showed a significant difference ($p_{lab} = 0.012$). The multiple comparisons showed the convergence trials of scenes 1 and 3 are significantly higher than this of scene 4 ($p_{1-4} = 0.036, p_{3-4} = 0.016$). For the 5 sessions, in the laboratory and the online study we did not find a significant difference for the convergence trial ($p_{lab} = 0.119, p_{online} = 0.090$). For the laboratory condition, we tested the effects of the scenes on decision times. The Kruskal-Wallis test did not show any significant difference either for the 5 scenes or for the 5 sessions ($p_{scenes} = 0.946, p_{sessions} = 0.696$). We investigated the effect of the scene on the threshold measurement. Figure 7 shows the intra-subject variability of thresholds' estimation during the five sessions and the five scenes of this study. Tables 7.1 and 7.2 present the Pearson correlations matrix for the five scenes and sessions in online conditions. For the online condition, the correlations were significant except for scenes 4-1 and 4-3. For the laboratory data, we report the correlation, but due to the small number of participants the statistical

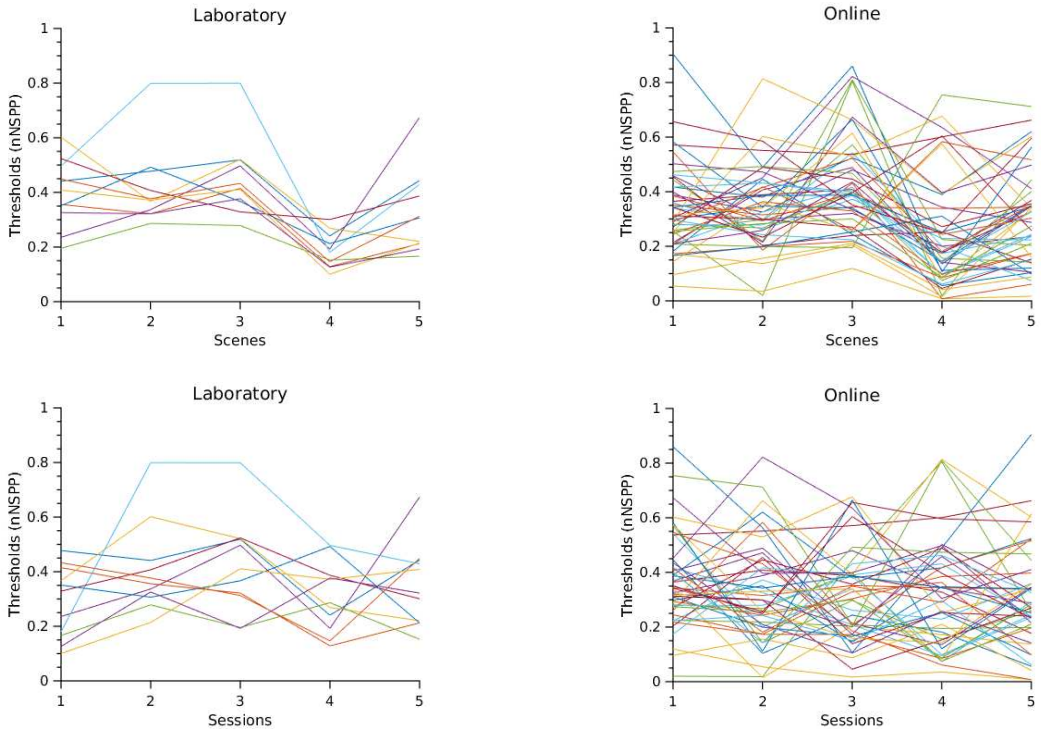


Figure 7.8: Raw data of thresholds from the five sessions and for the five scenes in both conditions. Each colour is a new participant and the line indicates the intra-subject variability.

power is low. We will just mention that we found a significant correlation between the scenes 2-3 ($R = 0.81, p < 0.05$) and between the sessions 2-3 ($R = 0.79, p < 0.05$).

7.8 Discussion

This research examined the perceptual threshold of visual noise in CGI with two implementations of the same task, one online and one in laboratory conditions. The aim of measuring perceptual threshold outside a controlled laboratory condition was to investigate noise perception under more ecologically realistic conditions. Online settings necessarily means that various monitors, environments and contexts may contribute to the actual psychophysical results. The objective of our first study was

Table 7.1: Correlation matrix for 5 scenes on online condition

Scene	1	2	3	4	5
1	-				
2	0.360*	-			
3	0.472*	0.284*	-		
4	0.002	0.343*	0.116	-	
5	0.356*	0.493*	0.337*	0.518*	-

Table 7.2: Correlation matrix for 5 sessions on online condition

Session	1	2	3	4	5
1	-				
2	0.350*	-			
3	0.172*	0.154*	-		
4	-0.130*	0.201*	0.289*	-	
5	0.315*	0.429*	0.318*	0.452	-

Chapter 7 Comparison of threshold measurements in laboratory and online studies using a Quest+ algorithm

to compare the measurements between the two conditions while the second study aimed at generalising the estimates of perceptual thresholds across different scenes with varying features. Before discussing some specific aspects of our data we will present some of the potential pitfalls associated with online studies.

Online research gives easy access to large participants pools via recruitment platforms in a short time (Gosling & Mason, 2015). As in the laboratory studies, it is necessary to develop some controls because quality assurance measures contribute to data quality (U. D. Reips, 2012; U.-D. Reips, 2000). There are several ways to improve data quality. For instance we have implemented ACQs during both online studies. An instructional manipulation check (IMC) is also mandatory for detecting participants who do not read carefully all the instructions (Oppenheimer et al., 2009), which might affect the final results. Also, the questions at the end of the study allowed us to probe the conditions under which the participants have completed the task. We excluded participants who self-reported low-quality answers or inappropriate conditions. We had also added a checkbox during the study to help participants to maintain their motivation while performing the task. In a multi-part online study is easier and more common for a participant to lose motivation. Our second study was more demanding as it was separated into 5 sessions so we offered a bonus payment when all 5 sessions were completed. Overall, we excluded 19.5% of the data from the first study and 18.3% from the second multi-part study.

The results of the first study showed equivalence in the measured thresholds online and in the laboratory, confirming that the data collected online robustly estimated the participants perceptual thresholds. The convergence trial was not significantly different between the two conditions, but participants in the online study seemed to converge faster than in the laboratory study and stayed stable until the end of the experiment. One may suppose that the online version of this study is easier and that participants were more consistent in detecting or ignoring a given amount of noise. This may happen because of the actual context under

which participants carried the experiments the study. The presence of ambient light, the distance from the screen and the screen brightness may influence the level of difficulty and it is possible that the increased inter-trials variations of these parameters may facilitate the overall noise discrimination.

The second study was a generalization of the first study using more varied indoor and outdoor scenes with different features. We noticed that the perceptual threshold of scene 4 was significantly lower than the thresholds of scenes 1, 2 and 3, indicating that the same general behavioural pattern may be observed in laboratory and online studies. The effect of the scenes on thresholds estimation was confirmed by the correlation analyses. For the online condition, we found a medium size correlation that is not uniform. However, it shows the tendency that a participant with a high threshold in one scene will have a high threshold in the other scenes. This relationship was not validated only for scenes 1-4 and 3-4, again pointing to a specificity of scene 4 in which the dark colours and the textures might constrain the visibility of the visual noise.

We confirmed that the level of difficulty depended on the scene and not on the order of the sessions as we did not find any significant difference when we tested the effects of the order. Moreover, the correlation between the thresholds of the five sessions did not show a regular correlation and we did not find a learning effect across the sessions.

The variable of decision time in the laboratory condition did not show any significant difference either for the 5 scenes or 5 sessions. It is noteworthy that the decision time was not considered for the online study as recorded the reaction time of each answer on the server's side. Even if the accuracy and precision of JavaScript in measuring reaction time is reliable, we could not benefit from this precision as we can not ignore the variability in the internet connections (de Leeuw & Motz, 2016; Semmelmann & Weigelt, 2017).

Results indicated that, given basic quality assurance measurements, we can collect

Chapter 7 Comparison of threshold measurements in laboratory and online studies using a Quest+ algorithm

reliable data from an online study, much as in a laboratory study. Moreover, we showed that the perceptual threshold for visual noise in CGI is strongly dependant on the features of the scene. This means that it is difficult to generalize the results from one scene to other scenes. Further work could explore more this effect by generating new scenes with different features to understand what are the properties in terms of luminance or textures that affects noise visibility. On the long term, these results could be used to implement an automatic stopping criterion for the generation of CGI.

8.1 Abstract

Rendering of computer generated images provides photorealistic images by simulating the light diffusion in a modelled 3D scene using Monte-Carlo estimation. This global illumination method is known to produce highly perceptible artefacts, i.e. visual noise if the number of samples per pixel is too low. A major issue is to decide when the computation can be stopped so that observers can not perceive the visual noise, but the characterization and measurement of this noise visible during the image rendering remain scientifically challenging. In this paper, we asked human observers to judge whether some noise was visible in the image and recorded their eye movements during the detection task to produce fixation maps. These maps were then compared to GBVS maps (Graph-Based Visual Saliency) to assess whether visible noisy parts of a scene might be predicted. We show that a GBVS map of a noise map obtained by differentiating two images with different noise levels might help to estimate the observers' fixation maps. Using these results, we propose an architecture that allows progressive rendering based on an approximation of human noise perception.

8.2 Introduction

Realistic image computation mimics the natural process of acquiring pictures by simulating the physical interactions of light between all the objects, lights and cameras lying within a modelled 3D scene. This process is known as global illumination and was formalised by Kajiya (Kajiya, 1986) with the rendering equation.

$$L_o(x, \omega_o) = L_e(x, \omega_o) + \int_{\Omega} L_i(x, \omega_i) \cdot f_r(x, \omega_i \rightarrow \omega_o) \cdot \cos \theta_i d\omega_i \quad (8.1)$$

where:

- $L_o(x, \omega_o)$ is the luminance travelling from point x in direction ω_o ;
- $L_e(x, \omega_o)$ is point x emitted luminance (it is null if point x does not lie on a light source surface);
- the integral represents the set of luminance L_i incident in x from the hemisphere of the directions Ω and reflected in the direction ω_o (two of these directions are shown in figure 1 as ω_{i1} and ω_{i2}). The reflected luminance are weighted by the materials reflecting properties (bidirectional reflectance function $f_r(x, \omega_i \rightarrow \omega_o)$) and the cosinus of the incident angle.

This equation cannot be analytically solved and Monte Carlo approaches are generally used to estimate the value of the pixels of the final image.

This equation is classically solved through Monte Carlo methods where sampling is performed through the build of random light paths between the camera and the light sources lying in the 3D scene: a ray is sent from the camera location through a pixel and is randomly reflected by the surface of the first object it encounters. The process is recursively performed until a light is reached or a Russian roulette process decides to stop the path. Numerous light paths are build per pixel and according to the law of large number and the Monte Carlo approach, the mean of the samples for each pixel converges to the solution following a $\frac{1}{\sqrt{N}}$ rate where N is the

number of samples per pixel (NSPP) (Shirley et al., 1996). Unfortunately, convergence often requires several hours (even days) of computation before a fully converged image is available, due to both the complexity of paths computation and the high number of samples that are required. Furthermore, a visual noise is perceivable if the generation of the image is stopped too early and the NSPP is too low (see figure 8.1), considerably affecting the perception of the image. Instead of increasing the number of samples per pixels it would therefore be economical to automatically stop the computation once the visual noise becomes undetectable by a human observer.

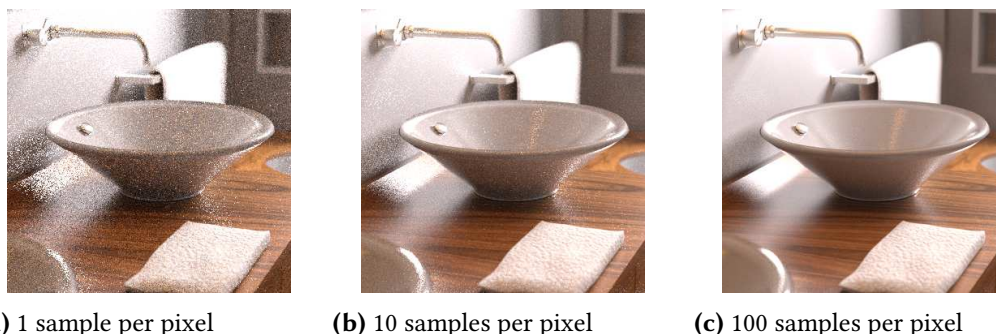


Figure 8.1: Effect of the number of light path estimation on visual noise.

In this article we propose a first step toward establishing a systematic method for optimized image computation by comparing where human observers look when searching visual noise in a CGI to a predicted visible noise map.

8.3 Previous works

Saliency maps were first introduced by Koch and Ullman (Koch & Ullman, 1985) as an attempt to account for the regularities of visual fixation patterns. Aspects of attentional control induced by the visual environment, i.e., bottom-up attention, are easier to grasp than those influenced by internal states as they mostly depend on the properties of the visual system. They proposed that the various visual

features contributing to attention allocation within an image (i.e. colour, orientation, movement etc.) may be combined into a single topographically oriented map, termed a saliency map, which integrates the normalized information from the individual feature maps into one global measure of conspicuity. Importantly, saliency maps usually attempt to reproduce key properties of human vision incorporating the specifics and limits of the human visual system.

A fairly large number of salience models currently exist, and Kümmerer et al. (Kümmerer et al., 2018) offered a saliency maps benchmark available online (Kümmerer et al., n.d.). The proposed saliency models have been designed for so-called *natural* images but saliency maps have also been applied to synthetic images as well. Hector Yee first used saliency maps to select the quality in global illumination render for animation (Y. Yee et al., 2002; Yee et al., 2001). In 2006, Longhurst (Longhurst et al., 2006) used GPU acceleration to compute real time saliency maps within a rendering system. Koulteris in 2014 proposed using a selective attention model named Level Of Detail (LOD) (Koulteris et al., 2014). Although the results appear promising, one challenge for applying saliency models to image rendering is that they generally account for estimating salient regions in an actual image, but they do not detect the Monte-Carlo artefactual noise generated during rendering. We propose to use noise maps obtained during rendering rather than the whole generated image itself when applying saliency maps to predict regions with perceptible visual noise. To this end we will use an extension of the Itti et al. influential model (Itti et al., 1998), the GBVS (Harel et al., 2007), which predicts human fixations with higher accuracy than the standard algorithms. These predicted regions will then be compared to regions that have been fixated by human observers to further probe whether saliency maps might account for noise perception.

8.4 Methods

We had 22 human observers performing a 2AFC task in which they had to decide whether they saw a single image with the same quality across the whole image, or whether they thought the picture was composed of two images with different qualities. The participants were separated in two groups. In the first group, we used one scene from the previous study 8.2 (a). We detected the areas more often (wall) or less often (floor) fixated and we varied either the texture of the wall 8.2 (b), or the texture of the floor 8.2 (c), or both 8.2 (d), to probe whether these changes in texture might affect where human observers gather visual information about the presence of visual noise within the image. The purpose of manipulating the texture is to attempt to control the location of the visible noise in the image. Indeed textures tend to visually hide the noise generated, known as the masking effect. In the second group, we presented 5 indoor and outdoor scene images 8.3. Each participant saw all 5 scenes in random order and during 5 different sessions. The five scenes were produced with different features (textures and colours).

For both experiments, in an actual trial, participants always saw a composite picture in which a random part of the picture had the highest NSPP and another part of the image had a lower NSPP. The NSPP of the lower quality image was individually set at 50% correct detection following a first experiment designed to estimate the individual psychometric detection functions (see figure 8.2). All experimental procedures received approval from the Ethical Committee in behavioural sciences of the University of Lille Agreement n°2019-392-S78) and conformed to the standards set by the Declaration of Helsinki. All participants gave written informed consent.

Participants were seated on an adjustable stool in a darkened quiet room, facing the center of the computer screen at a viewing distance of 60 cm. Head movements were restricted using a chin and forehead rest, so that the eyes in the primary gaze position were directed toward the centre of the screen. The experiment was separated in 2

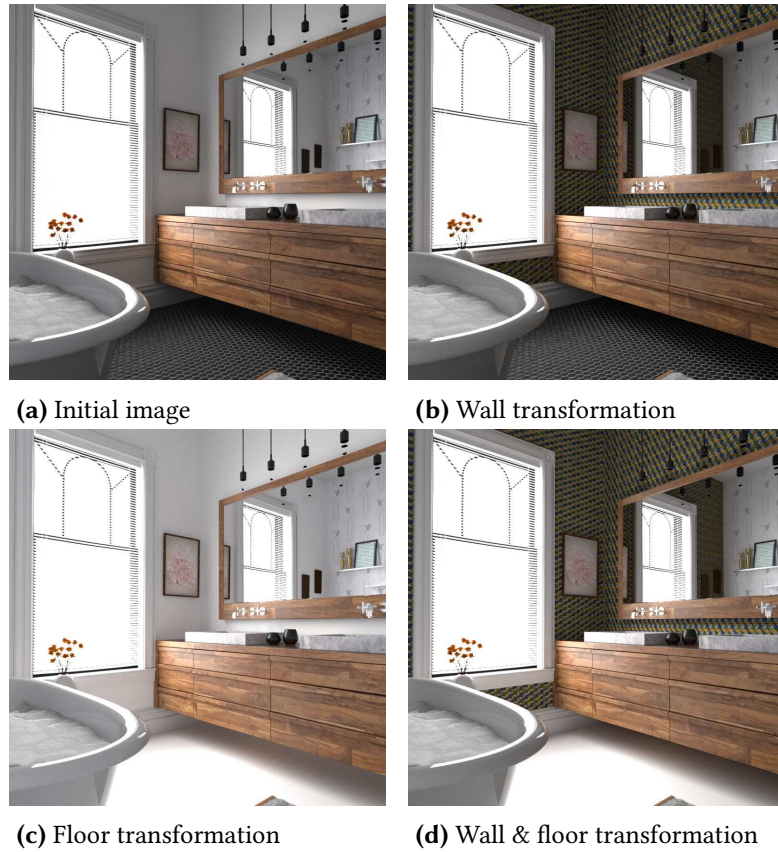
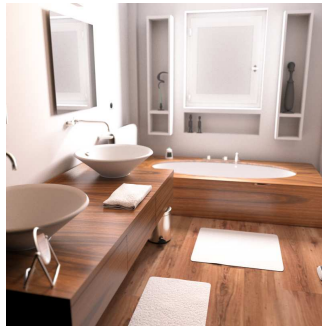


Figure 8.2: Same bathroom scene with some texture differences in order to change the location of the generated noise.



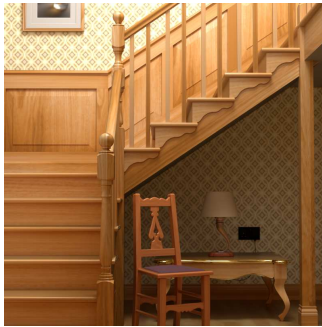
(a) Scene 1



(b) Scene 2



(c) Scene 3



(d) Scene 4



(e) Scene 5

Figure 8.3: Five different inside and outside scenes

tasks. The goal of the first task was to estimate the individual perceptual threshold for the visual noise. To assess the perception of the image quality, we asked observers whether they saw a single image with the same quality across the whole image, or whether they thought the picture was composed of two images with different qualities in a 2AFC task. Participants always saw a composite picture in which a part of the picture had the highest possible NSPP and another part of the image had lower NSPP. The participants reported their choice by pressing the left or right key of a joystick. This adaptive process was repeated until the NSPP reached a value corresponding to the 50% perceptual threshold of the observer.

In the second task, the goal was to define fixation locations when searching for visual noise. Participants were instructed to perform the same 2AFC task but now the NSPP of the comparison image was set according to the measured 50% perceptual threshold. Eye movements were measured continuously with an infrared video-based eye-tracking system (EyeLink, SR Research), sampled at 2,000 Hz. Viewing was binocular, but only the right eye position was recorded and digitized in both the vertical and horizontal axes. Eye movements were recorded while observers performed the 2AFC task so as to construct fixation maps quantifying the exploration of the scenes. These fixation maps may therefore be viewed as probability maps for extracting visual information regarding the presence of visual noise within each scene. In an actual experiment, each observer performed 288 trials for each scene.

8.5 Results

Fixation locations and durations were averaged across participants for each image, Z-transformed and smoothed with a two-dimensional Gaussian Kernel function:

$Kernel \sim \mathcal{N}(O, \sigma^2 I)$, where I is the input data matrix and $\sigma/2 = 1^\circ$ visual angle (Caldara & Mielliet, 2011; Lao et al., 2017).

Saliency maps were generated using GBVS (Harel et al., 2007). The GBVS algorithm extracts image feature vectors for each channel, forming an activation map using the feature vectors and then normalizing the activation map.

For each image we produced the saliency maps of the visual noise (i.e. noise saliency maps or NSM). We extracted a noise map by calculating the deviation between the first and the last image generated during the rendering process. With this method, we removed all the structure-related features of the image and kept only the visual noise in the different regions of the image. The noise map was then compared to the fixation map. Finally, we compared the fixation maps to the GBVS map computed over the first image of the dataset. To quantify the differences between the fixation maps and the noise maps or saliency maps we used the classic method of normalized root mean square error (nRMSE).

Figure 8.4 plots the resulting maps and the comparison scores between the fixation map and the 3 maps obtained from noise analysis and GBVS. The comparison between the fixation map and the NSM revealed a systematically lower error score than for the noise map or when the saliency map was computed over the whole image. In our task, participants had to search for visual noise and the saliency maps over the noise map capture part of the regions that were actually fixated. Importantly, the NSM also reveal regions where some visual noise is present that has not been fixated by the human observers. This indicate that the participants accurately looked at regions of the scene where the visual noise is easily detectable but ignored regions where some noise is present.

Figure 8.5 shows the fixation maps of human data, the 3 other maps and the comparison scores. The comparison between the fixation map and the NSM revealed a lower error score than for the noise map or when the saliency map was computed over the whole image in all scenes apart from Scene 3. We suppose that for this scene the saliency attracts more fixations than the noise because the saliency map

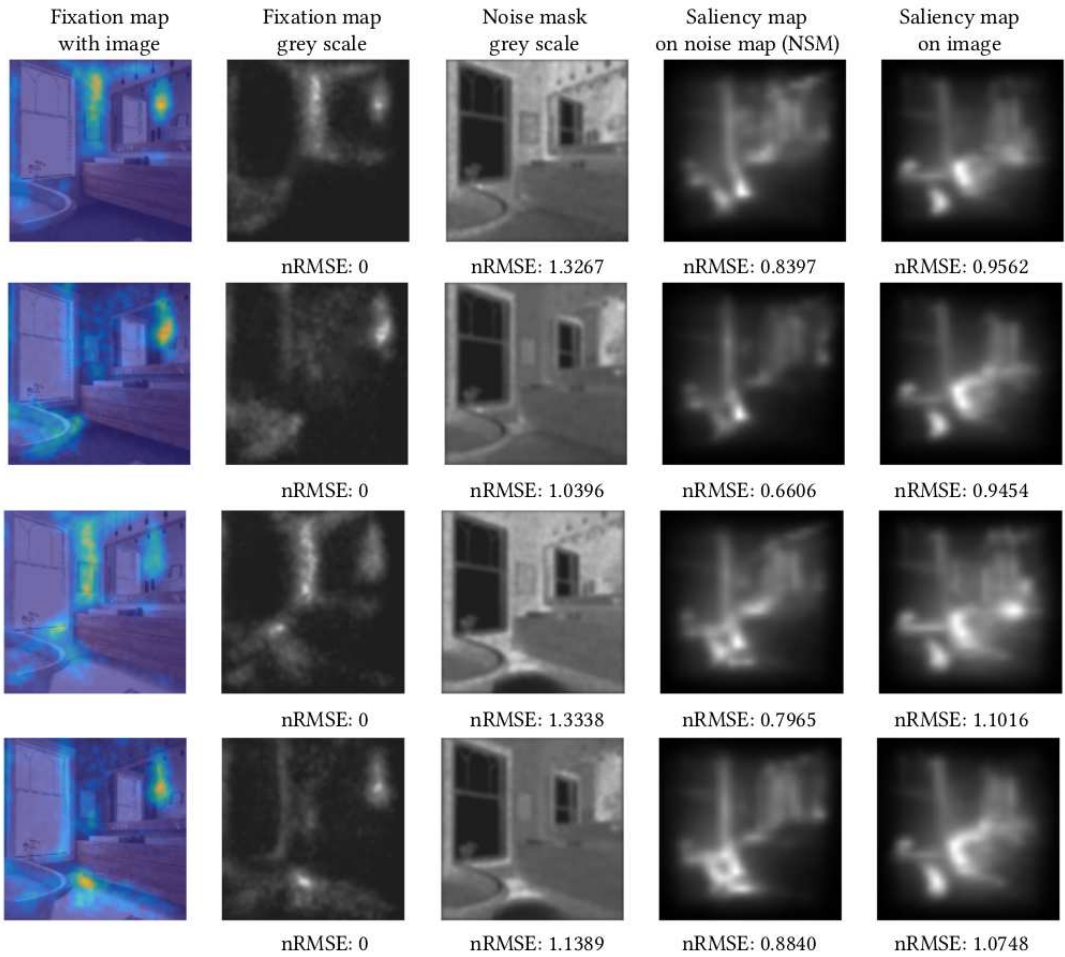


Figure 8.4: The four images used in the experiment (group 1). From left to right, a sample picture with fixations, the fixation map in grayscale, the noise map, the saliency map formed when using GBVS on the noise map (NSM) and the GBVS map computed over the image.

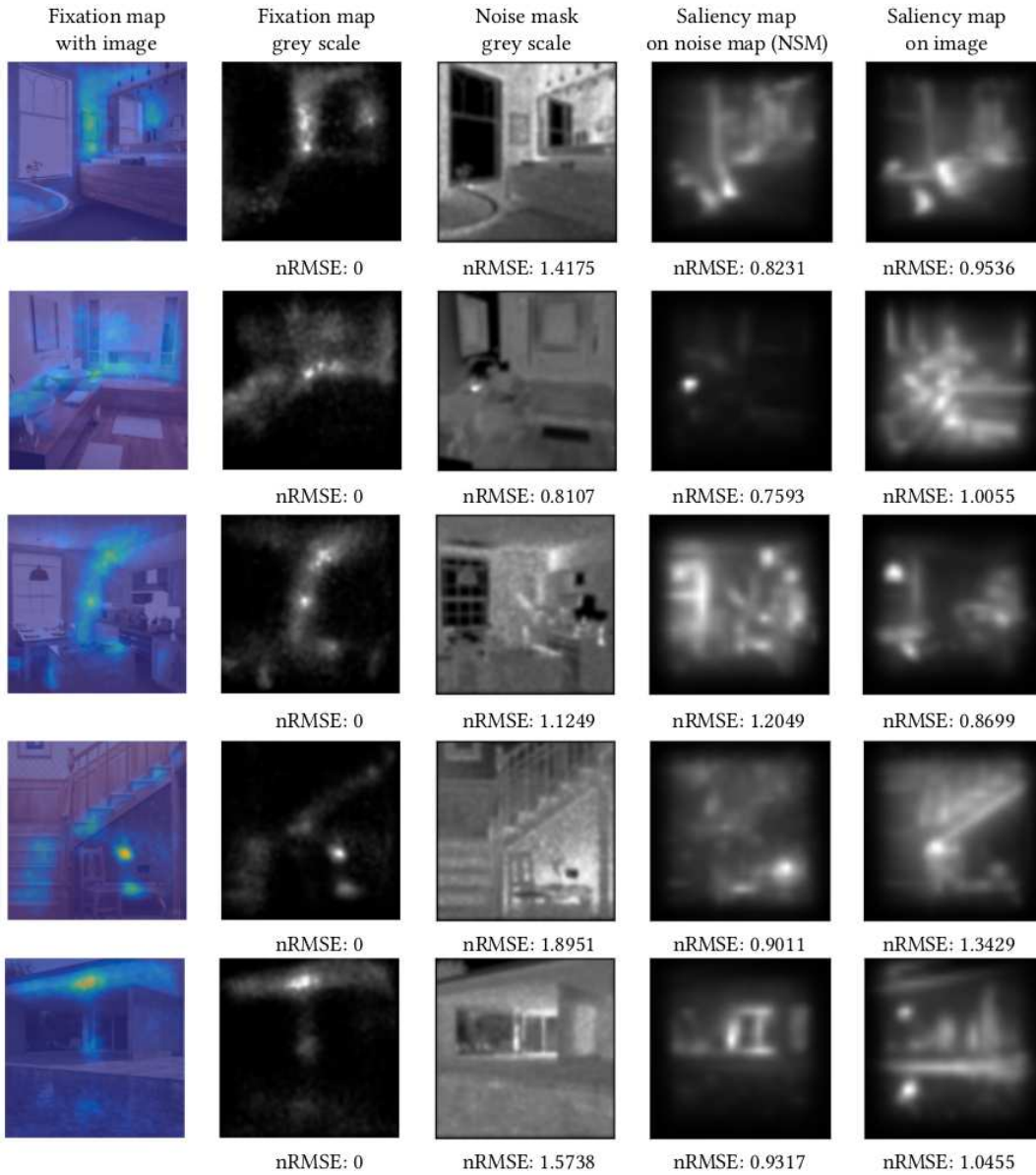


Figure 8.5: The five images used in the experiment (group2). From left to right, a sample picture with fixations, the fixation map in grayscale, the noise map, the saliency map formed when using GBVS on the noise map (NSM) and the GBVS map computed over the image.

which was computed over the whole image had a lower score than the other two maps.

8.6 Discussion

We have shown that applying the GBVS map model on the Monte Carlo noise estimated by the difference between two images allows us to better estimate the fixation maps obtained from human observers with a score of 88.8%. Also, we found that not all the noisy parts of a scene are fixated equally. We noticed that the noise depends on the textures and the colours of the scene but more tests are needed to define how each feature influences the perceived noise. As the fixation maps extracted during the experiment indicate where the noise is visible, we propose to implement an automatic stopping criterion adapted to human vision using these saliency models. The figure 9.1 illustrates how such a NSM could be implemented to attenuate human perceptible noise when rendering an image. It relies on an iterative process allowing to decide whether some noise is still visible within the image at any point during the rendering process. The proposed architecture relies on the comparison of the last two NSM, obtained from the last two computed difference images. A nRMSE score specifies how close the current NSM is to the previous one (obtained n samples earlier). nRMSE scores can be estimated as soon as $k + 2 \times n$ samples, i.e. after computing the initial image with a high noise level and the first two noise saliency maps, and compared to a stopping criterion.

Part III

Conclusions

In this thesis, we investigated diverse aspects of noise perception in CGI. The main objective aimed at better understanding the perception of visual noise and contributing to finding a stopping criterion for the stochastic computation algorithms for image synthesis by taking into account the human visual perception abilities. We will first summarize the main findings and, also, the limitations during these studies. Finally, we will present some perspectives for future work based on the results of our studies.

9.1 Main findings

As we mentioned before, the stochastic calculation algorithms generate noise in the image, which decreases when the NSPP increases. This process aims at optimal accuracy and precision but it may require a very long computing time to achieve high-quality images. The first task in all our experimental studies was always to estimate the perceptual threshold for noise perception in CGIs. Because Monte Carlo noise visual search is not conventionally studied we implemented and tested an original paradigm to estimate relative perceptual threshold. We assume that participants do not have a pre-existing knowledge of what visual noise is or the areas of a scene where noise is more likely to be detected. When searching for objects in real-world scenes, scene context may contribute to directing attention toward regions containing a target (M. L. H. Võ, 2021; M. L. H. Võ et al., 2019). In our case, we create a configuration of two images with a different level of noise from the

same rendering process. One part comes from the reference image (RI), i.e. the last image generated by the algorithm, and the other part comes from an image that we recuperated during the rendering, i.e. lower quality image than the RI. The two parts were cut randomly in order to create the whole scene. We asked our participants to evaluate whether they were capable of detecting a difference in quality or not, a task that should not depend on a subjective understanding on what constitute visual noise.

We examined also the suitability of online experiments on noise perception task (chapter 7). We conducted similar visual experiments in the laboratory and online using different scenes. The use of complex scenes from everyday life allowed us to study visual attention by implementing an ecological protocol and approach conditions that are natural as participants viewed the images in their natural environment. One of our contributions concerns our technique to develop the online version of our study (chapter 6). We implemented the same paradigm that we have used for laboratories studies using Python language and a python framework named Django. We used our laboratory method with only limited adaptation using known Python libraries. One great benefit of using a Python-based web framework for online studies is that it is possible to use any Python libraries or pre-existing functions such as the adaptive psychophysics methods, in our case we used Quest+. This implementation allows every researcher who wants to implement an online study to compare the results with the laboratory data with the same precision. We give free access to the necessary code and functions of the whole study through GitHub.

The second contribution concerns the obtained results. The results from this task showed that the selected scenes' boundaries include the threshold values and the thresholds are lower than the NSPP used to obtain the final image. These results open the way to integrating these thresholds into rendering algorithms with the final objective to a high-quality image (Buisine et al., 2021). We showed that the estimated thresholds are not significantly different between the online and the

laboratory condition. estimated thresholds are significantly equivalent online and in the laboratory. This is a significant finding for psychophysical studies as it implies that one can collect high-quality data from an online study. This finding could be very useful for rendering CGI that will be used in real life. We can optimize the rendering process without detectable loss of perceptual quality.

We found that the size of a gaze-contingent mask influenced how observers explored the scenes. Also, when the mask size increased and less noise was visible, the number of fixations increased. This finding could be considered as a contradiction to the results of Loschky et al. that showed that smaller windows led to longer fixations (Loschky & McConkie, 2002). However, one should realize that, in our experiments, the task of the participant was specifically to search for visual noise. This might explain why when the mask size was large enough to cover the noisy areas of a scene, the search and the duration of fixations were longer.

This study has an important implication. Real-time rendering is widely used, for instance in video game. The idea of computing until high quality is achieved for the minimal necessary area of a scene could improve the level of realism at the optimal rendering duration. Indeed our results showed that projecting a high-quality image on a small area of central vision is sufficient to impair the detection of differences in quality. We propose that this limit could be integrated into real-time rendering methods as a technique to adapt computation to preferentially target the areas that might be seen with central vision.

Furthermore, the analysis of gaze patterns showed that participants fixated the same areas regardless of the size of the mask. The areas more often or less often fixated depended on the features of each scene. This is why it is difficult to predict a general pattern of scan paths. We noticed that the areas without texture and with bright colours were more often fixated than the areas with textures and dark colours. To further explore these ideas and generalize our findings, we investigated the effect of the scene and the effect of the features of one scene in chapter 8.

In chapter 8, we recorded the participants' scan-path of scene exploration with five different scenes. The scenes showed indoor and outdoor natural scenes and every participant evaluated all five scenes. We found that all observers looked at the same areas and that these areas depended on the scene. More precisely, one of the scenes we used had few objects and dark textures and colours except for one object that was white. The results revealed that the average perceptual threshold of this scene was lower than the ones for the other scenes and that the colours and textures influenced the number of fixations. Even if it is difficult to generalize a pattern for all scenes, these findings could be helpful to categorize the scenes features and predict the fixations according to these visual features.

We selected one particular scene and we determined the areas more often (wall) and less often (floor) fixated. We modified these areas by adding or removing a texture to test how these transformations could affect fixations. We found that these changes influenced the fixation locations of our participants.

All these fixation maps were compared to GBVS maps computed on the noise map. We showed that applying the GBVS map model on the Monte Carlo noise estimated by the difference between two images better match the fixation maps obtained from human observers with a score of 88%. However, this study revealed that, in some scenes, salience attracts more fixations than noisy parts. Further testing is needed to understand why this occurs.

Finally, we found that observers did not fixate every salient noisy region of a scene. This could contribute to the prediction of the regions of interest for each scene and improve the computing time of production. In the next section, we will describe how this new idea of using a noise map as an input for computing a saliency map may be exploited to limit the presence of visible noise based on the HVS properties.

9.2 Perspectives

The results of this project may offer many perspectives for future studies. First, it seems necessary to test even more scenes with other features, apart from colour and texture, such as depth of view, orientation, and geometry of objects. This is necessary to better understand how noise perception may be affected by the image features and among them which are more or less likely to affect eye movements.

In the present work the estimated perceptual thresholds were obtained with naive human observers. The differences between expert and naive observers is rarely considered or quantified. The experts could belong to different domains such as graphic designer and photographers. Expertise can have a profound impact on how attention is allocated during search because it involves knowledge of multiple relations between features and image quality. Expertise should influence both the criteria for assessing the quality of a CGI and the strategy of visual search.

Two other aspects may also be considered for further studies. The first is to continue the analysis presented in chapter 8 to compare the fixation maps with the GBVS maps. The aim is to integrate the saliency of noise maps in the rendering process. This technique could be useful to define a stopping criterion of the rendering process by using a saliency map calculated on a noise map (i.e. deviation between two images with different qualities).

An hypothetical method to use stopping criterion during rendering process using a saliency map model applied to a noise map is described on figure 9.1. First an initial image is generated using k samples. Then, a fixed n number of samples are provided to output a new image (a more converged one). The absolute difference comparison is obtained from the initial image and the current one. A noise saliency map is then computed from the difference using a specific saliency model. For the first iteration, the rendering will continue to samples until at least two noise saliency maps are obtained. The nRMSE score between the last two obtained saliency maps

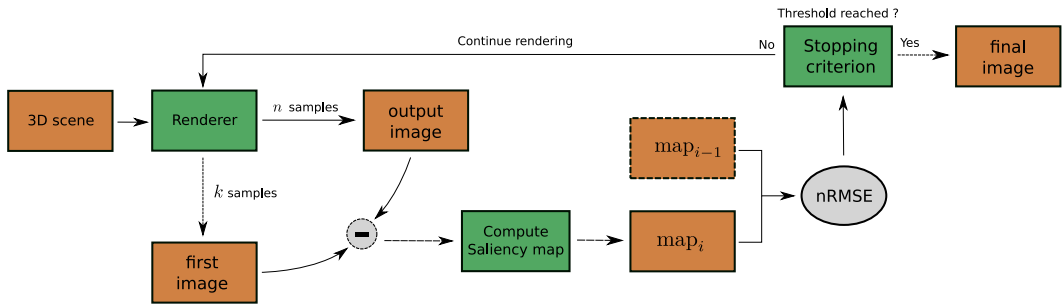


Figure 9.1: Stopping criterion applied during rendering process using noise saliency map model.

is then computed. Depending of a threshold score criterion the rendering process is stopped or otherwise iterated to further improve the image quality.

To validate the proposed rendering pipeline a good test would be to produce images with several defined nRMSE thresholds which may then be presented to human observers. Computing the saliency maps on noise mask is a new idea that has the benefit of taking in account the characteristics of human vision. Moreover, the stopping criterion would specifically target the perceptible visible noise and not the overall structure of the image. Our proposed rendering pipeline would therefore generalize regardless of the image characteristics such as the textures, luminance, colours, orientation, etc.

Finally, in the longer term, it seems interesting to expand our studies to different type of images such as 3D scenes and real-time rendering. Both of them are widely used nowadays, so the improvement of the quality of these images could have a significant impact for optimizing CGI rendering.

Bibliography

- Abrams, R. A., Meyer, D. E., & Kornblum, S. (1989). Speed and Accuracy of Saccadic Eye Movements: Characteristics of Impulse Variability in the Oculomotor System. *Journal of Experimental Psychology: Human Perception and Performance*, 15(3). <https://doi.org/10.1037/0096-1523.15.3.529>
- Al-najjar, Y. Y., & Soong, D. C. (2012). Comparison of Image Quality Assessment: PSNR, HVS, SSIM, UIQI. *International Journal of Scientific & Engineering Research*, 3(8), 1–5. <http://www.ijser.org/researchpaper%5CComparison-of-Image-Quality-Assessment-PSNR-HVS-SSIM-UIQI.pdf>
- Anderson, C. (2015). Docker [software engineering]. *Ieee Software*, 32(3), 102–c3.
- Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, 52(1). <https://doi.org/10.3758/s13428-019-01237-x>
- Appel, A. (1968). Some techniques for shading machine renderings of solids. *Proceedings of the April 30–May 2, 1968, spring joint computer conference on - AFIPS '68 (Spring)*, 37. <https://doi.org/10.1145/1468075.1468082>
- Arvo, J., & Kirk, D. (1990). Particle transport and image synthesis. <https://doi.org/10.1145/97879.97886>
- Ashby, J. (2006). Prosody in skilled silent reading: Evidence from eye movements. *Journal of Research in Reading*, 29(3). <https://doi.org/10.1111/j.1467-9817.2006.00311.x>
- Aust, F., Diedenhofen, B., Ullrich, S., & Musch, J. (2013). Seriousness checks are useful to improve data validity in online research. *Behavior Research Methods*, 45(2). <https://doi.org/10.3758/s13428-012-0265-2>
- Avcibas, I., Sankur, B., & Khalid, S. (2002). Statistical evaluation of image quality measures. *Journal of Electronic Imaging*, 11(2), 206. <https://doi.org/10.1117/1.1455011>

- Bahill, A. T., Clark, M. R., & Stark, L. (1975). The main sequence, a tool for studying human eye movements. *Mathematical Biosciences*, 24(3-4), 191–204. [https://doi.org/10.1016/0025-5564\(75\)90075-9](https://doi.org/10.1016/0025-5564(75)90075-9)
- Baloh, R. W., Sills, A. W., Kumley, W. E., & Honrubia, V. (1975). Quantitative measurement of saccade amplitude, duration, and velocity. *Neurology*, 25(11). <https://doi.org/10.1212/wnl.25.11.1065>
- Barbot, A., Landy, M. S., & Carrasco, M. (2011). Exogenous attention enhances 2nd-order contrast sensitivity. *Vision Research*, 51(9). <https://doi.org/10.1016/j.visres.2011.02.022>
- Bassi, S. (2007). A primer on python for life science researchers. *PLoS computational biology*, 3(11), e199.
- Becker, W. (1989). The neurobiology of saccadic eye movements. Metrics.
- Becker, W., & Jürgens, R. (1979). An analysis of the saccadic system by means of double step stimuli. *Vision Research*, 19(9). [https://doi.org/10.1016/0042-6989\(79\)90222-0](https://doi.org/10.1016/0042-6989(79)90222-0)
- Bertera, J. H., & Rayner, K. (2000). Eye movements and the span of the effective stimulus in visual search. *Perception and Psychophysics*, 62(3). <https://doi.org/10.3758/BF03212109>
- Biemer, P. P., Murphy, J., Zimmer, S., Berry, C., Deng, G., & Lewis, K. (2018). Using Bonus Monetary Incentives to Encourage Web Response in Mixed-Mode Household Surveys. *Journal of Survey Statistics and Methodology*, 6(2). <https://doi.org/10.1093/jssam/smx015>
- Binder, K. S., Pollatsek, A., & Rayner, K. (1999). Extraction of information to the left of the fixated word in reading. *Journal of Experimental Psychology: Human Perception and Performance*, 25(4). <https://doi.org/10.1037/0096-1523.25.4.1162>
- Birnbaum, M. H. (2004). Human research and data collection via the internet. *Annual Review of Psychology*, 55. <https://doi.org/10.1146/annurev.psych.55.090902.141601>
- Bolin, M. R., & Meyer, G. W. (1998). Perceptually based adaptive sampling algorithm. <https://doi.org/10.1145/280814.280924>
- Bonitz, V. S., & Gordon, R. D. (2008). Attention to smoking-related and incongruous objects during scene viewing. *Acta Psychologica*, 129(2). <https://doi.org/10.1016/j.actpsy.2008.08.006>

- Bosnjak, M., & Tuten, T. L. (2003). Prepaid and promised incentives in web surveys: An experiment. *Social Science Computer Review*, 21(2). <https://doi.org/10.1177/0894439303021002006>
- Boucart, M., Moroni, C., Szaffarczyk, S., & Tran, T. H. C. (2013). Implicit processing of scene context in macular degeneration. *Investigative Ophthalmology and Visual Science*, 54(3). <https://doi.org/10.1167/iovs.12-9680>
- Boyce, S. J., & Pollatsek, A. (1992). Identification of Objects in Scenes: The Role of Scene Background in Object Naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(3). <https://doi.org/10.1037/0278-7393.18.3.531>
- Boyce, S. J., Pollatsek, A., & Rayner, K. (1989). Effect of Background Information on Object Identification. *Journal of Experimental Psychology: Human Perception and Performance*, 15(3). <https://doi.org/10.1037/0096-1523.15.3.556>
- Boynton, G. M. (2009). A framework for describing the effects of attention on visual responses. *Vision Research*, 49(10). <https://doi.org/10.1016/j.visres.2008.11.001>
- Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. *Psychological Science*, 22(3). <https://doi.org/10.1177/0956797610397956>
- Brainard, D. (1997). The Psychophysics Toolbox Spatial Vision. *Spatial vision*, 10.
- Brooks, J. (2019). Peirce, J., & MacAskill, M. (Eds.). Building Experiments in PsychoPy. *Perception*, 48(2). <https://doi.org/10.1177/0301006618823976>
- Bruce, N. D., & Tsotsos, J. K. (2005). Saliency based on information maximization. *Advances in Neural Information Processing Systems*.
- Brysbaert, M., Drieghe, D., & Vitu, F. (2012). Word skipping: Implications for theories of eye movement control in reading. *Cognitive processes in eye guidance*. <https://doi.org/10.1093/acprof:oso/9780198566816.003.0003>
- Buisine, J., Bigand, A., Synave, R., Delepoulle, S., & Renaud, C. (2021). Stopping criterion during rendering of computer-generated images based on svd-entropy. *Entropy*, 23(1), 1–30. <https://doi.org/10.3390/e23010075>
- Buswell, G. T. (1935). *How people look at pictures: a study of the psychology and perception in art*. Univ. Chicago Press.

- Caldara, R., & Mielle, S. (2011). iMap: A novel method for statistical fixation mapping of eye movement data. *Behavior Research Methods*, 43(3), 864–878. <https://doi.org/10.3758/s13428-011-0092-x>
- Carrasco, M. (2011). Visual attention: The past 25 years. <https://doi.org/10.1016/j.visres.2011.04.012>
- Carrasco, M., Ponte, D., Rechea, C., & Sampedro, M. J. (1998). "Transient structures": The effects of practice and distractor grouping on within-dimension conjunction searches. *Perception and Psychophysics*, 60(7). <https://doi.org/10.3758/BF03206173>
- Castet, E., Jeanjean, S., Montagnini, A., Laugier, D., & Masson, G. S. (2006). Dynamics of attentional deployment during saccadic programming. *Journal of Vision*, 6(3). <https://doi.org/10.1167/6.3.2>
- Chandler, J., Mueller, P., & Paolacci, G. (2014). Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behavior Research Methods*, 46(1). <https://doi.org/10.3758/s13428-013-0365-7>
- Cheal, M. L., Lyon, D. R., & Hubbard, D. C. (1991). Does Attention have Different Effects on Line Orientation and Line Arrangement Discrimination? *The Quarterly Journal of Experimental Psychology Section A*, 43(4). <https://doi.org/10.1080/14640749108400959>
- Cook, R. L., Porter, T., & Carpenter, L. (1984). Distributed ray tracing. *Computer Graphics (ACM)*, 18(3). <https://doi.org/10.1145/964965.808590>
- Cook, R. L., & Torrance, K. E. (1981). Reflectance model for computer graphics. *Computer Graphics (ACM)*, 15(3). <https://doi.org/10.1145/965161.806819>
- Corsini, G., Mossa, A., & Verrazzani, L. (1996). Signal-to-noise ratio and autocorrelation function of the image intensity in coherent systems: Sub-Rayleigh and super-Rayleigh conditions. *IEEE Transactions on Image Processing*, 5(1). <https://doi.org/10.1109/83.481677>
- Crump, M. J., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a Tool for Experimental Behavioral Research. *PLoS ONE*, 8(3). <https://doi.org/10.1371/journal.pone.0057410>
- Dandurand, F., Shultz, T. R., & Onishi, K. H. (2008). Comparing online and lab methods in a problem-solving experiment. *Behavior Research Methods*, 40(2), 428–434. <https://doi.org/10.3758/BRM.40.2.428>

- David, E. J., Beitner, J., & Võ, M. L. H. (2021). The importance of peripheral vision when searching 3D real-world scenes: A gaze-contingent study in virtual reality. *Journal of Vision*, 21(7). <https://doi.org/10.1167/jov.21.7.3>
- de Leeuw, J. R., & Motz, B. A. (2016). Psychophysics in a Web browser? Comparing response times collected with JavaScript and Psychophysics Toolbox in a visual search task. *Behavior Research Methods*, 48(1). <https://doi.org/10.3758/s13428-015-0567-2>
- Django Software Foundation. (2013). Django: The Web framework for perfectionists with deadlines. *Djangoproject.Com*.
- Draschkow, D., & Võ, M. L. (2017). Scene grammar shapes the way we interact with objects, strengthens memories, and speeds search. *Scientific Reports*, 7(1). <https://doi.org/10.1038/s41598-017-16739-x>
- Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, 113(4). <https://doi.org/10.1037/0096-3445.113.4.501>
- Eckstein, M. P. (2011). Visual search: A retrospective. *Journal of Vision*, 11(5). <https://doi.org/10.1167/11.5.14>
- Eriksen, C. W., & Hoffman, J. E. (1972). Temporal and spatial characteristics of selective encoding from visual displays. *Perception & Psychophysics*, 12(2). <https://doi.org/10.3758/BF03212870>
- Findlay, J. M. (1997). Saccade target selection during visual search. *Vision Research*, 37(5). [https://doi.org/10.1016/S0042-6989\(96\)00218-0](https://doi.org/10.1016/S0042-6989(96)00218-0)
- Fine, E. M., & Rubin, G. S. (1999a). Reading with central field loss: Number of letters masked is more important than the size of the mask in degrees. *Vision Research*, 39(4). [https://doi.org/10.1016/S0042-6989\(98\)00142-4](https://doi.org/10.1016/S0042-6989(98)00142-4)
- Fine, E. M., & Rubin, G. S. (1999b). Reading with simulated scotomas: Attending to the right is better than attending to the left. *Vision Research*, 39(5). [https://doi.org/10.1016/S0042-6989\(98\)00208-9](https://doi.org/10.1016/S0042-6989(98)00208-9)
- Folk, J. R., & Morris, R. K. (2003). Effects of syntactic category assignment on lexical ambiguity resolution in reading: An eye movement analysis. *Memory and Cognition*, 31(1). <https://doi.org/10.3758/BF03196085>

- Foulsham, T., & Underwood, G. (2008). What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. *Journal of Vision*, 8(2), 1–17. <https://doi.org/10.1167/8.2.6>
- Gautier, V., O'Regan, J. K., & Le Gargasson, J. F. (2000). 'The-skipping' revisited in French: Programming saccades to skip the article 'les'. *Vision Research*, 40(18). [https://doi.org/10.1016/S0042-6989\(00\)00089-4](https://doi.org/10.1016/S0042-6989(00)00089-4)
- Georgeson, M. A., & Sullivan, G. D. (1975). Contrast constancy: deblurring in human vision by spatial frequency channels. *The Journal of Physiology*, 252(3). <https://doi.org/10.1113/jphysiol.1975.sp011162>
- Germine, L., Nakayama, K., Duchaine, B. C., Chabris, C. F., Chatterjee, G., & Wilmer, J. B. (2012). Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments. *Psychonomic Bulletin and Review*. <https://doi.org/10.3758/s13423-012-0296-9>
- Giordano, A. M., McElree, B., & Carrasco, M. (2009). On the automaticity and flexibility of covert attention: A speed-accuracy trade-off analysis. *Journal of Vision*, 9(3). <https://doi.org/10.1167/9.3.30>
- Girod, B. (2005). Psychovisual Aspects Of Image Processing: What's Wrong With Mean Squared Error? <https://doi.org/10.1109/mdsp.1991.639240>
- Gosling, S. D., & Mason, W. (2015). Internet research in psychology. *Annual review of psychology*, 66, 877–902.
- Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should We Trust Web-Based Studies? A Comparative Analysis of Six Preconceptions About Internet Questionnaires. *American Psychologist*, 59(2). <https://doi.org/10.1037/0003-066X.59.2.93>
- Grootswagers, T. (2020). A primer on running human behavioural experiments online. *Behavior research methods*, 52(6), 2283–2286.
- Guan, P., & Banks, M. S. (2016). Stereoscopic depth constancy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1697). <https://doi.org/10.1098/rstb.2015.0253>
- Harel, J., Koch, C., & Perona, P. (2007). Graph-based visual saliency. *Advances in Neural Information Processing Systems*. <https://doi.org/10.7551/mitpress/7503.003.0073>

- Harris, C. R., Kaplan, R., & Pashler, H. (2014). Alarming Events in the Corner of Your Eye: Do They Trigger Early Saccades? *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2542346>
- Hein, E., Rolke, B., & Ulrich, R. (2006). Visual attention and temporal discrimination: Differential effects of automatic and voluntary cueing. *Visual Cognition*, 13(1). <https://doi.org/10.1080/13506280500143524>
- Helmholtz, H. v. (1925). Treatise on physiological optics. *Washington, DC: The Optical Society of America. (Original work published 1866), 111.*
- Henderson, J. M. (2007). Regarding scenes. *Current Directions in Psychological Science*, 16(4), 219–222. <https://doi.org/10.1111/j.1467-8721.2007.00507.x>
- Henderson, J. M., Brockmole, J. R., Castelhana, M. S., & Mack, M. (2007). Visual saliency does not account for eye movements during visual search in real-world scenes. *Eye movements*. <https://doi.org/10.1016/B978-008044980-7/50027-6>
- Henderson, J. M., & Hollingworth, A. (1998). Eye Movements During Scene Viewing. *Eye Guidance in Reading and Scene Perception*, 269–293. <https://doi.org/10.1016/b978-008043361-5/50013-4>
- Henderson, J. M., McClure, K. K., Pierce, S., & Schrock, G. (1997). Object identification without foveal vision: Evidence from an artificial scotoma paradigm. *Perception and Psychophysics*, 59(3). <https://doi.org/10.3758/BF03211901>
- Henderson, J. M., Williams, C. C., Castelhana, M. S., & Falk, R. J. (2003). Eye movements and picture processing during recognition. *Perception and Psychophysics*, 65(5). <https://doi.org/10.3758/BF03194809>
- Herrmann, K., Montaser-Kouhsari, L., Carrasco, M., & Heeger, D. J. (2010). When size matters: Attention affects performance by contrast or response gain. *Nature Neuroscience*, 13(12). <https://doi.org/10.1038/nn.2669>
- Horé, A., & Ziou, D. (2010). Image quality metrics: PSNR vs. SSIM. *Proceedings - International Conference on Pattern Recognition*, 2366–2369. <https://doi.org/10.1109/ICPR.2010.579>
- Itti, L. (2005a). CHAPTER 94 - Models of Bottom-up Attention and Saliency. In L. Itti, G. Rees, & J. K. Tsotsos (Eds.), *Neurobiology of attention* (pp. 576–582). Academic Press. <https://doi.org/https://doi.org/10.1016/B978-012375731-9/50098-7>

- Itti, L. (2005b). Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. *Visual Cognition*, 12(6). <https://doi.org/10.1080/13506280444000661>
- Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Research*, 49(10). <https://doi.org/10.1016/j.visres.2008.09.007>
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40(10-12), 1489–1506. [https://doi.org/10.1016/S0042-6989\(99\)00163-7](https://doi.org/10.1016/S0042-6989(99)00163-7)
- Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11), 1254–1259. <https://doi.org/10.1109/34.730558>
- James, W. (1891). The Principles of Psychology. *The American Journal of Psychology*, 3(4). <https://doi.org/10.2307/1412102>
- Jonides, J. (1983). Further toward a model of the Mind's eye's movement. *Bulletin of the Psychonomic Society*, 21(4). <https://doi.org/10.3758/BF03334699>
- Juhasz, B. J., & Rayner, K. (2003). Investigating the Effects of a Set of Intercorrelated Variables on Eye Fixation Durations in Reading. *Journal of Experimental Psychology: Learning Memory and Cognition*, 29(6). <https://doi.org/10.1037/0278-7393.29.6.1312>
- Juhasz, B. J., & Rayner, K. (2006). The role of age of acquisition and word frequency in reading: Evidence from eye fixation durations. *Visual Cognition*, 13(7-8). <https://doi.org/10.1080/13506280544000075>
- Jun, E., Hsieh, G., & Reinecke, K. (2017). Types of motivation affect study selection, attention, and dropouts in online experiments. *Proceedings of the ACM on Human-Computer Interaction*, 1(CSCW), 1–15.
- Juric, I., Nedeljkovic, U., Novakovic, D., & Pincjer, I. (2016). Visual experience of noise in digital images. *Tehnicki vjesnik - Technical Gazette*, 23(5). <https://doi.org/10.17559/tv-20150317101822>
- Kajiya, J. T. (1986). The rendering equation. *Proceedings of the 13th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1986*, 20(4), 143–150. <https://doi.org/10.1145/15922.15902>
- Kietzmann, T. C., Geuter, S., & König, P. (2011). Overt visual attention as a causal factor of perceptual awareness. *PLoS ONE*, 6(7). <https://doi.org/10.1371/journal.pone.0022614>

- Kim, Y., Varshney, A., Jacobs, D. W., & Guimbretière, F. (2010). Mesh saliency and human eye fixations. *ACM Transactions on Applied Perception*, 7(2). <https://doi.org/10.1145/1670671.1670676>
- Klein, S. A. (2001). Measuring, estimating, and understanding the psychometric function: A commentary. <https://doi.org/10.3758/BF03194552>
- Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology*, 4(4). <https://doi.org/10.1007/978-94-009-3833-5>
- Koulieris, G. A., Drettakis, G., Cunningham, D., & Mania, K. (2014). C-LOD: context-aware material level-of-detail applied to mobile graphics. *Computer Graphics Forum*, 33(4), 41–49. <https://doi.org/https://doi.org/10.1111/cgf.12411>
- Krause, F., & Lindemann, O. (2014). Expyriment: a python library for cognitive and neuroscientific experiments. *Behavior Research Methods*, 46(2), 416–428.
- Kraut, R., Olson, J., Banaji, M., Bruckman, A., Cohen, J., & Couper, M. (2004). Psychological Research Online: Report of Board of Scientific Affairs' Advisory Group on the Conduct of Research on the Internet. *American Psychologist*, 59(2). <https://doi.org/10.1037/0003-066X.59.2.105>
- Kümmerer, M., Bylinskii, Z., Judd, T., Borji, A., Itti, L., Durand, F., Oliva, A., & Torralba, A. (n.d.). Mit/tübingen saliency benchmark (<https://saliency.tuebingen.ai/>, Ed.).
- Kümmerer, M., Wallis, T. S. A., & Bethge, M. (2018). Saliency benchmarking made easy: separating models, maps and metrics. In V. Ferrari, M. Hebert, C. Sminchisescu, & Y. Weiss (Eds.), *Computer vision – ECCV 2018* (pp. 798–814). Springer International Publishing.
- Lafortune, E. P., & Willems, Y. D. (1993). Bi-Directional Path Tracing. *Proc. SIGGRAPH*.
- Land, M. F., & Lee, D. N. (1994). Where we look when we steer. *Nature*, 369(6483). <https://doi.org/10.1038/369742a0>
- Land, M. F. (2012). Oculomotor behaviour in vertebrates and invertebrates. *The oxford handbook of eye movements*. <https://doi.org/10.1093/oxfordhb/9780199539789.013.0001>
- Lao, J., Mielle, S., Pernet, C., Sokhn, N., & Caldara, R. (2017). iMap4: An open source toolbox for the statistical fixation mapping of eye movement data with linear mixed modeling.

- Behavior Research Methods*, 49(2), 559–575. <https://doi.org/10.3758/s13428-016-0737-x>
- Leek, M. R. (2001). Adaptive procedures in psychophysical research. *Perception and Psychophysics*, 63(8). <https://doi.org/10.3758/BF03194543>
- Li, Z., & Itti, L. (2011). Saliency and gist features for target detection in satellite images. *IEEE Transactions on Image Processing*, 20(7). <https://doi.org/10.1109/TIP.2010.2099128>
- Lin, R. J., & Lin, W. S. (2014). A computational visual saliency model based on statistics and machine learning. *Journal of Vision*, 14(9). <https://doi.org/10.1167/14.9.1>
- Ling, S., & Carrasco, M. (2006). Sustained and transient covert attention enhance the signal via different contrast response functions. *Vision Research*, 46(8-9). <https://doi.org/10.1016/j.visres.2005.05.008>
- Longhurst, P., Debattista, K., & Chalmers, A. (2006). A GPU based saliency map for high-fidelity selective rendering. *Proceedings of the 4th international conference on Computer graphics, virtual reality, visualisation and interaction in Africa*, 21–29. <https://doi.org/10.1145/1108590.1108595>
- Loschky, L. C., & McConkie, G. W. (2002). Investigating spatial vision and dynamic attentional selection using a gaze-contingent multiresolutional display. *Journal of Experimental Psychology: Applied*, 8(2), 99–117. <https://doi.org/10.1037/1076-898X.8.2.99>
- Loschky, L. C., McConkie, G. W., Yang, J., & Miller, M. E. (2005). The limits of visual resolution in natural scene viewing. <https://doi.org/10.1080/13506280444000652>
- Lu, J., & Itti, L. (2005). Perceptual consequences of feature-based attention. *Journal of Vision*, 5(7). <https://doi.org/10.1167/5.7.2>
- Ma, Q., & Zhang, L. (2008). Saliency-based image quality assessment criterion. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5226 LNCS. https://doi.org/10.1007/978-3-540-87442-3_{ }139
- Macmillan, N. a. N., & Creelman, C. D. (2005). *Detection Theory: A User's Guide*.
- Mangun, G. R., & Hillyard, S. A. (1990). Allocation of visual attention to spatial locations: Tradeoff functions for event-related brain potentials and detection performance. *Perception & Psychophysics*, 47(6). <https://doi.org/10.3758/BF03203106>

- Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (1996). The relationship between the locations of spatial features and those of fixations made during visual examination of briefly presented images. *Spatial Vision*, *10*(3). <https://doi.org/10.1163/156856896X00123>
- Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (1997). Fixation sequences made during visual examination of briefly presented 2D images. *Spatial Vision*, *11*(2). <https://doi.org/10.1163/156856897X00177>
- Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, *44*(1), 1–23. <https://doi.org/10.3758/s13428-011-0124-6>
- Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. <https://doi.org/10.3758/s13428-011-0168-7>
- Martin, E. (1974). Saccadic suppression: A review and an analysis. *Psychological Bulletin*, *81*(12). <https://doi.org/10.1037/h0037368>
- Maunsell, J. H., & Treue, S. (2006). Feature-based attention in visual cortex. <https://doi.org/10.1016/j.tins.2006.04.001>
- Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? what does “failure to replicate” really mean? *American Psychologist*, *70*(6), 487.
- McConkie, G. W., Kerr, P. W., Reddix, M. D., Zola, D., & Jacobs, A. M. (1989). Eye movement control during reading: II. Frequency of refixating a word. *Perception & Psychophysics*, *46*(3). <https://doi.org/10.3758/BF03208086>
- McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics*, *17*(6). <https://doi.org/10.3758/BF03203972>
- McConkie, G. W., & Rayner, K. (1976). Asymmetry of the perceptual span in reading. *Bulletin of the Psychonomic Society*, *8*(5). <https://doi.org/10.3758/BF03335168>
- Miellat, S., Caldara, R., & Schyns, P. G. (2011). Local jekyll and global hyde: The dual identity of face identification. *Psychological Science*, *22*(12), 1518–1526. <https://doi.org/10.1177/0956797611424290>
- Miellat, S., O'Donnell, P. J., & Sereno, S. C. (2009). Parafoveal magnification: Visual acuity does not modulate the perceptual span in reading. *Psychological Science*, *20*(6). <https://doi.org/10.1111/j.1467-9280.2009.02364.x>

- Miellet, S., Zhou, X., He, L., Rodger, H., & Caldara, R. (2010). Investigating cultural diversity for extrafoveal information use in visual scenes. *10*, 1–18. <https://doi.org/10.1167/10.6.21.Introduction>
- Montagnini, A., & Castet, E. (2007). Spatiotemporal dynamics of visual attention during saccade preparation: Independence and coupling between attention and movement planning. *Journal of Vision*, *7*(14). <https://doi.org/10.1167/7.14.8>
- Montagnini, A., & Chelazzi, L. (2005). The urgency to look: Prompt saccades to the benefit of perception. *Vision Research*, *45*(27). <https://doi.org/10.1016/j.visres.2005.07.013>
- Moon, B., Carr, N., & Yoon, S. E. (2014). Adaptive rendering based on weighted local regression. *ACM Transactions on Graphics*, *33*(5). <https://doi.org/10.1145/2641762>
- Müller, N. G., Bartelt, O. A., Donner, T. H., Villringer, A., & Brandt, S. A. (2003). A physiological correlate of the "zoom lens" of visual attention. *Journal of Neuroscience*, *23*(9). <https://doi.org/10.1523/jneurosci.23-09-03561.2003>
- Murray, R. F., Beutter, B. R., Eckstein, M. P., & Stone, L. S. (2003). Saccadic and perceptual performance in visual search tasks II Letter discrimination. *Journal of the Optical Society of America A*, *20*(7). <https://doi.org/10.1364/josaa.20.001356>
- Myrodia, V., Buisine, J., & Madelain, L. Comparison of threshold measurements in laboratory and online studies using a quest+ algorithm. In: *Journal of vision*. Vision Science Society, 2021, May. <https://doi.org/https://doi.org/10.1167/jov.21.9.1959..>
- Myszkowski, K. (1998). The Visible Differences Predictor: applications to global illumination problems. https://doi.org/10.1007/978-3-7091-6453-2_{_}21
- Neider, M. B., & Zelinsky, G. J. (2006). Scene context guides eye movements during visual search. *Vision Research*, *46*(5). <https://doi.org/10.1016/j.visres.2005.08.025>
- Niebur, E., Itti, L., & Koch, C. (2002). Controlling the Focus of Visual Selective Attention, 247–276. https://doi.org/10.1007/978-0-387-21703-1_{_}6
- Ninassi, A., Le Meur, O., Le Callet, P., & Barba, D. (2007). Does where you gaze on an image affect your perception of quality? Applying visual attention to image quality metric. *Proceedings - International Conference on Image Processing, ICIP*, *2*. <https://doi.org/10.1109/ICIP.2007.4379119>
- Nothdurft, H. C. (2002). Attention shifts to salient targets. *Vision Research*, *42*(10). [https://doi.org/10.1016/S0042-6989\(02\)00016-0](https://doi.org/10.1016/S0042-6989(02)00016-0)

- Nuthmann, A. (2014). How do the regions of the visual field contribute to object search in real-world scenes? Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, 40(1). <https://doi.org/10.1037/a0033854>
- Nuthmann, A., Einhäuser, W., & Schütz, I. (2017). How Well Can Saliency Models Predict Fixation Selection in Scenes Beyond Central Bias? A New Approach to Model Evaluation Using Generalized Linear Mixed Models. *Frontiers in Human Neuroscience*, 11. <https://doi.org/10.3389/fnhum.2017.00491>
- Ohmer, S. O. (1997). Ray Tracers: Blue Sky Studios. *ANIMATIONWorld*. <https://www.awn.com/animationworld/ray-tracers-blue-sky-studios>
- Oliphant, T. E. (2006). *A guide to numpy* (Vol. 1). Trelgol Publishing USA.
- Oliphant, T. E. et al. (2007). Scipy: open source scientific tools for python. *Computing in Science and Engineering*, 9(1), 10–20.
- Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Journal of Experimental Social Psychology Instructional manipulation checks : Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, 45(4). <https://doi.org/10.1016/j.jesp.2009.03.009>
- Owen, L., Browder, J., Letham, B., Stocck, G., Tymms, C., & Shvartsman, M. (2021). Adaptive Nonparametric Psychophysics. <http://arxiv.org/abs/2104.09549>
- Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, 17, 22–27. <https://doi.org/10.1016/j.jbef.2017.12.004>
- Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon mechanical turk. *Judgment and Decision Making*, 5(5).
- Pappas, T. N., Michel, T. A., & Hinds, R. O. (1996). Supra-threshold perceptual image coding. *IEEE International Conference on Image Processing*, 1. <https://doi.org/10.1109/icip.1996.559477>
- Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, 42(1). [https://doi.org/10.1016/S0042-6989\(01\)00250-4](https://doi.org/10.1016/S0042-6989(01)00250-4)
- Parkhurst, D. J., & Niebur, E. (2003). Scene content selected by active vision. *Spatial Vision*, 16(2), 125–154. <https://doi.org/10.1163/15685680360511645>

- Pashler, H., & Wagenmakers, E.-J. (2012). Editors' introduction to the special section on replicability in psychological science: a crisis of confidence? *Perspectives on psychological science*, 7(6), 528–530.
- Pedersen, M., Bonnier, N., Hardeberg, J. Y., & Albrechtsen, F. (2009). Attributes of a new image quality model for color prints. *Final Program and Proceedings - IS and T/SID Color Imaging Conference*.
- Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1). <https://doi.org/10.3758/s13428-018-01193-y>
- Peirce, J. W. (2007). Psychopy—psychophysics software in python. *Journal of neuroscience methods*, 162(1-2), 8–13.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10(4). <https://doi.org/10.1163/156856897X00366>
- Pharr, M., Jakob, W., & Humphreys, G. (2016). *Physically based rendering: From theory to implementation: Third edition*.
- Posner, M. I. (1980). Orienting of attention. *The Quarterly journal of experimental psychology*, 32(1). <https://doi.org/10.1080/00335558008248231>
- Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, 109(2). <https://doi.org/10.1037/0096-3445.109.2.160>
- Privitera, C. M., & Stark, L. W. (2000). Algorithms for defining visual regions-of-interest: comparison with eye fixations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(9). <https://doi.org/10.1109/34.877520>
- Rajashekar, U., Cormack, L. K., & Bovik, A. C. (2002). Visual search: Structure from noise. *Eye Tracking Research and Applications Symposium (ETRA)*, 119–123.
- Ramasubramanian, M., Pattanaik, S. N., & Greenberg, D. P. (1999). A perceptually based physical error metric for realistic image synthesis. *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1999*, 73–82. <https://doi.org/10.1145/311535.311543>
- Rayner, K. (1979). Eye guidance in reading: Fixation locations within words. *Perception*, 8(1). <https://doi.org/10.1068/p080021>

- Rayner, K. (1998). Eye Movements in Reading and Information Processing: 20 Years of Research. *Psychological Bulletin*, 124(3). <https://doi.org/10.1037/0033-2909.124.3.372>
- Rayner, K., & Bertera, J. H. (1979). Reading without a fovea. *Science*, 206(4417). <https://doi.org/10.1126/science.504987>
- Rayner, K., Li, X., Williams, C. C., Cave, K. R., & Well, A. D. (2007). Eye movements during information processing tasks: Individual differences and cultural effects. *Vision Research*, 47(21). <https://doi.org/10.1016/j.visres.2007.05.007>
- Rayner, K., & McConkie, G. W. (1976). What guides a reader's eye movements? *Vision Research*, 16(8). [https://doi.org/10.1016/0042-6989\(76\)90143-7](https://doi.org/10.1016/0042-6989(76)90143-7)
- Rayner, K., & Morrison, R. E. (1981). Eye movements and identifying words in parafoveal vision. *Bulletin of the Psychonomic Society*, 17(3). <https://doi.org/10.3758/BF03333690>
- Rayner, K., & Pollatsek, A. (1981). Eye movement control during reading: Evidence for direct control. *The Quarterly Journal of Experimental Psychology Section A*, 33(4). <https://doi.org/10.1080/14640748108400798>
- Reddi, B. A., & Carpenter, R. H. (2000). The influence of urgency on decision time. *Nature Neuroscience*, 3(8). <https://doi.org/10.1038/77739>
- Reinagel, P., & Zador, A. M. (1999). Natural scene statistics at the centre of gaze. *Network: Computation in Neural Systems*, 10(4), 341–350. [https://doi.org/10.1088/0954-898X\(1999\)10\(4\)304](https://doi.org/10.1088/0954-898X(1999)10(4)304)
- Reips, U. D. (2002a). Internet-based psychological experimenting: Five dos and five don'ts. *Social Science Computer Review*, 20(3). <https://doi.org/10.1177/08939302020003002>
- Reips, U. D. (2002b). Standards for Internet-based experimenting. *Experimental Psychology*, 49(4). <https://doi.org/10.1026//1618-3169.49.4.243>
- Reips, U. D. (2012). The methodology of Internet-based experiments. *Oxford handbook of internet psychology*. <https://doi.org/10.1093/oxfordhb/9780199561803.013.0024>
- Reips, U.-D. (2000). The Web Experiment Method: Advantages, Disadvantages, and Solutions. *The effects of brief mindfulness intervention on acute pain experience: An examination of individual difference*, 1.
- Rosin, P. L. (2009). A simple method for detecting salient regions. *Pattern Recognition*, 42(11). <https://doi.org/10.1016/j.patcog.2009.04.021>

- Rutishauser, U., Walther, D., Koch, C., & Perona, P. (2004). Is bottom-up attention useful for object recognition? *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2. <https://doi.org/10.1109/cvpr.2004.1315142>
- Saida, S., & Ikeda, M. (1979). Useful visual field size for pattern perception. *Perception & Psychophysics*, 25(2). <https://doi.org/10.3758/BF03198797>
- Sasaki, K., Ihaya, K., & Yamada, Y. (2017). Avoidance of novelty contributes to the uncanny valley. *Frontiers in Psychology*, 8(OCT). <https://doi.org/10.3389/fpsyg.2017.01792>
- Sasaki, K., & Yamada, Y. (2019). Crowdsourcing visual perception experiments: a case of contrast threshold. *PeerJ*, 7, e8339. <https://doi.org/10.7717/peerj.8339>
- Sauter, M., Draschkow, D., & Mack, W. (2020). Building, hosting and recruiting: a brief introduction to running behavioral experiments online. *Brain sciences*, 10(4), 251.
- Schweitzer, R., & Rolfs, M. (2021). Intrasaccadic motion streaks jump-start gaze correction. *Science Advances*, 7(30). <https://doi.org/10.1126/sciadv.abf2218>
- Semmelmann, K., & Weigelt, S. (2017). Online psychophysics: reaction time effects in cognitive experiments. *Behavior Research Methods*, 49(4). <https://doi.org/10.3758/s13428-016-0783-4>
- Seo, H. J., & Milanfar, P. (2009). Static and space-time visual saliency detection by self-resemblance. *Journal of Vision*, 9(12). <https://doi.org/10.1167/9.12.15>
- Shirley, P., Wang, C., & Zimmerman, K. (1996). Monte Carlo Techniques for Direct Lighting Calculations. *ACM Transactions on Graphics*, 15(1), 1–36. <https://doi.org/10.1145/226150.226151>
- Sohn, W., Papathomas, T. V., Blaser, E., & Vidnyánszky, Z. (2004). Object-based cross-feature attentional modulation from color to motion. *Vision Research*, 44(12). <https://doi.org/10.1016/j.visres.2003.12.010>
- Spanier, J., Gelbard, E. M., & Bell, G. (1970). Monte Carlo Principles and Neutron Transport Problems. *Physics Today*, 23(9). <https://doi.org/10.1063/1.3022338>
- Sperling, G., & Melchner, M. J. (1978). The attention operating characteristic: Examples from visual search. *Science*, 202(4365). <https://doi.org/10.1126/science.694536>
- Stuart, S. (2014). Behind the Scenes of Disney's Tech-Centric 'Big Hero 6'. *PCmag*. <https://www.pcmag.com/news/behind-the-scenes-of-disneys-tech-centric-big-hero-6>

- Szirmay-kalos, L. (1998). Stochastic Methods in Global Illumination State of the Art Report. *Control Engineering*, 23(August).
- Takouachet, N., Delepouille, S., Renaud, C., Zoghlami, N., & Tavares, J. M. R. (2017). Perception of noise and global illumination: Toward an automatic stopping criterion based on SVM. *Computers and Graphics (Pergamon)*, 69, 49–58. <https://doi.org/10.1016/j.cag.2017.09.008>
- Tatler, B. W., Baddeley, R. J., & Gilchrist, I. D. (2005). Visual correlates of fixation selection: Effects of scale and time. *Vision Research*, 45(5). <https://doi.org/10.1016/j.visres.2004.09.017>
- Tatler, B. W., Baddeley, R. J., & Vincent, B. T. (2006). The long and the short of it: Spatial statistics at fixation vary with saccade amplitude and task. *Vision Research*, 46(12). <https://doi.org/10.1016/j.visres.2005.12.005>
- Taylor, M. M., & Creelman, C. D. (1967). PEST: Efficient Estimates on Probability Functions. *The Journal of the Acoustical Society of America*, 41(4A). <https://doi.org/10.1121/1.1910407>
- Theeuwes, J. (1991). Exogenous and endogenous control of attention: The effect of visual onsets and offsets. *Perception & Psychophysics*, 49(1). <https://doi.org/10.3758/BF03211619>
- Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception & Psychophysics*, 51(6). <https://doi.org/10.3758/BF03211656>
- Theeuwes, J. (1994). Stimulus-Driven Capture and Attentional Set: Selective Search for Color and Visual Abrupt Onsets. *Journal of Experimental Psychology: Human Perception and Performance*, 20(4). <https://doi.org/10.1037/0096-1523.20.4.799>
- Thibaut, M., Tran, T. H. C., Szaffarczyk, S., & Boucart, M. (2018). Impact of age-related macular degeneration on object searches in realistic panoramic scenes. *Clinical and Experimental Optometry*, 101(3). <https://doi.org/10.1111/cxo.12644>
- Thiele, A., Henning, P., Kubischik, M., & Hoffmann, K. P. (2002). Neural mechanisms of saccadic suppression. *Science*, 295(5564). <https://doi.org/10.1126/science.1068788>
- Thorpe, S. J., Gegenfurtner, K. R., Fabre-Thorpe, M., & Bülthoff, H. H. (2001). Detection of animals in natural images using far peripheral vision. *European Journal of Neuroscience*, 14(5). <https://doi.org/10.1046/j.0953-816X.2001.01717.x>

- Tran, T. H. C., Rambaud, C., Desprez, P., & Boucart, M. (2010). Scene perception in age-related macular degeneration. *Investigative Ophthalmology and Visual Science*, 51(12). <https://doi.org/10.1167/iovs.10-5517>
- Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, 14(1). [https://doi.org/10.1016/0010-0285\(82\)90006-8](https://doi.org/10.1016/0010-0285(82)90006-8)
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1). [https://doi.org/10.1016/0010-0285\(80\)90005-5](https://doi.org/10.1016/0010-0285(80)90005-5)
- Underwood, G., & Foulsham, T. (2006). Visual saliency and semantic incongruity influence eye movements when inspecting pictures. *Quarterly Journal of Experimental Psychology*, 59(11). <https://doi.org/10.1080/17470210500416342>
- Underwood, G., Foulsham, T., Van Loon, E., Humphreys, L., & Bloyce, J. (2006). Eye movements during scene inspection: A test of the saliency map hypothesis. *European Journal of Cognitive Psychology*, 18(3). <https://doi.org/10.1080/09541440500236661>
- Underwood, G., Templeman, E., Lamming, L., & Foulsham, T. (2008). Is attention necessary for object identification? Evidence from eye movements during the inspection of real-world scenes. *Consciousness and Cognition*, 17(1). <https://doi.org/10.1016/j.concog.2006.11.008>
- Van Diepen, P. M., Ruelens, L., & D'Ydewalle, G. (1999). Brief foveal masking during scene perception. *Acta Psychologica*, 101(1). [https://doi.org/10.1016/S0001-6918\(98\)00048-1](https://doi.org/10.1016/S0001-6918(98)00048-1)
- Van Diepen, P. M., & Wampers, M. (1998). Scene exploration with Fourier-filtered peripheral information. *Perception*, 27(10), 1141–1151. <https://doi.org/10.1068/p271141>
- Van Zoest, W., Donk, M., & Theeuwes, J. (2004). The role of stimulus-driven and goal-driven control in saccadic visual selection. *Journal of Experimental Psychology: Human Perception and Performance*, 30(4). <https://doi.org/10.1037/0096-1523.30.4.749>
- Veach, E. (1997). Robust Monte Carlo Methods for Light Transport Simulation. *Dissertation at the Department of Computer Science of Stanford University*, 134(December).
- Vergilino-Perez, D., Collins, T., & Doré-Mazars, K. (2004). Decision and metrics of refixations in reading isolated words. *Vision Research*, 44(17). <https://doi.org/10.1016/j.visres.2004.03.012>
- Vlaskamp, B. N., & Hooge, I. T. C. (2006). Crowding degrades saccadic search performance. *Vision Research*, 46(3). <https://doi.org/10.1016/j.visres.2005.04.006>

- Vö, M. L. H. (2021). The meaning and structure of scenes. *Vision Research*, 181. <https://doi.org/10.1016/j.visres.2020.11.003>
- Vö, M. L. H., Boettcher, S. E., & Draschkow, D. (2019). Reading scenes: how scene grammar guides attention and aids perception in real-world environments. <https://doi.org/10.1016/j.copsy.2019.03.009>
- Vö, M. L., & Henderson, J. M. (2011). Object-scene inconsistencies do not capture gaze: Evidence from the flash-preview moving-window paradigm. *Attention, Perception, and Psychophysics*, 73(6), 1742–1753. <https://doi.org/10.3758/s13414-011-0150-6>
- Volkman, F. C., Schick, A. M., & Riggs, L. A. (1968). Time course of visual inhibition during voluntary saccades. *Journal of the Optical Society of America*, 58(4). <https://doi.org/10.1364/JOSA.58.000562>
- Vorba, J., Hanika, J., Herholz, S., Müller, T., Krivánek, J., & Keller, A. (2019). *Path guiding in production*. <https://doi.org/10.1145/3305366.3328091>
- Vullings, C., & Madelain, L. (2018). Control of saccadic latency in a dynamic environment: Allocation of saccades in time follows the matching law. *Journal of Neurophysiology*, 119(2). <https://doi.org/10.1152/jn.00634.2017>
- Vullings, C., & Madelain, L. (2019). Discriminative control of saccade latencies. *Journal of Vision*, 19(3). <https://doi.org/10.1167/19.3.16>
- Wahid, M., Waris, A., Gilani, S. O., & Subramanian, R. (2019). The effect of eye movements in response to different types of scenes using a graph-based visual saliency algorithm. *Applied Sciences (Switzerland)*, 9(24). <https://doi.org/10.3390/app9245378>
- Wang, Z., Bovik, A., Sheikh, H., & Simoncelli, E. (2004). Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, 13(4), 600–612. <https://doi.org/10.1109/TIP.2003.819861>
- Ware, C., Turton, T. L., Bujack, R., Samsel, F., Shrivastava, P., & Rogers, D. H. (2019). Measuring and modeling the feature detection threshold functions of colormaps. *IEEE Transactions on Visualization and Computer Graphics*, 25(9). <https://doi.org/10.1109/TVCG.2018.2855742>
- Watson, A. B. (2017). QUEST+: A general multidimensional Bayesian adaptive psychometric method. *Journal of Vision*, 17(3), 10. <https://doi.org/10.1167/17.3.10>

- Watson, A. B., & Pelli, D. G. (1983). Quest: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, 33(2), 113–120. <https://doi.org/10.3758/BF03202828>
- Weigold, A., Weigold, I. K., & Russell, E. J. (2013). Examination of the equivalence of self-report survey-based paper-and-pencil and internet data collection methods. *Psychological Methods*, 18(1). <https://doi.org/10.1037/a0031607>
- Wellek, S. (2010). *Testing statistical hypotheses of equivalence and noninferiority* (Second edi). Boca Raton: Chapman & Hall/CRC Press.
- Whitted, T. (1980). An Improved Illumination Model for Shaded Display. *Communications of the ACM*, 23(6). <https://doi.org/10.1145/358876.358882>
- Williams, L. G. (1967). The effects of target specification on objects fixated during visual search. *Acta Psychologica*, 27(100). [https://doi.org/10.1016/0001-6918\(67\)90080-7](https://doi.org/10.1016/0001-6918(67)90080-7)
- Wolfe, J. M. (1994). Guided Search 2.0 A revised model of visual search. *Psychonomic Bulletin & Review*, 1(2). <https://doi.org/10.3758/BF03200774>
- Wolfe, J. M., & Bennett, S. C. (1997). Preattentive object files: Shapeless bundles of basic features. *Vision Research*, 37(1). [https://doi.org/10.1016/S0042-6989\(96\)00111-3](https://doi.org/10.1016/S0042-6989(96)00111-3)
- Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided Search: An Alternative to the Feature Integration Model for Visual Search. *Journal of Experimental Psychology: Human Perception and Performance*, 15(3). <https://doi.org/10.1037/0096-1523.15.3.419>
- Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? <https://doi.org/10.1038/nrn1411>
- Wolfe, J. M., Palmer, E. M., & Horowitz, T. S. (2010). Reaction time distributions constrain models of visual search. *Vision Research*, 50(14). <https://doi.org/10.1016/j.visres.2009.11.002>
- Wolfe, J. M., Võ, M. L., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and nonselective pathways. <https://doi.org/10.1016/j.tics.2010.12.001>
- Wong, A. L., Goldsmith, J., Forrence, A. D., Haith, A. M., & Krakauer, J. W. (2017). Reaction times can reflect habits rather than computations. *eLife*, 6. <https://doi.org/10.7554/eLife.28075>
- Woods, A. T., Velasco, C., Levitan, C. A., Wan, X., & Spence, C. (2015). Conducting perception research over the internet: A tutorial review. <https://doi.org/10.7717/peerj.1058>

- Xiao, G., Xu, G., Liu, X., Xu, J., Wang, F., Li, L., Itti, L., & Lu, J. (2014). Feature-based attention is independent of object appearance. *Journal of Vision*, 14(1). <https://doi.org/10.1167/14.1.3>
- Y. Yee, H., Dutré, P., & Pattanaik, S. (2002). Fundamentals of lighting and perception: the rendering of physically accurate images.
- Yang, J., Pitt, M. A., Ahn, W. Y., & Myung, J. I. (2021). ADOPy: a python package for adaptive design optimization. *Behavior Research Methods*, 53(2). <https://doi.org/10.3758/s13428-020-01386-4>
- Yarbus, A. L. (1965). The Role of Eye Movements in Vision Process.
- Yee, H., Pattanaik, S., & Greenberg, D. P. (2001). Spatiotemporal sensitivity and visual attention for efficient rendering of dynamic environments. *ACM Transactions on Graphics*, 20(1), 39–65. <https://doi.org/10.1145/383745.383748>
- Zwicker, M., Jarosz, W., Lehtinen, J., Moon, B., Ramamoorthi, R., Rousselle, F., Sen, P., Soler, C., & Yoon, S. E. (2015). Recent Advances in Adaptive Sampling and Reconstruction for Monte Carlo Rendering. *Computer Graphics Forum*, 34(2). <https://doi.org/10.1111/cgf.12592>