

CARDIFF UNIVERSITY

2019

**Data analysis of the impact of experience
on medical practitioners gaze behaviour
when viewing medical images**

SCHOOL OF COMPUTER SCIENCE AND
INFORMATICS

Author:

Sebastian GOLOS
C1546910

Supervisor:

Dr. Hantao LIU

Abstract

Understanding expert gaze behaviour in the field of medical imaging is a challenging topic as a result of the multitude of complex interacting factors existing in human biology and influencing human vision. This study performs a visual and statistical analysis on data gathered from an eye-tracking experiment performed in the University Hospitals KU Leuven in Belgium in 2018. The experiment recorded eye movements of medical practitioners with varying degrees of experience as they examined mammograms, with the goal of discerning differences between the groups, and identifying expert behaviour. The raw data was transformed into gaze information which represented saccade eye movements, and then an agglomerate visual analysis was performed, comparing and contrasting groups by applying scientific statistical techniques to reinforce the observations.

Acknowledgements

I would like to thank my supervisor Dr Hantao Liu for all of his help and guidance throughout the project. He has always been available to answer my questions and his help has been invaluable, especially his assistance for everything related to research in Human Vision, Data Analysis & Visualization and Python. Finally, I would like to thank my family for being supportive of me during my time at university and for their help proofreading this dissertation.

Contents

Abstract	i
Acknowledgements	ii
1 Introduction	1
1.1 Motivation	1
1.2 Studying and Modelling Vision	1
1.3 Aims and Objectives	3
2 Background	4
2.1 The Field of Medical Imaging	4
2.2 Purpose of Studying Gaze Behaviour	5
2.3 Experience as an Independent Variable	5
2.4 Eye-tracking Experiment	5
2.5 Previous Work	6
2.5.1 Identification of Practitioners Visual Search Patterns	7
2.5.2 Influence of Expertise and Experience on Performance	8
Surgery	8
Medical Diagnosis	9
2.5.3 Impact of Training on Viewing Behaviour	10
3 Approach	11
3.1 Extension of Previous Work	11
3.2 Limitations of Previous Work and Further Approach	12
3.3 Data Processing	13
3.3.1 Gaze Data Plotting	14
3.3.2 Saccade Analysis	15
3.4 Visual Analysis	17
3.5 Comparing Histograms	17
3.6 Statistical Analysis	18
4 Implementation	19
4.1 Environment	19
4.1.1 IPython and Jupyter Notebook	19
4.1.2 Libraries	19
NumPy	19

	SciPy	20
	Matplotlib	20
4.2	Algorithms	20
4.2.1	Data Processing	20
4.2.2	Plotting Fixation Points	20
4.2.3	Efficient Iteration	22
4.2.4	Calculating Saccade Values	22
4.2.5	Plotting Histograms and Lines of Best Fit	23
4.2.6	Euclidean and Intersection Distance Histogram Similarity	24
4.2.7	Plotting Mean Values and Error Margins	25
4.2.8	ANOVA and t-tests	25
4.2.9	Polar Plots	26
5	Results and Evaluation	28
5.1	Fixation Points	28
5.2	Saccade Amplitude Group Comparison	30
5.3	Physicists Saccade Amplitude Comparison	35
5.4	Expert Radiologists Saccade Amplitude Comparison	38
5.5	Trainee Radiologists Saccade Amplitude Comparison	41
5.6	Saccade Orientation Group Comparison	44
5.7	Joint Saccade Amplitude and Orientation Group Comparison	46
6	Future Work	48
7	Conclusion	49
8	Reflections on Learning	50
	References	51

List of Figures

1.1	Example Saliency Map	2
1.2	Example Saccadic Model	2
2.1	Experiment Procedure	6
3.1	Saccade Polar Plots	12
3.2	Eye-tracking Data Sample	13
3.3	Before and After Fixation Point Plotting on Mammograms	14
3.4	Example Saccade Amplitude Calculation	16
3.5	Example Saccade Orientation Calculation	17
5.1	Group Fixation Plots	29
5.2	Group Saccade Amplitude Distribution	30
5.3	Gaussian Line of Best Fit for Saccade Amplitude distribution	31
5.4	Group Mean Saccade Amplitude Graph	33
5.5	Physicists Saccade Amplitude Distribution	35
5.6	Physicists Mean Saccade Amplitude Graph	36
5.7	Expert Radiologists Saccade Amplitude Distribution	38
5.8	Expert Radiologists Mean Saccade Amplitude Graph	39
5.9	Trainee Radiologists Saccade Amplitude Distribution	41
5.10	Trainee Radiologist Mean Saccade Amplitude Graph	42
5.11	Group Saccade Orientation Distributioon	44
5.12	Saccade Polar Plots	46

List of Tables

5.1	Group histogram comparison using similarity measurements of Euclidean distance and Intersection distance.	32
5.2	Mean values of saccade amplitude for each participating group.	33
5.3	Results of the ANOVA test to evaluate the effect of participating group on the length of saccade amplitude.	33
5.4	Physicists histogram comparison using similarity measurements of Euclidean distance and Intersection distance.	36
5.5	Mean values for the saccade amplitudes performed by each Physicist.	36
5.6	Results of the independent t-test to determine statistical significance between physicists.	36
5.7	Expert radiologists histogram comparison using similarity measurements of Euclidean distance and Intersection distance.	39
5.8	Mean values for the saccade amplitudes performed by each expert radiologist.	39
5.9	Results of the ANOVA test to evaluate the effect of expert radiologist on the length of saccade amplitude.	40
5.10	Results of the independent t-test to determine statistical significance between expert radiologists.	40
5.11	Trainee radiologists histogram comparison using similarity measurements of Euclidean distance and Intersection distance.	42
5.12	Mean values for the saccade amplitudes performed by each trainee radiologist.	42
5.13	Results of the ANOVA test to evaluate the effect of trainee radiologists on the length of saccade amplitude.	43
5.14	Results of the independent t-test to determine statistical significance between trainee radiologists.	43
5.15	Group saccade orientation histogram comparison using similarity measurements of Euclidean distance and Intersection distance.	45

1 Introduction

1.1 Motivation

Human vision occurs as the eye focuses light onto the retina, where it is absorbed by a layer of photoreceptor cells; rod cells and cone cells, which convert light into electrochemical brain signals. Cone cells are responsible for high acuity tasks and colour vision, while rod cells enable sight at night and in dim light. This underlying mechanism is what enables visual attention and serves as an interface between the outside world and the decisions humans make as they navigate through their environments. It has been widely established that eye movements are a central feature of human vision [1], with broad functionality including enhanced extraction of features via microsaccadic motion [2] to high-level strategies for optimal information gathering [3]. The latter aspect is of significant interest in radiology, and particularly mammography screening, which are central themes of this report. Mammography screening has been estimated to be responsible for between an 8% to 23% decrease in breast cancer mortality rate during 1990-2002 [4]. Studying eye movements through eye-tracking technology has been demonstrated as an effective tool for recognizing causes of error [5], and advancing training methodologies for novice radiologists, since training techniques such as gaze training and pattern recognition can employ eye-tracking training data [6]. Discriminating and quantifying the characteristics and strategies present in the gaze behaviour of radiologists when viewing mammography images is therefore a fundamental step in establishing training techniques with greater efficacy and in developing sophisticated diagnostics equipment to increase the availability of effective patient care and tackle prevalent issues in global health..

1.2 Studying and Modelling Vision

The process of analysing human vision is performed through investigating eye-tracking data and constructing visual attention models as to deliver insight into the observer's mind. There exist a variety of visual attention models which focus on specific subclass' of eye movements like smooth pursuit [7], the optokinetic reflex [8] or gaze shifts [9]. This report focuses on fixations and saccades, along with a discussion on their related models; saliency and saccadic models. Fixations are points of interest where the fovea is concentrated on and where visual acuity is at its optimal. Saccades are rapid changes in eye position between two fixation points and

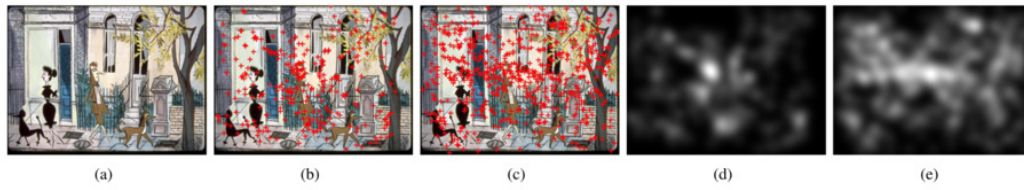


FIGURE 1.1: Illustrates original stimulus in (a); (b) and (c) demonstrate fixation maps and (d) and (d) display resulting saliency maps following the application of a saliency model.

can be measured using two variables; saccade amplitude and saccade orientation. Saccade amplitude refers to the distance between successive fixation points and saccade orientation signifies the angle between the two points. The order of fixation points and saccades which can be observed by a participant is labelled a visual scan path. Saliency models utilize fixation data to output 2D static saliency maps for understanding and predicting the salience of a visual scene.

However, saliency models have often been criticized for their failure to account for inherent oculomotor biases in human eye movements, such as predicating their conclusions on a hypothesis that eyes are likely to move in any direction [10]. To remedy the shortcomings of saliency models, saccadic models emerged. As demonstrated in Figure 1.2, which illustrates a saccadic model composed of three modules; a bottom-up saliency map, memory mechanism function and viewing biases [11]. By integrating stochastic attributes and including factors which are neglected by saliency research, saccadic models can integrate the intricacies of the human visual system to predict visual scan paths with improved accuracy and can be tailored to specific contexts such as varying scenes or tasks to mimic human behaviour and mitigate viewing biases [12].

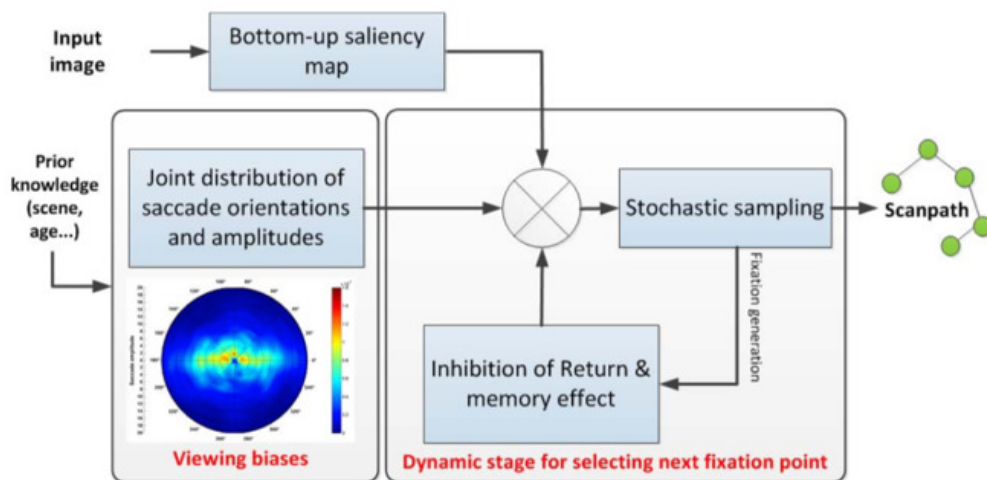


FIGURE 1.2: Le Meur's Saccadic Model. Source: O. Le Meur and Z. Liu, 2015.

1.3 Aims and Objectives

1. The goal of the project is to complete two technical tasks, expand on the second task by completing deeper analysis on the methodology used, as well as providing a comprehensive analysis of the results gathered through written investigation.
 - (a) Visualize the recorded fixation points of participants from the dataset onto the mammography images, and report on findings.
 - (b) Calculate saccade amplitude d and orientation ϕ from the dataset and construct visualizations including histograms and other formats to display differences between participant groups.
2. Using the resulting visualizations and new information, attempt to understand and construct further visual and verbal analysis of the results, elucidating and trying to understand the differences of results and their causalities as well as evaluating the methodology used to visualize the data and attempt to measure its utility for this sort of research.

2 Background

This chapter discusses the origins of the motivation towards the analysis completed in report, the experiment conducted to attain the data, and delivers comprehensive insight into the landscape and history of eye-tracking research in the field of medical imaging.

2.1 The Field of Medical Imaging

The discipline of medical imaging refers to the practice of producing images of the interiors of a body in order to be interpretable for the human eye and computers, with the purpose of aiding in diagnosis and developing treatments for diseases. Examples of medical imaging include x-ray, magnetic resonance imaging (MRI) and medical ultrasonography along with many others. As medical imaging techniques advance, they continue to support medical practitioners in detecting issues with greater success and the practice has become a necessary preliminary step in delivering patient care.

The current practice of examining medical images is constrained with regards to accuracy due to the reliance on human practitioners to observe the images, recent studies cite a 30% false negative diagnosis frequency with a similar false positive occurrence in particular fields of radiology [13]. This is largely due to the imperfect nature of the human visual system: the complexity of human physiology means accuracy is dependent on a number of variables. Some variables have not yet been entirely revealed while other factors, such as the practitioners experience, age or mood have been recognized by research, and their impact on performance has been widely evaluated, for instance, fixation duration (the time spent on a specific point on an image) decrease with age [14].

The risk of Inaccurate diagnosis poses a serious threat to global health. Patient's conditions can be mistreated or untreated, and this threat is further amplified by the shortage of radiographers in certain regions of the world. In England the number of patients having to wait over 6 weeks for an MRI or CT scan has risen from 9,000 to 29,000 between 2017 and 2018 [15]. It is therefore imperative to conduct a study which attempts to understand the differences in gaze behaviour of experts and novices in the field of radiography, in order to better and faster train people to experts and build sophisticated systems which assist practitioners in performing accurate and efficient diagnosis.

2.2 Purpose of Studying Gaze Behaviour

The field of eye tracking is a prominent scientific method for understanding gaze behaviour. Eye tracking provides a technique for recording subjects pupil movements when observing a scene, thus providing insight into how the subject evaluates the scene. Therefore, analyzing the observations from an eye tracking experiment has the potential to deliver insight into the technique and errors exhibited by subjects. A study by Voisin et al. [16] analysed quantitative metrics from eye tracking data in a study which monitored six radiologists while evaluating the probability of malignancy of forty mammographic images, utilising a Mirametrix S2 eye tracking device. Relevant data such as fixation points, their associated duration and fixation- saccade ratio was extracted, and they were able to demonstrate that gaze metrics were highly correlated with a radiologists diagnosis errors.

2.3 Experience as an Independent Variable

This report will deliberately focus on the variable of experience. We explore the impact of experience, controlling it as an independent variable, and observe its effect on the dependent variable: medical practitioners gaze behaviour when examining mammography images, which are low-energy X-rays of human breasts.

For this report, a dataset acquired from three groups of medical practitioners with different degrees of experience in mammography reading experience was employed. These three groups were comprised of Expert Radiologists ("Experts"), Trainee Radiologists ("Trainees") and Physicists. Experience as indicated by previous research in the medical field when examining medical images is a deterministic factor of performance. Krupinski et al. [17] studied the impact of training on viewing behaviour with four pathology residents throughout their training period. After completing the training, statistically significant improvements were seen through an increase in efficiency, caused by a decrease in the number of fixation points and re-visitations of already examined points.

2.4 Eye-tracking Experiment

The data was acquired from an experiment performed in the mammography reading room of University Hospitals of the KU Leuven. It is comprised of 198 multi-lateral oblique (MLO) views from 98 anonymised cases. The mammography images occur in the medio-lateral oblique (MLO) perspective which is an angled view from a lateral projection. The participants were presented the 98 cases in random order and asked to examine each image with their eye movements being tracked using a SensoMotoric Instrument (SMI) Red-m eye-tracking device with a 250 Hz sampling rate.

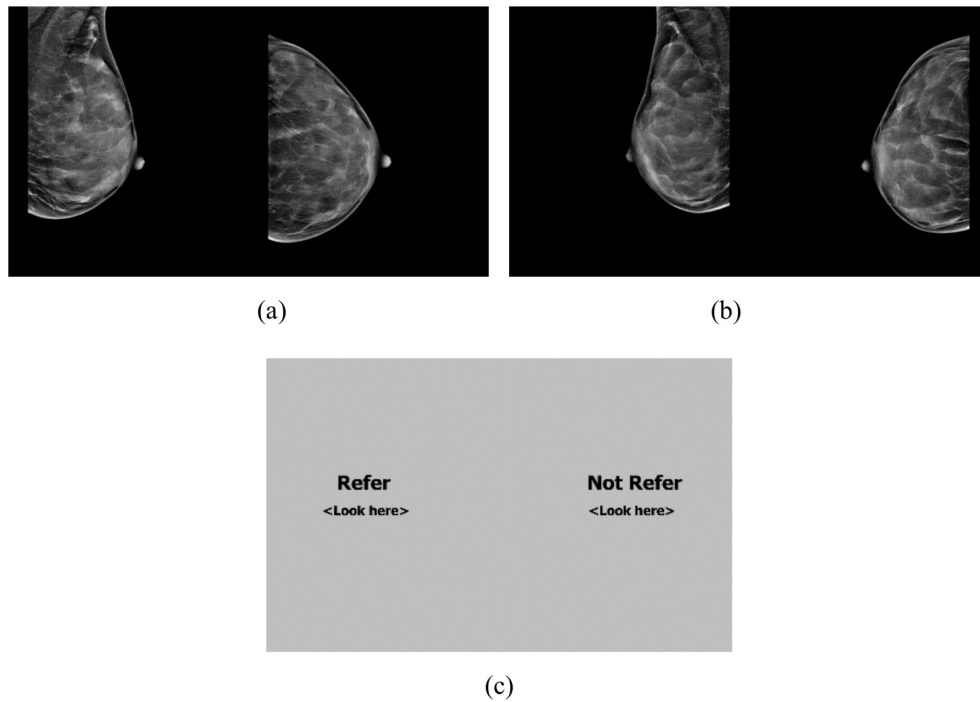


FIGURE 2.1: Illustration of the experimental procedure: (a) MLO and CC views of the left breast, (b) MLO and CC views of the right breast, and (c) question asked to the radiologists after viewing (a) and (b).

The aforementioned participants were three expert radiologists, comprising of eight, fifteen and twenty years of experience in mammography reading, referred to as R1, R2 and R3 henceforth. Three trainee radiologists T1, T2 and T3, and finally, two physicists, referred to as P1 and P2.

The experiment gathered gaze information from the raw-eye tracking data recorded during the procedure. The data points included number of fixations per stimulus, coordinates of fixations and duration periods. It is this catalogue of data which this report will be conducting analysis on.

2.5 Previous Work

Eye-tracking research within the field of medical imaging has been a regular topic of research topic in order to uncover how these medical practitioners perform visual search and recognition tasks. In this section, a thorough review is presented of the available literature in the domain of eye-tracking in medical imaging. Thus far, conducted research has revolved around three subject matters:

1. Identification of Practitioners Visual Search Patterns
2. Influence of Expertise and Experience on Performance
3. Impact of Training on Viewing Behaviour

2.5.1 Identification of Practitioners Visual Search Patterns

Studies of such sort begun as early as 1981. Carmody et al. [18] studied 4 radiologists as they performed diagnosis on 10 chest x-ray slides, 4 of which were normal and 6 abnormal. They were instructed to press a key upon finding a nodule and wore a pair of glasses with a corneal reflection to make their eye movements monitorable. The data recorded included eye movements and fixations, such as visual dwell times. The study concluded that false negative errors were influenced both by the visibility of a nodule as well as the scanning strategies employed by the radiologist.

Later Beard et al. [19] leveraged a more advanced Eye Mark Recorder Model V on 3 CT scans, comprised of a single chest and multiple abdominal scans, with the intention of understanding visual scan patterns used by radiologists when interpreting the given CT scans. 4 radiologists and 1 radiology resident participated, with each round of interpretations conducted by the radiologists, their scan patterns were rendered manually from tape records and systematic sequential visual scan patterns were discovered.

Suwa et al. [20] conducted an experiment using 20 CT scans, 10 being normal and 10 with pathologic lesions with 8 dentists leveraging an Eye-tracking system model 504. The intention of the study was to analyze whether dentists exhibit different behaviour when observing pathologic vs normal images. The variables recorded included time to discriminate, fixation points count, travel distance between fixation points and average time per fixation. The findings revealed that there was indeed a difference. Dentists tended to move sequentially when viewing a normal image, whereas in the case of an image with a pathologic lesion they tended to focus on suspected regions. Furthermore, the travel distance between fixations and minimum gaze fixation time was longer for pathologic images.

Previous literature has also included studies involving mammogram analysis. Kundel et al. [21] took gathered data in three independent institutions, employing experienced mammographers, mammography fellows and radiology residents harnessing an Applied Science Laboratories (ASL) eye-tracking device and tasked the participants with detecting cancers within the mammographs. The findings of the experiment found that 57% of the cancer locations were fixated within the first second of screening, leading to the conclusion that an initial detection of a cancer happens prior to visual scanning and that greater expertise may be comprised of a change to a scan-look-detect to look-detect-scan process.

Further studies revolving around mammography were conducted, one in particular involving Voisin et al. [16] devoted to analysing the association between gaze patterns and performance in diagnostics in lesion detection. The participants were given 40 mammograms, 20 being benign and 20 malignant and were tasked with giving the probability of a malignancy existing within a given image. A Mirametric S2 eye-tracking device was used and by tracking multiple gaze metrics including number of fixations, duration of fixations, fixation/saccade ratio and saccade length

they discerned that these variables were highly correlated with a radiologists diagnostic errors.

Studies have also branched into colonoscopy videos. Almansa et al. [22] using a ASL mobile eye-tracking device observed the relationship between gaze patterns and adenoma detection rate. Using 11 endoscopists to watch three high-definition videos from regular colonoscopies and tracking relevant data using the device, it was revealed that participants who detected the largest amount of adenomas displayed a tendency to focus on the centre of the screen, thus their central gaze time was significantly correlated with the amount of time the endoscopists spent on the centre of the screen.

2.5.2 Influence of Expertise and Experience on Performance

Surgery

Law et al. [23] hypothesised that there would be distinctive differences in the gaze behaviour between experts and non-experts of laparoscopic surgery. The group conducted an eye-tracking study using five expert surgeons and five students. The subjects were instructed to perform a virtual task: touch a small target with a virtual laparoscopic tool as quickly as possible and minimize error, this was done in 2 blocks of 5 trials using a ASL 504 remote eye-tracking device. The study demonstrated that the experts were significantly faster and more accurate and that novices spent longer periods of time observing the tool than experts.

Another study revolving around laparoscopic surgery was conducted by Kocak et al. [24] with eight experts, eight intermediates and eight novices, employing a Cyclops Eye Trak saccadometer. The participants were tasked with performing three common laparoscopic tasks, these included loops, rope and beans. The findings revealed experts exhibit a lower saccadic rate than novices and experts also had a higher peak velocity on average.

Ahmidi et al. [25] gathered 5 experts and 6 novice surgeons for an experiment involving endoscopic sinus surgery. The surgeons were instructed to locate and then touch using an endoscope a given anatomy in the sinus cavity. The experiment found that expert surgeons gaze information contained structure for eye level recognition which was not found in novices, and they also used a Hidden Markov Model to develop a strategy for determining the expertise levels of surgeons.

In the same year Richstone et al. [26] published a study which included twenty-one surgeons partaking in a laparoscopic surgery, eleven of the surgeons participated in the live surgery and 10 in the simulation. The participants wore a EyeLink II eye-tracking device which recorded their blink rate, fixation rate, pupil metric and vergence. They concluded that for both scenarios of live and simulation surgery, the metrics recorded can be used to distinguish experts and non-experts reliably.

Khan et al. [27] employed twenty-two surgeons, two of which were experts and twenty were novices in a study where the participants performed a surgical task and

later watched a video while wearing a Tobii X50 eye-tracking device. For the study, sixteen laparoscopic cholecystectomy cases were used and a statistically significant difference was revealed, demonstrating that a 55% overlap existed for experts between “doing” and “self-watching”, whereas only 43.8% for junior residents.

Medical Diagnosis

Nodine et al. [28] directed a study using 3 radiographers and 6 radiography trainees, leveraging an ASL 4000 SU eye-tracking device in evaluating forty mammography test cases, twenty containing lesions and twenty without. They participants were simply instructed to decide a given case was either normal or abnormal. The observations concluded that the experienced radiologists had significantly better performance regarding accuracy, although no significance in decision time was observed.

Similar to the study above, Tourassi et al. [29] observed in a study involving three breast imaging radiologists and three trainees when tasked with evaluating 20 mammograms, that residents accuracy was lesser than that of an expert, and the recall rate for both participants were similar on average.

There have been multiple studies involving CT images that focus on comparing experts and novices when conducting diagnosis such as Cooper et al. [30], Matsumoto et al. [31] and Mallet et al. [32]. Cooper et al. [30] discerned through a study of three experts, one trainee and four novices, when observing stroke images that, in the case of an acute stroke, trainee readers observed the region of interest with the 34th fixation in contrast to the expert which noticed it in the first fixation. Furthermore, for a chronic stroke case, novices spent a lesser time viewing the affected area than experts. These factors determined the statistical significance that was calculated, proving that experts performed better than trainees. The study performed by Matsumoto et al. [31] also revolved around stroke cases, completed two years later and involving twelve control subjects comprised of nurses, medical technologists, psychologists and medical students. Using twelve neurologists, they uncovered that neurologists and the control subjects all gazed at salient points in the images, however, it was only neurologists who gazed at visually low-salient areas with clinical importance. Mallet et al. [32] employed twenty-seven experienced and thirty-eight inexperienced radiologists to study twenty-three 3D CT colonography videos. Their results revealed that experienced participants were better at polyp identification, also there was no difference between the groups in the percentage of pursuits and total examination period.

Further techniques of medical imaging were utilized for discerning differences in performance between experts and novices, Manning et al. [33], Vaidyanathan et al. [34] and Leong et al. [35] employed radiographs, dermatological images and chest images. Manning et al. [33] with three distinct groups of eight experienced radiologists, five radiographers and eight novice radiography students leveraging an ASL 504 remote eye-tracker and various gaze metrics displayed that radiologists and radiographers post training were better at performing diagnosis than novices.

Vaidyanathan et al. [34] contrasted twenty-two dermatology experts and twelve novices eye movements when observing thirty-four dermatological images. Their findings were that experts employ weighing importance to regions following a brief fixation whereas novices require several re-fixations. Furthermore, they reported that saccade amplitude and median fixation duration were significantly higher for experts than novices. Leong et al. [35] employed twenty-five participants from different specializations to examine and identify any fractures in thirty-three skeletal radiographs. Using a Tobii 11750 eye-tracking device, they demonstrated that although there was no statistical significance in the time spent between the disciplines on viewing the radiographs, experts led a larger number of true positives.

2.5.3 Impact of Training on Viewing Behaviour

To determine the effectiveness of training on gaze behaviour in laparoscopic surgery Wilson et al. [36] employed an ASL mobile eye-tracking device and thirty trainee surgeons in an experiment which divided the participants into three groups, each receiving a different training. The first group was displayed an experts eye movements when performing a coordination task. The second group was given the same option as the first but without a gaze cursor. The third group was only allowed to examine their own performance and not given any video like the predecessors. A preliminary statistical analysis validated that there existed no statistical significance in the difference of performance of the groups. However, following the training, it was proven that the first groups performance was significantly faster than the other groups. The first group also spent significantly more time on target locking fixations.

A similar study was completed by Vine et al. [37] which intended to evaluate the impact of gaze training in laparoscopic surgery without informing the participants on the purpose of the training. Twenty-seven naïve participants were dispersed into two groups, one a discovery learning group and the other a gaze training group. The groups were tasked with moving foam balls into a cup using with only a single instrument. A preliminary analysis revealed no statistical significance prior to training. However, afterwards the gaze training group was able to complete the task significantly faster and displayed more accuracy than the discovery learning group.

Krupinski et al. [38] conducted a study in pathology with a ASL SU4000 eye-tracking device. It monitored four pathology residents viewing behaviour over their training period i.e. once a year for four years. In each experiment, the participant was instructed to choose their top three locations they would like to magnify in twenty breast core biopsy surgical pathology cases. Using fixation positions and dwell times as metrics, they determined that residents had become more efficient each year, displaying fewer fixations and fewer revisited locations.

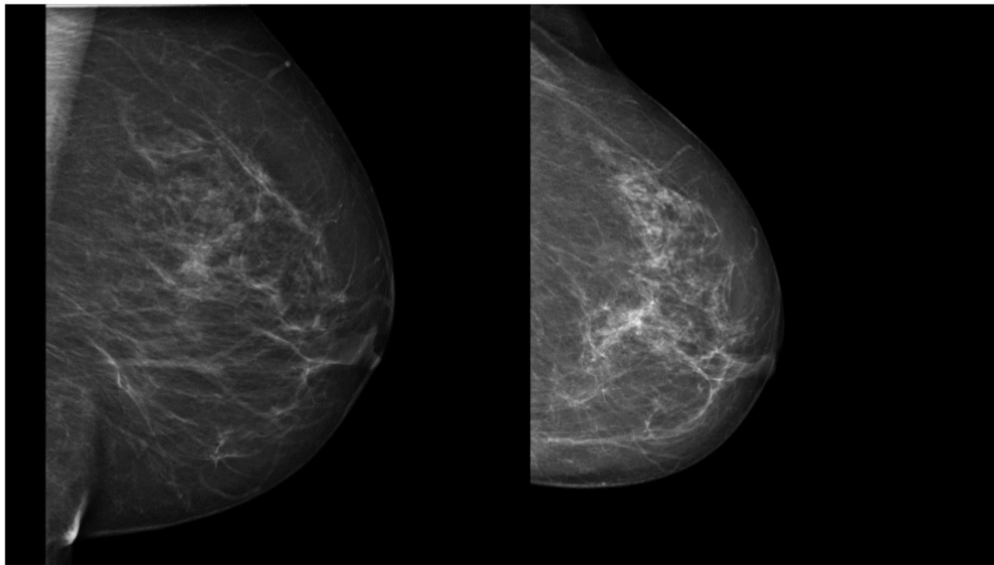
3 Approach

This section of the report discusses the circumstances of the performed analysis, details of the mathematical methods applied to achieve the objectives in the deliverables, the suitability of the approach and justification of the techniques applied.

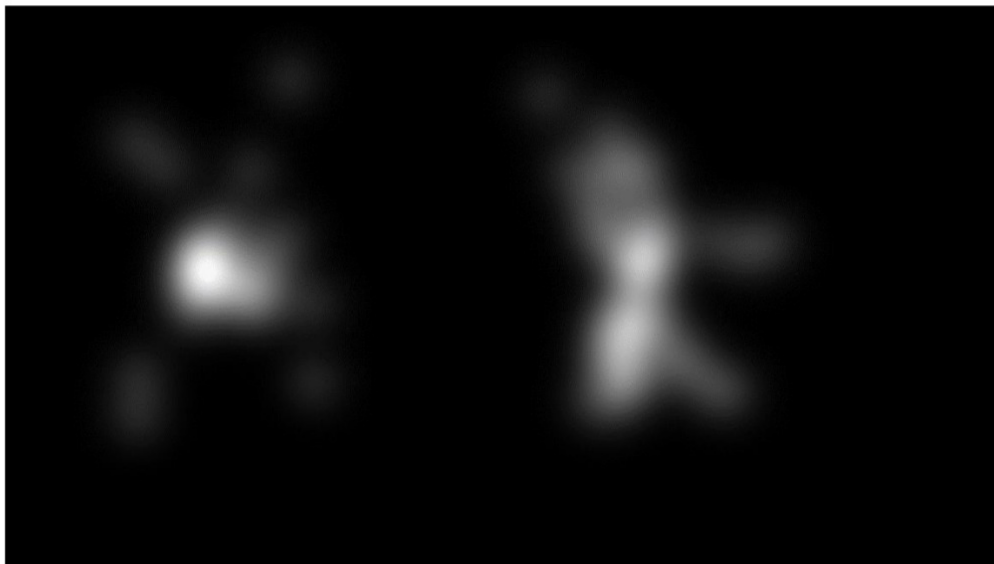
3.1 Extension of Previous Work

In order to discuss the approach used in this report, it is necessary to highlight that the work conducted in this report is an extension of a prior analysis performed by Lucie Lévêque in her PhD thesis ‘Analysing and quantifying visual experience in medical imaging’ [39].

In her thesis, Lucie employed the same dataset as in this report and completed an investigation of the impact of expertise and experience on gaze allocation, and evaluated the performance of state-of-the-art saliency models. Through analyzing the mean duration of fixation points across all the stimulus, Lucie generated computational saliency maps to illustrate the directed attention of the participants and to understand imperative areas of the stimuli. Additionally, a visual and statistical analysis was included, deducing the presence of statistical significance between various participant groups and within participants in discrete groups, and argued that gaze metrics can be utilized to quantify the extent of which trainees or physicists are in agreement with experts in regards to where to focus on in images.



(a)



(b)

FIGURE 3.1: Illustration of the saliency maps constructed for two patient' cases: (a) patient 1, MLO and CC views of the left breast.

3.2 Limitations of Previous Work and Further Approach

While Lucie's work was useful and the necessary preliminary data analysis of the data, as she explicated in her conclusion; the analysis does not offer a complete perspective on the information hidden within the data, and computer-generated

saliency maps were judged insufficient to generate accurate scan paths emulating human gaze behaviour [40].

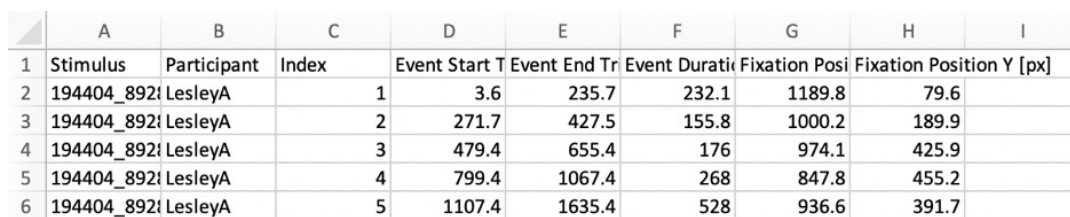
To deal with these limitations, this report focuses on examining saccades. As mentioned in Chapter 1, saccades are rapid, ballistic movements of eyes between fixation points, and can be reduced to two components which this report will use to perform statistical analysis on: saccade amplitude and saccade orientation.

3.3 Data Processing

The raw data itself was not undecipherable by a human reader. The complete catalogue of data was scattered over multiple files and folders. Folders containing csv files referenced the stimulus (mammograms) observed by the participants and contained eye-tracking data gathered throughout the experiments (See Figure 3.2). The stimuli were stored as bmp files in a separate folder from the csv files. Python was employed with several libraries to assemble the images and data, and transform it into accessible and manipulable formats.

The original data in csv files were as follows:

- Stimulus ID
- Participant
- Index
- Event Start Trial Time [ms]
- Event End Trial Time [ms]
- Fixation Point X [px]
- Fixation Point Y [px]



	A	B	C	D	E	F	G	H	I
1	Stimulus	Participant	Index	Event Start Time	Event End Time	Event Duration	Fixation Position X	Fixation Position Y	
2	194404_892	LesleyA	1	3.6	235.7	232.1	1189.8	79.6	
3	194404_892	LesleyA	2	271.7	427.5	155.8	1000.2	189.9	
4	194404_892	LesleyA	3	479.4	655.4	176	974.1	425.9	
5	194404_892	LesleyA	4	799.4	1067.4	268	847.8	455.2	
6	194404_892	LesleyA	5	1107.4	1635.4	528	936.6	391.7	

FIGURE 3.2: Screenshot of the sample csv file containing eye-tracking data acquired in the original experiment performed at University Hospitals of the KU Leuven.

3.3.1 Gaze Data Plotting

The initial approach taken to perform the analysis was to visualize the gaze data. This helped in developing a better understanding of the raw data and of the problem. Fixation points with their according index points were plotted superimposed onto the stimuli, with legends to illustrate which participant performed the fixation.

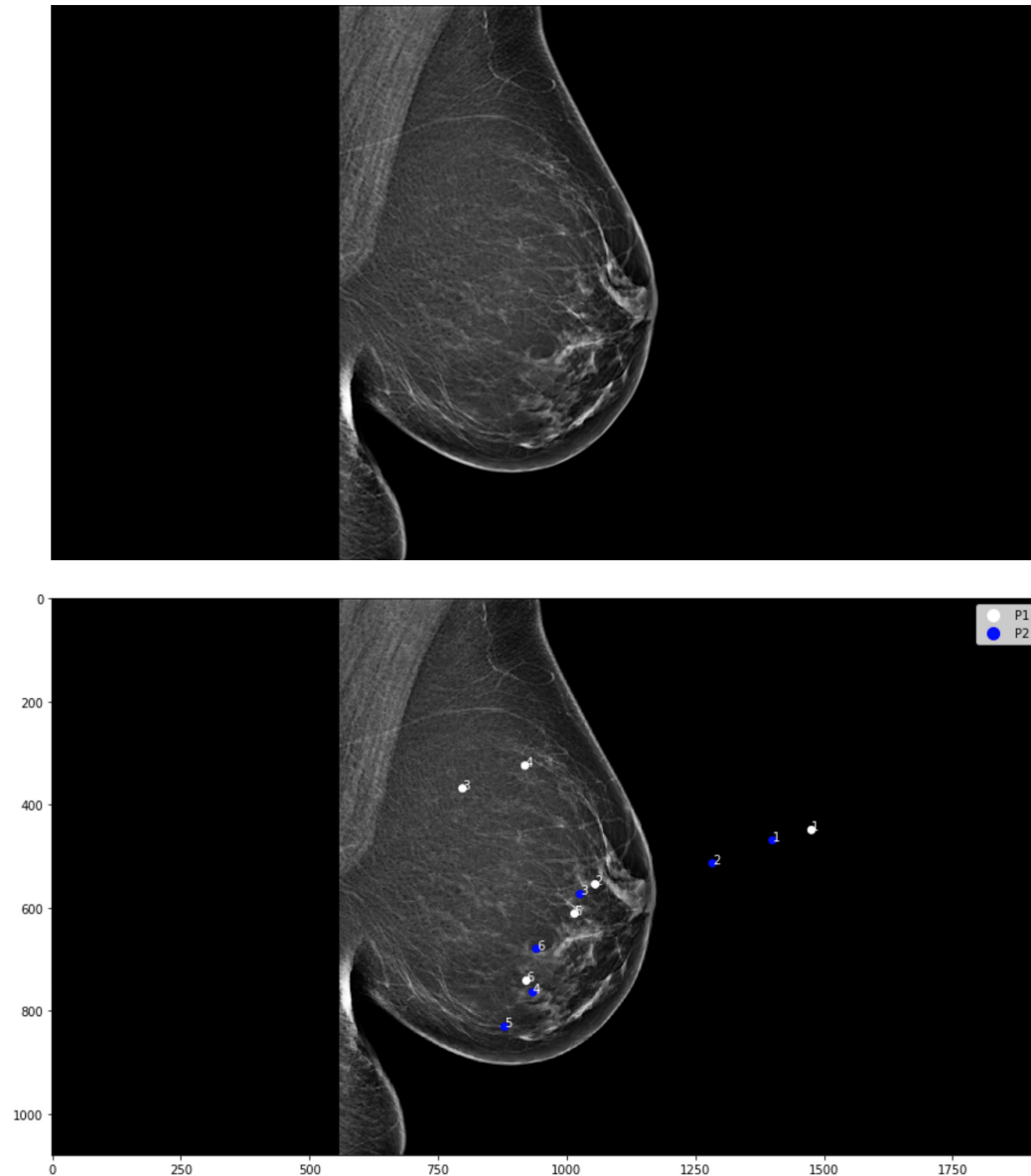


FIGURE 3.3: Top: Original mammogram sample bmp image. Bottom: Result of visualization of fixation points performed by Physicists P1 and P2 on the original mammogram.

By overlaying fixation points onto the stimuli, it helped in observing whether any potential outliers or oddities existed within the data. Additionally, it aided in contrasting gaze behaviour between participants as seen in Figure 3.3. However,

as mentioned in section 3.1, Lucie L  v  que did extensive work with the fixation points and stimuli, generating saliency maps and statistically analysing the means of fixation time durations, therefore this report focused on another approach.

3.3.2 Saccade Analysis

The second approach was to derive quantitative values for the components of a saccade eye movement: saccade amplitude and saccade orientation, and afterwards perform visual and additional statistical analysis on the values across each participant group.

All the columns in the raw data was used apart from duration of fixations for this task. Fixation Point X and Y along with their appropriate index points were vital to calculate saccade amplitude and saccade orientation. The index points displayed the order in which each fixation had occurred, and saccade amplitude and orientation exist between two fixation points. A detailed explanation is given below:

- x, y coordinate fixation points: The coordinates of a single location where a participant had fixated his gaze on
- Indexes of fixation points: The index values indicate the sequence in which the fixation points were performed by a participant
- Saccade Amplitude d , expressed in degree of visual angle, is the Euclidean distance between two consecutive fixation points P_t and P_{t-1}
- Saccade Orientation ϕ is the angle, expressed in degree, between two consecutive fixation points P_t and P_{t-1}

Saccade Amplitude d was calculated using the Euclidean Distance between sequential fixation points $P_t(x_2, y_2)$ and $P_{t-1}(x_1, y_1)$, defined as:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (3.1)$$

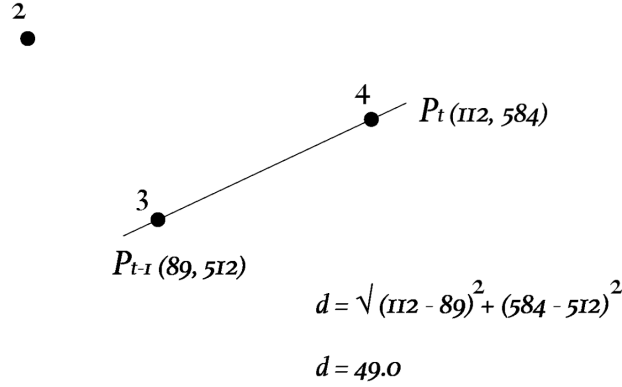


FIGURE 3.4: Example of a calculation for saccade amplitude between two sequential fixation points.

Saccade Orientation ϕ was calculated using the angle between two fixation points $P_t(x_2, y_2)$ and $P_{t-1}(x_1, y_1)$; defined as the arc tangent of theta θ :

$$\phi = \tan^{-1}(\Theta) \quad (3.2)$$

Where θ :

$$\Theta = \frac{dy}{dx} \quad (3.3)$$

And where delta dy and delta dx :

$$dy = (y_2 - y_1) \quad dx = (x_2 - x_1) \quad (3.4)$$

Therefore, to derive the angle from two fixation points:

$$\phi = \tan^{-1}\left(\frac{y_2 - y_1}{x_2 - x_1}\right) \quad (3.5)$$

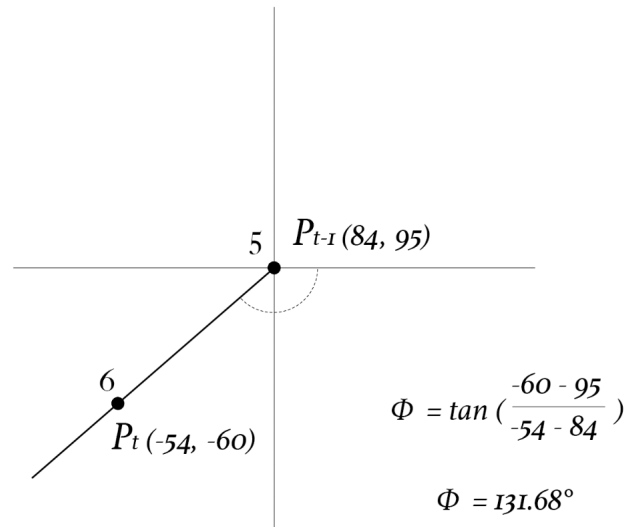


FIGURE 3.5: Example illustration of saccade orientation calculation between two consecutive fixation points.

3.4 Visual Analysis

The resulting quantities of saccade amplitude and saccade orientation data for each participant group occurred in large amounts, rendering individual analysis implausible. Therefore, to understand the data an agglomerate analysis was conducted, employing visualization techniques to comprehend and distinguish differences between samples. The visualization techniques included histograms, mean plots with error bars and polar plots.

The use of visual analysis methods e.g. histograms enabled the data to be interpretable by the human eye and permitted the use of additional statistical techniques to discern differences. Saccade amplitude and saccade orientation were both compared separately for each group and individual, and then together on a polar plot with a gaussian density function applied.

3.5 Comparing Histograms

Histograms were a reoccurring visual analysis technique leveraged frequently as seen in Chapter 5: Results and Evaluation. In order to maximize the utility of histograms, two histogram similarity comparison techniques were used to derive quantitative metrics which measure how similar were two histograms to each other and to discern the performance and gaze behaviour differences between participant groups. The formulas are defined as:

1. Histogram Intersection

$$D_{\cap} = 1 - \frac{\sum_i (\min(h_1(i), h_2(i)))}{\min(|h_1(i)|, |h_2(i)|)} \quad (3.6)$$

This method simply takes (for each pair of bin values) the minimum value of two histograms and sums them across all bins. The larger the measure is, the more the two histograms are similar.

2. L=2 or Euclidean distance

$$D_{L2} = \sqrt{\sum_i (h_1(i) - h_2(i))^2} \quad (3.7)$$

This method simply squares (for each pair of bin values) the difference of two histograms and sums them across all bins. Finally, the square root of the sum reveals the similarity measure. The smaller the measure is, the more the histograms are similar.

A caveat surrounding histogram similarity comparisons is there is an array of available methods, each with different advantages, and performances vary depending on conditions. Therefore, it can be a complex task to select the optimal method. One way around this, as seen in this report is through applying more than one approach and observing whether there is consistency in the results.

3.6 Statistical Analysis

For identifying statistical significance in the differences discovered between participant groups, one-way Analysis of Variance (ANOVA) tests were employed. ANOVA tests are a technique for comparing the means of two or more independent samples with the assumption of normality and homogeneity of the variance. In addition, the report uses independent t-tests throughout to further support the arguments presented and to distinguish whether the observed analyzed valued between participant groups are significantly different or whether they could have occurred by chance.

Other than group-by-group comparison, for each group, individuals in groups were compared with identical techniques as mentioned above. This was an important step in order to evaluate whether any member of each group had displayed any significant difference in their observation recordings which could of led to misguided conclusions on the differences between parties.

4 Implementation

This chapter discusses the technologies, libraries and techniques applied to implement the methods mentioned in the approach.

4.1 Environment

The scope of the project encompassed performing an in-depth visual and statistical analysis on a dataset which had been collected in a previous experiment. Therefore, no additional data collection was necessary, and the data had already been approved for any ethical standards. The implementation was completed on a MacBook Pro 2016 2 GHz Intel Core i5 with 8 GB memory, running Python 3.7.1.

4.1.1 IPython and Jupyter Notebook

Python is a well-known general-purpose computer language with a respectable reputation. However, I will justify the decision to rely on it for this project over other languages prominent in this field of research for instance R. The primary reason for selecting Python was the flexibility, extensibility and portability it offered, particularly when used with the Jupyter notebook tool, which offers an interactive experience for exploring data in one environment and facilitates seamless sharing of results and code.

4.1.2 Libraries

The core libraries used to complete the bulk of the technical aspects of this research project were the python libraries Numpy, Scipy and Matplotlib.

NumPy

NumPy provided an extensive library for manipulating and working with large multi-dimensional arrays as well as matrices, accompanied by a vast amount of integrated mathematical functions to operate on the arrays which was imperative for working with the dataset provided for this project. Alongside a providing a larger number of tools and functions to manipulate data, NumPy's arrays are more compact than Python lists, therefore it was a crucial component for ensuring performance of the implementation.

SciPy

SciPy is an extension of NumPy in the sense that it provides additional tools for the scientific computing which was performed and offered more nuanced and specific functions required to operate on NumPy array objects implemented during the project.

Matplotlib

Matplotlib provided a plotting library which was built as an extension of NumPy and SciPy libraries. It facilitated methods for implementing coherent, intelligible and easy to manipulate data visualization models. The matplotlib library offered an over decade old library with a vast collection of tools suited for deriving meaningful data visualizations which was ideal for the technical aspects of the project which largely consisted of plotting data in various formats.

4.2 Algorithms

This section discusses the Python implementation of the algorithms used to achieve the agreed deliverables.

4.2.1 Data Processing

The algorithm for accessing and iterating through the data was implemented in such a manner as to facilitate the possibility of future additions to the data set e.g. additional eye-tracking data, without having to modify the code. In the listing below a for-loop was used which iterated through the data folders, extracted and stored each participant groups data into discrete dataframes representing independent groups i.e. *df_PP* for Physicists, *df_RRR* for Expert Radiologists and *df_TTT* for Trainee Radiologists.

```
1 for im_name in glob.glob('Stimuli eye-tracking MLO/*.bmp'):
2
3     mammograph = im_name.replace('Stimuli eye-tracking MLO/', '')
4     df_PP = pd.read_csv('MLO-PP/' + mammograph + '.csv')
5     df_RRR = pd.read_csv('MLO-RRR/' + mammograph + '.csv')
6     df_TTT = pd.read_csv('MLO-TTT/' + mammograph + '.csv')
```

LISTING 4.1: Accessing the dataset

4.2.2 Plotting Fixation Points

In order to visualize the fixation points and get an initial understanding of each participants gaze behaviour, inside of the loop which accessed the raw data and stored it in dataframes, the dataframes were indexed to retrieve and plot the fixation points. Additionally, to make it obvious who performed the fixation, the fixations

were colour coded by participant in each group with index labels and a legend was added to add clarify.

```

1 #Mapping colours to participant names e.g. 'w' or white to the first
  Physicist
2 colors = ['w', 'b', 'y']
3 col_PP = df_PP.Participant.map(dict(zip(participants_PP, colors)))
4 col_RRR = df_RRR.Participant.map(dict(zip(participants_RRR, colors)))
5 col_TTT = df_TTT.Participant.map(dict(zip(participants_TTT, colors)))
6
7 plt.figure(figsize=(50,100))
8 img = plt.imread(im_name)
9 plt.subplot(1,3,1), plt.imshow(img, cmap='gray'), plt.scatter(x=df_PP['
  Fixation Position X [px]'], y=df_PP['Fixation Position Y [px]'], c=
  col_PP)
10 plt.subplot(1,3,2), plt.imshow(img, cmap='gray'), plt.scatter(x=df_RRR['
  Fixation Position X [px]'], y=df_RRR['Fixation Position Y [px]'], c=
  col_RRR)
11 plt.subplot(1,3,3), plt.imshow(img, cmap='gray'), plt.scatter(x=df_TTT['
  Fixation Position X [px]'], y=df_TTT['Fixation Position Y [px]'], c=
  col_TTT)
12
13
14 dfs = [df_PP, df_RRR, df_TTT]
15 plots_array = [131, 132, 133]
16 participant_groups = [participants_PP, participants_RRR, participants_TTT]
17
18 n = 0
19 s = 0
20 for x in dfs:
21     n = 0
22     ix_ind = x['Index']
23     for i in ix_ind:
24         subplot(plots_array[s]), plt.annotate(i, (x.iloc[n,6], x.iloc[n
25         ,7]), color='w')
26         n = n + 1
27         s = s + 1
28
29 b = 0
30 for group in participant_groups:
31     legend_objects = []
32     i = 0
33     for individual in group:
34         legend_objects.append(Line2D([0], [0], marker='o', color=colors[i
35         ], lw=0, label=name_replacement(group[i]), markersize=10))
36         i = i + 1
37     plt.subplot(plots_array[b]), plt.legend(handles=legend_objects)
38     b = b + 1

```

LISTING 4.2: Visualizing the fixation points onto mammograms

4.2.3 Efficient Iteration

While performance was not imperative in this project since the code only has to be executed once to display all the visualizations and statistical analysis, there was a noticeable difference in performance between iterative functions accessing the constructed dataframe objects. Using `get_value(arg1,arg2)` delivered significant improvements in speed over more common methods such as `iterrows()` or `loc[arg1,arg2]`. Initially this information was acquired through internet research [41], which documented `get_value(arg1,arg2)` as 20 times faster than `iterrows()` and twice as fast as `loc[arg1,arg2]`, and when tested in the analysis conducted in the report, similar performance was observed. Using `get_value(arg1,arg2)` was useful in this scenario since the difference was noticeable.

```
1 x = l.get_value(l.index[i-1], 'Fixation Position X [px]')
2 y = l.get_value(l.index[i-1], 'Fixation Position Y [px]')
3 x_2 = l.get_value(l.index[i], 'Fixation Position X [px]')
4 y_2 = l.get_value(l.index[i], 'Fixation Position Y [px]')
```

LISTING 4.3: Improved performance for accessing and indexing dataframes

4.2.4 Calculating Saccade Values

To derive saccade values of saccade amplitude and saccade orientation methods, libraries were used therefore this simplified the overall process; SciPy provides a Euclidean distance function which was employed to calculate saccade amplitude and the `atan2` function available in the popular math library was utilized to calculate saccade orientation.

The more difficult segment of the work was mapping indexes to fixation points and storing them in arrays which represented the person who performed the fixation points, as the data in its initial state would contain dataframes for multiple participants therefore it was difficult to distinguish which participant performed which fixation or saccade. To tackle this, arrays were defined which captured each unique individual in a group i.e. the Expert Radiologists in RRR ('R1', 'R2', 'R3'), then, for each of the extracted individuals a new dataframe was constructed which stored only the gaze tracking information contained in the csv which corresponded to the individual for that particular stimuli.

```
1 participants_PP = df_PP.Participant.unique()
2 participants_RRR = df_RRR.Participant.unique()
3 participants_TTT = df_TTT.Participant.unique()
4
5 participant_groups = [participants_PP, participants_RRR, participants_TTT]
6
7 for group in participant_groups:
8     for individual in group:
9
10         if individual in participants_PP:
```

```

11     l = df_PP.loc[df_PP['Participant'] == individual]
12     max_row = l['Index'].max()
13     l.set_index('Index')
14
15     if individual in participants_RRR:
16         l = df_RRR.loc[df_RRR['Participant'] == individual]
17         max_row = l['Index'].max()
18         l.set_index('Index')
19
20     if individual in participants_TTT:
21         l = df_TTT.loc[df_TTT['Participant'] == individual]
22         max_row = l['Index'].max()
23         l.set_index('Index')
24
25     max_row = l['Index'].max()
26     l.set_index('Index')
27
28     for i in l['Index']:
29         if (i < max_row):
30             x = l.get_value(l.index[i-1], 'Fixation Position X [px]')
31             y = l.get_value(l.index[i-1], 'Fixation Position Y [px]')
32             x_2 = l.get_value(l.index[i], 'Fixation Position X [px]')
33             y_2 = l.get_value(l.index[i], 'Fixation Position Y [px]')
34             if individual in participants_PP:
35                 amplitude = distance.euclidean((x,y),(x_2,y_2))
36                 saccade_amplitude_PP.append(amplitude)
37                 saccade_orientation_PP.append(math.atan2(y_2-y, x_2-x)
38             )
39             individual_arrays_PP[individual].append(amplitude)
40             if individual in participants_TTT:
41                 amplitude = distance.euclidean((x,y),(x_2,y_2))
42                 saccade_amplitude_TTT.append(amplitude)
43                 saccade_orientation_TTT.append(math.atan2(y_2-y, x_2-x
44             ))
45             individual_arrays_TTT[individual].append(amplitude)
46             if individual in participants_RRR:
47                 amplitude = distance.euclidean((x,y),(x_2,y_2))
48                 saccade_amplitude_RRR.append(amplitude)
49                 saccade_orientation_RRR.append(math.atan2(y_2-y, x_2-x
50             ))
51             individual_arrays_RRR[individual].append(amplitude)

```

LISTING 4.4: Methods for calculating saccade values

4.2.5 Plotting Histograms and Lines of Best Fit

The most prominent method used in the analysis was data visualization using histograms. Histograms are simple to implement with Python, and the Matplotlib library as they come as a built-in function. The histograms were normalized to avoid visual distortions and maintain consistency throughout groups, with y and x axis limits set as well. For each plot a line of best fit was plotted, this was really helpful

in evaluating the histograms, as it made the disparities between different histograms clear.

```

1 subplot(4,1,1)
2 density = stats.gaussian_kde(saccade_amplitude_PP)
3 n1, x, _ = plt.hist(saccade_amplitude_PP, bins=100, density=True,
4                     edgecolor='k', color='#38597e', range=(0, 1400))
5 plt.ylim(0.000, 0.006)
6 plt.title('Physicists', fontsize=14), plt.xlabel('Saccade Amplitude',
7         fontsize=14), plt.ylabel('Probability', fontsize=14)
8 plt.plot(x, density(x), c='y')
9 groups_best_ax.plot(x, density(x), c='b')
10
11 subplot(4,1,2)
12 density = stats.gaussian_kde(saccade_amplitude_RRR)
13 n2, x, _ = plt.hist(saccade_amplitude_RRR, bins=100, density=True,
14                     edgecolor='k', color='#38597e', range=(0, 1400))
15 plt.ylim(0.000, 0.006)
16 plt.title('Expert Radiologists', fontsize=14), plt.xlabel('Saccade
17     Amplitude', fontsize=14), plt.ylabel('Probability', fontsize=14)
18 plt.plot(x, density(x), c='y')
19 groups_best_ax.plot(x, density(x), c='y')
20
21 subplot(4,1,3)
22 density = stats.gaussian_kde(saccade_amplitude_TTT)
23 n3, x, _ = plt.hist(saccade_amplitude_TTT, bins=100, density=True,
24                     edgecolor='k', color='#38597e', range=(0, 1400))
25 plt.ylim(0.000, 0.006)
26 plt.title('Trainee Radiologists', fontsize=14), plt.xlabel('Saccade
27     Amplitude', fontsize=14), plt.ylabel('Probability', fontsize=14)
28 plt.plot(x, density(x), c='y')
29 groups_best_ax.plot(x, density(x), c='r')
30
31 groups_best_ax.legend(('Physicists', 'Expert Radiologists', 'Trainee
32     Radiologists'), loc='upper right')
33 subplot(4,1,4)
34 plt.ylabel('Probability', fontsize=14), plt.xlabel('Saccade Amplitude',
35     fontsize=14), plt.title('All Groups', fontsize=14)
36 histogram_figure.show()

```

LISTING 4.5: Plotting Histograms and Gaussian Distribution plot

4.2.6 Euclidean and Intersection Distance Histogram Similarity

For deriving the Intersection distance of two histograms a simple function was implemented which quantifies and compares the overlap or intersection between two histograms. For Euclidean distance the SciPy method for calculating Euclidean distance was utilized on the histogram arrays.

```

1 def return_intersection(hist_1, hist_2):
2     minima = np.minimum(hist_1, hist_2)
3     intersection = np.true_divide(np.sum(minima), np.sum(hist_2))

```

```

4     return intersection
5
6 #Histogram Intersection Distance
7 int_dis_1 = return_intersection(n1, n2)
8 int_dis_2 = return_intersection(n2, n3)
9 int_dis_3 = return_intersection(n3, n1)
10 print(int_dis_1, int_dis_2, int_dis_3)
11
12 #Euclidean Intersection Distance
13 euc_dist_1 = distance.euclidean(n1, n2)
14 euc_dist_2 = distance.euclidean(n2, n3)
15 euc_dist_3 = distance.euclidean(n3, n1)
16 print(euc_dist_1, euc_dist_2, euc_dist_3)

```

LISTING 4.6: Measuring the similarity of histograms

4.2.7 Plotting Mean Values and Error Margins

The mean values for saccade amplitude were calculated using NumPy and plotted on a graph against other candidate means with a 95% error bar for showing the confidence intervals.

```

1 #plotting means
2 plot(0, np.mean(saccade_amplitude_PP), 'o', markersize=10)
3 errorbar(0, np.mean(saccade_amplitude_PP), yerr=(1.96*np.std(
    saccade_amplitude_PP))/np.sqrt(len(saccade_amplitude_PP)), linewidth
    =0.7, capsize=7, capthick=0.5)
4
5 plot(1, np.mean(saccade_amplitude_RRR), 'o', markersize=10)
6 errorbar(1, np.mean(saccade_amplitude_RRR), yerr=(1.96*np.std(
    saccade_amplitude_RRR))/np.sqrt(len(saccade_amplitude_RRR)), linewidth
    =0.7, capsize=7, capthick=0.5)
7
8 plot(2, np.mean(saccade_amplitude_TTT), 'o', markersize=10)
9 errorbar(2, np.mean(saccade_amplitude_TTT), yerr=(1.96*np.std(
    saccade_amplitude_TTT))/np.sqrt(len(saccade_amplitude_TTT)), linewidth
    =0.7, capsize=7, capthick=0.5)

```

LISTING 4.7: Mean value analysis

4.2.8 ANOVA and t-tests

For both one-way ANOVA and independent t-tests functions from the `skipy` library were used. The data was printed out and captured in a results table in this report then compared using verbal analysis.

```

1 # ANOVA test comparing each group
2 print('ANOVA test result: ', stats.f_oneway(saccade_amplitude_PP,
    saccade_amplitude_RRR, saccade_amplitude_TTT))
3
4 # Independent t-tests for group-by-group comparison

```



```

5 stat, p = ttest_ind(saccade_amplitude_RRR, saccade_amplitude_PP)
6 print('Independent t-test(Expert Radiologists | Physicists) p-value = ',p)
7 stat, p = ttest_ind(saccade_amplitude_PP, saccade_amplitude_TTT)
8 print('Independent t-test(Physicists | Trainee Radiologists) p-value = ',p
9      )
9 stat, p = ttest_ind(saccade_amplitude_TTT, saccade_amplitude_RRR)
10 print('Independent t-test(Trainee Radiologists| Expert Radiologists) p-
      value = ',p)

```

LISTING 4.8: Statistical analysis implementation

4.2.9 Polar Plots

Mapping and plotting saccade amplitude with its according saccade orientation was a difficult feat to accomplish, particularly due to implementing the density distribution of the plot.

The saccade orientation and saccade amplitude values had to be transformed, as can be seen on lines 3, 6 and 9. The original issue was that with the usage of polar plots resulted in incorrect distance values as the functions used were not compatible with polar plots, therefore the densities were computed on xy cartesian coordinates and applied sin and cos transformations prior to calculating the z values.

```

1 plt.figure(figsize=(30,30))
2
3 xy = np.vstack([np.sin(saccade_orientation_PP)*saccade_amplitude_PP,np.cos
4                 (saccade_orientation_PP)*saccade_amplitude_PP])
4 z = gaussian_kde(xy)(xy)
5
6 x2y2 = np.vstack([np.sin(saccade_orientation_RRR)*saccade_amplitude_RRR,np
7                  .cos(saccade_orientation_RRR)*saccade_amplitude_RRR])
7 z2 = gaussian_kde(x2y2)(x2y2)
8
9 x3y3 = np.vstack([np.sin(saccade_orientation_TTT)*saccade_amplitude_TTT,np
10                  .cos(saccade_orientation_TTT)*saccade_amplitude_TTT])
10 z3 = gaussian_kde(x3y3)(x3y3)
11
12 # Sorts the points by density, so that the densest points are plotted last
13 idx = z.argsort()
14 x, y, z = np.array(saccade_orientation_PP)[idx], np.array(
15                     saccade_amplitude_PP)[idx], z[idx]
16
17 idx2 = z2.argsort()
18 x2, y2, z2 = np.array(saccade_orientation_RRR)[idx2], np.array(
19                     saccade_amplitude_RRR)[idx2], z2[idx2]
20
21 idx3 = z3.argsort()
22 x3, y3, z3 = np.array(saccade_orientation_TTT)[idx3], np.array(
23                     saccade_amplitude_TTT)[idx3], z3[idx3]

```

```
23 ax1 = plt.subplot(311, polar=True), plt.scatter(x, y, c=z, edgecolor='',  
    alpha = 0.75), plt.title('Physicists', fontsize=14), plt.colorbar()  
24 ax2 = plt.subplot(312, polar=True), plt.scatter(x2, y2, c=z2, edgecolor='',  
    , alpha = 0.75), plt.title('Expert Radiologists', fontsize=14), plt.  
    colorbar()  
25 ax3 = plt.subplot(313, polar=True), plt.scatter(x3, y3, c=z3, edgecolor='',  
    , alpha = 0.75), plt.title('Trainee Radiologists', fontsize=14), plt.  
    colorbar()  
26  
27 plt.show()
```

LISTING 4.9: Plotting saccade amplitude with orientation onto polar plots

5 Results and Evaluation

This section of the paper will consist of an analysis of saccade values i.e., saccade amplitude and saccade orientation, which were derived from the dataset for each participant for every group: Physicists (P1, P2), Expert Radiologists ("Experts": R1, R2, R3) and Trainee Radiologists ("Trainees": T1, T2, T3).

Calculating and presenting saccade amplitude and orientation proved to be of value. The methodologies used in this section to present the differences between the participating groups provided insight into the possible characteristics which influence accuracy and overall performance of an individual when performing a diagnosis on a mammograph. While many of the reasons behind differences are only possible speculations, they do coincide with current research in gaze behaviour.

5.1 Fixation Points

Using coordinates for fixation points and indexes present in the dataset, the analysis for each of the 196 mammograph images included plotting for all participants their fixation points, this was done for every group as demonstrated in Figure 5.1. The fixation points were labelled with their corresponding indexes, the indexes recorded the order in which the fixation points occurred in the sequence, additionally, the name of the observer with the relevant colour was presented in the legend on the top-right of the image.

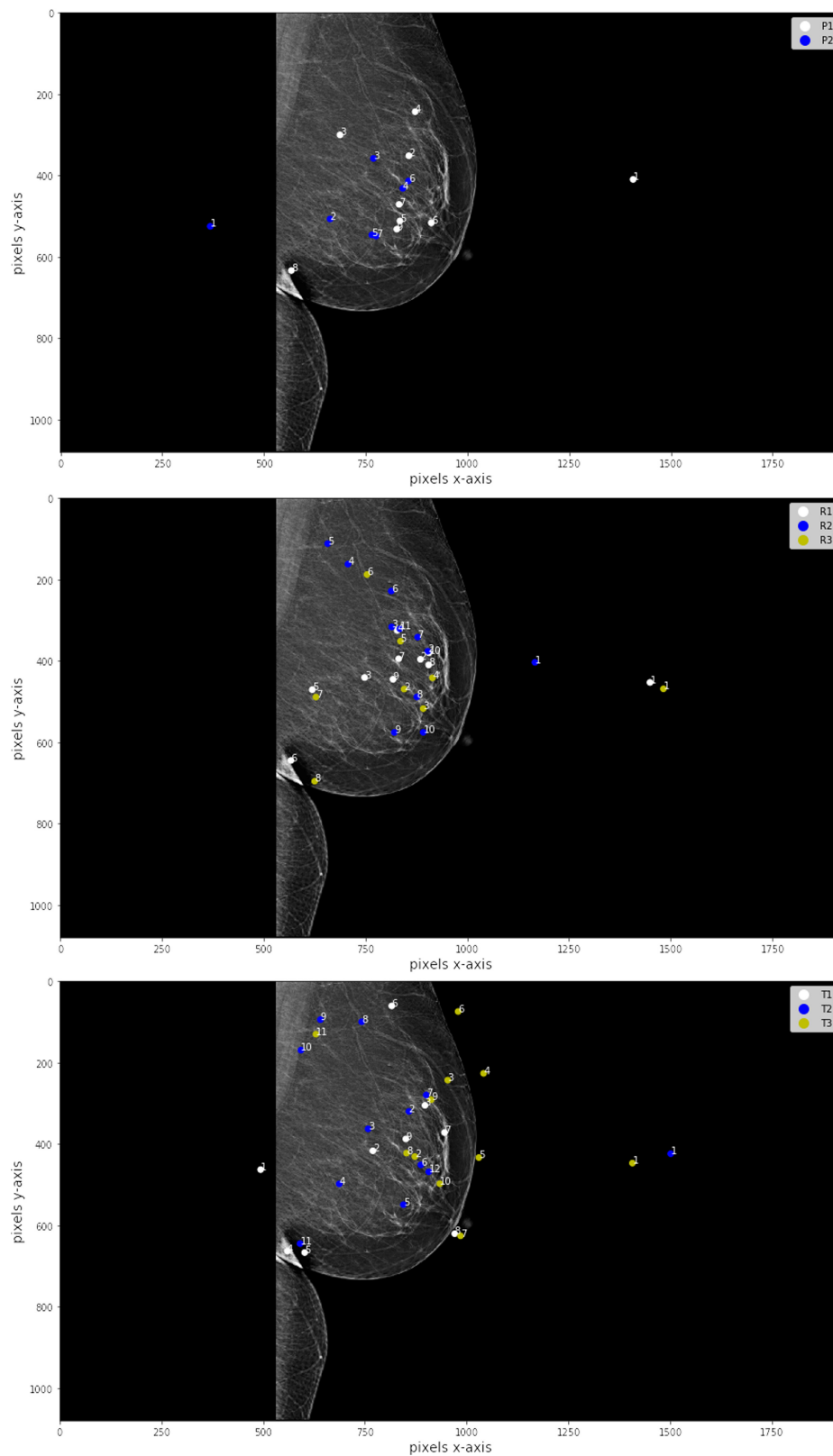


FIGURE 5.1: Examples of fixation plotting completed on images. (Top) Physicists, (Middle) Expert Radiologists and (Bottom) Trainee Radiologists. Done for all 196 images.

5.2 Saccade Amplitude Group Comparison

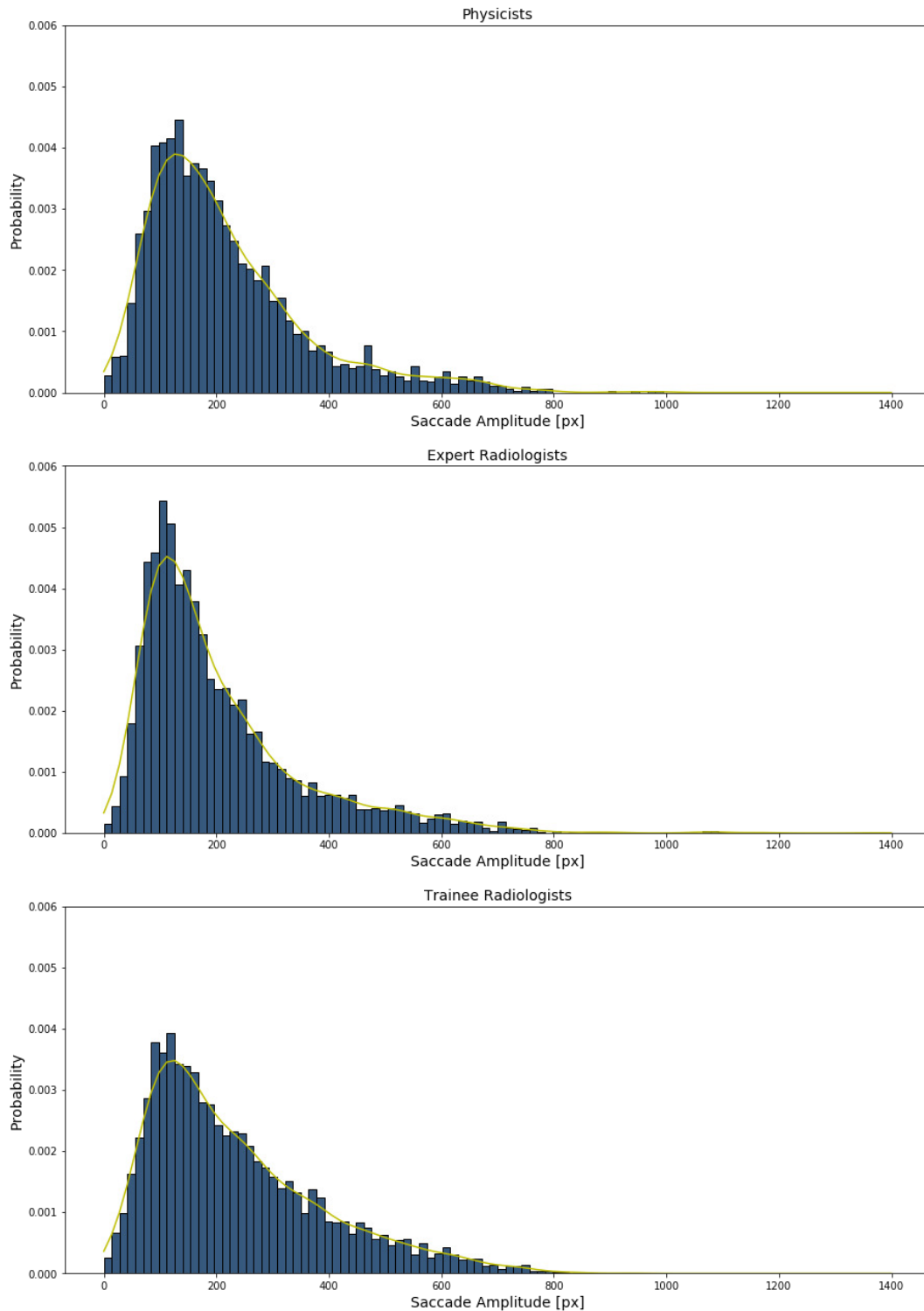


FIGURE 5.2: Distribution of saccade amplitude values for each participating group.

All groups to display a similar general behaviour, which is a positive sign and can be explained by the shared domain of expertise in medical imaging between expert radiologists, trainee radiologists and physicists. For all groups, the distributions of saccade amplitude display a right skew, and a majority of their values lying between 100-150px. There is an initial low frequency of saccades beginning at the lowest x-axis value (0px), followed by a sharp rise with a peak occurring between 70-140px. Following the peak which consistently occurs at a less than 200px, there occurs a steady decrease, therefore we can infer from this data that regardless of medical speciality or expertise, there exists between parties a preference for shorter saccades than longer, and that the majority of saccades will occur within a concentrated finite range in the lower end of possible saccades.

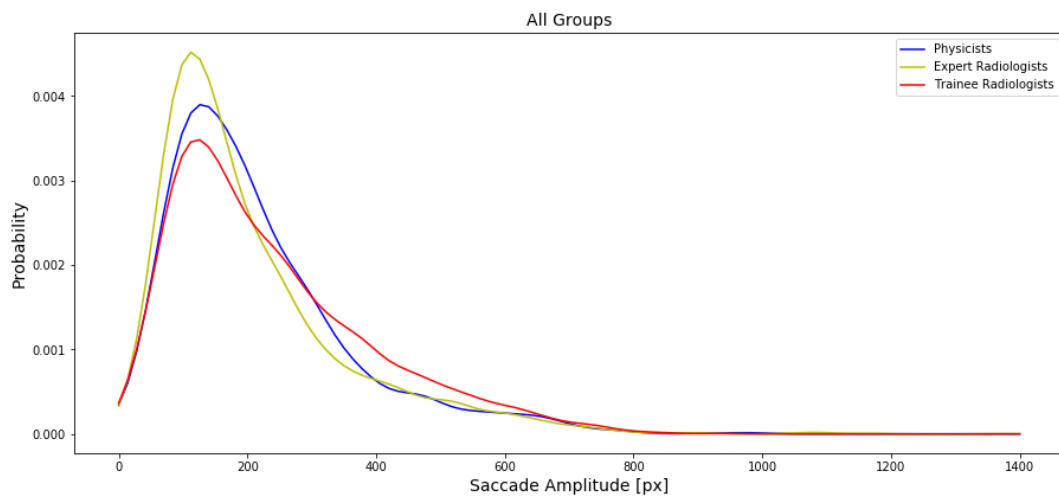


FIGURE 5.3: Overlay of the Gaussian line of best fit derived from each participant group.

The experts display a distinguishing difference in gaze behaviour. Figure 5.3 presents the set of lines of best fit in an overlay. Experts performed a higher frequency of shorter saccades relative to other groups, as can be seen in the fact that they have the highest peak in a most concentrated area. This can be seen as an indicator of expert behaviour and can likely be accounted by the difference in confidence between the participants. Experts are likely to know where to look instinctively in a consistent manner, and avoid making unnecessary saccades, whereas trainees and physicists may speculate and make ineffectual movements with low efficacy. It is also worth mentioning, that experts line of best fit extends further than other groups indicating that they also performed few saccades which were noticeably longer than other groups, however these were low in frequency therefore likely had little influence on the overall statistics.

Physicists and trainees exhibited a wider spread of saccades, their cumulative saccades form a line of best fit which is both lower in frequency in the range of saccades which experts performed (100-150px), and performed a larger number of

saccades in the surrounding, but outside of the suggested expert range. Physicists have a wider spread than trainees and experts, which suggests that trainees showed more similar behaviour to that of an expert than physicists.

To strengthen the analysis further, quantitative metrics were calculated using the histograms. Groups were compared in couples using Euclidean distance and Intersection distance for deriving a similarity measure of the distributions as seen in Table 5.1.

TABLE 5.1: Group histogram comparison using similarity measurements of Euclidean distance and Intersection distance.

Group A	Group B	Euclidean Distance	Intersection Distance
Physicists	Experts	0.00321	0.88747
Experts	Trainees	0.00367	0.86979
Trainees	Physicists	0.00249	0.90334

From Table 5.1 we can see that all groups exhibit similar behaviour, which supports the previous statement of a general pattern existing between groups. Trainees and physicists display the largest similarity (Intersection distance closest to 1 and Euclidean distance is the smallest), this could be explained by their shared novice level of experience in diagnosis. Whereas experts display the greatest contrast when compared against both groups which signifies their aptitude.

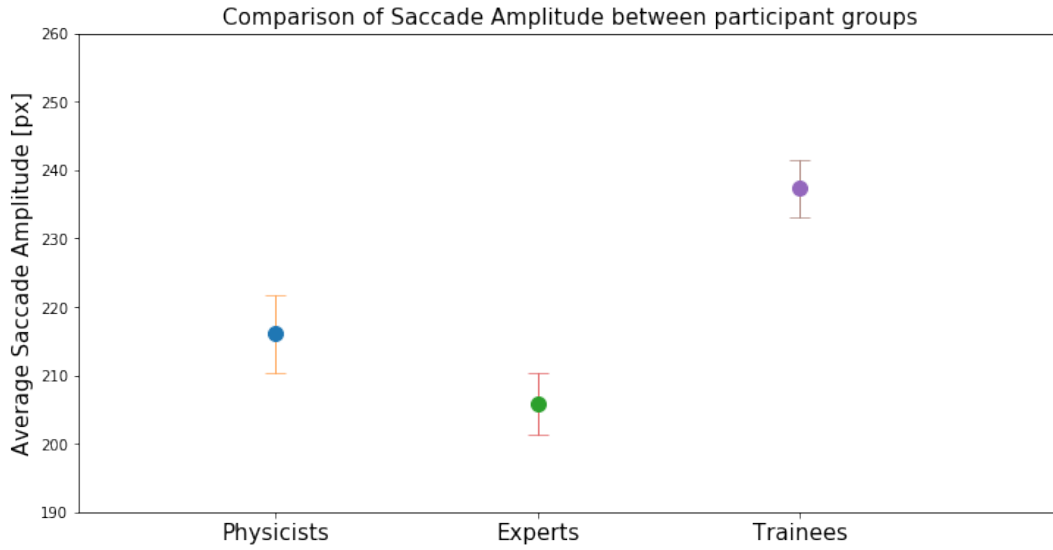


FIGURE 5.4: Graph of the mean values of saccade amplitude for each participating group, averaged over all recorded saccades for all test stimuli. Error bars indicate a 95% confidence interval.

TABLE 5.2: Mean values of saccade amplitude for each participating group.

Participant Group	Mean Values	Standard Deviation
Physicists	216.0	142.3
Experts	205.9	149.3
Trainees	237.3	156.1

Figure 5.4 illustrates the mean saccade amplitude of each participant groups given the dataset. The visualization indicates that there is difference between groups, with experts displaying the lowest mean value for amplitude, followed by physicists and finally trainees. The differences were further statistically analysed using an ANOVA test, with saccade amplitude set as the dependent variable, and groups as independent variable. The results of the ANOVA test are summarised in Table 5.3, where the F-statistic (i.e., F) and its associated significance (i.e, p-value) and degree of freedom (i.e., df) are contained.

TABLE 5.3: Results of the ANOVA test to evaluate the effect of participating group on the length of saccade amplitude.

Factor	df	F	P-value
Group	2	53.81	$5.24e^{-24}$

These results demonstrate that the difference in the means of saccade amplitude are statistically significant between the participating groups (i.e, $p < 0.05$). The result is desirable as it indicates that there are notable variances in the gaze behaviour of groups with different levels of experience, therefore, this proposes the opportunity for further analysis to determine the factors influencing performance and validates

the initial utility of gathering the data. In addition, the demonstrated differences suggest that improved performance is correlated with shorter saccades as experts displayed substantially shorter average saccades over the other groups.

Interestingly, trainees exhibit a longer resulting mean value than physicists and experts. This may be partially because trainees dataset contained a larger quantity of saccade amplitudes values than any other group, therefore, while they demonstrated a similar frequency of saccade amplitudes within the shared focal range, they also performed a larger amount of saccade amplitudes in the range beyond 150px, which would have influenced the overall mean. On the other hand, experts low mean value can be explained potentially by a lesser total number of saccades in the higher ranges of saccade amplitude, therefore not influencing dramatically the mean to spike away from the concentrated range.

Saccade amplitude comparison was also performed against individuals in their assigned group. This was done in order to determine whether any significant differences existed between members of groups, and if so, hypothesize the causality behind the difference and analyze whether it could have impacted the analysis done in this section.

5.3 Physicists Saccade Amplitude Comparison

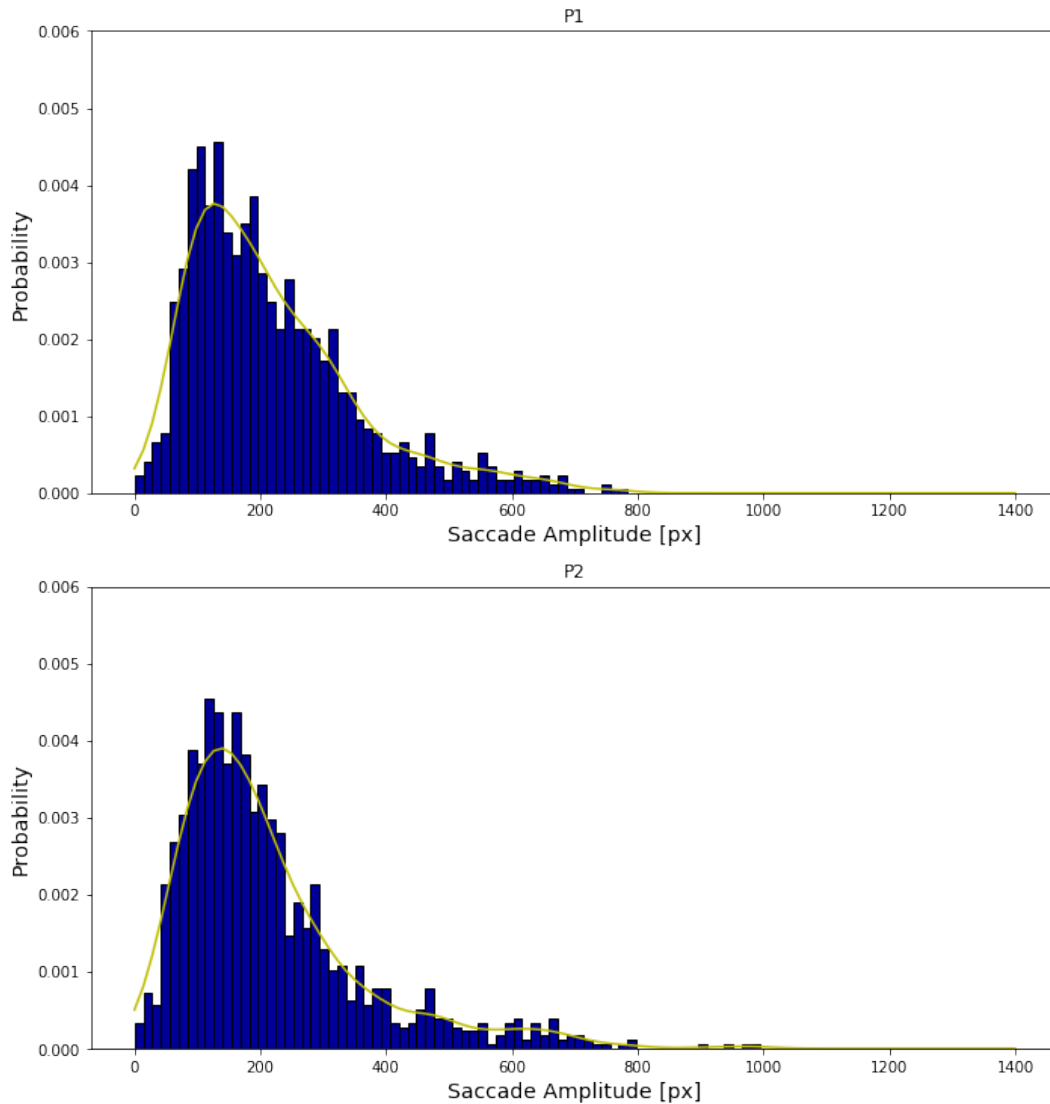


FIGURE 5.5: Distribution of the recorded saccade amplitudes of physicists P1 and P2 individually.

The histograms and lines of best fit suggest homogeneity and consistency between the physicists. Both physicists display a wider spread, while like all other participant groups, their predominant saccade amplitudes are concentrated around 100-150px. Both P1 and P2, and more often P1 had made regular large saccades i.e., in the region of 400px to 1000px, this is particularly noticeable when contrasted with experts. These large movements can be potentially accounted for by the lack of expertise and experience of the physicists, therefore due to this uncertainty they traverse larger parts of the stimulus.

The resulting Euclidean distance and Intersection distance values from the histogram similarity analysis in Table 5.4 below support the verbal analysis above. A low Euclidean distance value, and a high Intersection distance value demonstrates

TABLE 5.4: Physicists histogram comparison using similarity measurements of Euclidean distance and Intersection distance.

Physicist	Physicist	Euclidean Distance	Intersection Distance
P1	P2	0.00344	0.87551

that the physicists exhibited comparable gaze behaviour and no significant differences exist.

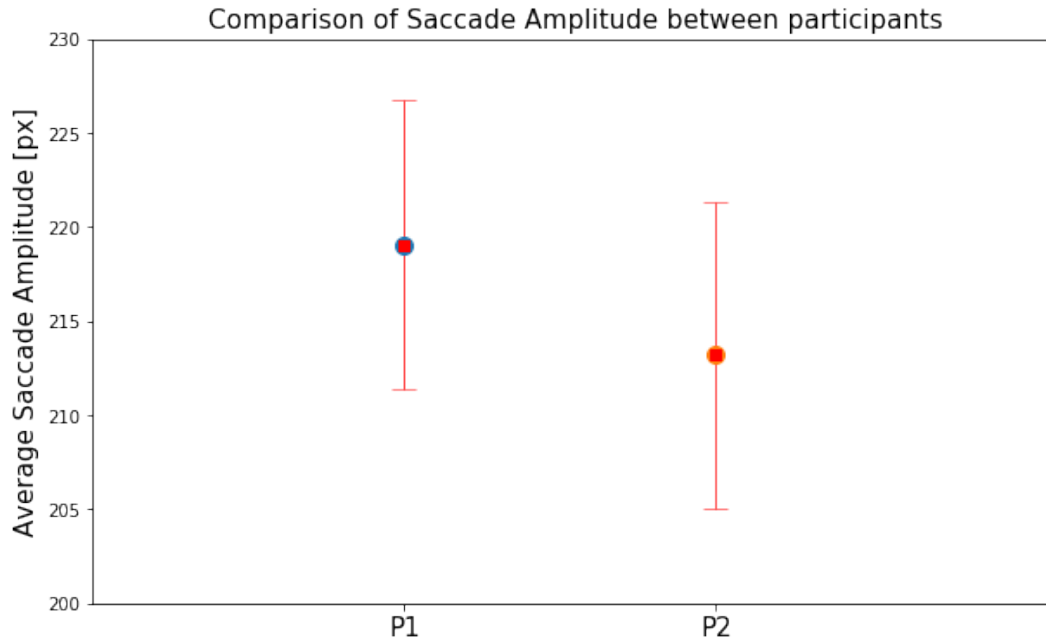


FIGURE 5.6: Graph of the mean saccade amplitude for each physicist, averaged over all recorded saccades for all test stimuli. Error bars indicate a 95% confidence interval.

TABLE 5.5: Mean values for the saccade amplitudes performed by each Physicist.

Physicist	Mean Values	Standard Deviation
P1	219.04	135.97
P2	213.20	148.00

The mean values of the saccade amplitudes performed by the physicists when reviewing all 196 images were calculated to help interpret the data and histogram. The means at initial glance suggest that the physicists have similar gaze behaviour as suggested above.

TABLE 5.6: Results of the independent t-test to determine statistical significance between physicists.

Physicist	Physicist	P-value
P1	P2	0.5953

Table 5.6 presents the results of an independent a t-test for determining statistical significance, the results gave a p-value > 0.05 , therefore no statistical significance exists between the physicists gaze behaviour. This may be because medical physicists use physics to deliver clinical services in diagnosis and manage the technology used in radiology. Therefore, while they also interact with mammograms as a primary interest in their work, they do it for different purposes than radiologists and may follow a protocol, or have had similar preliminary training for observing medical images rather than relying on experience or intuition.

5.4 Expert Radiologists Saccade Amplitude Comparison

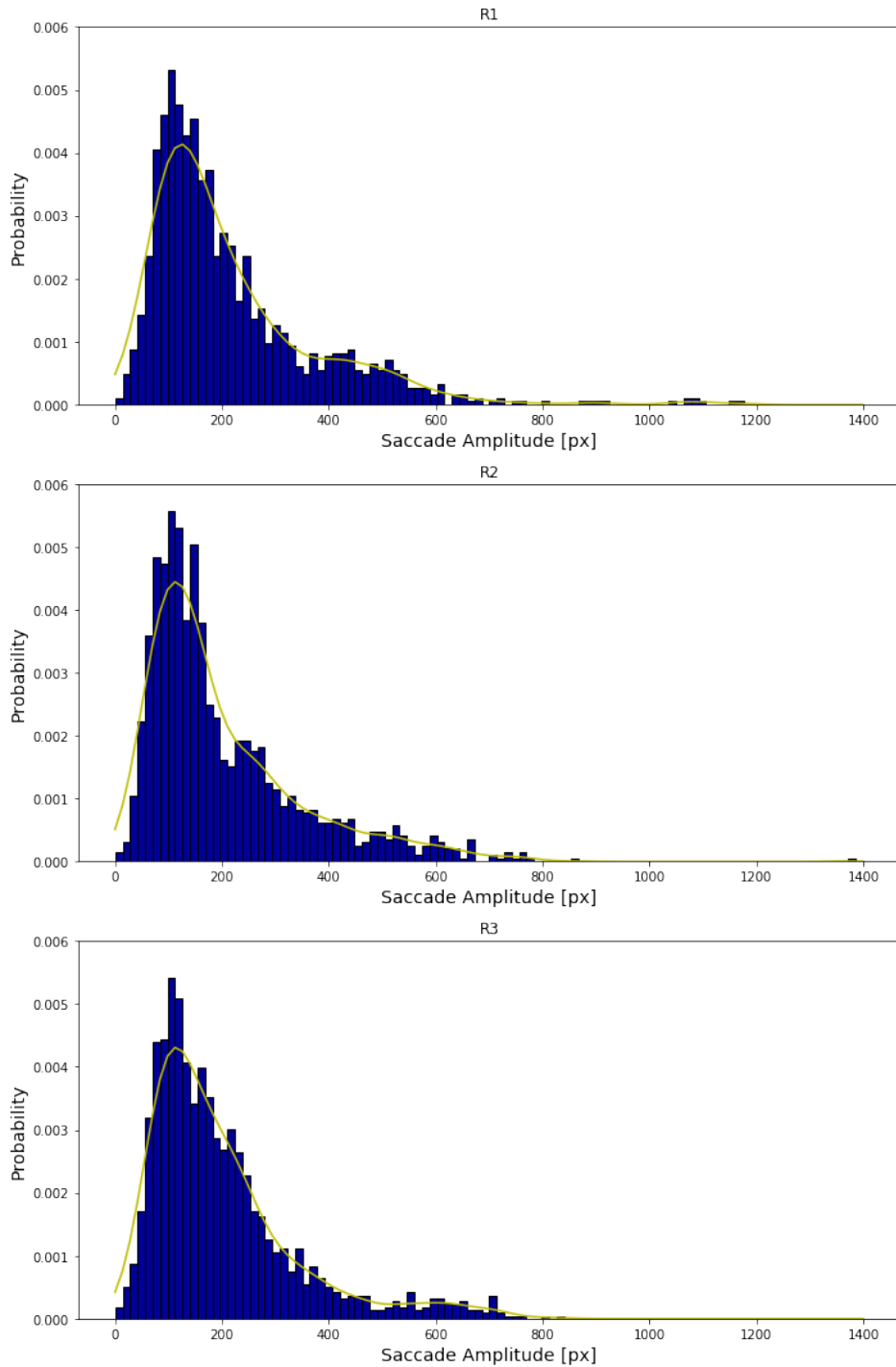


FIGURE 5.7: Distribution of the recorded saccade amplitudes of expert radiologists R1, R2 and R3 individually.

The expert radiologists display a higher frequency in the concentrated area 100-150px than physicists as mentioned in the group comparison, with a higher peak than both physicists and trainees. Experts saccade amplitudes predominantly span

the range of 100-150px. We can speculate that this consistency in behaviour between all experts can be explained by the experts having developed a reliable observation strategy through extensive experience, as well as similar lengths and conditions of training in mammography screening.

TABLE 5.7: Expert radiologists histogram comparison using similarity measurements of Euclidean distance and Intersection distance.

Expert	Expert	Euclidean Distance	Intersection Distance
R1	R2	0.00301	0.89925
R2	R3	0.00320	0.89207
R3	R1	0.00257	0.89991

From evaluating the values present in Table 5.7 there does not appear to exist any apparent disparities between radiologists, the Euclidean distance and Intersection distance metrics both returned values which signify similarity of gaze behaviour between the experts. This therefore supports the above analysis of an existing consistency present in expert behaviour.

TABLE 5.8: Mean values for the saccade amplitudes performed by each expert radiologist.

Expert	Mean Values	Standard Deviation
R1	216.30	160.39
R2	202.69	148.95
R3	199.86	139.19

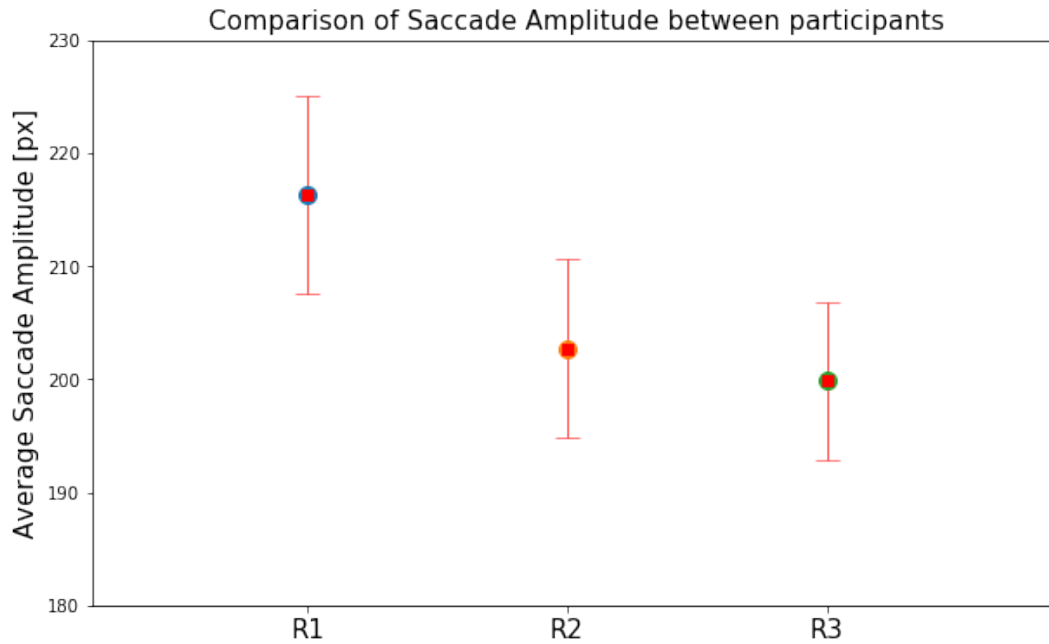


FIGURE 5.8: Graph of the mean saccade for each expert radiologist, averaged over all recorded saccades for all test stimuli. Error bars indicate a 95% confidence interval.

TABLE 5.9: Results of the ANOVA test to evaluate the effect of expert radiologist on the length of saccade amplitude.

Factor	df	F	P-value
Experts	2	4.75	0.0087

The means of saccade amplitude values for each expert radiologist was calculated. From observing Figure 5.8 and Table 5.9 there is a suggestable difference in radiologist R1 from the remainder of the radiologists. An ANOVA test was conducted to validate whether any statistically significant differences were found.

A statistically significant p-value (i.e., $p < 0.05$) was found, therefore this indicates that there was at least one participant in the group which influenced the overall mean in a significant manner. However, since the ANOVA report does not explicitly state which member caused the statistical significance outcome, independent t-tests were performed to compare individual radiologists.

TABLE 5.10: Results of the independent t-test to determine statistical significance between expert radiologists.

Expert	Expert	P-value
R1	R2	0.0228
R2	R3	0.5952
R3	R1	0.0034

Table 5.10 above indicates that R1 displayed statistically significant different behavior when compared with R2 and R3 (i.e., $p\text{-value} < 0.05$). Whereas the difference between R2 and R3 was negligible, therefore their performance was similar and may be explained by them having a more similar background and catalogue of experience and training.

5.5 Trainee Radiologists Saccade Amplitude Comparison

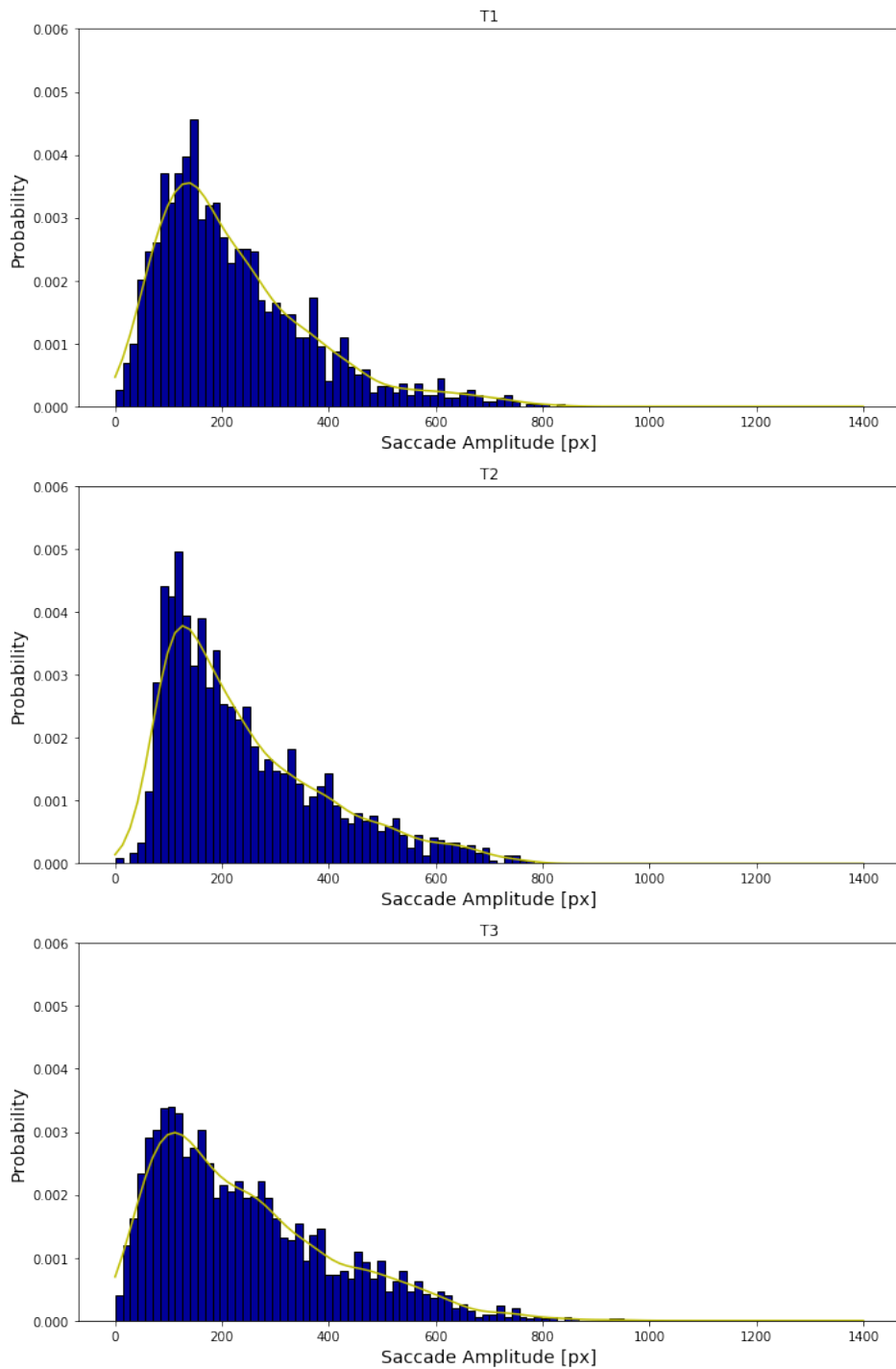


FIGURE 5.9: Distribution of the recorded saccade amplitudes of trainee radiologists T1, T2 and T3 individually.

The histograms reveal that trainees display a larger frequency of saccades outside of the concentrated 100-150px range, and particularly outside the experts range. Additionally, the line of best fit in each trainee histogram displays more of a linear

decrease from the peak frequency saccade amplitude, in contrast the histograms of experts displays a more negative exponential decrease from the peak frequency saccade amplitude. This is particularly applicable to T3, who also from observation, displayed a considerably lesser number of saccades in the concentrated 100-150px range than T1 and T2.

TABLE 5.11: Trainee radiologists histogram comparison using similarity measurements of Euclidean distance and Intersection distance.

Trainee	Trainee	Euclidean Distance	Intersection Distance
T1	T2	0.00393	0.86451
T2	T3	0.00483	0.83335
T3	T1	0.00360	0.86350

Inspecting Table 5.11 displays no significant disparities in the histogram similarities between any of the trainees and validates highly consistent and similar gaze pattern behaviour. For each comparison there resulted a low Euclidean distance result value indicating similarity, and likewise a high Intersection distance value.

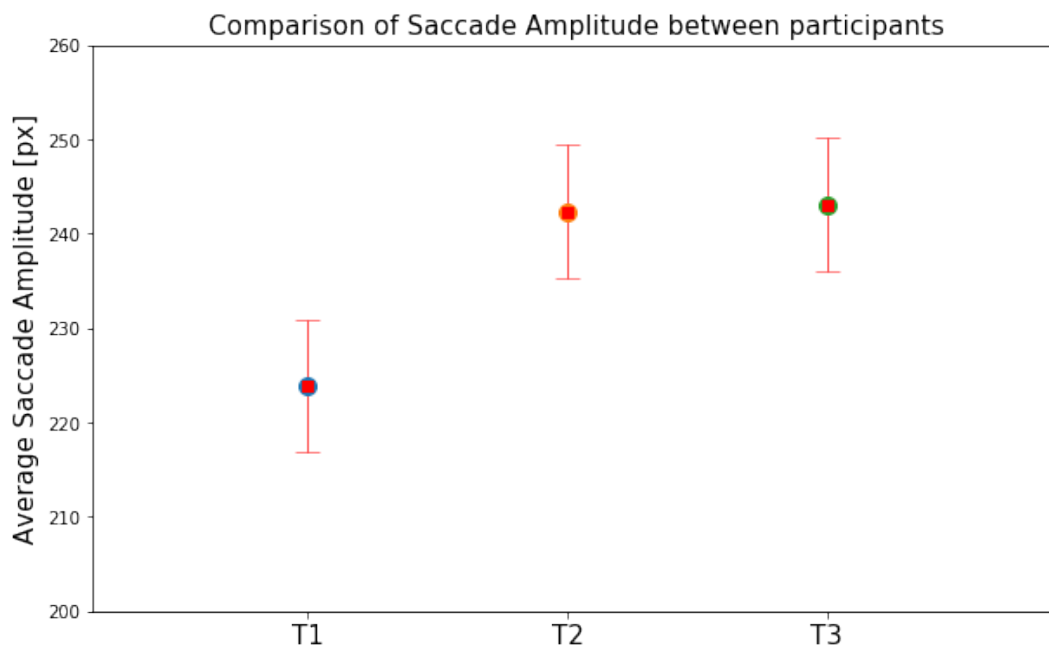


FIGURE 5.10: Graph of the mean saccade for each trainee radiologist, averaged over all recorded saccades for all test stimuli. Error bars indicate a 95% confidence interval.

TABLE 5.12: Mean values for the saccade amplitudes performed by each trainee radiologist.

Trainee	Mean Values	Standard Deviation
T1	216.90	145.65
T2	242.34	149.45
T3	243.08	167.54

TABLE 5.13: Results of the ANOVA test to evaluate the effect of trainee radiologists on the length of saccade amplitude.

Factor	df	F	P-value
Trainee	2	8.15	0.00029

Similar to the event with experts, from observing Graph 5.10 and Table 5.12 there is a suggestible noteworthy difference in the mean saccade amplitude values of trainee T1 from T2 and T3.

To discern whether any of the trainees behaviour was significantly different from the mean an ANOVA test was conducted, shown in Table 5.13. A statistically significant p-value was computed (i.e., $p < 0.05$). This was not enough to arrive at a conclusion that T1 was the cause of the significance, therefore, independent t-tests were performed to validate this.

TABLE 5.14: Results of the independent t-test to determine statistical significance between trainee radiologists.

Trainee	Trainee	P-value
T1	T2	0.00038
T2	T3	0.88632
T3	T1	0.00027

The resulting p-value confirmed that T2 and T3 were very similar i.e., p-value = 0.88632 ($p > 0.05$), this indicates not only that there exists no statistically significant difference between them, but also its nearness to 1 suggests that their behavior was substantially similar. Potentially explained by having similar levels of training and experience. In the contrary, T1 displayed statistically significant different behavior from both T2 and T3 (i.e., p-values < 0.05). T1 had a much lower mean of saccade amplitudes, liking to that of an expert radiologist therefore the difference could be speculated to be a result of T1 having more training than the other trainees and that with more training T2 and T3 would expect their means to decrease.

5.6 Saccade Orientation Group Comparison

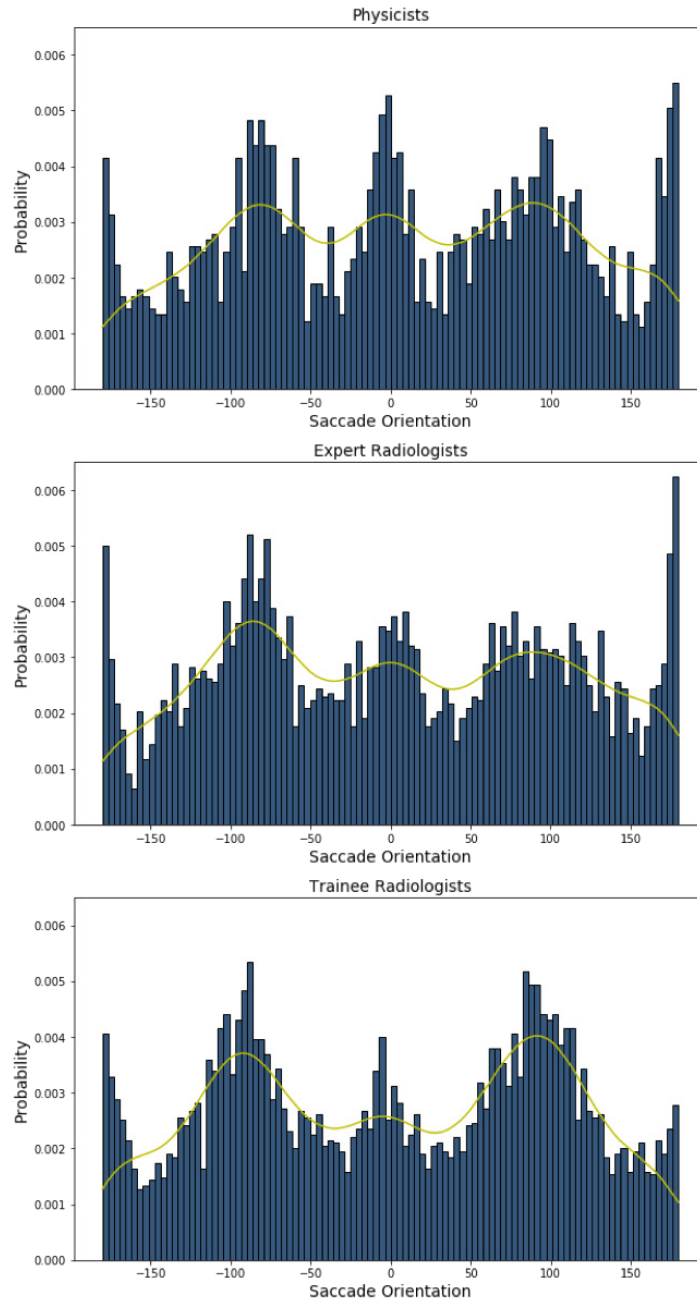


FIGURE 5.11: Saccade orientation histogram distribution plots for each group.

From an initial observation of Figure 5.11 there exists a shared pattern in the orientation of each groups saccades, which can be more evidently seen by analyzing the lines of best fit. A wave-like shape is formed for each group, with peaks at around the 90, -90, 180, -180 and 0-degree angles. This is consistent with current research on the human visual systems habitual tendencies for eye movements. Humans display natural preferences for performing short saccades in a horizontal

TABLE 5.15: Group saccade orientation histogram comparison using similarity measurements of Euclidean distance and Intersection distance.

Group A	Group B	Euclidean Distance	Intersection Distance
Physicists	Experts	0.00801	0.88642
Experts	Trainees	0.00867	0.88452
Trainees	Physicists	0.01000	0.86149

direction compared to vertical saccades [42] [43]. However, while vertical saccades occur less than horizontal ones, they are still highly frequent, whereas diagonal saccades occur at the least frequency. This can be observed in the saccade orientation distribution of all groups, each group display a clear preference for saccades with orientations nearing the strict vertical or horizontal axis, although it is worth noting that some groups display a more significant frequency in particular mentioned directions than others i.e., experts saccade orientations at 180 degrees are much more frequent than in the trainees cases.

The symmetries of the graphs also point to insight. It is interesting to observe that in the scenario of the experts their distribution of saccades is flatter than that of other members, particularly when compared with trainees, which may be explained by a lesser amount of fixation points in the dataset of experts. Furthermore, experts form a much more asymmetrical distribution than trainees and physicists who form a comparatively symmetrical distribution. This may be partially explained by experts using previous fixation points as a lead for deciding on the next area to observe, whereas lesser experienced individuals may make sharper uninformed schematic movements as well as more frequent re-visitations.

To substantiate the above evaluation, Euclidean distance and Intersection distance analysis was conducted for the saccade orientation histograms as seen in Table 5.15

The results support the above discussion: all groups displayed largely similar behaviour i.e., very small Euclidean distance values and Intersection distance values close to 1. This is likely due to the common behaviours humans exhibit in making saccades. There does not exist a group couple which displays striking and decisively different behaviour.

5.7 Joint Saccade Amplitude and Orientation Group Comparison

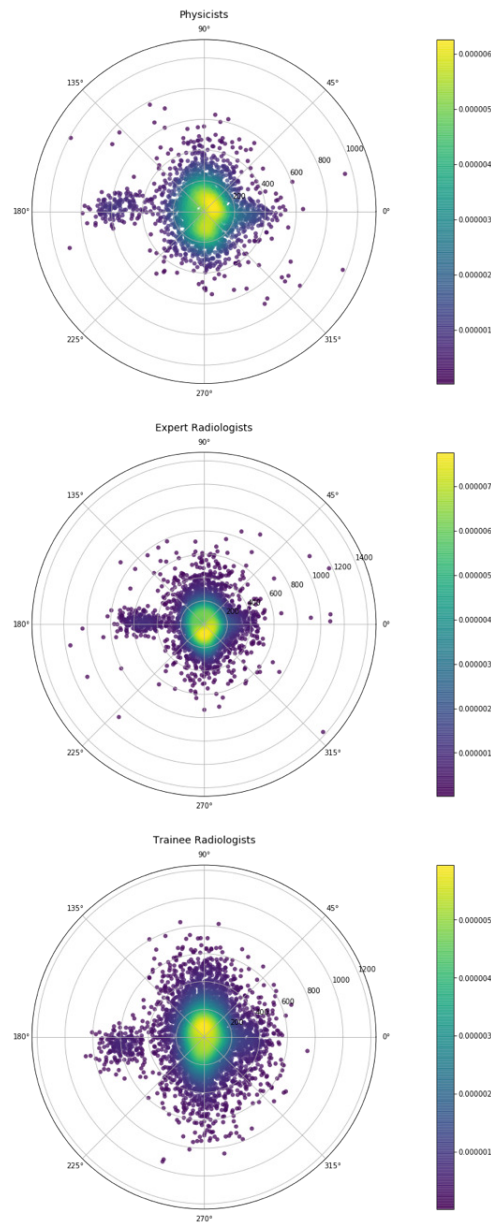


FIGURE 5.12: Polar plots which display saccades, by mapping their orientation angle with the according saccade amplitude, plotted as the distance from the centre.

A key factor worth noting when examining differences in the polar plot is that there is substantial difference in the amount of fixation points in the plots from group to group. This is due to there being a varying number of participants in the groups i.e., two participants in the Physicist group, and three in both Expert and Trainee radiologists. Furthermore, each participant completed a different number of fixations cumulatively over the entire study of the 196 images.

The groups display differences in the concentration of saccades. For instance, the trainees exhibit a more dispersed graph of mapped values, especially when contrasted with experts, whose points are noticeably more concentrated around the centre of the polar plot with a higher density, meaning experts performed shorter saccades more often, potentially because experts perform sharper and more consistent gaze behaviour. The physicists display the most spread, with their distribution forming a more gradual decrease in density from the centre, indicating that their saccades occurred in frequency of amplitude and orientation in a more evenly distributed manner than the other participating groups.

The plot displays a notable pattern of conspicuous long saccades at a horizontal 180-degree angle across all groups, this is potentially explained by human preference of making horizontal gaze movements mentioned above in Section 5.6. However, it is also interesting to note that there exists a gap of medium length saccades in this degree angle, particularly for Physicists and somewhat Expert Radiologists. Another similarity is there is a concentration of saccades in the centre, which again can be potentially accounted for by the research stating humans preference for shorter saccades.

6 Future Work

While the sample size of expert radiologists used in this study is seen as adequate because of the high degree of consistency amongst expert readers [44], expanding and improving on the sources, quantity and quality of data would be beneficial. Additionally, this report focused on investigating mammograms, however, other artefacts from different medical imaging modalities, or practitioners from other medical disciplines could be added, such as nurses or general practitioners. Expanding the fields of medical specialties encompassed in the study could reveal further insight into the problem. Most importantly, variables such as gender [45], task at hand [46] or cultural heritage [47] have been suggested to impose dissimilarities in human visual perception. Observing additional controlled variables would have been highly constructive to the study, particularly to validate and reinforce conclusions which were hypothesized, or an examination of the precise influence of these variables could be performed.

The central theme of this report focused on statistical and visual analysis from calculating saccade amplitude and orientation. However, more information still lies to be learned through employing more sophisticated methods, in particular, the utility of these values could be extended in future work through developing saccadic models. Sophisticated saccadic models have been developed which are capable of mitigating human biases, incorporating intricacies and calculating scan paths with high performance. Le Meur et al. [48] were able to discern differences in gaze behaviour between participants from independent age groups i.e. adults, 8-10, 6-8, 4-6 and 2-year old's, and then were able to leverage the observations to build an age-dependent saccadic model for generating scan paths which accurately emulate age-specific visual scan paths. The consequences of such sophisticated models would be significantly beneficial, it would assist in the development of intelligent systems that could aid radiologists, or independently perform diagnosis on medical images in an expert-like accuracy. Progress in this form would not only support health care facilities in general, but provide crucial support in areas where there is a shortage of expertise or medical practitioners such as developing and remote areas.

Finally, this report substantiates the utility of employing saccade information in understanding gaze behaviour and individuals characteristics. Therefore, since this report can attest to the value of observing saccades, it can then be employed to as a recommendation for applying gaze behaviour analysis in further research, not only in different fields of medicine but expand into entirely different disciplines.

7 Conclusion

Overall, the results attained from this project were positive and in-line with the initial objectives and the desired deliverables were achieved. This dissertation intended to investigate the gaze behaviour, with a focus on saccade eye movements, of medical practitioners when viewing mammography images with experience as the independent variable. The utility of studying the impact of experience on gaze behaviour has been validated in the past in previous experiments, however, the research conducted in this report reinforces its utility and additionally demonstrates the differences in saccades agents exhibit when conducting diagnosis on the stimuli.

To improve the practice of diagnosis in clinical settings it is imperative to understand how medical experts function and how their gaze tendencies differ from novices or inexperienced observers. The analysis in this report documents the saccades experts perform when conducting diagnosis on the stimuli over a large sample size and contrasts them to trainee radiologists and physicists through applying visual analysis techniques and the findings were reinforced by quantitative results gathered from statistical evaluation. Therefore, the information discovered in the analysis presented may assist in the establishment of improved practices in diagnostics and may be utilized to develop improved training resources as well as diagnostics system which can aid the discussed shortage of radiologists in areas around the globe.

8 Reflections on Learning

Throughout the period of completing this project I have acquired an innumerable amount of skills which I will continue to hone and develop throughout my career. A critical approach towards data and statistics has been enveloped in me, statistics can at times seem intuitive yet without proper appraisal we can make incorrect assumptions and enact mislead decisions. Furthermore, a fond attitude towards data analysis as a choice of career, and a sincere interest in data science has been established in me that I want to explore further. From my experience at university studying Computer Science I have been involved in building complex systems and applications, data analysis contrasts these activities in that it is more of an exploratory field, seeking to develop understanding and uncover mysteries lying within fragmented artefacts. I have experienced growth in my professional practices including an understanding in the correct application of experimental techniques, report writing and communicating with stakeholders. Finally, the development of a more critical and analytical perspective will translate into other facets of life beyond career and support me in making wiser decisions.

References

- [1] E Kowler. *Eye movements: The past 25 years*. Vision Research, 2011, 51:1457–1483.
- [2] J. Najemnik and W. S. Geisler. *Optimal eye movement strategies in visual search*. Nature, 2005.
- [3] M. Rucci. *Temporal encoding of spatial information during active visual fixation*. Current Biology, 2012, pp. 510–514.
- [4] Oxford Journals. *The impact of mammography and adjuvant therapy on U.S. breast cancer mortality(1975-2000): Collective results from the Cancer Intervention and Surveillance Modeling Network*. :/ /jncimono.oxfordjournals.org/content/2006/36.toc. 2006.
- [5] S. Weinstein. *Holistic component of image perception in mammogram interpretation: Gaze-tracking study*. Vol. 242. Radiology, 2007, pp. 396–402.
- [6] *Eyetracking for skills assessment and training: A systematic review*. Vol. 191. Elsevier Journal of Surgical Research, 2014, pp. 169–178.
- [7] J. Pola and H. J. Wyatt. *Smooth pursuit: response characteristics, stimuli and mechanisms*. CRC Press, 1991.
- [8] C. Distler and K.-P. Hoffmann. *The Oxford Handbook of Eye Movements: The optokinetic reflex*. Oxford University Press, 2011, pp. 65–83.
- [9] J.Crawford and E.Klier. *The Oxford Handbook of Eye Movements: Neural control of three-dimensional gaze shifts*. Oxford University Press, 2011, pp. 339–356.
- [10] B.T. Vincent B. Tatler. *The prominence of behavioural biases in eye guidance*. 2009, pp. 1029–1059.
- [11] O. Le Meur and Z. Liu. “Saccadic model of eye movements for free- viewing condition”. In: *Vision Research* 116 (2015), pp. 152–164.
- [12] O. Le Meur and A. Coutrot. *How saccadic models help predict where we look during a visual task? Application to visual quality assessment*. Vol. 13. Electron. Imag., 2016, pp. 1–7.
- [13] A. Brady. “Error and discrepancy in radiology: Inevitable or avoidable?” In: *Insights Imaging* 8 (2017), pp. 171–182.
- [14] P. Räämä. “The maturation of eye movement behavior: Scene viewing characteristics in children and adults,” in: *Visual Research* 103.8 (2014), pp. 83–91.

- [15] BBC. *Patients facing NHS tests 'bottleneck'*. <https://www.bbc.co.uk/news/health-46103014>. Accessed: 01/04/2019. 2018.
- [16] G. Tourassi. "Investigating the association of eye gaze pattern and diagnostic error in mammography". In: *Proc. SPIE* 8673 (2013).
- [17] A. Graham E. Krupinski and R. Weinstein. "Characterizing the development of visual search expertise in pathology residents viewing whole slide images". In: *Human Pathology* 44 (2013), pp. 357–364.
- [18] C. F. Nodine D. P. Carmody and H. L. Kundel. "Finding lung nodules with and without comparative visual scanning". In: *Perception Psychophys.*, 29 (1981), pp. 594–598.
- [19] C. Wilcox. "A study of radiologists viewing multiple computed tomography examinations using an eyetracking device". In: *J. Digit. Imag.*, 3 (1990), pp. 230–237.
- [20] T. Yosue. "Analyzing the eye movement of dentists during their reading of CT images". In: *Odontology* 89 (2001), pp. 54–61.
- [21] "Using gaze-tracking data and mixture distribution analysis to support a holistic model for the detection of cancers on mammograms". In: *Academic Radiology* 15 (2008), pp. 881–886.
- [22] M. B. Wallace. "Association between visual gaze patterns and adenoma detection rate during colonoscopy: A preliminary investigation". In: *Amer. J. Gastroenterol* 106 (2011), pp. 1070–1074.
- [23] "Influence of experience and training on dental students' examination performance regarding panoramic images". In: *J. Dental Educ.*, 80 (2016), pp. 156–164.
- [24] W. Melvin. "Eye motion parameters correlate with level of experience in video-assisted surgery: Objective testing of three tasks". In: *J. Laparoendosc. Adv. Surgical Techn.*, 6 (2005).
- [25] M. Ishii. "Surgical task and skill classification from eye tracking and tool motion in minimally invasive surgery". In: *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 13 (2010), pp. 295–302.
- [26] L. Kavoussi. "Eye metrics as an objective assessment of surgical skill," in: *Ann. Surgery*, 252 (2010), pp. 277–182.
- [27] A. T. Meneghetti. "Analysis of eye gaze: Do novice surgeons look at the same location as expert surgeons during a laparoscopic operation?" In: *Surgical Endoscopy* 26 (2012), pp. 3536–3540.
- [28] S. Weinstein. "Time course of perception and decision making during mammographic interpretation". In: *Amer. J. Roentgenol.*, 179 (2002), pp. 917–923.

- [29] E. Krupinski. "Investigating the link between radiologists' gaze, diagnostic decision, and image content". In: *J. Amer. Med. Inform. Assoc.*, 20 (2013), pp. 1067–1075.
- [30] J. Saada. "Radiology image perception and observer performance: How does expertise and clinical information alter interpretation? Stroke detection explored through eye-trackingL." In: *Proc. SPIE* 7263 (2009), 72630K.
- [31] H. Matsumoto et al. "Where do neurologists look when viewing brain CT images? An eye-tracking study involving stroke cases." In: *PLoS ONE* 6 (2011), e28928.
- [32] S. Mallett et al. "Tracking eye gaze during interpretation of endoluminal three-dimensional CT colonography: Visual perception of experienced and inexperienced readers," in: *Radiology* 273 (2014).
- [33] T. Crawford. "How do radiologists do it? The influence of experience and training on searching for chest nodules". In: *Radiography* 12 (2006), pp. 134–142.
- [34] "Recurrence quantification analysis reveals eye-movement behavior differences between experts and novices". In: *Proc. Symp. Eye Tracking Res. Appl.*, (2014), pp. 303–306.
- [35] "Visual search behaviour in skeletal radiographs: A cross-speciality study". In: *Clin. Radiol.*, 62 (2007), pp. 1069–1077.
- [36] J. McGrath. "Gaze training enhances laparoscopic technical skill acquisition and multi-tasking performance: A randomized, controlled study". In: *Surgical Endosc.*, 25 (2011), pp. 3731–3739.
- [37] M. R. Wilson. "Cheating experience: Guiding novices to adopt the gaze strategies of experts expedites the learning of technical laparoscopic skills". In: *Surgery* 152 (2012), pp. 32–40.
- [38] A. Graham E. Krupinski and R. Weinstein. "Characterizing the development of visual search expertise in pathology residents viewing whole slide images". In: *Hum. Pathol.*, 44 (2013), pp. 357–364.
- [39] Lucie Lévesque. *Analysing and quantifying visual experience in medical imaging*. Cardiff University, 2018.
- [40] HANTAO LIU. "State of the Art: Eye-Tracking Studies in Medical Imaging". In: (2018).
- [41] Roland Jeannier. *Optimum approach for iterating over a DataFrame*. <https://medium.com/@rtjeannier/pandas-101-cont-9d061cb73bfc>. Accessed: 18/02/2019. 2017.
- [42] G. Underwood T. Foulsham A. Kingstone. "Turning the world around: Patterns in saccade direction vary with picture orientation". In: *Vision Research* 48 (2008), pp. 1777–1790.

- [43] W. Geisler J. Najemnik. "Eye movement statistics in humans are consistent with an optimal strategy". In: *Journal of Vision* 8 (2008), pp. 1–14.
- [44] E. Krupinski. "Diagnostic accuracy and visual search efficiency: Single 8 MP vs. dual 5 MP displays." In: *Journal of Digital Imaging* 30 (2017), pp. 144–147.
- [45] H. Maeda. "Gender differences and reproducibility in exploratory eye movements [44] of normal subjects." In: *Psychiatry Clin. Neurosci.* 54 (2000), pp. 31–36.
- [46] A. Yarbus. "Eye Movements and Vision". In: *Plenum* (1967).
- [47] R. Nisbett. "The Geography of Thought: How Asians and Westerners Think Differently...And Why". In: *Free Press* (2003).
- [48] Andrea Helo. "Visual Attention Saccadic Models Learn to Emulate Gaze Patterns From Childhood to Adulthood." In: *IEEE TRANSACTIONS ON IMAGE PROCESSING* 25 (2017).
- [49] Andrew D Short. *Cognitive and psychological science insights to improve climate change data visualization*. Great Britain: John Wiley; ISBN 0-471-96570-7, 1999.
- [50] John Taylor. *Introduction to error analysis, the study of uncertainties in physical measurements*. 1997.
- [51] D. Turgeon and E. Lam. "Influence of experience and training on dental students' examination performance regarding panoramic images," in: *J. Dental Educ.*, 80 (2016), pp. 156–164.