EXPLOITING AN UNDERSTANDING OF HUMAN VISUAL PERCEPTION TO FACILITATE HUMAN-MACHINE ELECTROCARDIOGRAM INTERPRETATION OF DRUG-INDUCED LONG QT SYNDROME

A THESIS SUBMITTED TO THE UNIVERSITY OF MANCHESTER
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
IN THE FACULTY OF SCIENCE AND ENGINEERING

2021

By
Alaa Alahmadi
Department of Computer Science

Contents

Word Count: 52,651

List of Tables

List of Figures

Abstract

Electrocardiograms (ECGs), which capture the electrical activity of the human heart, are widely used in clinical practice for detecting cardiac pathologies, and notoriously difficult to interpret. Many commonly prescribed medications, including antihistamines, antibiotics and antidepressants, can produce a complication known as 'drug-induced long QT syndrome (diLQTS)', characterised by a prolongation of the QT-interval on the ECG. It is associated with a life-threatening arrhythmia known as Torsade de Points (TdP)—the leading cause of sudden cardiac death in young, otherwise healthy people. Self-monitoring for diLQTS could therefore save many lives, but detecting it on the ECG is difficult, particularly at high and low heart rates, even for clinicians who routinely read ECGs. Whilst there have been attempts to automate ECG interpretation for several decades, the accuracy of these methods remains limited. In particular, automated QT measurement algorithms have proved unsatisfactory for detecting LQTS. A major challenge for automated QT algorithms is identifying the precise end of the T-wave (the terminal point), especially when the T-wave's morphology is abnormal. This thesis is the first work to examine the ability of laypeople to interpret an ECG for drug-induced QT-prolongation monitoring, devise a novel ECG visualisation technique to enable them to interpret it accurately, and exploit an understanding of the human visual perceptual process to improve automated QT-prolongation detection. The approach draws from the field of pre-attentive processing theory in human vision, showing through several experiments that using pseudo-colour to expose QT-interval duration on the ECG significantly improves laypeople's accuracy in detecting diLQTS at risk of TdP regardless of heart rate and T-wave morphology. An understanding of how humans use pseudo-colour to interpret ECG data is combined with clinical knowledge that considers the morphology of the T-wave to develop a novel, rule-based algorithm that reliably automates the detection of diLQTS, thus facilitating an explainable, shared human-machine ECG interpretation.

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright

Acknowledgements

I wish to express my deepest gratitude to my supervisors, Dr Caroline Jay and Dr Markel Vigo, for their invaluable advice, continuous support, and patience during my PhD. Caroline is an exceptionally great supervisor who guided me convincingly and inspired me immensely to be a scientist and do the right thing even when the road got tough. I went through many challenges during my PhD, but she was always supportive and made me look at things from a different positive perspective—I'll miss her word "step back". Her flexibility and ability to see the big picture have always opened up a limitless, free space for me to think outside the box and be creative. Markel has a keen eye that notices the smallest important details that make a big difference in communicating research contributions to multiple audiences. His treasured support was really influential in shaping my experiments and critiquing my results. Thanks to both of you.

A special thank you to Dr Alan Davies for helping me learn ECG interpretation and for reviewing my papers. I would also like to thank all the members of the Interaction Analysis and Modelling lab and the members of the Digital Experimental Cancer Medicine Team for their continuous feedback throughout the research project.

I wish to acknowledge the tremendous support and love of my husband, Ahmed Almuraee, and my little son, Odai Almuraee. They gave me strength and kept me going on, and I highly appreciate their patience, especially when I was so busy at the beginning of the PhD. In all honesty, it is my husband's spirit of fun and adventure that has made my study and my life in the UK a wonderful, unforgettable time. I would also like to thank my parents, Adel Alahmadi and Norah Alahmadi, for believing it would (one day) come to an end and my uncle, Dr Hasan Alahmadi, who encouraged me to follow this path. I would also like to thank all my friends and colleagues, especially Deemah Alqahtani, Afrah Alkhalaf and Manal Binkhonain, for their support during the PhD trip. I am grateful for the moments that brought us together to chat, laugh, and sometimes cry because of our PhD journey.

Chapter 1

Introduction

1.1 Research motivation

Sudden cardiac death in young, apparently healthy people is a catastrophic event—often caused by a wide range of hidden asymptomatic diseases [Thi18]. Considerable research attention has been devoted to a specific cardiac disorder known as 'long QT syndrome' (LQTS)—the leading cause of sudden cardiac death in young adults [GEB18, SWA$^+$19, Sch20]. LQTS is associated with a life-threatening arrhythmia known as *Torsades de Pointes* (TdP), a form of polymorphic ventricular tachycardia that is often precipitated by specific triggers including auditory stimuli (*e.g.* sudden noise or alarm), a slow heart rate during sleep, stress-related emotions and strenuous physical activity (notably swimming) [PK01, SWA$^+$19, BSW14, VTVAV15]. Athletes with LQTS are thus particularly at risk of TdP arrhythmia attacks and sudden cardiac death [JA13, BSW14, GPR16, BP09].

LQTS can be congenital—a result of genetic mutations in cardiac ion channels (channelopathies)—or acquired, resulting from the clinical administration of pharmacological drugs [Sch20, YC03, CMY08, DAF$^+$10, Tis16]. Whilst TdP can result from both congenital and acquired LQTS, acquired drug-induced long QT syndrome (diLQTS) is by far the most common cause of TdP [NAS12, YC03, CMY08, ESTB20, WS20]. A steadily increasing number of commonly prescribed medications have been reported to cause diLQTS, TdP and sudden cardiac death, even in people with a structurally normal heart [Kha02, CMY08, ESTB20]. This, in turn, is a significant issue for clinicians, the pharmaceutical industry and regulatory authorities, mainly because many of these medications, including antihistamines, antibiotics, antidepressants and antiarrhythmic drugs [WS20, CMY08, YC03], are widely used, often for self-limited

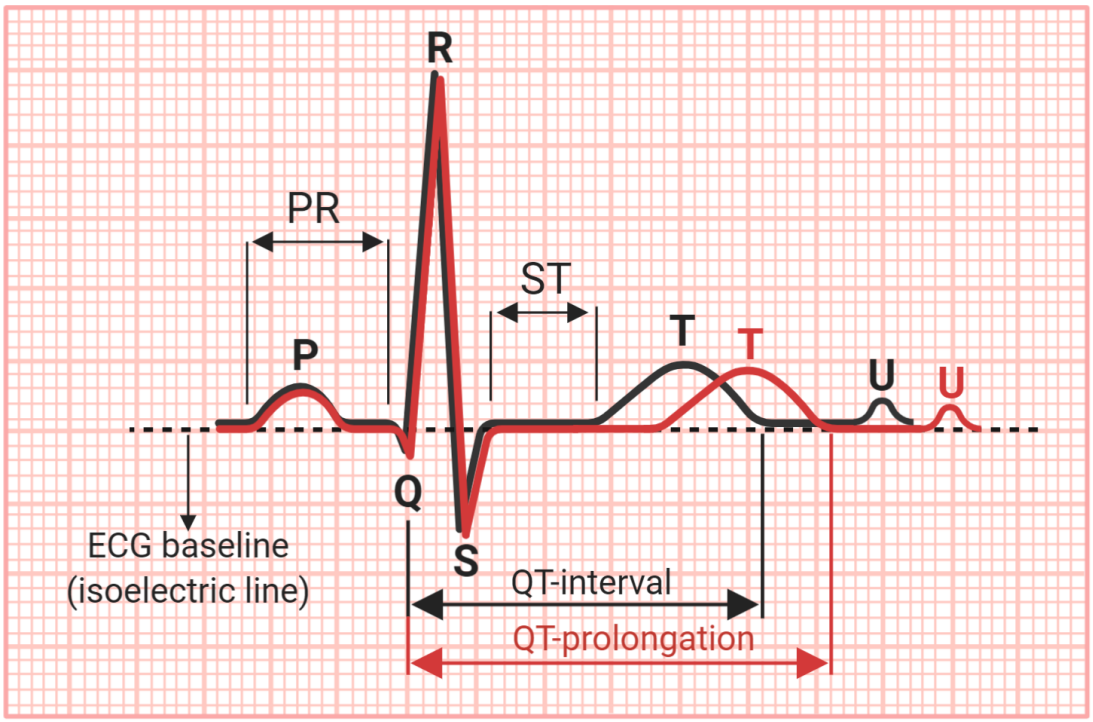


Figure 1.1: An illustration of the electrocardiogram (ECG) waveform, the QT-interval and QT-prolongation. Different 'waves' (peaks and troughs) are labelled with letters and represent different stages of the heartbeat, known as the 'ECG complex'. The QT-interval is measured from the beginning of the Q-wave to the end of the T-wave.

diseases. A major difficulty with identifying LQTS (both congenital and acquired), however, is that it is often asymptomatic; sudden cardiac death can be the first clinical manifestation, and therefore it may go undiagnosed, or underdiagnosed, without an electrocardiogram (ECG) [Kha02, SC08, RV05, CMY08].

The ECG, which represents the electrical activity of the human heart, is a powerful diagnostic tool widely used in clinical practice for assessing cardiac function and detecting pathologies. Its result is displayed as a graphical signal, where the 'waves' (peaks and troughs) are labelled with letters and represent different stages of the heartbeat, known as the 'ECG complex'. LQTS is characterised by a prolongation of the QT-interval on the ECG, representing a delay in the ventricular repolarisation activity of the heart, as shown in Fig 1.1.

Clinical research has shown that even a small ($\approx 10ms$) QT-interval increase from a drug-free ECG baseline is considered a significant side effect of a QT-prolonging drug [RFF$^+$09, Bre10, FDA$^+$05], and the Committee for Proprietary Medicinal Products (CPMP) guidelines state that increases in the QT-interval greater than 60ms from

the drug-free baseline should raise concerns about potential TdP [fPMP+97, CMY08]. Moreover, a single dose of a QT-prolonging drug could dramatically prolong the QT-interval within 24 hours for some patients, with the risk of TdP increasing with continued use [LVL+18, JVM+14]. Recent innovations in healthcare wearable technologies have made it possible to record high-quality, clinically reliable ECG data outside of the clinical environment [FBG+19, CKA18, SGH+17, SPS+19, WRY19, RFP19]. Self-monitoring ECGs for drug-induced QT-interval prolongation could therefore save many lives, especially for people at high risk, including patients on a known QT-prolonging drug and patients participating in a clinical trial testing a new drug [Sch20, Sha02, KW12, DAF+10, NAS12]. However, there are currently a number of challenges that limit the potential of ECG self-monitoring of this critical condition:

1. **To date, there is no evidence that it is possible for lay people to interpret an ECG and detect QT-prolongation.**

2. **ECG interpretation is known to be complex, and LQTS detection is particularly difficult even for clinicians. Small drug-induced increases in the QT-interval from a drug-free ECG baseline can be clinically significant but are difficult to perceive visually without rigorous measurement.**

   Although the ECG test is cheap and easy to repeat—making it widely available in clinical practice even outside of clinical cardiology specialisms, including in general wards, in general practitioner (GP) surgeries, and in other clinical settings [DS14]—its interpretation is known to be complex, and learning how to accurately interpret it can take many years [WBA+14, OCT+09, SAW03]. Detecting LQTS, in particular, is known to be difficult, even for clinicians who routinely read ECGs [VRS+05, STN08]. Research has shown that delayed diagnosis of LQTS is common, and LQTS patients are often misdiagnosed with other conditions (particularly epilepsy or seizure disorder; in these cases ECGs were frequently requested, but interpretation errors occurred) [MMC+09, SHL19, KK09, OMK10, MBA14]. Small increases in the QT-interval can be clinically significant but are difficult to perceive visually without precise measurement [CMY08]. From a perceptual-cognitive perspective, this may be related to the fact that people are poor at perceiving quantity represented along a horizontal scale [LSZ+09, WA84, PK11].

3. **Changes in the T-wave morphology and heart rate (often caused by QT-prolonging drugs) are known to further complicate visual LQTS detection and TdP risk assessment.**

   Research has also shown that changes in the T-wave morphology and artefacts in the ECG signal can cause misinterpretation of the QT-interval length [CMY08, GPAW14, Mor01]. The effect of heart rate on the QT-interval is another challenge, as it is the proportionate rather than absolute length that is important, and it is common to misinterpret the QT-interval at heart rates that differ from the 'standard' 60 bpm [LMJM04, CMY08]. Accordingly, it is common in clinical practice to apply a QT correction formula (QTc) to correct QT-interval to heart rate, and then use a 'cut off' value to identify at-risk QT-prolongation. However, recent research has shown that these correction formulae underestimate or overestimate QT-prolongation (depending on the formula used); and are thus inaccurate in identifying patients at risk of drug-induced TdP for fast and slow heart rates [DLD$^+$03, LMJM04]. This is particularly problematic, as QT-prolonging drugs often alter the heart rate and affect the morphology of the T-wave to a greater extent, causing distorted and bizarre T-waves [VJM$^+$15].

4. **There are currently no reliable algorithms able to automate LQTS detection. Current automated approaches to ECG interpretation rely on recognizing different ECG waves, identifying the beginning of the Q-wave and the end of the T-wave to measure the QT-interval—a process that quickly becomes challenging in the presence of anomalies, artefacts or non-standard ECG waves.**

   Automated ECG interpretation methods were introduced in the 1950s to assist clinicians who had less training in ECG interpretation [Rau07]. These methods can identify a normal sinus rhythm with reasonable accuracy but are much poorer at reliably detecting abnormalities [SW17, EI13, RSG09]. Therefore, whilst most modern ECG machines in hospitals provide an automated measurement of the QT-interval, these automated QT values are usually correct only for a noise-free, normal sinus rhythm, in which the ECG waveforms (particularly the T-wave morphology) are well defined [SW17]. However, ECG wave characteristics (*i.e.* the height and width of different waves) are known to differ substantially across individuals, and are affected by factors including age, race, sex and health status [GMZ06, MMDY94, HST$^+$16]. At present, there

are no standard definitions for the ECG waves [Wil80, Par85, SW17]. As a result, differences in current signal processing measurements persist [SW17]. In addition to the challenge of correctly recognising the different ECG waves, identifying the precise end of the T-wave (the terminal point) is particularly difficult to make automatically, especially when the T-wave's morphology is abnormal [GPAW14, HC94, GMZ06, Mor01].

As long QT syndrome detection currently depends on accurate QT-interval measurement, automated QT measurement algorithms have proved unsatisfactory for detecting LQTS [KBD$^+$18, TAS$^+$15, EI13, GL13, TABW11, RSG09, CMY08, MCbA01]. Garg and Lehmann [GL13] found that even a widely used computerised ECG interpretation system was not able to detect QT-interval prolongation in 52.5% of patients affected. In a recent study, seven ECG interpretation programs commonly used in clinical practice were assessed for their accuracy in measuring different ECG intervals including the QT-interval [DBDM20]. The results showed that substantial differences in ECG interval measurements were found between programs, and the largest differences were for the QT-interval confirming previous research [DBDM20]. A major issue with understanding the limitations of automated QT-interval measurement algorithms is that the majority are proprietary or unavailable, thus formally bench-marking the performance of different algorithms, and hence advancing research in this area, is not possible at present [DBDM20, SW17, VDPL$^+$18, MSSS17, CMY08].

5. **Drug-induced QT-prolongation, in particular, is known to invalidate automated QT-interval measurement algorithms due to the substantial changes it causes to T-wave morphology and heart rate, thus causing underestimation of the risk of TdP.**

   Research has shown that drug-induced long QT syndrome, in particular, can be underestimated and under-reported by computerised methods in patients on Methadone, a drug that is infamous for prolonging the QT-interval and increasing the risk of TdP [TAS$^+$15]. At present, clinicians do not regard automated QT measurement algorithms as sufficiently reliable for use in the clinical TdP risk assessment of a QT-prolonging drug [SW17, TAS$^+$15, RSG09, CMY08, TMH$^+$13]. Accordingly, clinical recommendations state that QT-interval should be measured and corrected to heart rate manually by a clinician specialising in ECG interpretation [ITD$^+$20, IP13, RSG09, CMY08, FDA$^+$05].

This thesis uses knowledge from human visual perception to re-visualise ECG data, such that QT-interval prolongation can be easily perceived by laypeople, and uses this same visualisation as the basis of an explainable, automated interpretation algorithm. The rationale for using a science-of-perception-based approach to inform the design and development of ECG self-monitoring interventions is as follows:

- Empirical evidence has demonstrated that accurate ECG interpretation depends primarily on the human perceptual skill of pattern recognition [WBA+14], with experts in ECG interpretation relying on this first, and resorting to a more systematic approach (*i.e.* measuring the duration and amplitude of different waves and intervals precisely) if they run into barriers [DMH+19].

- Reasoning is often an interaction with cognitive tools including computer systems, a concept described as 'distributed cognition' emphasising that little intellectual work is done with our eyes and ears closed [Hut00]. Data visualisation techniques therefore have an important role to play in supporting rapid and accurate interpretation of complex information [War12, HBE95, HBE96]. One of the most powerful aspects of visualisation is that it exploits the highest bandwidth channel between the computer and the human brain, as we acquire more information through vision than through all of the other senses combined [War12].

- To support an intuitive ECG interpretation for laypeople, without extensive prior training required, we need to understand how people perceive ECG data, and the extent to which using knowledge of human perception in the design of ECG visualisation techniques can aid the interpretation process and make it easier to perceive clinically significant signal patterns.

- Using an understanding of human visual perception of the ECG as the basis for machine interpretation has the potential to endow machines with a 'human-like' perceptual ability to recognise clinically significant signal patterns, and thus facilitate an explainable, shared human-machine ECG interpretation.

1.2 Aim and research questions

The overarching aim of this research was to examine laypeople's ability to perceive drug-induced long QT syndrome (diLQTS) on the ECG and explore whether exploiting knowledge of human visual perception to re-visualise the ECG signal could facilitate diLQTS detection by humans and machines simultaneously. Based on the aforementioned research motivation and challenges, the specific research questions were as follows:

1. Can laypeople identify a drug-induced QT-prolongation on the standard ECG visualisation method? Does the presentation of the ECG signal trace (as a single complex representing one heartbeat or a 10-second rhythm strip showing more than one complex) affect this ability?

2. Can the use of science-of-perception-based visualisation techniques to re-visualise the ECG signal improve laypeople's ability to visually detect drug-induced QT-interval increases from a drug-free ECG baseline at a regular heart rate?

3. Can the visualisation technique be developed to support laypeople in differentiating between prolonged QT-intervals at risk of developing TdP and 'normal' QT-intervals showing no risk of TdP at varying heart rates, with the presence of drug-induced T-wave morphological changes, and without comparison with a drug-free ECG baseline?

4. Can the perceptual heuristics used by humans when interpreting the visualisation be modeled to improve automated QT-prolongation detection algorithms?

5. Can the algorithm be developed to reliably detect drug-induced QT-prolongation at risk of TdP regardless of heart rate and T-wave morphology across a wide range of ECG cases?

1.3 Overview of this thesis

This thesis is submitted, with permission from the supervisory team from the Faculty of Science and Engineering, in the journal format (formerly known as alternative format). The main chapters within the thesis (Chapter 4 to 8) are therefore in the form of research papers. These chapters follow a sequential approach in addressing the research questions, and the section titled 'Thesis context' at the beginning of each chapter is dedicated to explain its unique contribution to the overall research problem the thesis aims to address. The content of each chapter and the corresponding publication are outlined below:

- **Chapter 2** presents the theoretical and practical context of this research using an interdisciplinary review approach—across the fields of cardiac physiology, computer science and cognitive psychology—to gain a broader understanding of the research problem, discuss the challenges of current ECG interpretation approaches and suggest potential, science-of-perception-based solutions for detecting long QT syndrome on the ECG.

- **Chapter 3** summarises and appraises the published evidence on computer-based ECG interpretation and visualisation methods proposed in the current literature to support LQTS detection and/or risk assessment. It discusses recent advances in and limitations of these methods and identifies the current knowledge gaps in the field.

- **Chapter 4** addresses the first research question through a psychophysical and eye-tracking experiment that quantifies laypeople's ability to detect drug-induced QT-interval prolongation at a regular heart rate on the standard ECG, and determines whether the presentation of the ECG signal trace (as a single heartbeat or ten second lead) affects this ability. The content of this chapter is adapted from:

  > Alaa Alahmadi, Alan Davies, Markel Vigo, and Caroline Jay. Can lay people identify a drug-induced QT-interval prolongation? A psychophysical and eye-tracking experiment examining the ability of non-experts to interpret an ECG. *Journal of the American Medical Informatics Association*, 2018.

- **Chapter 5** addresses the second research question with a psychophysical and eye-tracking experiment that evaluates whether using a pseudo-colouring visualisation technique to highlight QT-interval duration on different coordinate systems (Cartesian *vs.* Polar) can support laypeople in identifying increases in the QT-interval from a drug-free ECG baseline at a regular heart rate. The content of this chapter is adapted from:

  > Alaa Alahmadi, Alan Davies, Jennifer Royle, Markel Vigo, and Caroline Jay. Evaluating the impact of pseudo-colour and coordinate system on the detection of medication-induced ECG changes. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2019.

- **Chapter 6** addresses the third research question with a multi-reader, multi-case (MRMC) receiver operating characteristic (ROC), psychophysical and eye-tracking experiment that evaluates whether using the pseudo-colouring visualisation technique and changing the coordinate system (Cartesian *vs.* Polar) can support laypeople in differentiating between prolonged QT-intervals at risk of developing TdP and 'normal' QT-intervals showing no risk of TdP at varying heart rates and across different QT-prolonging drugs affecting T-wave morphology. The content of this chapter is adapted from:

  > Alaa Alahmadi, Alan Davies, Markel Vigo, and Caroline Jay. Pseudo-colouring an ECG enables lay people to detect QT-interval prolongation regardless of heart rate. *PLoS One*, 15(8):e0237854, 2020.

- **Chapter 7** addresses the fourth research question by exploring how modelling the visual perceptual process—that we hypothesise the human is using to interpret the pseudo-colouring visualisation—can be used as a basis for an automated ECG interpretation algorithm. A pilot version of an explainable algorithm that uses a 'human-like' signal representation approach is presented, where its accuracy is compared with human interpretation data and current signal processing techniques in detecting QT-prolongation at risk of TdP across several heart rates and QT-prolonging drugs. The content of this chapter is adapted from:

  > Alaa Alahmadi, Alan Davies, Katherine Dempsey, Markel Vigo, and Caroline Jay. Human-machine perception of complex signal data. In *Human-Like Machine Intelligence*. Oxford University Press, 2021.

- **Chapter 8** addresses the fifth research question and presents an enhanced version of the explainable 'human-like' algorithm tested on a larger number of ECGs ($n = 5050$) across four known QT-prolonging drugs and placebo. We empirically compares two approaches to building the enhanced algorithm: a manually-curated 'expert' algorithm that incorporates knowledge from the clinical literature considering the T-wave morphology; and a statistical machine learning decision tree, which automates the generation of the rules from the same set of data. The effect of drug type on the algorithm's sensitivity to increases in the QT-interval was also modelled to evaluate the effectiveness of the algorithm. The chapter also reports the results of two focus groups—one consisting of patients, the other of clinicians—which explored the relevance of our approach to clinical practice. The content of this chapter is adapted from:

  > Alaa Alahmadi, Alan Davies, Jennifer Royle, Leanna Goodwin, Katharine Cresswell, Zahra Arain, Markel Vigo, and Caroline Jay. An explainable algorithm for detecting drug-induced QT-prolongation at risk of torsades de pointes (TdP) regardless of heart rate and T-wave morphology. *Computers in Biology and Medicine*, 2021.

- **Chapter 9** concludes by summarising and synthesising the main findings of this thesis, and recommending directions for future work.

1.4 Contributions

This research provides a new perspective on the challenging topic of electrocardiogram (ECG) interpretation and drug-induced long QT syndrome detection. It is the first work to examine the ability of laypeople to detect drug-induced QT-prolongation on an ECG; devise a novel ECG visualisation technique to enable them to interpret it accurately; and exploit an understanding of visual perception to improve automated QT-prolongation detection—facilitating an explainable, shared human-machine ECG interpretation. Empirically addressing the five research questions reported in Section 1.2 yielded the following contributions:

1. **A demonstration that laypeople are able to detect drug-induced QT-prolongation on the standard ECG, and that they are able to do this more reliably when viewing multiple ECG complexes than when they are viewing a single complex.**

   A psychophysical and eye-tracking study was used to quantify laypeople's ability to detect QT-prolongation on the ECG, and demonstrated that the majority of people can perceive a clinically significant difference in QT-interval length (drug-induced QT $>$ 500ms with a potential risk of TdP) when compared with a 'normal' drug-free ECG baseline, when both ECGs have a heart rate of 60 bpm. The study also demonstrated that the rhythm strip, which shows ten seconds of the ECG, and therefore more than one ECG complex/heartbeat, is a better form of presentation than a single complex, as it is less likely to be misinterpreted due to artefacts in the signal. This is the first study to examine the ability of laypeople to detect QT-prolongation, and the results provide the evidence that self-monitoring is possible.

2. **A novel ECG visualisation technique that improves laypeople's accuracy in visually detecting small drug-induced increases in the QT-interval from a drug-free ECG baseline at a regular heart rate—eliminating the need for QT-interval measurement.**

   A novel ECG visualisation technique which uses pseudo-colour to highlight QT-interval duration on the ECG was designed, implemented, and evaluated using a science-of-perception-based design approach. To understand the impact of the coordinate system on ECG data interpretation, presentation of the ECG signal on Cartesian and Polar coordinates was compared with and without

pseudo-colouring. Psychophysical and eye-tracking methods were used to systematically evaluate the technique, and the results demonstrated that introducing pseudo-colour to the ECG significantly improves interpretation accuracy, and that the coordinate system interacts with colour. Compared with the standard ECG visualisation method, the pseudo-colour significantly improves people's ability to detect small increases in the QT-interval ($\approx$ 19ms from a drug-free baseline was estimated to be a noticeable increase for 75% of people) when the ECG is displayed on a standard Cartesian coordinate system, but the greatest accuracy is achieved when pseudo-colour is combined with Polar coordinates ($\approx$ 9ms QT-interval increase from a drug-free baseline was estimated to be a noticeable increase for 75% of people). This shows that pseudo-colour can improve sensitivity to QT-interval changes such that people can perceive increases that are much smaller than a 1mm square on the standard ECG grid (which represents 40ms), even when T-wave morphology is abnormal.

3. **An extension of the ECG visualisation technique that works at any heart rate, and which improves laypeople's accuracy in visually distinguishing between 'normal' and prolonged QT-intervals at risk of TdP regardless of heart rate and T-wave morphology—eliminating the need for QT-interval measurement, heart rate correction and a drug-free baseline ECG comparison.**

   We extended the development of pseudo-colouring technique to be automatically adjusted according to heart rate using a clinically reliable TdP risk assessment method known as the QT-nomogram [CIKD07]. We used a multi-reader, multi-case (MRMC) receiver operating characteristic (ROC), psychophysical and eye-tracking methods to systematically evaluate the technique, and we found that applying pseudo-colouring to ECGs according to the QT-nomogram supports laypeople in visually detecting QT-prolongation at risk of TdP, as well as identifying 'normal' QT-intervals showing no risk of TdP, regardless of heart rate, T-wave morphology and coordinate system. The technique was tested with QT-intervals across several heart rates for multiple patients on different QT-prolonging drugs; no comparison baseline ECG was provided. Pseudo-colour also helped reduce reaction times and increased satisfaction when reading the ECGs. Eye movement analysis indicated that pseudo-colour helped to focus visual attention on the areas of the ECG crucial to detecting QT-prolongation. The results provide further evidence that self-monitoring ECGs for drug-induced

LQTS is feasible.

4. **An algorithm that exploits an understanding of the human perceptual process to improve automated QT-prolongation detection without prior identification and detection of the Q-wave or T-wave—outperforming current signal processing techniques.**

   We conduced an exploratory study modelled the visual perceptual process used by humans when interpreting the pseudo-coloured ECG to develop a simple, explainable rule-based algorithm, which was compared with human interpretation and current signal processing techniques used to measure the QT-interval. The logic behind this novel human-like algorithm differs considerably from that used by standard signal processing methods, as it takes a human-like perceptual perspective, calculating the percentage area of pseudo-colours under the ECG signal curve, without prior identification and detection of the Q-wave or T-wave. The results of this study demonstrate that a model of the human perceptual process can be used as a basis for an automated interpretation algorithm, yielding more accurate results than current signal processing techniques, and similar performance to humans, which indicates that similar 'human-like' processes may be at work.

5. **An extension of the algorithm to incorporate clinical expertise about T-wave morphology. This enhanced 'expert' algorithm can reliably detect drug-induced QT-prolongation at risk of TdP regardless of heart rate and T-wave morphology across a wide range of ECG cases—facilitating an explainable, shared human-machine ECG interpretation.**

   We conducted an extended evaluation study introduced an enhanced version of the explainable algorithm (called the 'expert' algorithm), which uses a 'human-like' approach, where human perception of the pseudo-coloured ECG signal is used to determine features, and diagnostic rules that are determined according to the clinical literature. The study also explored whether the algorithm can be improved by automating the rule-generation with a decision tree, and reported the results of two focus group evaluations—one with patients and one with clinicians—which explored the relevance of our approach to clinical practice.

The results of the study showed that the 'expert' algorithm significantly improves machine detection of drug-induced QT-prolongation at risk of TdP regardless of heart rate and T-wave morphology, across a wide range of ECG cases. This combined approach was more accurate than the averaged human participants. Therefore, we hypothesise that a human-like approach to the development of diagnostic algorithms—incorporating both human pattern recognition ability and clinical expertise—may have wide utility. Automating rule generation based on these features has potential, but here it was not sufficient to produce a reliable, trustable algorithm. A human-in-the-loop approach, where machine learning is a tool used to surface potential rules, but these are validated empirically before being implemented, maybe the best approach to use in clinical practice.

The focus group evaluations confirmed that the explainability of the algorithm is important, supporting an 'expert' rule-based approach, rather than a fully automated approach. All patients and clinicians had a positive attitude towards using the pseudo-colouring technique and perceived the explainable 'expert' algorithm as a supportive tool that may help to overcome the potential challenges associated with human interpretation of the ECG. Clinicians commented that our approach would be particularly useful in reducing common errors associated with manual QT-interval measurement, and in resolving issues with inter-observer variability, particularly in clinical trial settings testing a new drug where accurate, frequent QT-interval monitoring is crucial.

Overall, as well as having the potential to revolutionise self-monitoring ECGs for LQTS (potentially saving many lives), this thesis lays the theoretical foundations for how we can combine human perception and clinical knowledge to produce explainable algorithms that are robust and transparent enough for use in clinical practice.

1.5 External recognition of the contributions

- The work presented in Chapter 5 received the following awards:
  - The University of Manchester Carole Goble Medal for outstanding doctoral paper in computer science (2019).
  - An excellence award of £1000 from Taibah University, Saudi Arabia (2019).
  - It was Highly Commended in the The Institution of Engineering and Technology (IET) Healthcare Technologies Awards (2020).
- The work presented in Chapter 6 received an excellence award of £1000 from Taibah University, Saudi Arabia (2020).
- The author was selected as the only Computer Science finalist in the Scientific and Parliamentary Committee/Royal Academy of Engineering Exhibition of STEM for BRITAIN for the work presented in Chapter 8 (2021). This work also received an excellence award of £1000 from Taibah University, Saudi Arabia (2021), and accordingly the author was awarded a congratulatory, distinction certificate for receiving three consecutive excellence awards during her PhD.

Chapter 2

Background

2.1 Chapter overview

This chapter presents the theoretical and practical foundations of this research using an interdisciplinary review approach—across the fields of cardiac physiology, computer science and cognitive psychology—to gain a broader understanding of the research problem, discuss the challenges of current ECG interpretation approaches and suggest potential, science-of-perception-based solutions for detecting long QT syndrome on the ECG.

To understand what an ECG represents, we begin with a brief introduction to the basic physiological components of the normal and prolonged (during long QT syndrome) cardiac cycles and how they are reflected on the ECG. Next, the chapter provides an overview of the cellular action potential mechanism that drives the cardiac cycle. This addresses critical questions relevant to this research including: How can pharmacological drugs cause long QT syndrome (LQTS)?; What are the most significant drug-induced ECG changes (associated with LQTS) that could serve as major risk predictors of Torsades de Pointes (TdP) arrhythmia and sudden cardiac death? The chapter then introduces the standard ECG interpretation method currently used in clinical practice, with the aim of understanding why LQTS, in particular, is difficult to recognise clinically and computationally. The main challenges associated with human and machine interpretation of LQTS reported in the literature are also summarised and discussed.

One of the challenges associated with ECG interpretation is that it visualises the raw signal of the heart electrical activity directly, which can be dynamic (where different cardiac pathologies interfere) and highly variable across individuals due to many

factors—including age, gender and specific ethnic genes [GMZ06, MMDY94, HST+16]—without any pre-processing of that visualised signal. This complicates the signal representation in general and the perception of a difference or abnormal change within the signal in particular. The human brain has the ability to filter the accumulated, vast sensory information from the surrounding complex environment and process what is important [War21]. Thus, to explore potential approaches to improving LQTS detection on the complex ECG signal, the chapter introduces a fundamental theory in human vision known as pre-attentive processing. We describe how introducing pre-attentive attributes such as colour (hue and intensity) to complex visual information can increase stimulus salience and modulate visual processing speed in human perception. We then discuss how such visual attributes can move from pre-attentive to attentive processing when influenced by an individual's intentions and visual task goals. [1]

2.2 ECG interpretation of long QT syndrome: from basic physiology to clinical practice

2.2.1 The cardiac cycle: structure and function of the heart

The heart is a vital organ responsible for pumping blood throughout the body, supplying tissues with oxygen and nutrients, and removing carbon dioxide and other waste products [Kat10, Opi04]. The internal cavity of the heart is divided into four functional chambers: the left and right atrium for receiving blood and the left and right ventricle for pumping it [Kat10, DS14].

A single cardiac cycle (*i.e.* one heartbeat) can be divided into two primary phases: the diastole and the systole, which occur in the atria first and then in the ventricles [Kat10, Opi04]. Initially, all the chambers of the heart are in the diastole phase, where the atria and ventricles are relaxed (not contracting), allowing both atria to fill with blood [DS14, Kat10]. The deoxygenated blood from the body is emptied into the right atrium via the superior vena cava (SVC) and inferior vena cava (IVC) [DS14, Kat10]. The left atrium receives oxygenated blood from the lungs via four pulmonary veins. Then, while the ventricles are still in the diastole phase, the atria systole phase begins where both atria contract, allowing blood to flow passively from the left atrium and right atrium into the left ventricle and right ventricle, respectively [DS14, Kat10]. The mitral valve regulates the movement of blood from the left atrium to the left ventricle,

[1] The illustrations/figures in this chapter were created using BioRender.com.

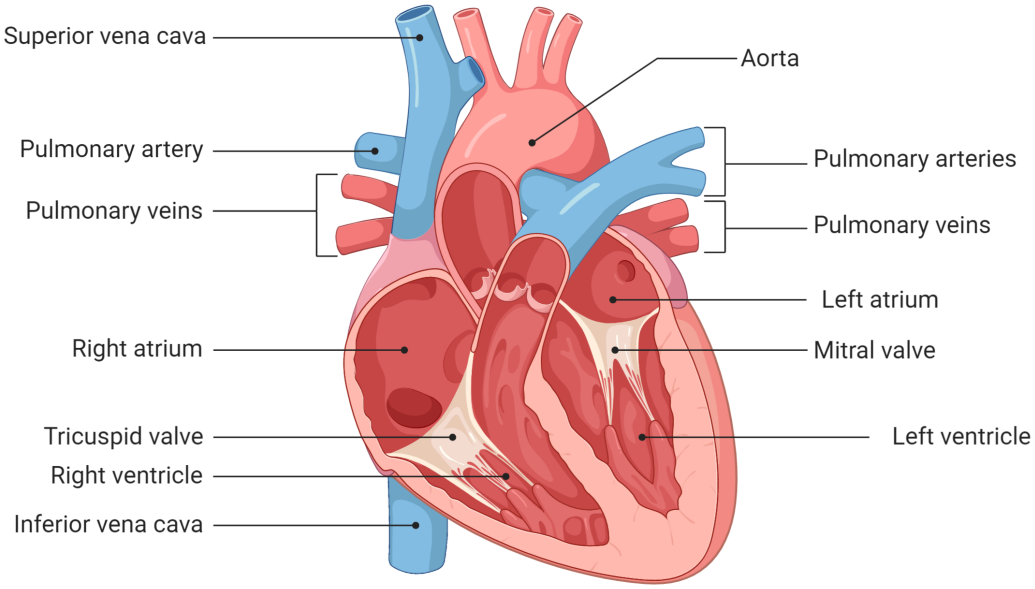


Figure 2.1: An illustration of the heart's chambers and vessels.

while the tricuspid valve allows blood to flow from the right atrium to the right ventricle, preventing blood from flowing backwards [DLRH19]. The ventricular systole phase then begins where both ventricles contract [DS14, Kat10]. The deoxygenated blood in the right ventricle is pumped into the lungs through the pulmonary artery to be oxygenated, while the oxygenated blood in the left ventricle is pumped into the rest of the body via the aorta [DS14, Kat10]. Figure 2.1 shows an illustration of the chambers and vessels of the heart.

How does the ECG relate to the cardiac cycle?

The cardiac cycle mechanism is regulated by an electrical conduction system that coordinates the contraction of the four chambers of the heart and governs the rate at which the heart beats [IMZ12, Kat10, Kha08, DS14]. This electrical conduction system consists of several components that are responsible for generating electrical impulses and passing them through the heart including (1) the sinoatrial node (SAN or SA node) for generating the heart's electrical impulses; (2) the atrioventricular node (AV node) for intentionally delaying impulses from the atria giving the ventricles enough time to finish filling with blood; (3) the bundle of His for allowing the impulses to travel from the atria to the ventricles; and (4) the Purkinje fibres, which play a significant role in

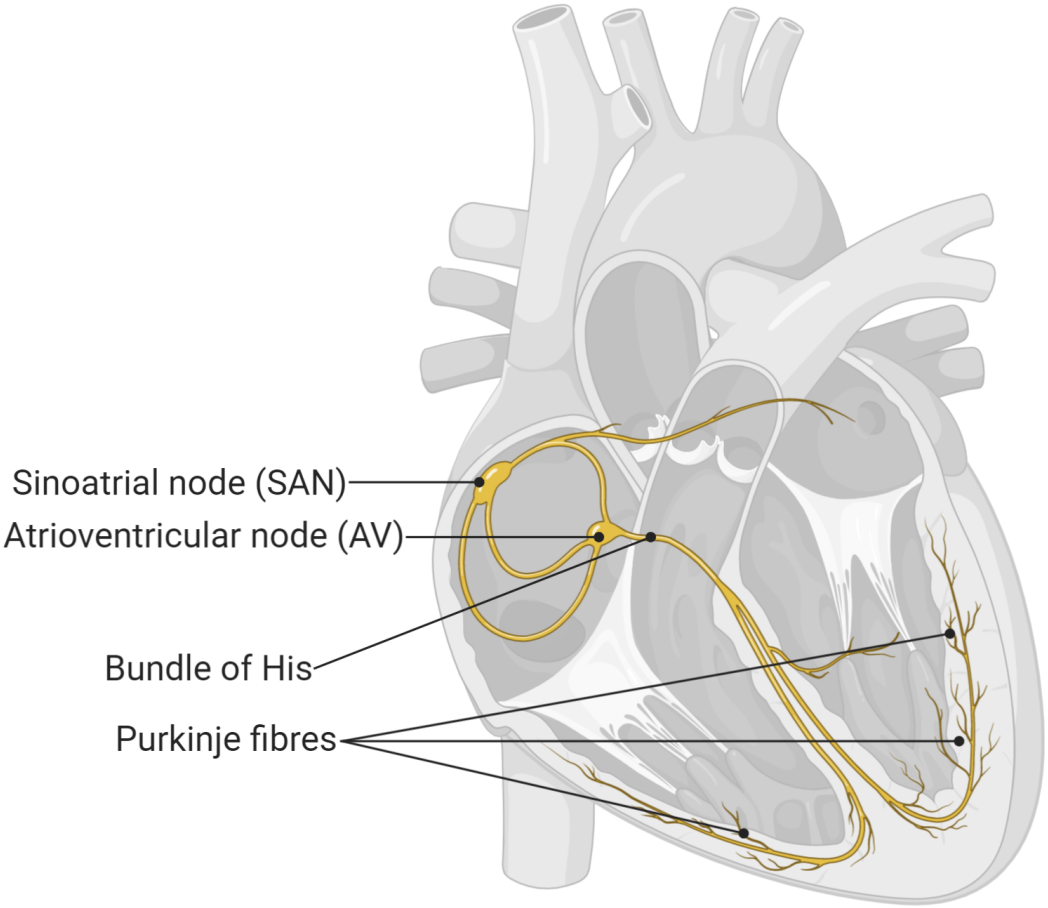


Figure 2.2: An illustration of the electrical conduction system of the heart.

activating the electrical conduction and impulse propagation during the ventricular systole phase [IMZ12, Kat10, Kha08, DS14]. Figure 2.2 illustrates the heart's electrical conduction system.

The heart's electrical activity is measured and recorded by an electrocardiography device using electrodes placed on the skin of the limbs and chest [Kha08, DS14, CAM$^+$06]. The result of this recording is shown on an electrocardiogram (ECG) as a graphical signal consisting of different waves, which are labelled with letters to represent different stages of the cardiac cycle [Kha08, DS14, CAM$^+$06], as shown in Figure 2.3. A combination of these waves, representing a single cardiac cycle, is known as an 'ECG complex' [DS14]. The heart's electrical activity during the cardiac cycle's diastole and systole phases can be shown on the ECG, where the P-wave

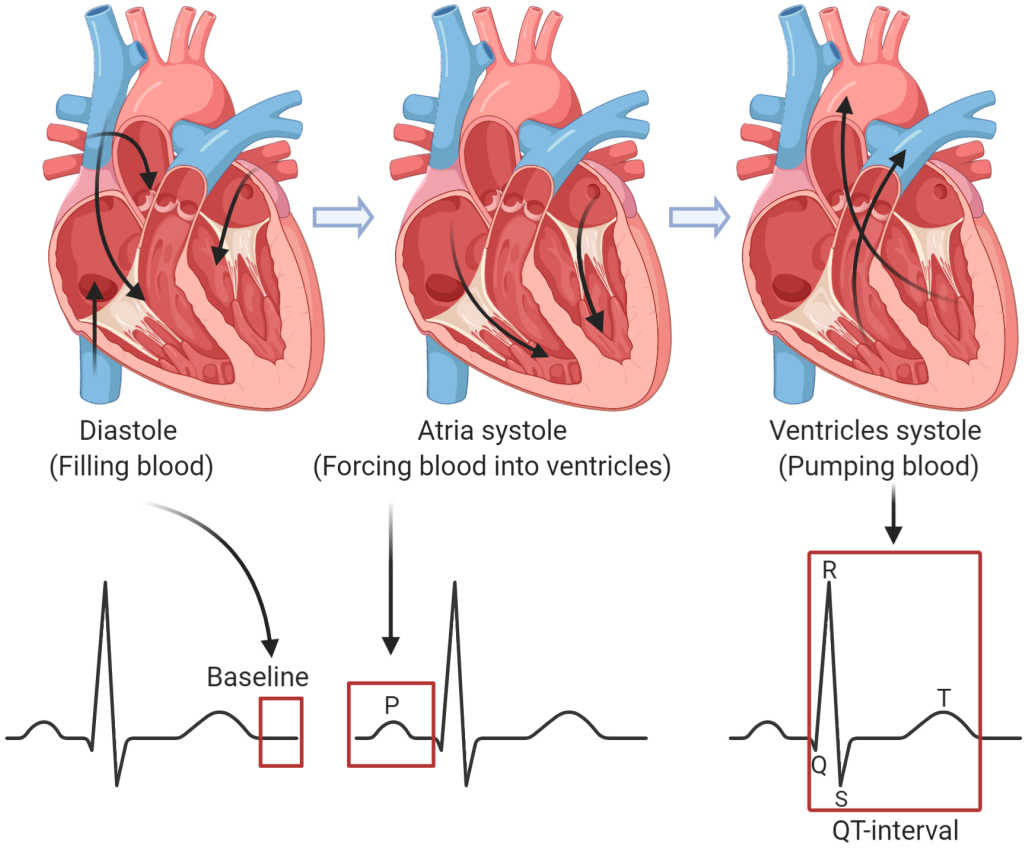


Figure 2.3: An illustration of how the diastole and systole phases of the cardiac cycle are reflected on an electrocardiogram (ECG).

represents the atrial systole, while the Q through T waves represent the ventricles systole [IMZ12, Kha08, DS14, CAM$^+$06]. Figure 2.3 illustrates the diastole and systole phases during the cardiac cycle on the ECG.

Although the heart's electrical activity signal displayed on the ECG may not reflect all the cardiac cycle's mechanical events [BNNK18, Opi04], changes in this signal can be strong indicators of cardiac pathologies [TRB16, IMZ12, DS14]. Therefore, the ECG is considered a useful tool in diagnosing and detecting many types of cardiovascular disease (CVD) [Kha08, DS14, CAM$^+$06], particularly in asymptomatic patients who may suffer from a clinically silent cardiac conduction disorder or arrhythmia—an abnormal heart rhythm caused by an irregular activity of the heart's electrical signal [TRB16, CMY08].

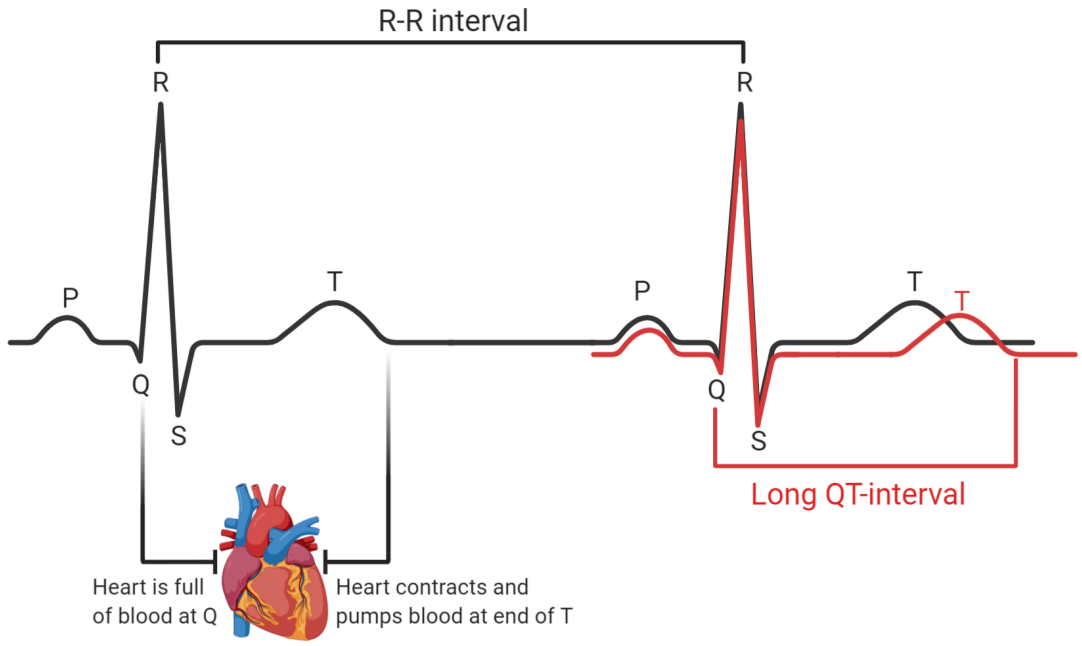


Figure 2.4: An illustration of electrocardiogram (ECG) waveforms and the long QT-interval. The RR-interval represents the time between heartbeats and is used to calculate heart rate.

What happens to the cardiac cycle during long QT syndrome?

As shown in Figure 2.3, the ventricles systole occupies the largest part of the heart's electrical activity during the cardiac cycle [IMZ12]. This electrical activity coordinates the contraction-relaxation cycle of the ventricles' muscles, and its duration is measured on the ECG from the onset of the Q-wave to the end of the T-wave, which is known as the QT-interval [YC03, CMY08]. In long QT syndrome (LQTS), the ventricles' electrical activity takes longer than normal to coordinate this cycle; thus, the heart is not pumping enough blood [YC03, CMY08, BPR+17]. Such delay prolongs the QT-interval on the ECG (as illustrated in Figure 2.4) and can result in syncope (fainting due to the insufficient blood flow to the brain) and sudden cardiac death [TRB16]. LQTS can be congenital (presented at birth and caused by specific genetic mutations) or acquired as a result of taking commonly prescribed medications including antibiotics, antidepressants, antihistamines and antiarrhythmic drugs [YC03, CMY08].

2.2.2 Cellular mechanisms of cardiac electrical activity

One of the characteristic features of a living biological cell is the ability to conduct electrical signals through the cell membrane [Spe12, LZ01]. The membrane potential, or membrane voltage, is the difference in voltage between the interior and the exterior of a biological cell [Spe12, LZ01]. This is accomplished by controlling the concentration of differently charged ions between the intracellular fluid within the cell, and the extracellular fluid outside the cell [IMZ12, Spe12]. The membrane potential is a vital aspect of the cardiac cycle and cell-to-cell electrical signalling, and ultimately the heart function [IMZ12, Spe12]. This section describes the basic principles with respect to ions and membranes' electrical and chemical properties and how these forces combine to induce the heart muscles to contract and relax during the cardiac cycle. These principles are essential for interpreting the ECG, as well as understanding the causes and mechanisms underlying drug-induced long QT syndrome, and sudden cardiac death [B$^+$18, TRB16].

Cardiac action potential

In physiology, an action potential is a mechanism by which certain types of body cells, known as electrically excitable cells, rapidly change in voltage to propagate an electrical signal [CAK07, Spe12, LZ01]. It induces the cell's resting membrane potential—where the voltage is mostly negative inside the cell—to depolarise and become gradually more positively charged when triggered by external electrical impulses [CAK07, Spe12, LZ01]. This process is accomplished by regulating the flow of differently charged ions, particularly sodium (Na^+), potassium (K^+) and calcium (Ca^{++}), between the inside and outside of the cell, through proteins called ion channels [IMZ12, Spe12]. At the end of the depolarisation phase, a repolarisation phase begins where the cell returns to its resting state (*i.e.* it becomes mostly negatively charged again) following a brief recovery period [IMZ12, Spe12]. An action potential occurs in nerve cells to communicate and conduct information and in muscle cells (also known as myocytes) to induce the contraction process [IMZ12, Spe12].

The heart is essentially a muscular organ that consists of cardiac muscle cells, known as cardiomyocytes or contractile cells, and cardiac pacemaker cells [IMZ12, Spe00, Spe12]. The cardiomyocytes make up the heart atria and ventricles [IMZ12, DS14]. The cardiac pacemaker cells can spontaneously generate the heart's intrinsic

electrical activity, triggering the cardiac action potential mechanism in the cardiomyocytes and causing the heart muscles to contract and pump blood during a heartbeat [Spe00, IMZ12]. The sinoatrial node (SA node), atrioventricular node (AV node) and the bundle of His are all types of cardiac pacemaker cells [Spe00, IMZ12, DS14] (Figure 2.2). The SA node is the primary pacemaker of the heart and has the highest rate of spontaneous depolarisation electrical impulses [Spe00, IMZ12, DS14].

The action potential occurs in both the cardiomyocyte and the pacemaker cells [Spe00, IMZ12, DS14]. The action potential in ventricular cardiomyocyte cells is stronger and lasts longer than in the pacemaker cells [IMZ12, Spe00, Spe12]. It plays a significant role in maintaining the function of the cardiac cycle, mainly at the stage of pumping blood from the heart to the rest of the body [Kat10, Opi04]. It is generally divided into 5 stages (stages 0-4) [IMZ12, Spe00, Spe12]. Phase 4 acts as the baseline at which membrane potential begins and ends. Phase 0 is the phase of fast depolarisation; Phase 1 to 3 are when repolarisation occurs; Phase 4 is the resting phase [IMZ12, Spe00, Spe12, PTH07]. The action potential mechanism during each phase is outlined below.

Phase 4 (Resting phase) When the heart ventricles are relaxed, the potential inside the cardiomyocyte cell is negative compared with the outside, and the cell membrane is at rest [PTH07, Spe12]. The ions outside the cell are primarily sodium (Na^+) and calcium (Ca^{++}), whereas inside the cell they are mainly potassium (K^+) [IMZ12, Spe00, Spe12]. During this phase, the voltage inside the cell is more or less constant, and it has a stable resting membrane potential of roughly -90 mV [PTH07, CMY08, Spe12]. This is accomplished by maintaining the concentration gradient of K^+ ions across the cell membrane. [IMZ12, Spe00, Spe12].

Phase 0 (Fast depolarisation) The SA node cells generate a positive depolarisation electrical impulse, which is typically fired at a rate of between 60 and 100 beats per minute (BPM) [DS14]. This electrical impulse stimulates the resting membrane potential to depolarise, by increasing the positive voltage within the cell slightly until it reaches a certain threshold (roughly -70mv) [PTH07, CMY08, Spe12]. At this threshold, membrane permeability to potassium decreases and fast sodium channels open, causing a rapid influx of inward sodium currents (I_{Na}) to the cell, and a fast and steep depolarisation in ventricular cells from -70mV to +30mV [PTH07, CMY08, Spe12],

reflected on the ECG as a combination of Q, R and S waves, known as the QRS complex (Figure 2.6).

Phase 1 (Notch, Early repolarisation) This phase begins by rapidly closing the sodium channels such that they become inactivated. [IMZ12, Kat10]. This is combined with rapidly opening and closing the potassium channels, allowing for a short transient efflux of K^+ [IMZ12, Kat10]. The decrease in sodium permeability makes the membrane potential slightly more negative [IMZ12, Kat10]. The short diffusion of outward potassium currents (I_{to}) leads to a partial repolarisation, known as 'notch' or 'early repolarisation' [PTH07, CMY08]. This phase is often very short and is included in the depolarisation (The QRS complex) representation on the ECG [Kat10].

Phase 2 (Plateau) The calcium channel opens in this phase, allowing a greater influx of inward L-Type Ca^{++} currents ($ICa-L$) to flow into the cell, which eventually balances the efflux of K^+ currents [IMZ12, Kat10]. This phase creates a constant voltage in the membrane potential, known as 'plateau' phase [PTH07, CMY08, KBM20], reflected on the ECG as the ST-segment (Figure 2.6).

Phase 3 (Raid repolarisation) Rapid repolarisation follows in phase three, where calcium channels close whilst potassium channels remain open, allowing for a more rapid efflux of K^+ [IMZ12, Kat10]. This rapid repolarisation requires more potassium ions to flow out of the cell, driven predominately by the interaction of two types of potassium ion channels: the rapid (I_{Kr}) and slow (I_{Ks}) outward currents of the delayed rectifier potassium channels [PTH07, CMY08]. The I_{Kr} is activated before the I_{Ks}, and both of them are crucial for ventricular repolarisation [GWY$^+$11, CMY08]. This phase is reflected on the ECG as the T-wave (Figure 2.6).

To summarise, the cellular action potential driving the cardiac cycle is regulated by a specific series of depolarising and repolarising ion currents mediated by cardiac ion channels. During the ventricles' systole phase, the influx of sodium ions (I_{Na} inward currents) are responsible for depolarising the cell, and the diffusion of potassium ions out of the cell (including I_{to}, I_{Kr} and I_{Ks} outward currents) maintains the repolarisation and prepares the ventricular muscles to relax for the subsequent diastole phase of the cardiac cycle. The influx of calcium (I_{Ca-L}) ions helps balance the outward potassium currents, regulating the repolarisation time. Figure 2.5 illustrates the cardiac ion mechanisms during the ventricular depolarisation and repolarisation cycle and how they are

shown on the ECG. Figure 2.6 shows the reflection of each action potential phase on the ECG.

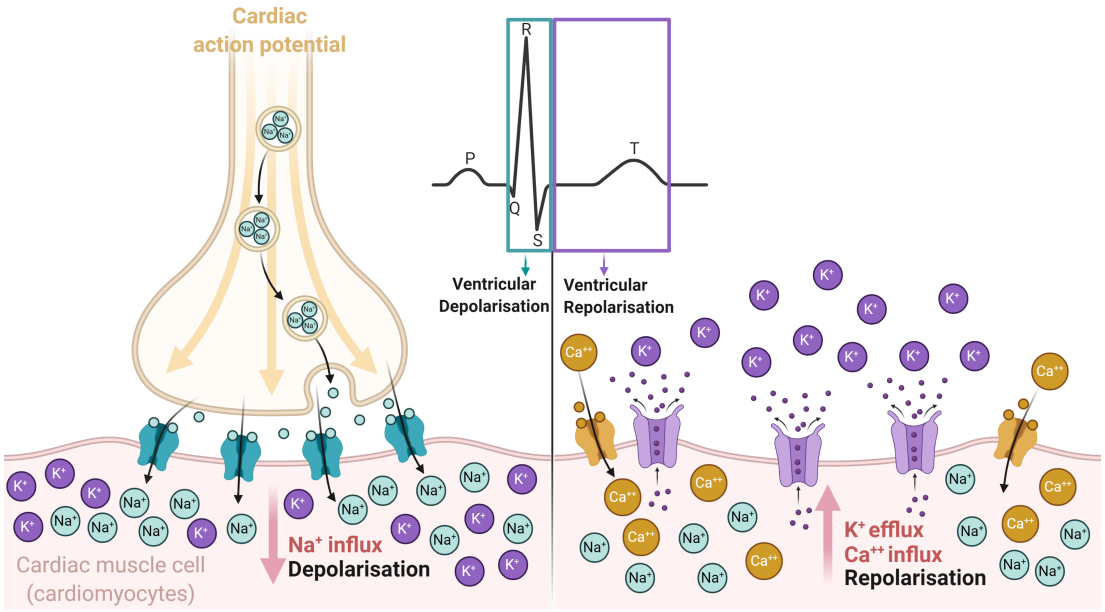


Figure 2.5: An illustration of the cellular action potential mechanisms during the ventricular depolarisation and repolarisation cycle. Depolarisation (represented on the ECG by the QRS complex) is induced by the flow of Na^+ ions into the cell, while repolarisation (represented by the T-wave) is maintained by the diffusion of K^+ ions out of the cell—balanced by the influx of Ca^{++} ions.

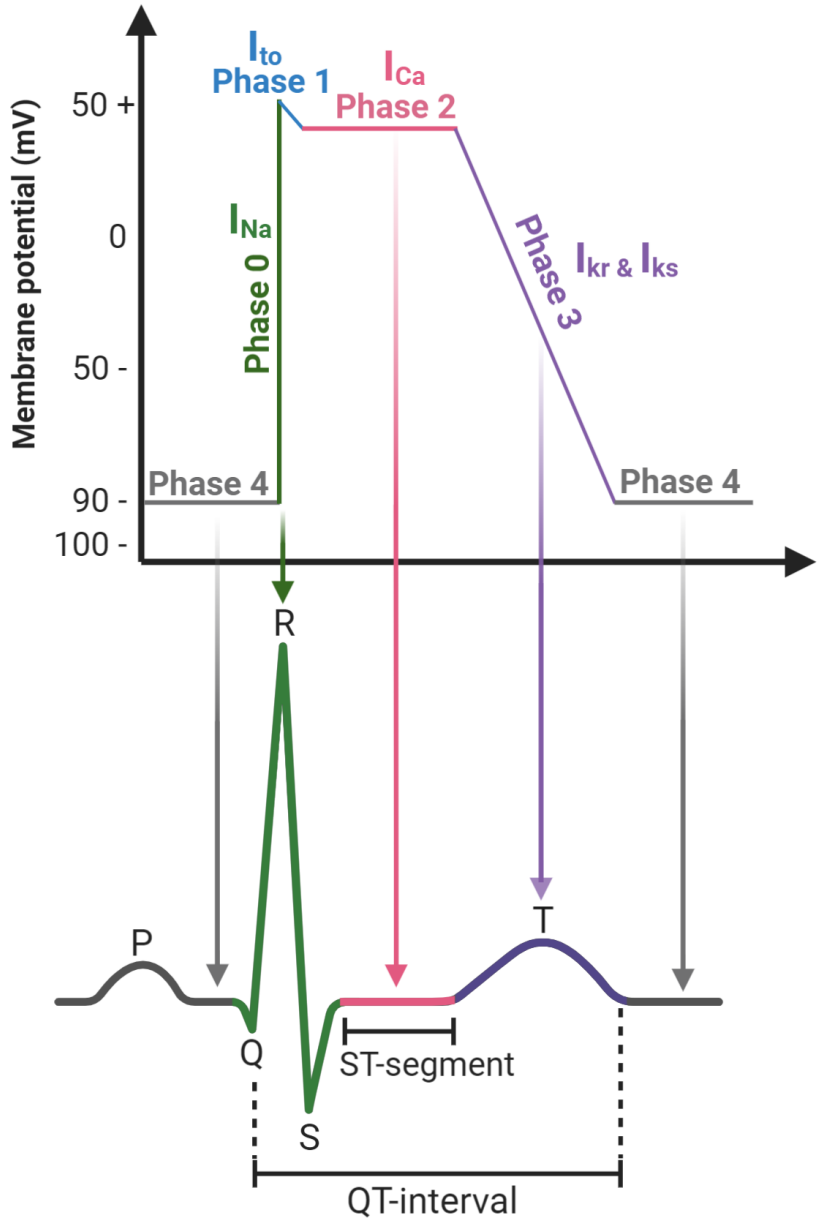


Figure 2.6: The depiction of each phase of the cardiac action potential on the ECG. I_{Na} = inward sodium currents. I_{Ca} = inward calcium currents. I_{to}, I_{Kr} and I_{Ks} are types of outward potassium currents.

2.2.3 'Down the rabbit hole': ion channelopathies, drug-induced long QT syndrome and sudden cardiac death

Prolongation of the cardiac action potential—caused by alterations in the cardiac ion channels—can lead to a life-threatening heterogeneous group of disorders known as channelopathies [SWA$^+$19]. These can be caused by genetic mutations or pharmacological drugs including ion channel blockers, which target one or more of the cardiac ion channels to alter the flow of ions within the heart [CMY08, RV05]. For example, people with high blood pressure can be treated with calcium channel blockers (CCBs) to slow the flow of calcium ions and widen blood vessels, thereby increasing blood flow [EBB04]. Research in precision medicine has shown that individuals may have mutations in one or more of the genes known to cause channelopathies, but little or no prolongation of the cardiac action potential at baseline (*i.e.* before taking the ion channel blocker) [Sch20, DDB$^+$97, Sha81, ASS$^+$99]. In these patients, the mutations of the genes render certain ion channels more susceptible to ion channel blocking drugs [CMY08, Sch20].

First-principles in the cardiac cellular mechanisms show that the potassium channel mutations result in decreased outward currents, while the calcium and sodium channel mutations result in increased inward currents [Ant07, SWA$^+$19, MK05]. Hence, prolonging the cardiac action potential is caused by either an increase in sodium and calcium inward currents or a reduction in the potassium outward currents [SWA$^+$19, MK05].

As described previously, the primary role of potassium channels during the cardiac action potential is to repolarise the cardiac cells [PTH07, CMY08, Spe12]. A reduction in the net outward potassium currents, particularly the rapid (I_{Kr}) and slow (I_{Ks}) currents, can therefore slow down the repolarisation process in the ventricular cells of the heart [B$^+$18, BPR$^+$17]. This, in turn, prolongs the cardiac action potential and the QT-interval on the ECG, which is clinically known as long QT syndrome (LQTS)(Figure 2.7). The human ethera-go-go-related gene (hERG), also known as KCNH2, encodes the rapid I_{Kr} currents, whereas the KCNQ1 and KCNE1 genes together encode the slow I_{Ks} currents [RV05, GWY$^+$11]. These genes account for the majority of LQTS cases [SWA$^+$19]. However, LQTS has also been shown to be caused by mutations in 13 other different cardiac ion-channel genes including SCN5A (encodes for the sodium channels) and ANK2 that encodes ankyrin-R protein affecting sodium channel activation [SWA$^+$19, RV05]. Hence, each specific gene mutation leads to a particular type

of LQTS [SWA+19, RV05, CMY08], which has a unique morphological signature on the ECG [Zar06].

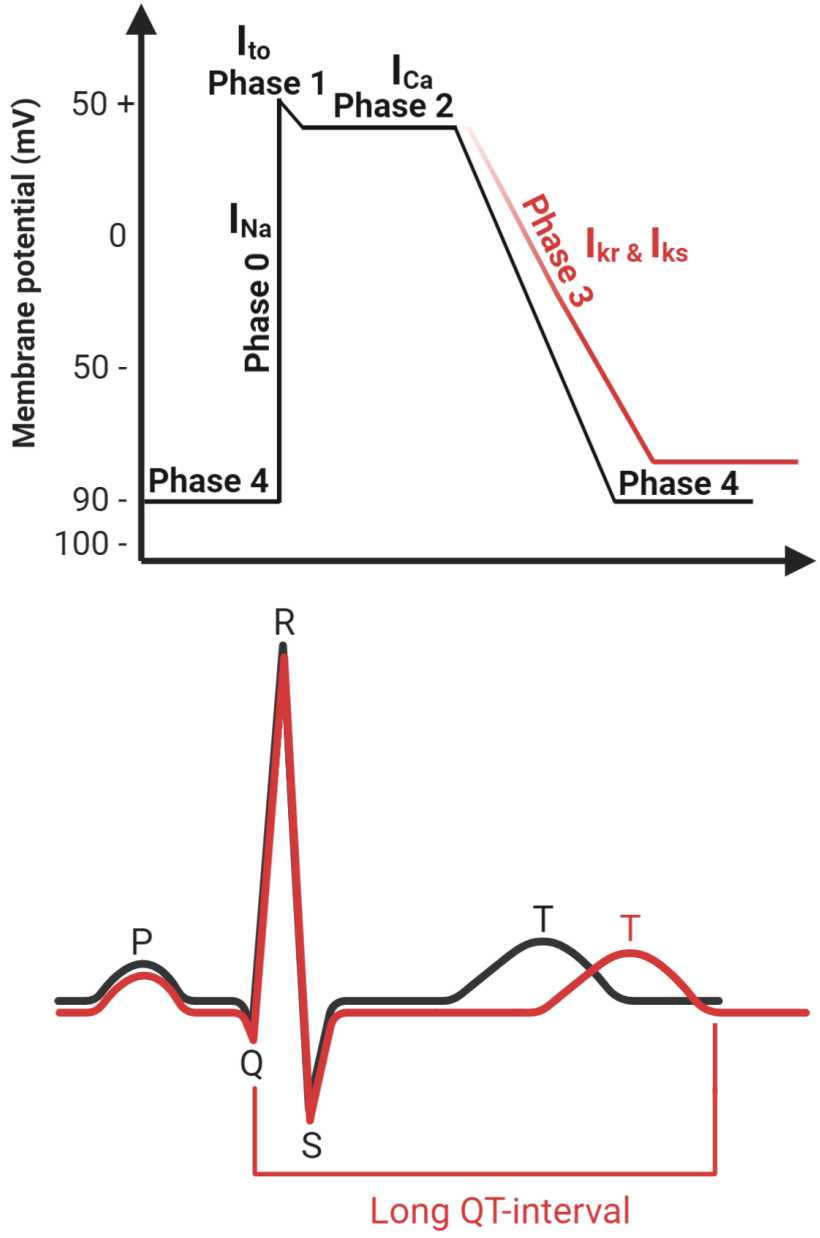


Figure 2.7: Blocking of the cardiac potassium channels (particularly I_{Kr} and I_{Ks} outward currents) prolongs phase 3 of the cardiac action potential, which is reflected on the ECG as a long QT-interval.

The morphology of the QT-interval can thus help to determine the type of long QT syndrome [CMY08, Zar06]. For example, long QT syndrome type 1 (where the affected gene is KCNQ1, blocking I_{Ks} potassium currents) is shown on the ECG by a broad T-wave; type 2 (where the affected gene is hERG, blocking I_{Kr} potassium currents) is characterised by a notched T-wave; while type 3 (where the affected gene is SCN5A, blocking I_{Na} sodium currents) is recognised by a peaked and delayed onset T-wave [CMY08, Zar06]. Figure 2.8 shows these common types of long QT syndrome distinguished on the ECG by the T-wave morphology.

Although long QT syndrome is often caused by genetic mutations, drugs that block cardiac ion channels can also cause an acquired, drug-induced long QT syndrome in healthy people without any genetic abnormality [NAS12, YC03, CMY08, ESTB20, WS20]. The rapid (I_{Kr}) outward potassium current is most susceptible to pharmacological effect [CMY08]. Subsequently, drugs that block the human ether-a-go-go-related gene (hERG) potassium channel are the most common cause of drug-induced long QT syndrome (diLQTS) [TKK06, RPM$^+$05, Bro04]. Since the mutation in this gene (hERG or KCNH2) causes congenital long QT syndrome type 2 (LQTS2), the drug-induced long QT syndrome shares very similar clinical characteristics to congenital LQTS2. Examples include notched T-waves on the ECG, both may cause syncope, or sudden cardiac death (often precipitated by auditory stimuli such as the ringing of an alarm clock or a telephone) [CMY08, Bro04, SWA$^+$19].

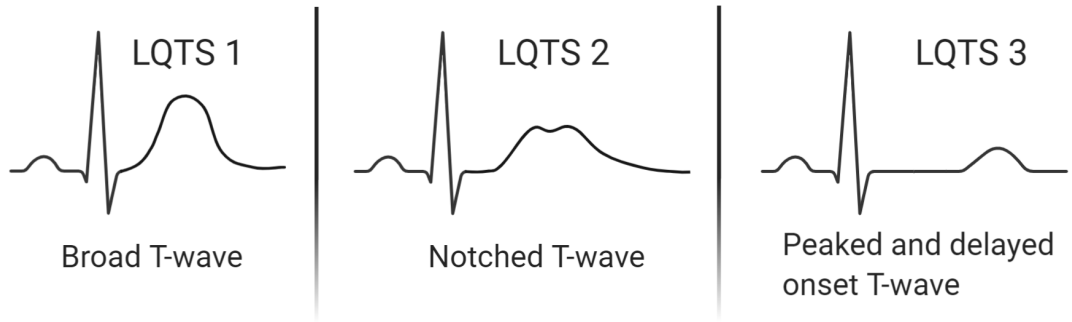


Figure 2.8: An illustration of three distinct common types of long QT syndrome (LQTS) based on the T-wave morphology on the ECG.

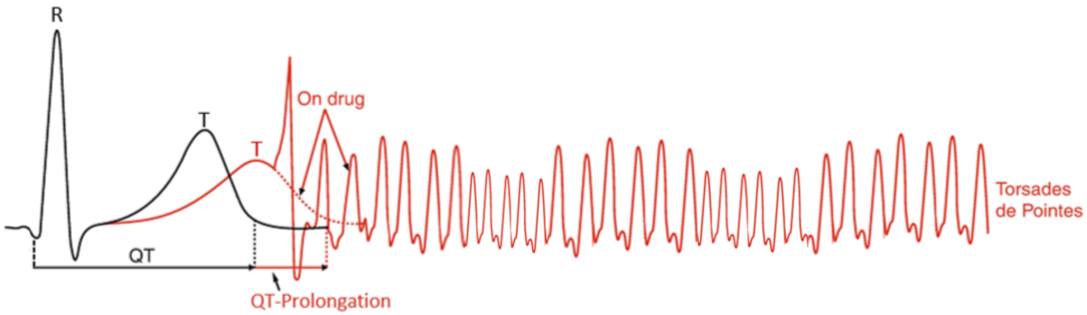


Figure 2.9: An illustration of drug-induced Torsades de Pointes (TdP) pattern on the ECG.

How can long QT syndrome lead to sudden cardiac death?

A malfunction in cardiac ion channels delays cell-to-cell electrical signal transmission, impairing the ventricular repolarisation phase during the cardiac cycle [SWA+19]. This delay can lead to electrical signal re-entry or abnormal automaticity into the subsequent cardiac cycle, resulting in a rapid and abnormal heart rate known as ventricular tachycardia (VT), leading to syncope, dizziness, shortness of breath and chest pain [TRB16, GEB18, Ant07, KS09]. If it is sustained (*i.e.* lasts longer than 30 seconds) and untreated, it can lead to a life-threatening abnormal heart rhythm called ventricular fibrillation (VF), which is commonly known as cardiac arrest [FA19b, CMY08, KS09].

The pattern of ventricular tachycardia (VT) on the ECG can be divided into two types: monomorphic and polymorphic [FA19b]. Monomorphic VT shows symmetrical cardiac cycle patterns, while polymorphic VT shows non-symmetrical cardiac cycle patterns [FA19b]. A unique form of the polymorphic ventricular tachycardia (PVT) that is associated with long QT syndrome is known as Torsades de Pointes (TdP)—the leading cause of sudden cardiac death in young, apparently healthy people [YC03, PK01, WS20]. In clinical practice, TdP is shown to be often short-lived and self-terminating; however, it may degenerate into ventricular fibrillation (VF), and sudden cardiac death [CMY08]. Figure 2.9 illustrates the drug-induced Torsades de Pointes (TdP) pattern on the ECG.

Whilst TdP can be part of both congenital and acquired long QT syndrome (LQTS), acquired LQTS caused by pharmacological drugs is by far the most common cause of TdP [NAS12, YC03, CMY08, ESTB20, WS20]. Moreover, there is a clinically silent type of long QT syndrome known as ‘forme-fruste LQTS’, in which individuals have mutations in one of the cardiac ion channels genes but no QT-prolongation at their ECG

baseline (*i.e.* prior to taking medication) [DDB+97, Sha81, ASS+99]. In these patients, undetected LQTS can manifest itself as drug-induced TdP and subsequent sudden cardiac death upon receiving the QT-prolonging drug [DDB+97, Sha81, CMY08].

The risk of drug-induced long QT syndrome is potentially high, and the TdP risk of many widely used pharmacological drugs has only been recognised many years after they were marketed [CMY08]. As such, a steadily increasing number of medications are now being reported to cause drug-induced QT-prolongation, TdP and sudden cardiac death [Kha02, CMY08, ESTB20]. This, in turn, has troubled both clinicians and drug regulatory authorities, particularly because many of these QT-prolonging drugs are commonly prescribed, often for self-limited diseases [YC03, WS20]. In recent years, there has therefore been considerable renewed research in the risk assessment, understanding, and prevention of drug-induced TdP [ESTB20, CMY08]. A particular focus has been on understanding the role of cardiac ion channels in mediating ventricular repolarisation and the pathogenesis of TdP that can be predicted from the ECG [CNC15, SKQ+15, KFBW09, TRB16, RF18, GEB18, SWA+19].

Significant ECG risk predictors of sudden cardiac death associated with drug-induced long QT syndrome

Pure hERG-blocking drugs and multi-channel blocking drugs (blocking the hERG potassium channel with calcium or sodium channels) are strongly associated with QT-interval prolongation [JVM+14, JG14, CMY08]. However, not all QT-prolonging drugs cause TdP [BBH+84, CMY08, Tis16, KGS+12]. There is a whole spectrum of pathological mechanisms (possibly sub-clinical genetic abnormalities in ventricular repolarisation) and drugs differing in their propensity to cause drug-induced TdP and sudden cardiac death [CMY08, KGS+12, DAF+10]. The prolongation of the QT-interval is thus a significant risk predictor of TdP, but it is not the sole or optimal determinant, and recent research has shown that some morphological ECG changes within the prolonged QT-interval can provide a more precise determination of the risk of TdP [DAF+10, TRR+07, SKQ+15]. Thus, identifying the significant ECG markers associated with drug-induced TdP arrhythmia attacks and sudden cardiac death may help to better assess the TdP risk of different QT-prolonging drugs [ESTB20, JG14, YC03, CMY08].

Several studies have linked specific T-wave morphological changes in patients with long QT syndrome to certain abnormalities in ventricular repolarisation that can lead to drug-induced TdP and sudden cardiac death [KFBW09, HES92, TRR+07, SKQ+15,

CHA$^+$15]. One of these ECG abnormalities is an increase in the T-wave duration, particularly from its peak to its end (known as the T^{peak}-T^{end} interval, or the right slope of the T-wave, as illustrated in Figure 2.10A) [SKQ$^+$15, CNC15, Ant19]. It reflects a significant reduction in potassium outward currents, specifically in the rapid I_{Kr} potassium currents during phase 3 of the cardiac action potential [Ant07]. Clinical research has shown this ECG marker in particular to be a significant risk predictor of TdP [CZSZ08, JVM$^+$14, KGS$^+$12, BSR$^+$16], and many sudden cardiac death incidences have been reported where there is a prolonged T^{peak}-T^{end} interval [CNC15, PRUE$^+$11, XHC$^+$19]. A notched T-wave (commonly caused by pure hERG-blocking QT-prolonging drugs [VJM$^+$15], and illustrated in Figure 2.10B) combined with a prolonged QT-interval is also shown to increase the risk of TdP [TRR$^+$07]. A cardiac modelling simulation study revealed that the origins of notched T waves in patients with long QT syndrome are due to a defect in the potassium and sodium channels in particular [SDPF$^+$14]. The first and second peaks of a notched T-wave are due to a decrease in the outward potassium currents, while the abnormal re-activation of the inward sodium currents causes a 'trough' between the two T-wave peaks [SDPF$^+$14].

In normal ECGs, a small wave known as the U-wave may appear after the T-wave, which may represent the Purkinje fibers (shown in Figure 2.2) repolarisation [Wat75], but this is not always observed because of its small size [DS14]. As such, another TdP risk predictor is when the T-wave duration increases and merges with a prominent U wave, forming either a partial or complete/giant T-U wave fusion (which can be a positive or negative deflection) [KNK$^+$10, NSK$^+$12, KFBW09], as illustrated in Figure 2.10C, D and E. This ECG abnormality is also known as the T-U complex [vEPvH$^+$07]. It reflects a very prolonged ventricular repolarisation duration, where the T-wave has a slow-moving deflection that fails to return to the ECG baseline, fused with an emerging prominent U-wave [CMY08, YA98, vEPvH$^+$07]. The giant, complete T-U wave fusion in particular is a critical marker on the ECG, as it has been shown to precede TdP episodes directly [KFBW09, NSK$^+$12]. Furthermore, ST-segment elevation—a sign of myocardial infarction (MI) that can be caused by structural heart disease or pharmacological drugs including chemotherapy [KBS$^+$13, TBM$^+$04]—is also a critical risk predictor for sudden cardiac death when associated with drug-induced QT prolongation [KE14, SWA$^+$19]. In addition to the T-U wave morphological changes, bradycardia or bradyarrhythmias—a slow heart rate that is less than 60 beats per minute—also increases the risk of drug-induced TdP and sudden cardiac death in patients with long QT syndrome [CNK$^+$15, Nam10, TRR$^+$07].

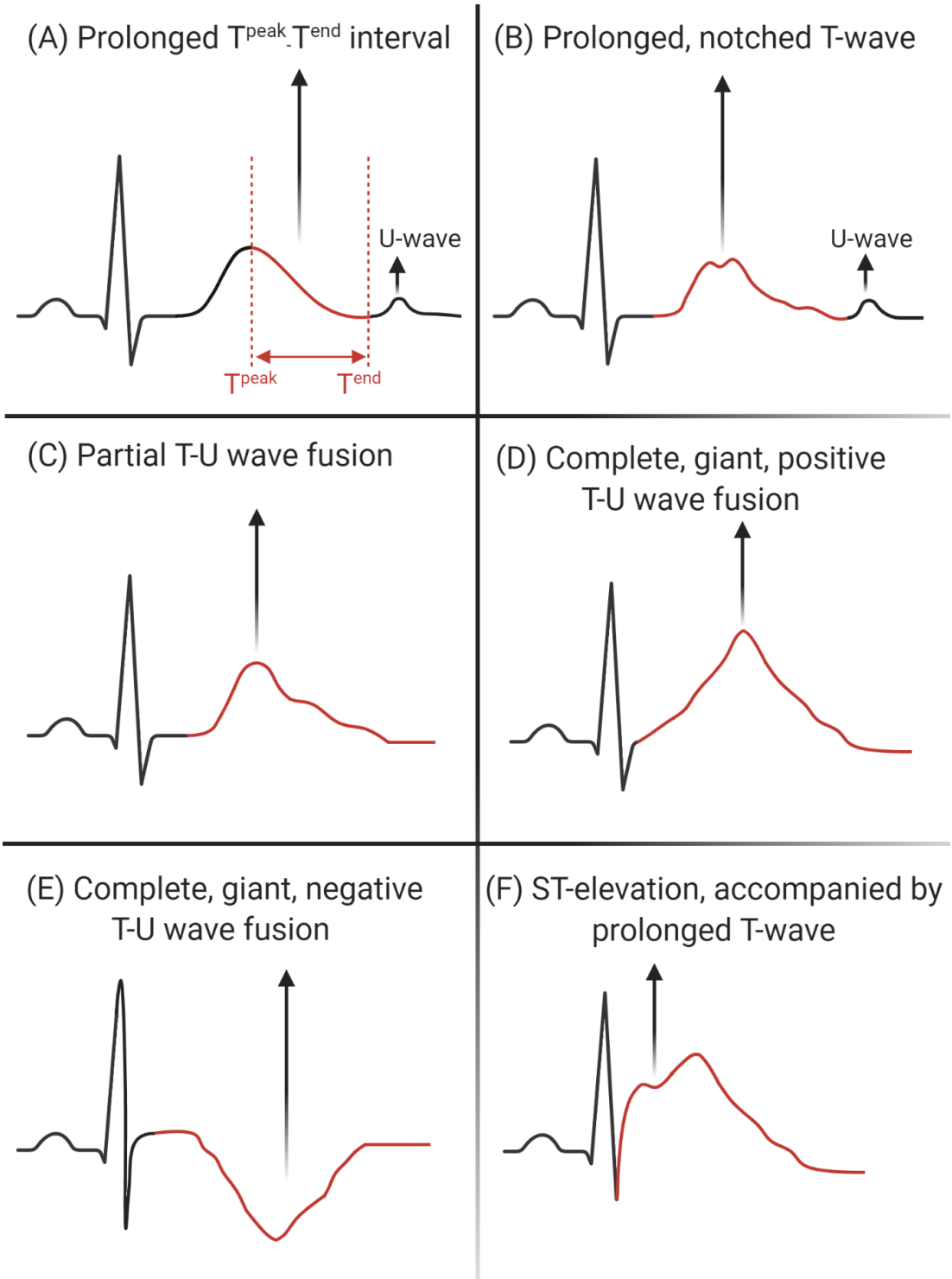


Figure 2.10: An illustration of the ventricular repolarisation morphological changes associated with drug-induced QT-prolongation, acting as significant ECG risk predictors of Torsades de Pointes (TdP) and sudden cardiac death.

2.2.4 ECG interpretation of long QT syndrome in clinical practice

The ECG is one of the most important diagnostic tools widely used in clinical practice to assess the heart function and detect cardiac pathologies [Kus20, DS14, Sch71]. It is cheap and easy to repeat, making it widely available outside of clinical cardiology areas including general wards, general practitioner (GP) surgeries, and other clinical settings [DS14]. In addition, modern wearable technologies for cardiac monitoring have made it possible to record and interpret an ECG outside the clinical environment [Pod17, CKA18, FBG+19]. This wide availability of ECG technologies has the potential to increase the rapid detection of drug-induced long QT syndrome and improve assessment of Torsades de Pointes (TdP) risk, preventing sudden, unexpected cardiac death. However, ECG interpretation is known to be complex, and long QT syndrome in particular is regarded as difficult to detect, even for clinicians who routinely read ECGs [VRS+05]. This section presents an overview of the current standard ECG interpretation method and discusses the primary human and machine challenges associated with interpreting long QT syndrome on the ECG in clinical practice, as reported in the literature.

Overview of the standard method

In hospitals, clinicians commonly interpret a short (10-second strip) ECG via 12 leads, where electrodes are placed in different positions on the human body to measure the direction of the heart's electrical activity from different views and angles [Kus20, DS14]. These electrodes detect the heart's electrical depolarisation and repolarisation changes on the patient's skin during each heartbeat [DS14]. The 12 leads can be divided into two categories: the chest leads (V1 to V6 leads where electrodes are placed on the chest) and the limb leads (electrodes are placed on the wrists and legs) including lead I, II, III, aVR, aVL and aVF [DS14]. These multiple leads provide the most comprehensive view of the heart, but useful ECG information can be obtained from a single lead 'view' (often used by wearable ECG technologies) [CKA18, FBG+19]. Clinically, lead II has conventionally been used to measure the QT-interval as the T-wave is often clearly shown in this lead [CMY08]. However, if the T-wave is not visible in the lead II, then lead V2 or V5 can be used instead [CMY08].

The standard method for visualising the ECG signal is with a line graph showing the voltage on the vertical Y-axis and time in milliseconds (ms) on the horizontal X-axis [DS14, BPW+38]. This graphical representation of the signal is recorded on pre-printed grid ECG paper consisting of small and large squares to support the reader in measuring the ECG waveform characteristics and intervals [Kus20, DS14]. Each small square is equal to 1 millimetre (mm) in height and width, representing 0.1 millivolts (mV) along the Y-axis and 40 milliseconds (ms) along the X-axis; five small squares represent a large square in a single dimension, which is thus equal to 0.5 mV along the Y-axis and 200 milliseconds (ms) along the X-axis [DS14]. Figure 2.11 shows an illustration of the 12-lead ECG and standard background grid.

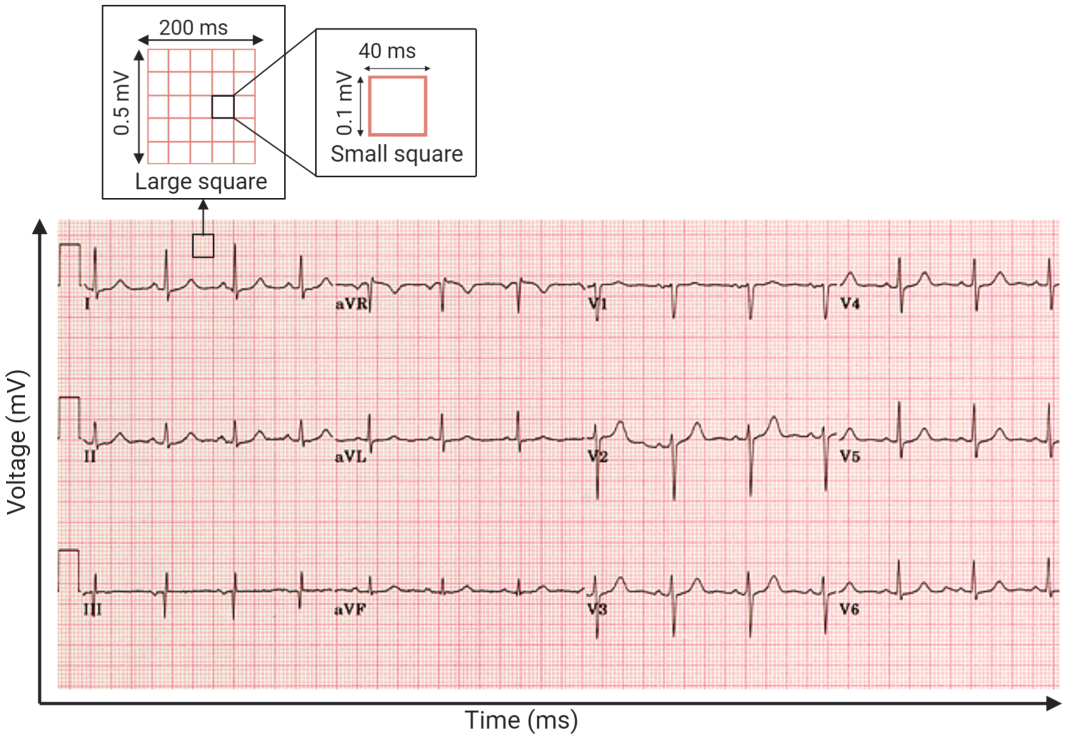


Figure 2.11: An illustration of the standard 12-lead ECG and background grid.

The ECG waveform

As mentioned in Section 2.2.1, the ECG signal consists of different waves labelled with letters which represent different stages of the heartbeat (Figure 2.3). This pattern is known as the standard ECG waveform, shown in Figure 2.12, in which each wave has specific 'normal' characteristics (*i.e.* amplitude and duration). The distance between

waves—including the PR-interval, ST-segment and QT-interval—also has a specific 'normal' range used as a clinical reference to identify an abnormal range in the ECG signal [Kus20, DS14, CAM+06]. Changes in the morphology or characteristics of the standard ECG waveform can indicate the presence of a cardiac pathology [DS14]. As such, the standard background grid supports the ECG interpreter in measuring the characteristics of these different waves and intervals [Kus20, DS14]. For example, an increase in the amplitude of the ST-segment by one small square in a limb lead, or two small squares in a chest lead, shows a condition known as ST-elevation, which indicates a sign of myocardial infarction (MI) that could lead to a heart attack [DB14].

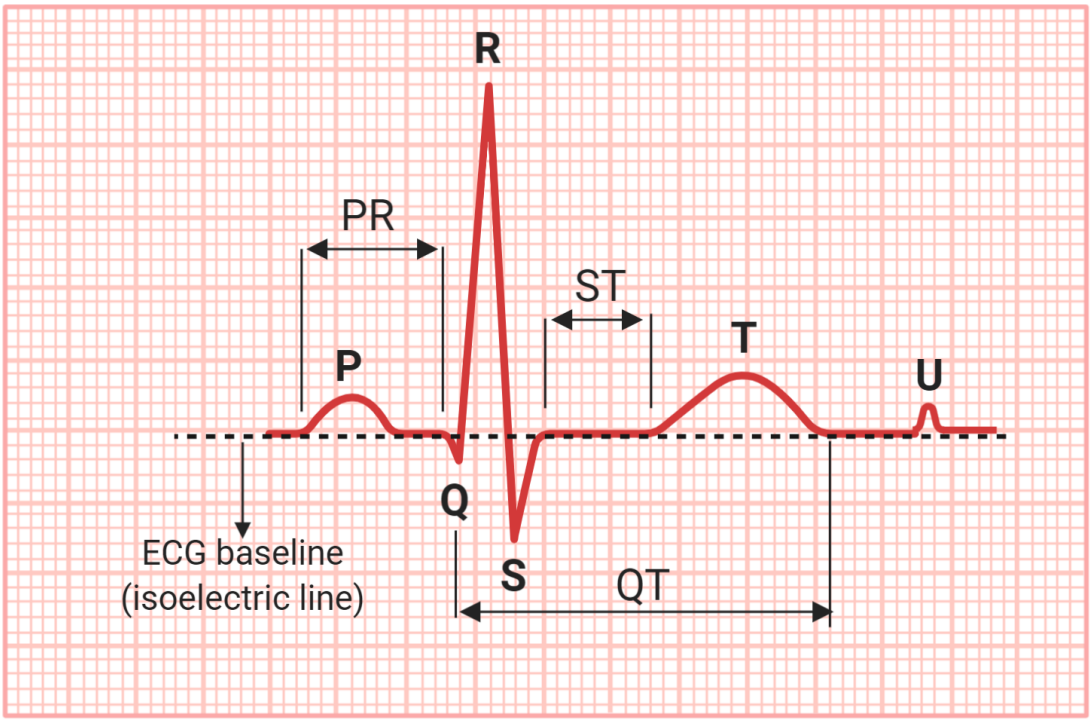


Figure 2.12: An illustration of the standard ECG waveform for a single heartbeat.

Measurement of the QT-interval

Clinical recommendations for standardizing and interpreting the ECG have indicated that the QT-interval should be measured manually by a clinician with experience in ECG interpretation [ITD+20, IP13, RSG09, CMY08, FDA+05], as current automated QT measurements are not considered reliable for use in the clinical assessment of the cardiac safety of a QT-prolonging drug [CMY08, SW17, TAS+15, RSG09]. The reasons for this are many, and are discussed in detail in Section 2.2.4.

In clinical practice, the traditional way to measure the QT-interval is to count the small squares (each representing 40ms) on the standard ECG background grid, from the beginning of the Q-wave to the end of the T-wave [Pri10, AAKR+02, GMZ06], as shown in Figure 2.12. Whether or not the QT-interval is considered to be prolonged depends on a number of factors, particularly heart rate, as the QT-interval varies inversely with heart rate [GMZ06, LMJM04, CMY08]. There are many ways to calculate heart rate from the ECG [Dub00, DS14]. One common method is to measure the RR-interval, the time elapsed between two successive R waves representing heartbeats on the ECG (Figure 2.4), and then apply a formula to calculate beats per minute (bpm) heart rate including:

$$HR = \frac{60}{RR-interval\,(seconds)} \tag{2.1}$$

$$HR = \frac{60,000}{RR-interval\,(milliseconds)} \tag{2.2}$$

Thus, it is common in clinical practice to correct the QT-interval to the heart rate (according to the RR-interval) using a QT correction (QTc) formula, and then (depending on the age and sex of the patient) use a 'cut off' value to identify QT-interval prolongation [CMY08, Dub00, DS14]. Many QTc formulas have been suggested in the past [CMY08], but the most widely used in clinical practice are Bazett's square-root formula and Fridericia's cube-root formula [Sha02, LMJM04, CMY08], which are described as follows:

$$Bazett\,(QTc) = \frac{QT}{RR-interval^{1/2}} \tag{2.3}$$

$$Fridericia\,(QTc) = \frac{QT}{RR-interval^{1/3}} \tag{2.4}$$

Table 2.1 shows the suggested clinical rating of QTc interval values (corrected with Bazett's formula) for normal, borderline, and prolonged QT-interval [GMZ06].

Table 2.1: Suggested QTc interval values (corrected with Bazett's formula) for diagnosing QT-prolongation [GMZ06].

	Adult men	Adult women	Children (1-15 year)
Normal	<430ms	<450ms	<440ms
Borderline	430 to 450ms	450 to 470ms	440 to 460ms
Prolonged	>450ms	>470ms	>460ms

However, measuring the QT-interval is challenging, and the relationship of the QT-interval to the RR-interval (QT dynamicity) is complicated, particularly when assessing the effect of drugs on the QT-interval duration [VRS$^+$05, STN08, CSDB$^+$09, DLD$^+$03, CMY08].

Challenges associated with human interpretation

A numerous number of studies have shown that delayed diagnosis of long QT syndrome is frequent, and patients with long QT syndrome are commonly misdiagnosed with other conditions (particularly epilepsy or seizure disorder) [MMC$^+$09, SHL19, KK09, OMK10, MBA14]. Khouzam et al. (2009) discuss the case of a 19-year-old woman with long QT syndrome who was misdiagnosed with a seizure disorder for eight years [KK09]. Another recent study showed similar errors in diagnosing LQTS of a 17-year-old boy [SHL19]. Empirical evidence also appears to confirm the notion that the majority of clinicians lack the skills to measure the QT-interval accurately and recognise long QT syndrome [VRS$^+$05, BY05, STN08], despite ongoing efforts to improve training in and recommendations for ECG interpretation and QT-interval measurement in clinical practice [OCT$^+$09, RSG09, RHA$^+$16, SPA19, ITD$^+$20]. It should be noted here that although a considerable body of research has focused on reporting long QT syndrome misdiagnosis [SHL19, KK09, MSC$^+$91, MMC$^+$09, OMK10, MBA14], or highlighting issues with QT-interval measurement in clinical practice [VRS$^+$05, STN08, CSDB$^+$09, DLD$^+$03, CMY08, GPAW14, LPL04, IP13, LS52], less attention has been paid to investigating why long QT syndrome is difficult to recognise from a cognitive-perceptual perspective. Only a few studies have investigated the perceptual-cognitive mechanisms underpinning ECG interpretation [Aug03, WBA$^+$14, BZF$^+$14, DBV$^+$16], and these have focused mainly on the interpretation of generally common ECG abnormalities such as myocardial infarction, with scant attention to long QT syndrome.

The critical issues and proposed solutions to address them (if any) presented in the literature regarding human recognition of long QT syndrome on the ECG are classified and discussed below.

1. **Difficulty with visual recognition of long QT syndrome**

   Some cardiac pathologies are reflected on the ECG by specific signal patterns, which often act as unique ECG signatures that can be observed with the naked eye [Kus20, DS14]. However, this is not the case with long QT syndrome, which is characterised primarily by a prolongation of the QT-interval that varies inversely with heart rate (HR); this makes it difficult to identify visually without actually measuring the interval and properly correcting it for heart rate [CMY08, Kus20, DS14]. Because of the QT/HR relationship, the faster the heart rate, the shorter the QT-interval, which is why recognising QT-prolongation is particularly difficult at fast heart rates exceeding the 'standard' of 60 bpm [Mal01, LPL04, LMJM04, MGH$^+$19]. From a perceptual-cognitive perspective, difficulties detecting the increased interval may be related to the fact that people are poor at perceiving quantity represented along a horizontal scale [LSZ$^+$09, WA84, PK11]. This issue is discussed in detail in section 2.3.2.

2. **Challenges associated with identifying the end of the T-wave**

   Accurate QT-interval measurement is currently viewed as crucial to recognising long QT syndrome in clinical practice [Kus20, ITD$^+$20]. There appears to be general agreement in the current literature that determining the beginning of the Q-wave (or the R-wave if there is no Q-wave) is relatively easy, but identifying the end of the T-wave is the most challenging aspect of measuring the QT-interval. [ITD$^+$20, CMY08, LS52, LPL04, SPA19, Sch20]. This problem is further complicated by morphological changes of the T-wave, the presence of a U-wave and/or ECG baseline noise [Mor01, GMZ06, CMY08, GPAW14]. The greatest challenge lies in the fact that the morphology of the T-wave itself can be very variable across individuals in general, and may become distorted and fail to end properly with drug-induced QT-prolongation in particular [MMDY94, GMZ06, CMY08, HST$^+$16].

As described in section 2.2.2, the T-wave represents the ventricular repolarisation on the ECG [Kus20, DS14]. Since pharmacological drugs that prolong the QT-interval often affect several cardiac ion channels, including potassium, sodium and calcium, during the ventricular repolarisation phase, this causes severe morphological changes in the T-wave [VJM$^+$15, Kha02, CMY08, ESTB20]. These changes include flat, biphasic and notched T-waves, and the slow-moving deflection of the T-wave end, which is the drug-induced T-wave change that is most problematic for QT measurement, particularly if it is accompanied by a prominent U-wave [VJM$^+$15, CMY08], as illustrated in Figure 2.10C. These morphological changes in the ECG may indicate an increased risk of developing TdP arrhythmia [DAF$^+$10, TRR$^+$07, SKQ$^+$15], as shown in section 2.2.3 (Figure 2.10), and they also obscure the end of the T-wave, thus affecting the accuracy of the QT-interval measurement [CMY08, LPL04]. This issue complicates both manual and automated QT-interval measurement, and the best method to use remains a subject of debate [CMY08, KOO$^+$11, SW17].

Several approaches have been proposed to identify the end of the T-wave under these circumstances [LS52, GMZ06, PDJVdBW08, CMY08, IP13, GPAW14, PKN$^+$09]. Among these the tangent and threshold methods are the most widely used in clinical practice [PDJVdBW08], but all methods have shown to be associated with potential inaccuracies [CMY08, Mor01]. In the threshold method, the end of the T-wave is the point where the descending limb of the T-wave meets the isoelectric baseline; in the tangent method, the end of the T-wave is the interception point of a tangent line drawn at the maximum downslope of the T-wave peak and the isoelectric baseline [PKN$^+$09]. Figure 2.13 illustrates these methods under normal and abnormal T-wave morphology conditions. Recent research has shown that the tangent method provides consistently shorter QT-interval results than the threshold method by up to 10 milliseconds [PKN$^+$09].

3. **Underestimation or overestimation of QT-prolongation caused by heart rate correction**

   As the length of a normal QT-interval varies inversely with heart rate, it is common in clinical practice to apply a QT correction formula (QTc), and then use a ‘cut off’ value to identify at-risk QT-prolongation. Examples of QTc formulae include Bazett’s, Fredericia’s, Framingham’s and Hodge’s [LMJM04]. However, recent research has shown that these correction formulae underestimate or

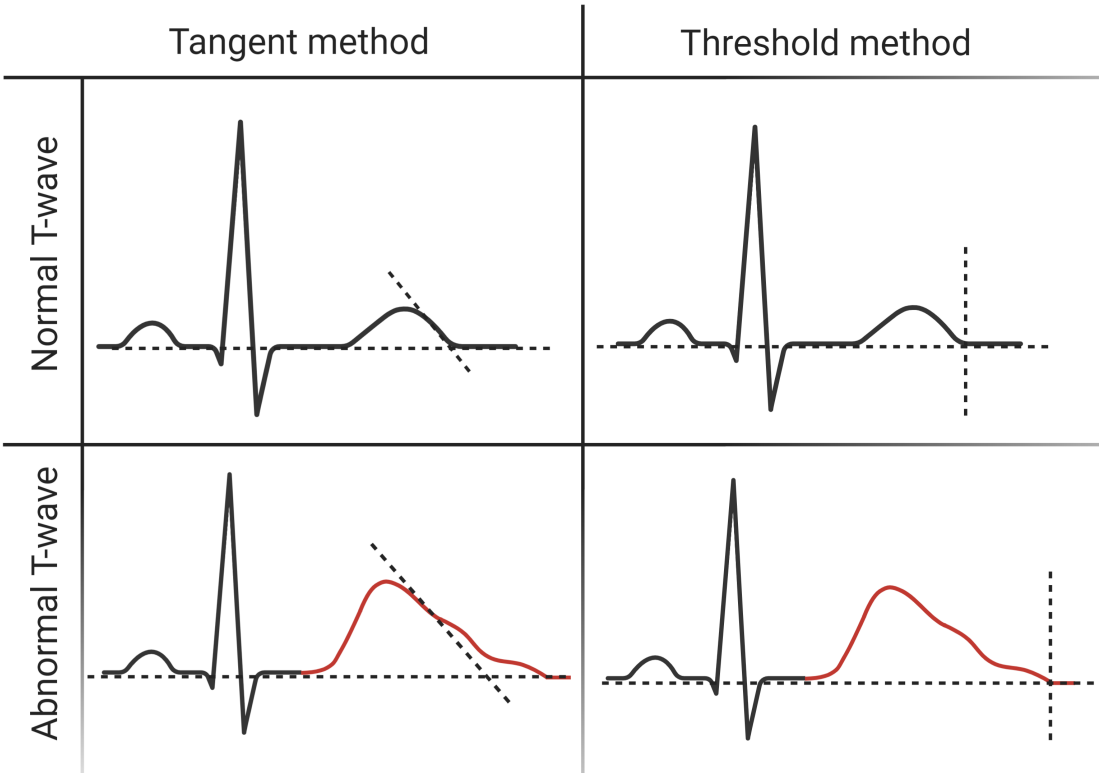


Figure 2.13: An illustration of the tangent and threshold methods commonly used in clinical practice to determine the end of the T-wave. In the tangent method, the end of the T-wave is the interception point of a tangent line drawn at the maximum downslope of the T-wave peak and isoelectric baseline. In the threshold method, the end of the T-wave is the point where the descending limb of the T-wave meets the isoelectric baseline.

overestimate QT-prolongation (depending on the formula used); and are thus inaccurate in identifying patients at risk of drug-induced TdP for fast and slow heart rates [DLD+03, LMJM04]. This issue has been addressed with empirical study of TdP risk, resulting in the 'QT nomogram', a risk assessment method designed specifically for identifying patients at risk of drug-induced TdP according to heart rate [FWM+05, CIKD07]. The nomogram plot can be seen in Figure 2.14.

It was developed after screening several ECG cases that reported drug-induced TdP and comparing these with control cases (*i.e.* when no TdP was reported). The actual QT-interval value (not the QTc) and the heart rate for each case were plotted as coordinates on a graph; this produced a line showing the upper bound of the QT-interval value at risk for TdP as a function of heart rate [CIKD07]. Evaluation of the QT nomogram demonstrated that it had higher sensitivity and specificity than widely accepted QTc formulas [WGG+10]. Therefore, we adopted the QT nomogram in our research as a reliable evidence-based TdP risk assessment method and used it as a basis for designing and evaluating our proposed solutions.

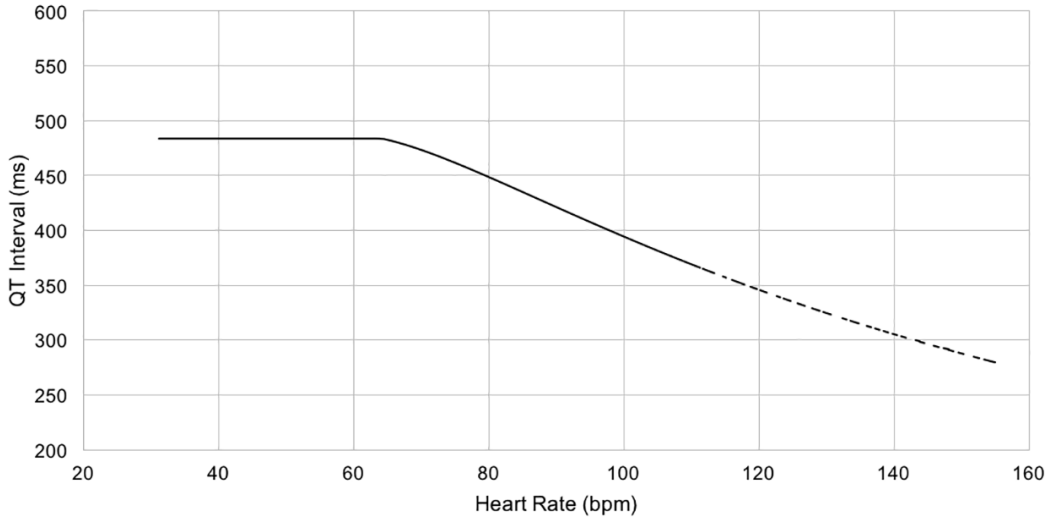


Figure 2.14: The QT-nomogram for identifying QT-prolongation at risk of TdP over heart rate. If the QT/HR value falls on or above the risk line, the patient is at risk of TdP; below the line the patient is not considered at risk of TdP [CIKD07].

Challenges associated with machine interpretation

Automated ECG interpretation methods were introduced in the 1950s to assist clinicians who had less training in ECG interpretation [Rau07]. These methods can identify a normal sinus rhythm with reasonable accuracy but are much poorer at reliably detecting abnormalities [SW17, EI13, RSG09]. Therefore, whilst most modern ECG machines in hospitals provide an automated measurement of the QT-interval, these automated QT values are usually correct only for a noise-free, normal sinus rhythm, in which the ECG waveforms (particularly the T-wave morphology) are well defined. Changes in the ECG waveform characteristics (*i.e.* the height and width of different waves), noise and artefacts in the signal, and changes in the T-wave morphology can easily invalidate automated QT-interval measurement [SW17, CMY08]. As long QT syndrome detection currently depends on accurate QT-interval measurement, automated QT measurement algorithms have proved unsatisfactory for detecting LQTS [KBD$^+$18, TAS$^+$15, EI13, GL13, TABW11, RSG09, CMY08, MCbA01]. Garg and Lehmann [GL13] found that even a widely used computerised ECG interpretation system was not able to detect QT-interval prolongation in 52.5% of patients affected. Research has also shown that drug-induced long QT syndrome in particular can be underestimated and under-reported by computerised methods in patients on Methadone, a drug that is infamous for prolonging the QT-interval and increasing the risk of TdP [TAS$^+$15].

At present, clinicians do not regard automated QT measurement algorithms as sufficiently reliable for use in the clinical assessment of the cardiac safety of a QT-prolonging drug [SW17, TAS$^+$15, RSG09, CMY08, TMH$^+$13]. Accordingly, clinical recommendations state that QT-interval should be measured manually by a clinician specialising in ECG interpretation [ITD$^+$20, IP13, RSG09, CMY08, FDA$^+$05].

Automation has made great strides forward in terms of interpreting many kinds of complex data, so one might expect similar progress to have been made with automated QT-interval interpretation. The issues that make automated recognition of long QT syndrome on the ECG particularly difficult, and the solutions proposed to address them (if any) in the current literature regarding machine interpretation are classified and discussed below.

1. **Challenges in bench-marking QT-interval measurement algorithms**

   The majority of automated QT-interval measurement algorithms are proprietary or unavailable, thus formally bench-marking the performance of different algorithms, and hence advancing published research in this area, is not possible at present [DBDM20, SW17, VDPL$^+$18, MSSS17, CMY08]. Kligfield (2010) pointed out that the field lacks a universally accepted definition of the end of the T-wave, and QT-interval measurement algorithms have become proprietary engineering solutions that vary substantially between core ECG laboratories and algorithm developers.

2. **Challenges in recognising different ECG waves**

   Automated QT-interval measurement algorithms are based on the precise determination of the onset and offset of the different waves and complexes (P-wave, QRS complex, T-wave) [SW17]. This process is relatively straightforward if the ECG signal has a normal sinus rhythm, but it quickly becomes challenging in the presence of anomalies, artefacts, or non-standard ECG waves [SW17]. A major issues is that ECG wave characteristics are known to differ substantially across individuals, and are affected by factors including age, ethnicity, sex and health status [GMZ06, MMDY94, HST$^+$16]. At present, there are no standard definitions for the ECG waves [Wil80, Par85, SW17, CMY08], thus differences in the automated QT measurement results between the algorithms persist [EI13, HG06, SW17, CMY08].

3. **Challenges in identifying the end of the T-wave**

   Ample evidence exists to demonstrate that precise identification of the T-wave end is frequently inaccurate when performed manually and particularly problematic if supported by a computer [DBDM20, SW17, CMY08, TAS+15, TABW11]. Based on the two primary approaches used to identify the end of the T-wave in clinical practice (discussed previously in section 2.2.4) automated measurement algorithms in the literature fall into two categories: threshold-based methods and slope-based methods [PKN+09, MCM96, MCM95b, CMY08].

   Threshold-based methods identify the end of the T-wave based on threshold levels, defined as a fraction ranging from 5% to 15% of the amplitude or differential of the T-wave [Hun05, CMY08]. There are threshold levels because whilst it is relatively easy for humans to perceive when the signal returns to the isoelectric line (indicating the T-wave end based on the threshold method), it is actually challenging to detect this automatically [CMY08]. This is because the ECG isoelectric line is not constant in terms of voltage (i.e. the baseline can have any random low voltage value approaching zero), and as such the end of the T-wave by the automated threshold methods needs to be estimated as the time point of proportionate signal amplitude to the T-wave amplitude [Hun05, CMY08].

   In slope-based methods, the end of the T-wave is defined as the maximum T-wave slope's interception with the isoelectric line [HVB+17, Hun05, CMY08]. As the steepest part of the descending portion of the T-wave is affected by changes to morphology, various methods have been proposed to identify the maximum T-wave slope, including the tangent method, peak slope method, and least-square fitting method [MCM96, MCM95b, Hun05, CMY08, WLX+16]. Generally, slope-based methods provide a better estimation of the QT-interval than threshold-based methods, as the results of the latter vary according to the threshold level and the T-wave morphology [HVB+17, XR98, CMY08]. However, the accuracy of slope-based methods can be significantly reduced when the T-wave has a slow-moving deflection that fails to return to the ECG baseline, making the steepest slope hard to define [CMY08]. In addition, these methods depend on accurate identification of the isoelectric line, which is known to be influenced by noise and ECG baseline wander (a type of artifact) [SW17, CMY08]. Bizarre automated QT measurements are also sometimes reported, even for normal, noise-free ECGs [CMY08].

4. **Abstracting the QT-interval numerically risks masking other morphological ECG predictors of TdP risk**

   Assessing the risk of drug-induced TdP also benefits from consideration of the ventricular repolarisation morphology more generally [TRR+07, HES92]. As such, abstracting the QT-interval numerically risks masking other potentially clinically significant changes in ventricular repolarisation. As shown in Section 2.2.3 and Figure 2.10, partial/complete T-U wave fusion, where the T and U waves cannot be distinguished, are significant risk predictors of TdP [KNK+10, KFBW09]. In most cases, this is shown to be, in fact, a prolonged biphasic, or notched T-wave [YA98, VJM+15]. For certain T-U morphologies commonly reported in drug-induced long QT syndrome, measuring the QT-interval as a single value may lack diagnostic validity, and the presence of prolonged fused T-U waves should be viewed as a more accurate predictor of TdP [CMY08, HES92]. As such the T-wave morphology continues to provide the richest information for recognising LQTS [EI13, KBD+18, GL13, MCbA01, RSG09, TAS+15, SW17]. Recent research has therefore investigated how to apply computational analysis to identify T-wave morphology changes when assessing the effects of drugs on ventricular repolarisation [VJM+15, SNK+17]. These approaches are discussed further in Chapter 3.

Consequently, in the last few years, researchers have become increasingly interested in exploring new computer-based approaches to improve LQTS detection and/or risk assessment without relying solely on QT-interval measurement, which is discussed in more detail in Chapter 3.

2.3 Perception and pre-attentive processing in human vision

2.3.1 Overview

Perception, broadly speaking, is the process of recognising and interpreting sensory information [HW97]. It is generally agreed that human visual perception involves complex parallel processing mechanisms, which have been studied across different scientific disciplines including neuroscience, cognitive science and vision science [HE85, War21, RFTT02]. Comprehensive theoretical foundations of human visual perception can be found in [SW12, WFG+12, YA82, HE85].

In his 1979 study on the affordance theory, the great perception theorist Gibson J.J. argues that humans perceive possibilities for action in order to operate in the environment [Gib79]. This theory is attractive from a data visualisation perspective, as the goal of most data visualisation is decision-making [Kir16, DCKH+19, War21], and as such considering visual perception as a process designed for action can support acknowledging the fact that it has become an increasingly important part of designing and evaluating data visualisations. For example, translating the affordance theory of visual perception into the visualisation design, we might construct a science-of-perception-based design principle stating that a good visualisation has affordances that make the users' task easy by intuitively showing the actions they can take. Great scholars who incorporate visual perception knowledge with visualisation design include Colin Ware, who explored in his book *'Information Visualization: Perception for Design'* [War21] the role of the science of perception and vision in establishing key design principles for improving clarity, utility and persuasiveness of data visualisations.

Here we present an overview of visual perception and pre-attentive processing theory in a human vision, considering ECG interpretation of LQTS and visualisation design as a theoretical lens. Methodological advances proposed in the current literature for visualising the ECG signal data are discussed in Chapter 5, Section 5.2.2.

2.3.2 Perceiving quantity represented along a horizontal scale

Several studies in human perception that have attempted to correlate the results of behavioural science studies with neuroscience brain studies have found that whilst the adult human brain can easily perceive quantity along a vertical scale, it is poor at judging it along a horizontal scale [PK11, Lib91, WA84]. This may explain why clinicians struggle to accurately perceive QT-interval increases along the horizontal time scale of the standard ECG visualisation [VRS$^+$05, STN08]. A natural question arising here is how we can change the horizontal length visual encoding required for LQTS detection to one that is more intuitive and easier to perceive.

Prior work seeking to establish a scientific foundation for graphical data visualisation showed that types of visual encoding involved in perceiving difference in quantity include: (1) position along a common scale, (2) position along non-aligned scales, (3) length, (4) area, (5) volume and curvature, and (6) shading and colour [CM84]. According to cognitive fit theory—which states that the correspondence between visual encoding type and task type results in superior task performance for humans [VG91, Ves91]—determining how to improve LQTS detection via visual encoding requires first defining the type of visual task underlying LQTS detection.

> *Cognitive fit theory:*
> *"Internal perceptual representation of the problem + External, matched visual representation of the problem → Superior task performance." [VG91, Ves91].*

Clinical research has shown that small increases in the QT-interval can be clinically significant but are difficult to perceive visually without precise measurement [CMY08, VRS$^+$05, STN08]. As such, a science-of-perception-based approach may be more appropriate to optimise the visual detection of small differences in the QT-interval duration (*i.e.* few milliseconds), that exist in a large quantity scale within a single heartbeat ($\approx$ 400 to 600 milliseconds at 60 bpm for example). Thus, this type of visual task requires encoding the differences in duration within the ECG signal data in a manner that can be perceptually separated over a large horizontal scale.

2.3.3 Perceiving quantity represented with Cartesian *vs.* Polar coordinates

In geometry, a coordinate system is a graphical representation method that uses two or more numerical dimensions to identify data points in a given space, such as Euclidean space [GPS04]. Several types of coordinate systems have been used to represent quantitative data including Cartesian and Polar coordinates—the most common coordinates for visualising time-series data representing changes in quantity over time [Kir16, AJB16]. Figure 2.15 shows an illustrative example of representing different time series data with Cartesian and Polar coordinates using a line graph and histogram.

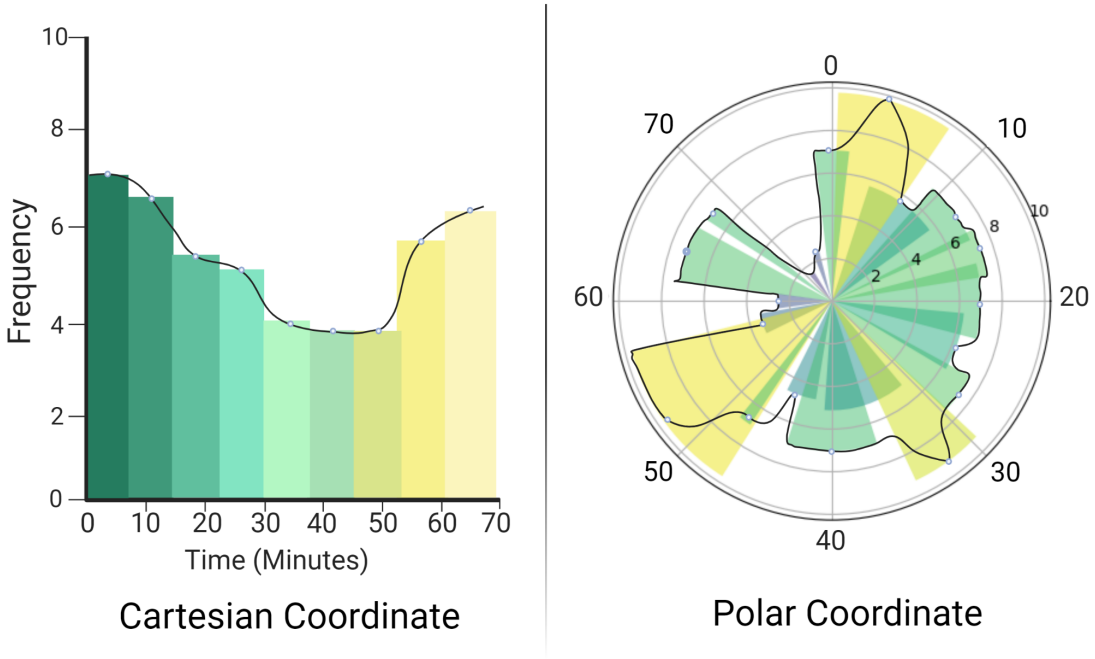


Figure 2.15: An illustrative example of representing different time series data with Cartesian and Polar coordinates using a line graph and histogram.

An ECG signal can be described as a variable time series that presents changes in the waveform over time [TLT+20]. The Cartesian coordinate system is the standard means of presenting the ECG signal in clinical practice [DS14, BPW+38], and to date, no study has specifically investigated the effect of different coordinate systems on ECG interpretation of LQTS.

A number of studies have evaluated the effect of the coordinate system on the perception of changes in quantity; some focused on comparing Cartesian with Polar coordinates when representing nominal data (e.g. comparing bar charts with their polar counterparts, pie charts) [SP08, CM86]; some focused on time-series data in particular [AJB16, FFM+13], while others focused on the technical aspects of developing Cartesian *vs.* Polar visualisation tools rather than evaluating their impact on visual perception [BBBD08, SCGM00].

Given that the human visual system is very sensitive to whether positions of visual elements are in their exact location [GPBC10, SCGM00], one study evaluated the effect of Cartesian vs. Polar coordinates on human visual recognition memory [DBB10]. The results showed that people can memorise and relocate positions of visual elements slightly better on Cartesian coordinates than Polar coordinates, with accuracy of 60% and 55% for Cartesian and Polar coordinates respectively [DBB10].

Adnan et al. [AJB16] examined perception of quantity in time-series visualisations. They showed Cartesian coordinates to be most effective for detecting trends and identifying maximum and minimum values when used with positional and colour visual encodings, and Polar coordinates to be most effective for finding minimum values when using area visual encoding.

The circular layout used in the Polar coordinate system has also been employed to expose changes in data collected over a long period of time. Page et al. [PSCA15, PAS+16] proposed an 'ECG Clock' generator, to visualise the changes in QT-interval values automatically generated by a 24-hour Holter ECG monitor. Circular layouts have been also used to help with detection of symmetrical patterns in data [HGM+97] and to measure symmetry in graphs [WK17].

2.3.4 Visual perception in light of ECG interpretation

One of the challenges associated with ECG interpretation is that it visualises the raw complex signal of the heart electrical activity directly, which can be dynamic (where different cardiac pathologies interfere) [CMY08], and highly variable across individuals due to many factors—including age, gender and specific ethnic genes [GMZ06, MMDY94, HST+16]—without any pre-processing of that visualised signal. This complicates the signal representation in general and the perception of a difference or abnormal change within the signal in particular. The human brain has the ability to filter the accumulated, vast sensory information from the surrounding complex environment and process what is important [War21]. This appears to be because of the massively parallel visual processes whereby the brain segments the world into regions and finds links and patterns within them (bottom-up processing), which is highly influenced by what we are looking for (top-down processing) [War21, TOB10]. ECG interpretation is thought to be dependent primarily on top-down processing, which uses contextual medical knowledge to aid the perception of signal patterns [WBA+14, SW17]. From a human perception perspective, re-visualising the ECG signal by harnessing bottom-up processing has the potential to draw visual attention to the critical information contained within the ECG signal without prior medical knowledge and extensive training.

The phenomenon of bottom-up processing first noted in the Gestalt principles of visual perception, which articulate factors that regulate perceptual grouping, including proximity, similarity, closure, continuation and symmetry [Wer23, Sta19, War21]. These primitive perceptual organisation processes derive relevant groupings and patterns from an image or scene without prior knowledge of its contents [Low12, vdH17, Sta19]. In addition, a more recent principle of bottom-up processing is the principle of simplicity, which states that people tend to perceive the simplest possible interpretation of any given visualised information [Fel16]. It is generally agreed that prior low-level perceptual mechanisms relating to ease of search is first performed, where the visual image is broken into pre-attentive elements of colour, form, and motion [War21, WU19]. Zavagno et al. (2014) have shown that colour to be a relatively strong grouping factor that functions according to the principles of Gestalt theory, and can override other types of pre-attentive elements including form (shape and size) [ZD14]. The research within this thesis focus on this fundamental theory in a human vision known as pre-attentive processing, which outlines a set of visual attributes such as colour (hue and intensity) known to be detected rapidly and accurately by the

human eye [Not93]. Introducing pre-attentive attributes to complex visual information can increase stimulus salience and modulate visual processing speed in human perception [War21, WU19, HBE96, HBE95, Not93, Tre85]. The process of using pre-attentive attributes to visualise data is also known as sensory representation processing, as these attributes are effective in stimulating early stages of the neural processing of the human visual sensory system [War21]. That is, the pre-attentive attributes derive their expressive power from their ability to use the perceptual organisation processing power of the brain without prior learning, which can then move from pre-attentive to attentive processing when influenced by an individual's intentions and visual task goals [The94, The95, EY97, AN09, TZGM10, TOB10, War21]. The pre-attentive processing theory is discussed in more detail in the section below.

2.3.5 Pre-attentive processing and ease of visual search

Human vision and neuroscience studies that use direct psychophysical experiments—studying human perceptual responses to physical stimuli by varying the properties of a stimulus along one or more physical dimensions [Ste17]—-have established a theory of pre-attentive processing that is fundamental to how we understand visual distinctiveness [Tre85]. This theory states that a particular set of visual properties are detected rapidly and accurately by the human eye [War21, WU19, The13, Not93]. For instance, to count the four digits shown in Figure 2.16A, we need to scan all numbers sequentially, whereas in Figure 2.16B we only need to scan the red numbers. This is because colour is pre-attentively processed [War21].

The theoretical perceptual mechanism underlying the 'pop-out' effect of colour in separating the target pattern (number 4) from their surroundings was termed pre-attentive processing, as early researchers believed that it must occur prior to conscious attention [Tre85, Not93, HBE96]. Other examples of pre-attentive properties include orientation, shape, size and spatial positioning, as illustrated in Figure 2.17.

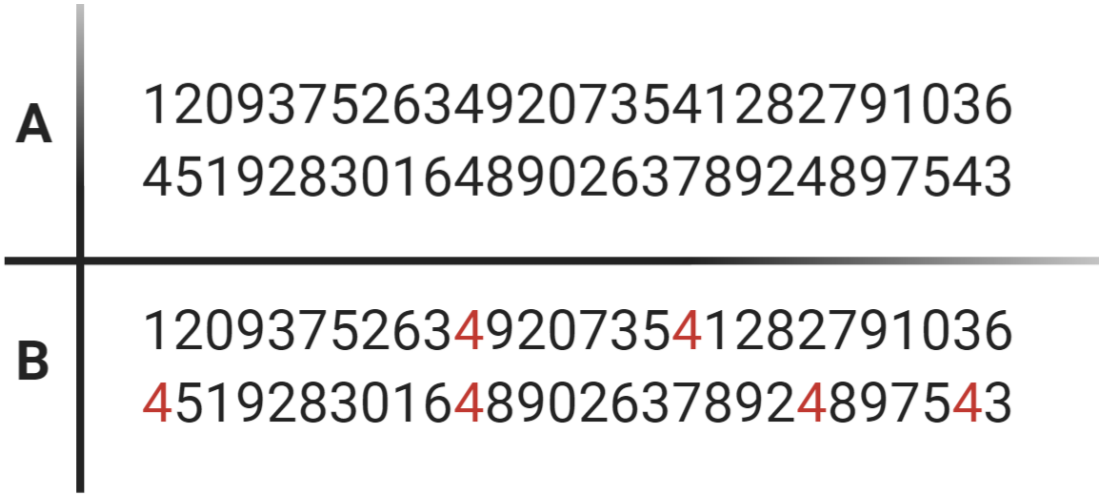


Figure 2.16: An example of a pre-attentive processing task. (A) To count the four digits, we need to scan all numbers sequentially. (B) To count the four digits, we only need to scan the red numbers, as they are pre-attentively processed due to the colour pop-out effect.

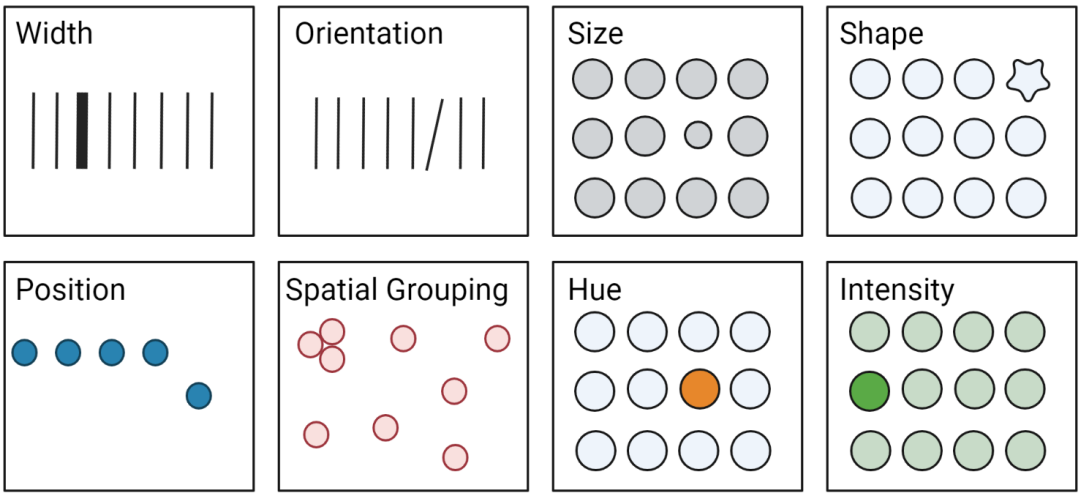


Figure 2.17: Examples of pre-attentive properties.

Empirical evidence appears to confirm the notion that a complex visual image is deconstructed via a primitive perceptual grouping process into pre-attentive elements of colour, form, and motion, and then a 'searchlight' model of visual attention is applied using eye movements to sweep for target patterns [War21]. It is generally agreed that the visual search process requires obtaining information in bursts, *i.e.* a snapshot of a stimulus per fixation [War21]. As such, a more complex image that contains non-preattentive elements is scanned as a thread, one by one, at a rate of approximately 40 visible elements per second [War21]. This means that we can usually perceive between three to six visible elements before the eye moves to another fixation [War21]. Colour is a useful pre-attentive attribute that saves the observer from carrying out a continuous jump in eye movements during visual search, thus helping to focus fixations and increase the perceptual efficiency of visual elements [War21, GCC17, Hea96, PGLS95].

2.3.6 Colour perception: an overview of basic theories

Trichromacy theory

A number of theories have been proposed to explain how humans and some other mammals have evolved to perceive colours. One of the earliest and most well-known is trichromatic theory [Bow83, Mal11, War21]. This states that there are three distinct colour receptors in our retinas (known as 'cones'), which are sensitive to different wavelengths of light in the visible spectrum [Bow83]. One receptor is sensitive to blue (short-wavelength cone sensitivity), the other to green (medium-wavelength cone sensitivity), and the third to red (long-wavelength cone sensitivity) [Mal11, War21]. The combinations of these three colours produce all the colours we can perceive [Mal11, War21].

Figure 2.18 shows how different wavelengths of light are absorbed by the three different receptors. The short-wavelength cone sensitivity peaks at 450 nanometers, while the medium and long wavelength cones overlap considerably and have a higher sensitivity that peaks at 540 and 580 nanometers respectively [Bow83, Mal11, War21]. The much lower sensitivity of the short-wavelength cone is why it is not recommended to present text or other detailed information in pure blue on a black background [War21].

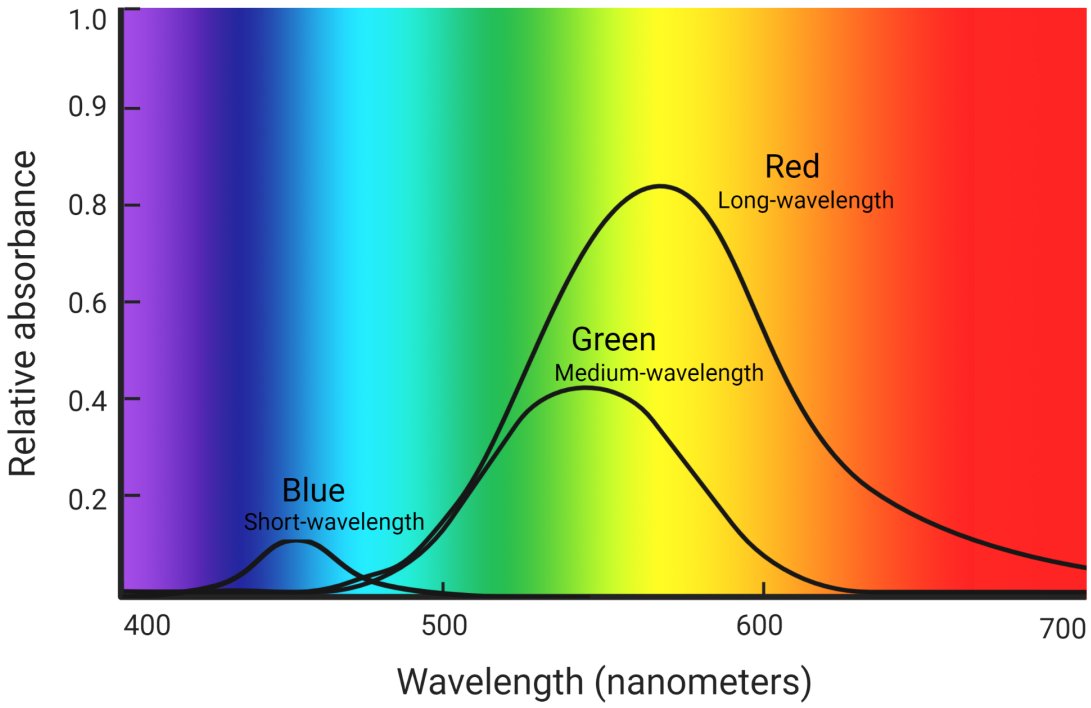


Figure 2.18: Sensitivity of human vision cones.

Colour blindness

Colour blindness is caused by the lack of either the short wavelength-sensitive cones (tritanopia, blue-blind), the medium-wavelength-sensitive cones (deuteranopia, green-blind), or long-wavelength-sensitive cones (protanopia, red-blind) [War21, BELL17]. Tritanopia results in the inability to distinguish between blue and yellow, while protanopia and deuteranopia, which are more common, result in the inability to distinguish between red and green [War21, BELL17]. Since colour-blind people can perceive colours with a mixture of no more than two pure cones sensitive to the spectral wavelength, their condition is called 'dichromacy' instead of 'trichromacy'. The prevalence of colour blindness is around 10% in males, while in females it is only 1% [War21].

Green is the most difficult colour to perceive by the majority of colour-blind individuals [HTWW07, War21]. A mixture of blue and orange/red variations, however, can be distinguished by most colour-blind people; where tritanopia readers will perceive it as teal and pink, deuteranopia readers will perceive it as blue and brown and protanopia readers will perceive it as blue and olive [HTWW07], as illustrated in Figure 2.19.

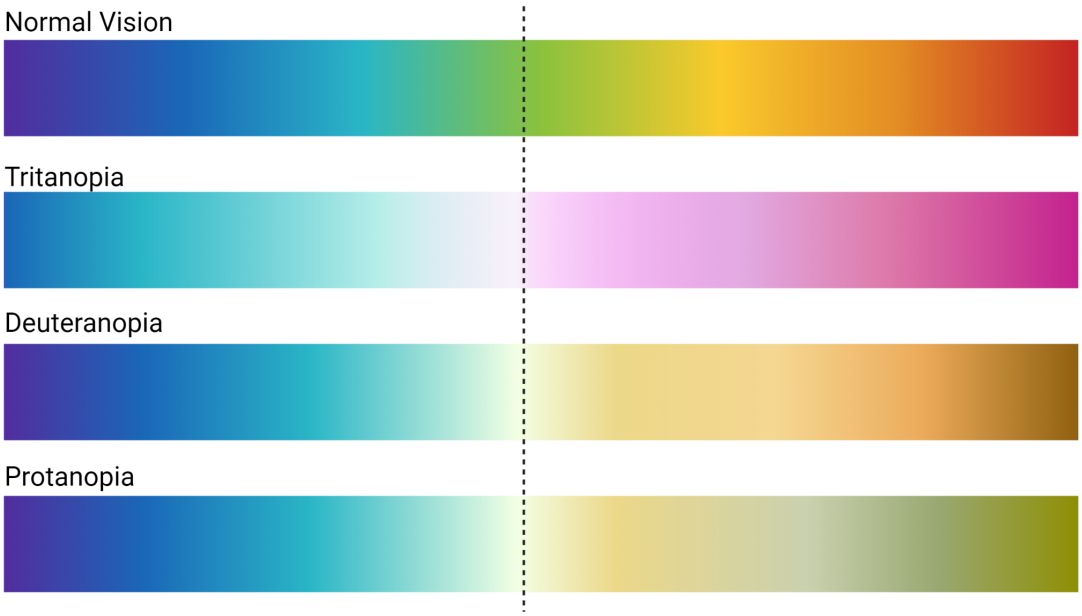


Figure 2.19: Perception of a spectrum-approximation colour sequence with normal vision *vs.* colour blindness.

Any sequence of colours, apart from the green-to-red variations, are distinguishable by most colour-blind people, including blue-to-yellow, blue-to-green, blue-to-red and black-to-white variations [HTWW07]. Another solution is to vary colours by lightness (adding white dimension) or contrast (adding black dimension), and the effect can be tested on a gray-scale printout of the colours [War21]. As a rule of thumb, if colours are distinguishable on a gray-scale variations, then they are more likely to be perceived by colour-blind individuals [War21].

Computer simulation models are used to transform colour vectors of digital images from normal vision spectrum into the spectrum of colour-blindness, based on previous physiological and psychological studies of colour vision [BVM97, MOF09]. This can help to quantify and understand how colour-blind individuals will perceive certain images [BVM97, MOF09]. Other researchers utilise similar simulation methods to improve information presentation for colour-blind people, for example by automatically adjusting colours according to the simulation results [KOF08].

The role of colour in attracting attention and increasing visual salience

The attention process is centred around the fovea—a small depression in the retina of the eye that is around 1.5mm in diameter corresponding to 5° of the visual field—where vision is most acute [JBCW91, Hen92, FG01]. The 'searchlight' beam of eye movements can be expanded and contracted based on the visual task type, the observer's stress level and the size of the search area [War21]. For instance, if the eye is fixating a small search area, then the 'searchlight' beam is the size of the fovea, but if it is fixating a larger search area, then the 'searchlight' beam expands. This concept is known as the useful field of view (UFOV), which defines the visual search area through which information can be perceived in a brief glance [WO14, War21]. The UFOV is considered particularly useful when presenting visual information for observers working under extreme stress, who may experience a phenomenon known as 'tunnel vision' [Dir83, War21]. Prior research shows that this phenomenon usually occurs during a high cognitive load where the UFOV is narrowed; thus, only the most prominent information is perceived, which is usually at the centre of the field of view [Wil85, War21].

Colour, in particular, is known to aid and influence the attention process in the fovea, as well as to some extent redirect the attention to the coloured objects that are away from the fovea (*i.e.* within peripheral vision), thus expanding the UFOV [Põd07, FG01, War21]. Figure 2.20 shows an illustration of how colour hue can help attract attention within the UFOV and separate visual elements (represented by red circles here) from their surroundings, thus increasing their salience so that they can be perceived in foveal vision.

Most of the research on colour perception appears to support the notion that colour can proceed from pre-attentive to attentive processing when influenced by an individual's intentions and visual task goals [The94, The95, EY97, AN09, TZGM10, TOB10, War21]—notably the change detection task [HCLF12, War21]. This usually occurs interchangeably between bottom-up processing, which is driven by the pre-attentive attribute known as 'stimulus-driven capture' [TOB10]; and top-down processing, where the brain directs an individual's attention towards the target pattern that fits in with their task goal, a concept known as 'contingent capture' [TOB10, TZGM10].

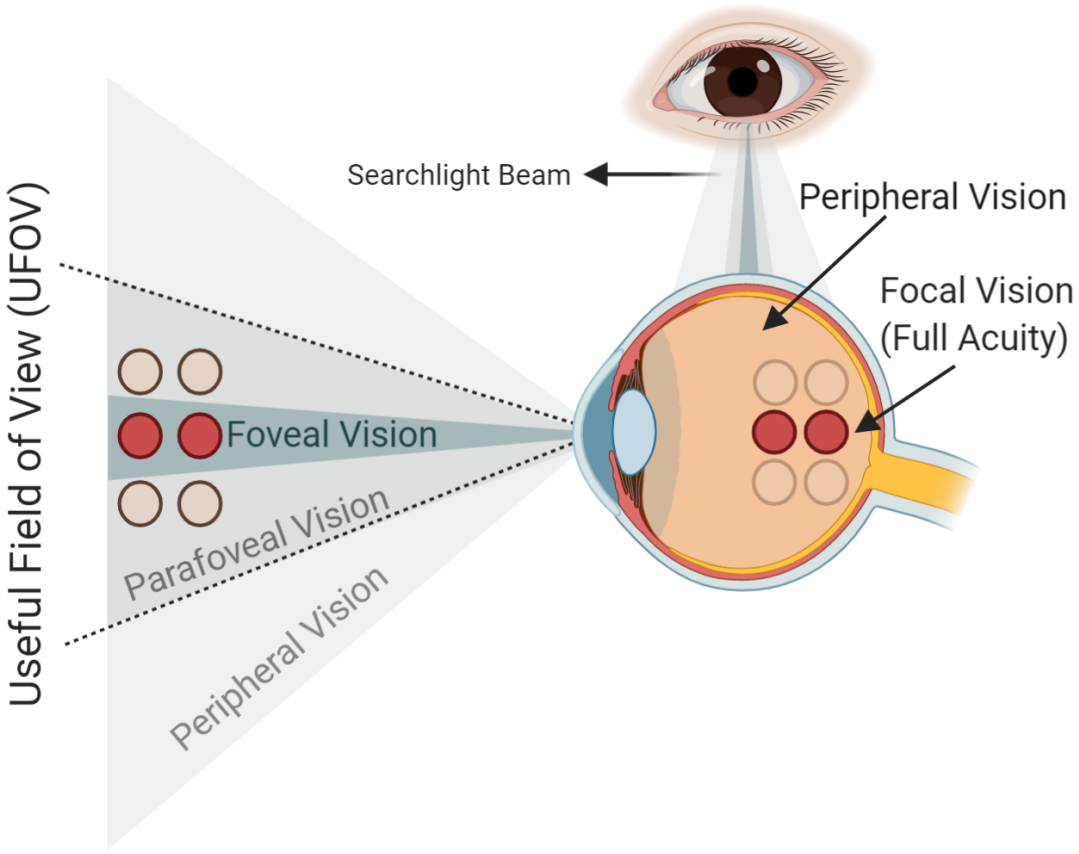


Figure 2.20: An illustration of how colour hue can help attract attention within the useful field of view (UFOV) and separate visual elements (represented by red circles here) from their surroundings, thus increasing their salience so that they can be perceived in foveal vision.

As such, we assume that colour can be used as an additional visual encoding to the horizontal length to detect the change in the QT-interval. However, the question that arises here is how to apply colour to large, continuously varying values along the ECG horizontal scale. A useful technique is pseudo-colouring, which is commonly used to represent continuously varying values using a gradient sequence of colours [War12].

2.3.7 Applications of pseudo-colouring

Colour vision exploits the highest bandwidth channel between the physical stimulus and the human brain, as we acquire more information through vision than through all of the other senses combined [War12]. Pseudo-colouring, also known as colour-mapping, is the process of translating data values into colours to generate easily perceptible patterns and visual elements within the data [ZH15, War88, War21]. It is widely used in scientific visualisations including astronomical radiation charts and medical imaging applications, as well as in many information visualisations such as geographic and time series visualisations, where applications include encoding elevation in the data (*e.g.* forecast temperatures) or showing changes over time [War21, SMY$^+$05]. Pseudo-colouring is also used in other computer science domains, including computer vision and image processing, to support machine interpretation and classification of images [GK13, RVL$^+$07, KMS00].

The application of pseudo-colouring requires not only knowledge of the type and nature of the data, but also an understanding of the task and the application domain [ZH15, War88, War21]. When applied appropiately, pseudo-colour can reveal insights from the data that may be difficult to obtain without colour [ZH15]. A good example is the application of pseudo-colouring in medicine, where it is widely used to support clinical diagnosis, for example of breast diseases [ZSST11], and to highlight details in organs and bones structures that would otherwise be difficult to perceive [Quw09, Sem18]. Figure 2.21 shows an example of using a spectrum-approximation pseudo-colour sequence to support an inexperienced clinician in perceiving different tissue types in the knee on an MRI scan result. We hypothesise that applying pseudo-colouring to the ECG image has the potential to expose QT-interval duration changes, thus supporting LQTS detection.

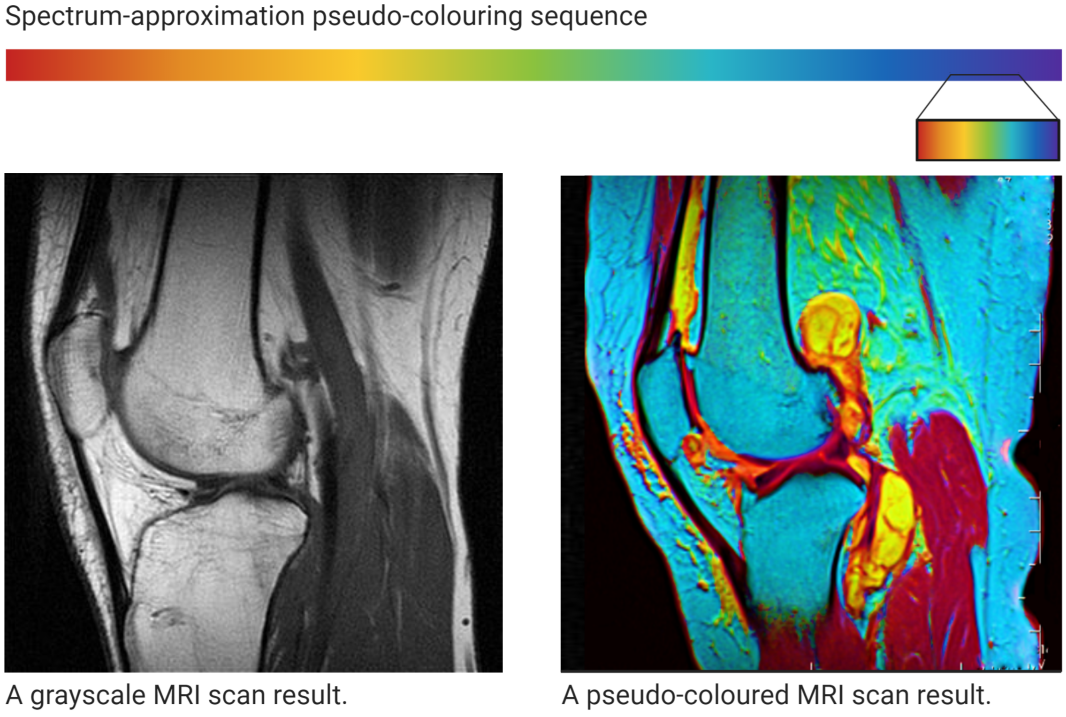


Figure 2.21: An example of using the pseudo-colouring technique to support the distinction between different tissue types of a knee on an MRI scan result.

Perceptual pseudo-colour spaces

Several studies have investigated what makes a pseudo-colouring sequence effective, particularly when presented on a computer screen. Early research developed different colour spaces, where each colour space represents an arrangement of colours in a three-dimensional space, and evaluated their effect on human perception [Sto16, Wri67]. The basic colour space is the RGB space, which utilises the combination of the three human vision cone channels (red, green and blue) to represent any colour [Sto16]. As this space is limited to the primary RGB light wavelengths, it is unable to reproduce colours from other lights such as yellow, blue, and purple [War21]. Cylindrical colour spaces including hue, saturation and value (HSV) and hue, saturation and lightness (HSL) have been developed to improve the perception of the basic RGB colour space [ZH15]. The 'Commission Internationale de l'Eclairage' (CIE) colour space uses a set of abstract human vision cones channels (labeled as as XYZ) to establish quantitative links between wavelengths of colour light in the visible spectrum and physiologically perceived colours in human vision [SG31]. More details on how this system works can be found in [War21]. Many studies have shown that uniform colour spaces (first proposed by Munsell [Mun15] to specify colours based on hue, intensity and lightness) can be used to determine the degree of perceived difference between similar colour sequences (*i.e.* having similar wavelength frequencies on the electromagnetic spectrum) [HLCL12, ZH15, War21].

Task-driven pseudo-colouring

A number of studies have explored pseudo-colour representation methods for specific tasks. For example, one study proposed using the 'task-colour-cube' method, which presents different pseudo-colouring scales (*e.g.* unsegmented and segmented colour sequences) to support different tasks including comparison, localization, and identification of data values [TFS08]. Figure 2.22 shows an example of this method.

Other research has investigated the effect of colour *vs.* no-colour on the perceptual task of reading values and forms, comparing several pseudo-colour sequences with greyscale [War88]. This work has provided evidence-based design guidelines for information presentation and shown the spectrum-approximation pseudo-colour sequence to be particularly effective in conveying form and value information, and the greyscale sequence to be partially effective in conveying form (as shown in Figure 2.23) [War88].

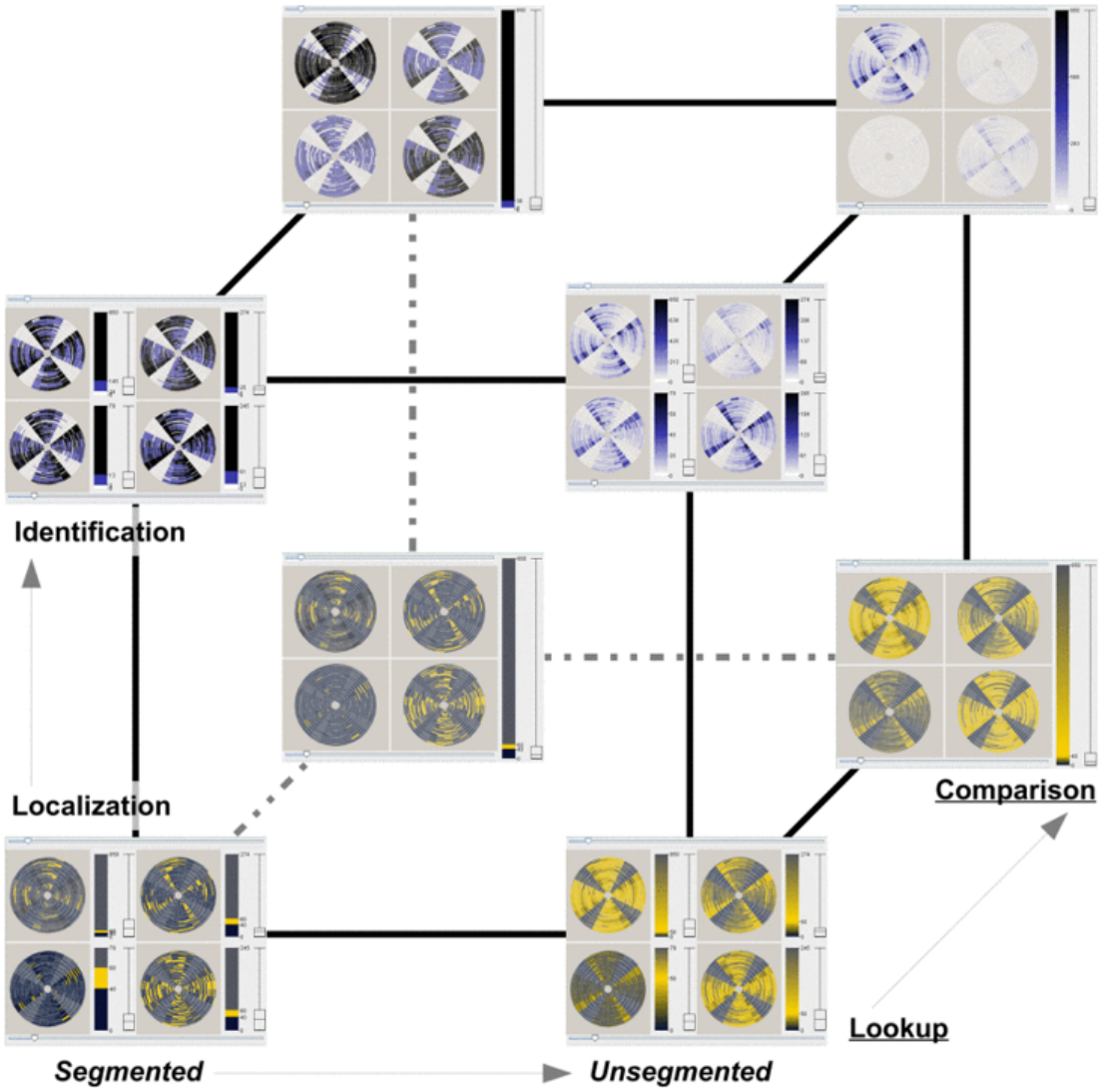


Figure 2.22: The task-colour-cube that spans along the pseudo-colouring scale (segmented *vs.* unsegmented), as well as the task type (identification *vs.* localization, and lookup *vs.* comparison) represented at each corner of the cube [TFS08]. Copyright ©2008, IEEE.

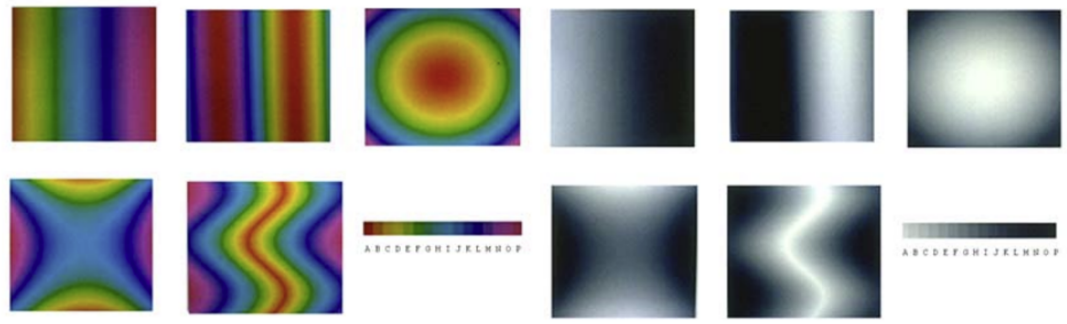


Figure 2.23: A comparison of a spectrum-approximation pseudo-colour sequence *vs.* grayscale sequence in showing form and value. The pseudo-colour improves the representation of both form and value information, while the grayscale partially helps in representing form [War88]. Copyright ©1988, IEEE.

2.3.8 Methods for measuring and modelling visual perception

In 1860, Gustav Theodor Fechner proposed the term 'psychophysics', a scientific methodology for studying the relationship between the intensity of a physical stimulus and human perception and sensation, by systematically varying the properties of the stimulus along one or more physical dimensions [P$^+$16]. In a classical psychophysical experiment, the parameter of interest is typically the difference threshold, or the just noticeable difference (JND), which estimates the smallest unit or change in a stimulus a person can detect [P$^+$16]. Psychophysics also refers to the general study of a perceptual system, and as such it has widespread and significant theoretical and practical applications across different scientific disciplines, including cognitive science and computer science [P$^+$16]. Psychophysical methods are widely used to systematically evaluate science-of-perception-based visualisations [War21] and human-like perceptual algorithms, which make the algorithms classification inference more explainable (see examples in [RYKC$^+$18, GJS$^+$17, GWB13, ECS16, HSSB98]).

In conjunction with psychophysics, eye-tracking methodology is widely used to quantify visual behavior when performing a given task, to understand differences in locus and level of attention [Duc07]. In cardiology, eye-tracking has been used to study the visual behavior of medical practitioners reading an ECG [AT06, DVHJ16, DBV$^+$16]. To date, studies have neither applied psychophysical methods to understand ECG interpretation nor applied eye-tracking methods to investigate the ability of laypeople to perceive differences in ECG morphology.

2.4 Chapter summary

This chapter used an interdisciplinary review approach across the fields of cardiac physiology, computer science and cognitive psychology to address important questions related to the research problem. First, we investigated how pharmacological drugs can induce LQTS, and found that drugs that mainly block the potassium ion channel can prolong the cardiac action potential leading to LQTS, and (hERG) potassium blockers are the most common cause of drug-induced long QT syndrome (diLQTS). We also found that some drugs could block multiple ion channels affecting different phases of the cardiac action potential, which are reflected on the ECG by substantial T-wave morphological changes that are commonly accompanied by additional changes in the ST-segment and U-wave. We also identified the significant ECG risk predictors of TdP arrhythmia attacks that are associated with diLQTS and sudden cardiac death. We investigated why this critical condition (despite its clinical importance) is currently challenging to recognise on the ECG clinically and computationally. We found that prolonged, large and fused T-U wave morphology changes indicate an increased risk of developing TdP arrhythmia, but they obscure the end of the T-wave and affect the QT-interval measurement accuracy. This issue complicates both manual and automated QT-interval measurement, and the best method to use remains a subject of debate. Consequently, current methods used for detecting drug-induced LQTS on the ECG are sub-optimal for both humans and machines, and investigating other potential approaches to ECG interpretation may save many lives. The chapter finally introduced a fundamental theory in a human vision known as pre-attentive processing, which outlines a set of pre-attentive attributes such as colour (hue and intensity) known to be detected rapidly and accurately by the human eye. These attributes can proceed from pre-attentive to attentive processing when influenced by an individual's intentions and visual task goals. Using these attributes in design can improve both the effectiveness and efficiency of visualising complex data. Therefore, we hypothesise that introducing a pseudo-colouring technique to the complex ECG image has the potential to expose QT-interval duration changes, thus supporting LQTS detection intuitively.

Chapter 3

Computer-Based ECG Interpretation and Visualisation Methods for Supporting Long QT Syndrome Detection and/or Risk Assessment: a Summary and Appraisal of Published Evidence

3.1 Chapter overview

Chapter 2 described how that measuring the QT-interval is a challenging task, as precise identification of the T-wave end is difficult, whether performed manually or automatically. Over the last few years, research has thus increasingly focused on exploring computer-based approaches to improving LQTS detection and/or risk assessment without relying solely on QT-interval measurement. This chapter summarises and appraises the published evidence on computer-based ECG interpretation and Visualisation methods proposed in the current literature for supporting LQTS detection and/or risk assessment. It discusses recent advances and limitations of these methods, identifies the current knowledge gaps in the field and formulates the research hypothesis.

3.2 Review objectives

This review sought to explore computer-based methods reported in the current literature intended to support long QT syndrome (LQTS) detection or risk assessment on the ECG, including both semi-automated and visualisation methods supporting human interpretation, and fully-automated ECG interpretation approaches. The review addressed three questions as follows:

- RQ1: What kinds of computer-based approaches are used to support LQTS detection and/or risk assessment on the ECG? How effective are they?
- RQ2: Do the approaches focus on congenital or drug-induced LQTS, or both?
- RQ3: Do the approaches support risk assessment of potential TdP?

3.3 Review method

3.3.1 Eligibility criteria

Study focus

Studies that focused primarily on the use of computer-based methods to support long QT syndrome (LQTS) detection and/or risk assessment (whether congenital or acquired LQTS) on the ECG were included. Studies that focused on other cardiac conditions or focused only on QT-interval measurement without any particular focus on detecting or assessing QT-prolongation or LQTS were excluded.

Populations

Only studies focusing on detecting/assessing LQTS in adult humans were included. Studies that focused on detecting/assessing LQTS in children or animals were excluded.

Study design

Only empirical research studies that explored computer-based detection and/or assessment methods were included. Clinical case reports or papers that did not report any computer-based methodological contributions were excluded. Studies that reported the

results of a proprietary commercial ECG analysis system without explicitly explaining how the system works were excluded.

Publication type

Only full text, English language papers were included. Conference abstracts and non-English papers were excluded.

3.3.2 Information sources

Four electronic databases— PubMed, Springer, Science Direct and IEEE—were searched for relevant publications up to the end of January 2021. We also searched Google Scholar to avoid bias in favour of any specific database.

3.3.3 Search query construction

The three key terms of this review are 'computer-based', 'long QT syndrome' and 'ECG'. The search query was therefore constructed as ('Long QT syndrome' OR 'Long QT' OR 'LQTS' OR 'QT-prolongation' OR 'QT-interval prolongation' OR 'Prolonged QT') AND ('Computer-based' OR 'Computer' OR 'Computerised' OR 'Computerized' OR 'Computational' OR 'Automated' OR 'Algorithm' OR 'Algorithms') AND ('ECG' OR 'EKG' OR 'Electrocardiogram').

3.3.4 Study selection process

The study selection process followed the guidelines in the PRISMA statement [LAT+09]. Search terms were entered into databases, and all studies captured by the search query were first screened by title and abstract. Selected studies were then screened for inclusion by reading the full-text.

3.3.5 Quality appraisal

The selected papers were assessed for inclusion in the review using an adapted version of the quality appraisal checklist for reporting research studies in the field of healthcare intervention proposed by Downs Sara H and Nick Black [DB98], which is described as follows:

Table 3.1: Criteria used for quality appraisal.

ECG sample
Q1: Does the study report the number of ECGs used in evaluating the approach?
Q2: Does the study describe how the ECG sample was selected?
Q3: Does the ECG sample represent the target group to which the results will be generalised?
Evaluation measures used
Q4: Does the study explain the proposed approach explicitly?
Q5: Does the study describe the main outcomes to be measured?
Q6: Is the method for analysing the results appropriate and sufficiently explained?

Two responses were used when scoring the criteria: 'yes' and 'no'. Papers were scored eligible if they addressed all criterion.

3.4 Results

The electronic database search retrieved 825 papers. These were initially screened by title to remove duplicates, leading to 520 papers. The titles and abstracts were then reviewed for relevance based on the inclusion criteria. This stage revealed that the vast majority of work in this area has focused on using computer-based methods to analyse phenotype and genotype differences in LQTS patients (%33 of the retrieved studies); predict potential drug-induced TdP from hERG inhibition or multiple ion channel blocking drugs (%16 of the retrieved studies); manage QT-prolonging drug doses in a clinical scenario (%14 of the retrieved studies); evaluate the effects of different drug concentrations on QT-prolongation without any particular focus on ECG interpretation (%7 of the retrieved studies); and model gender differences in LQTS patients (%3 of the retrieved studies). The primary focus of these studies was not to assist in distinguishing LQTS patients from healthy control subjects based on ECG features, nor to support LQTS detection or risk assessment on the ECG, but rather to analyse clinical

and genetic LQTS characteristics independently of ECG features. Some studies were completely irrelevant to the review objectives (%22 of the retrieved studies). Thus, relatively few studies were found that have investigated the impact of using computer-based ECG interpretation methods to support LQTS detection and/or risk assessment (only %5 of the retrieved studies). These relevant papers ($n = 26$) were thoroughly examined by reading the full-text, and the reference lists were manually searched for further eligible papers. The final number of papers selected for inclusion in this review was 10. Figure 3.1 illustrates the study selection process.

3.4.1 Overview of the characteristics of the included studies

Table 3.2 provides a summary of the included studies. The studies' time frame ranges from 2005 to 2020, with most of them ($n = 7$, 70%) conducted recently (2017-2020). Most studies were empirical experiments that used ECGs of congenital LQTS patients and healthy control subjects ($n = 7$), and relatively few studies ($n = 3$) used ECGs from drug-induced LQTS patients without including healthy control ECGs.

Half of the studies focused on improving the detection of LQTS using ECG features (particularly T-wave morphology). The other half focused on enhancing the risk assessment of LQTS (one study was based on the QT/RR relationship, two studies were based on T-wave morphology, one study was based on QTc values and one study used a combination of clinical characteristics and multiple ECG features).

All studies described their computational approach explicitly, but only one study made their software open source. Two studies reported the use of their own commercial ECG analysis software but explained the computational methods involved sufficiently.

Five studies used all 12 leads of a 10-second ECG recording, then calculated the average QT-interval from a number of manually selected heartbeats (also known as representative heartbeats), while three studies selected only lead II data from a 12-lead 10-second ECG. Two studies used continuous ECG datasets (24-hour recording) from a Holter monitor.

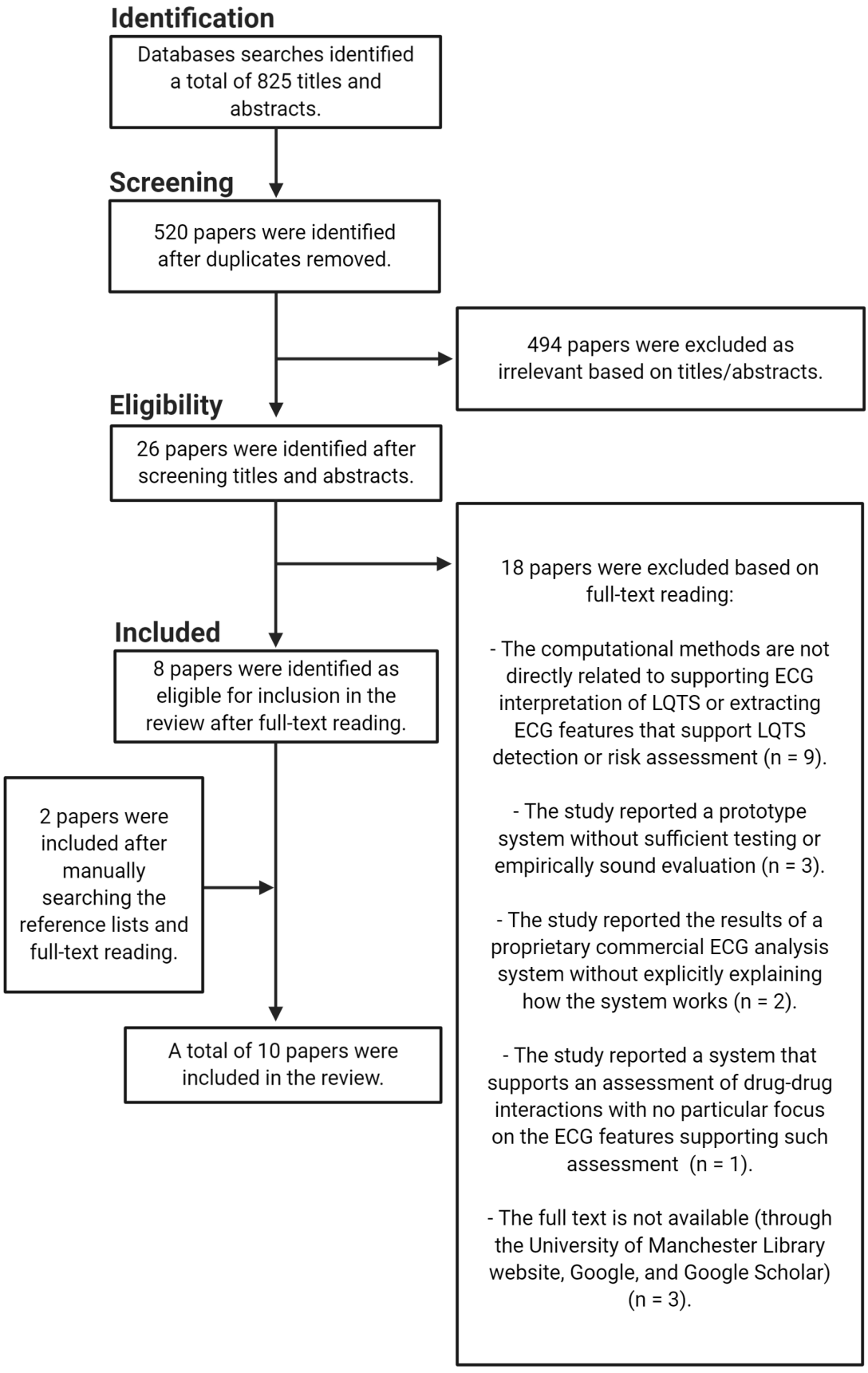


Figure 3.1: The PRISMA flowchart for the study selection process.

Table 3.2: Overview of the included studies characteristics.

Study reference	Author, year	Study focus	LQTS detection/ risk assessment	Approach type	Availability	Clinical implications
[CZM05]	Couderc et al., 2005	Discriminating hERG (LQTS2) carrier from non-carrier on the ECG.	Detection of congenital LQTS.	A fully-automated signal processing approach and logistic regression statistical modelling.	Proprietary, commercial software.	T-wave morphology can help detect the hERG mutation in LQTS2 patients at borderline QT-interval.
[CKH+07]	Couderc et al., 2007	Quantifying T-wave morphology to identify individuals who developed drug-induced TdP.	Risk assessment of drug-induced LQTS.	A mixed approach of fully-automated and semi-automated signal processing methods, along with logistic regression statistical modelling.	Proprietary, commercial software.	The propensity of an individual to drug-induced TdP can be recognised from the T-wave morphology on the ECG baseline (before taking the medication).
[PAS+16]	Page et al., 2016	Visualising the changes in QT-interval values automatically generated by a 24-hour Holter ECG monitor.	Risk assessment of congenital LQTS.	A QT clock visualisation technique showing automated QTc values on a circular plot representing a 24-hour clock.	Open source.	The QT clock helps uncover dynamic QT-interval changes during 24-hour between different types of LQTS, as well as between men and women with LQTS.
[PMX+17]	Page et al., 2017	Investigating beat-to-beat QT-interval and heart rate relationship from Holter ECG monitor.	Risk assessment of congenital LQTS.	A web-based tool uses a semi-automated QT measurement approach and heat map visualisation technique showing the QT/RR changes relationship.	Source-code unavailable.	A heat map QT/RR visualisation technique helps detect if the QT is adapting to sudden heart rate changes.

Table 3.2: Overview of the included studies characteristics.

Study reference	Author, year	Study focus	LQTS detection/ risk assessment	Approach type	Availability	Clinical implications
[PSSH+17]	Porta-Sánchez et al., 2017	Analysing the diagnostic accuracy of automated, quantitative analysis of T-wave morphology to distinguish congenital LQTS patients from healthy control subjects.	Detection of congenital LQTS.	A fully automated signal processing approach quantifying the T-wave morphology including flatness, asymmetry, and notching, producing a morphology combination score (MCS).	Proprietary, commercial software.	An automated, quantitative analysis of T-wave morphology helps detect congenital LQTS from healthy subjects.
[HSB+18]	Hermans et al., 2018	Investigating the added value of T-wave morphology in the detection of LQTS patients from healthy control subjects.	Detection of congenital LQTS.	A mixed approach of fully automated signal processing and support vector machine learning classifier using different ECG measurements as inputs, including heart rate, QT-interval and T-wave morphology.	Source-code unavailable.	A support vector machine learning approach that adds T-wave morphology to other ECG measurements significantly improves the detection of congenital LQTS.

Table 3.2: Overview of the included studies characteristics.

Study reference	Author, year	Study focus	LQTS detection/ risk assessment	Approach type	Availability	Clinical implications
[HRK19]	Hajimolahoseini et al., 2019	Examining the accuracy of a deep learning approach using a convolutional neural network (CNN) in classifying 12-lead ECGs (each ECG image represents a single heartbeat) as 'normal' or LQTS.	Detection of congenital LQTS.	A mixed approach of signal processing and deep learning CNN classifier. The 12-lead ECGs were segmented into individual heartbeats using a signal processing QRS peak detector.	Source-code unavailable.	A deep learning approach can improve detection of congenital LQTS based on the ECG morphology of a single heartbeat.
[MPS+19]	Morettini et al., 2019	Examining the accuracy of an artificial neural network (ANN) in classifying the hERG potassium-channel block as 'high block' or 'low block' based on the T-wave morphology, after a QT-prolonging (dofetilide) administration.	Risk assessment of drug-induced LQTS.	A mixed approach of clinical characteristics, manually measured T-wave morphology and artificial neural network (ANN) classifier.	Source-code unavailable.	An artificial neural network (ANN) classifier that uses manually measured T-wave morphology characteristics can help assess the risk of hERG potassium-channel block levels caused by a QT-prolonging drug.

Table 3.2: Overview of the included studies characteristics.

Study reference	Author, year	Study focus	LQTS detection/ risk assessment	Approach type	Availability	Clinical implications
[CZY+20]	Chen et al., 2020	Predicating all-cause mortality of drug-induced long QT syndrome based on clinical and ECG characteristics using random survival forests and non-negative matrix factorisation.	Risk assessment of drug-induced LQTS.	Combined predictive modelling statistical approaches including random survival forests (RSF) and non-negative matrix factorisation (NMF) that use clinical characteristics and manually measured ECG features as inputs.	Source-code unavailable.	A combined approach of RSF-NMF statistical models can predict all-cause mortality from drug-induced LQTS including a prolongation of J–T^{peak} interval corrected to heart rate and QRS complex on the ECG, along with cancer, serum potassium and calcium levels.
[HBV+20]	Hermans et al., 2020	Investigating the diagnostic accuracy of T-wave morphology, along with the corrected QT-interval (QTc), age and sex in detecting LQTS.	Detection of congenital LQTS.	A mixed approach of signal processing quantifying T-wave morphology and multiple support vector machine learning classifiers.	Source-code unavailable.	Adding T-wave morphology characteristics to age, sex, and QTc features in a support vector machine model improves LQTS detection.

3.4.2 Synthesis of the studies' findings

RQ1: What kinds of computer-based approaches are used to support LQTS detection and/or risk assessment on the ECG? How effective are they? All studies focused on LQTS detection or risk assessment, and no study focused on both. As such, we divided the studies into these two broad categories: 'LQTS detection' and 'LQTS risk assessment'.

LQTS detection: All studies that aimed to improve LQTS detection using computerised methods have relied mainly on statistical analysis approaches to predict the key ECG features that differentiate LQTS patients from healthy control subjects ($n = 5$) [CZM05, PSSH$^+$17, HSB$^+$18, HRK19, HBV$^+$20]. In most of these studies signal processing techniques have been used to measure ECG features automatically—notably the T-wave morphology—and then use these automated numerical values as inputs to train a machine learning model ($n = 3$) [CZM05, HSB$^+$18, HBV$^+$20] or to compute a T-wave morphology combination score (MCS = 1.6 $\times$ flatness + asymmetry + notch) ($n = 1$) [PSSH$^+$17].

The study by Couderc et al. (2005) sought to detect LQTS as early as possible (that is when the QT-interval is at the borderline level and not yet considered prolonged) [CZM05]. They focused their study on patients with congenital LQTS type 2, who have gene (hERG) mutation, and attempted to discriminate between hERG mutation carriers from non-carriers based on the QT-interval, RR-interval and T-wave morphology of 46 carriers and 49 non-carriers (total number of ECGs used = 1583) [CZM05]. The ECG features were measured by an automated signal processing commercial software application called 'The COMPAS' (University of Rochester Medical Center, NY, USA) [CZM05]. After several experiments manipulating the inclusion of these ECG features into a logistic regression model, they found that the ascending slope of the T-wave morphology (clinically known as the J–T^{peak} interval), along with the QT-interval and RR-interval, can effectively distinguish the hERG mutation carrier from non-carrier at a borderline QT-interval, with a sensitivity and specificity equal to 92.7% and 90% respectively [CZM05].

Two studies by the same authors investigated the effectiveness of a support vector machine learning (SVM) classifier in distinguishing LQTS patients from healthy control subjects [HSB$^+$18, HBV$^+$20]. The first study empirically compared two models: a baseline model that included patient demographic characteristics and ECG features including age, sex, RR-interval, QT-interval, and corrected QTc-interval, and an

extended model that added T-wave morphology measurements to the same baseline model's inputs [HSB$^+$18]. Using a total of 688 12-lead ECGs recorded from congenital LQTS patients and healthy subjects, the results showed that the extended model including T-wave morphology characteristics significantly improved the detection accuracy of LQTS patients. The area under the receiver operating characteristic (ROC) curve (AUC) for the comparison extended model was 0.901, while for the baseline model it was 0.821 [HSB$^+$18]. The second study used a similar approach, but compared three SVM classifiers rather than two: a baseline model including age, gender and corrected QTc-interval as inputs; a T-wave morphology model using pre-calculated T-wave morphology characteristics measured by an automated signal morphology fitting approach as inputs; a model combining the baseline and T-wave morphology models [HSB$^+$18]. The ECG data used were from 333 patients with congenital LQTS and 345 healthy subjects. The results confirmed the results of the first study, showing that the combined approach that added T-wave morphology characteristics to age, gender, and QTc values as inputs improves LQTS detection, with an overall accuracy of 84% [HBV$^+$20].

A study by Porta-Sánchez et al. (2017) used a highly sophisticated signal processing system— proprietary, commercial QT Guard Plus software (GE Healthcare, Milwaukee, Wisconsin)—to measure different characteristics of T-wave morphology including flatness, asymmetry, and notching [PSSH$^+$17]. They assessed the overall normality of the T-wave morphology using a combination score calculated as 'MCS = 1.6 $\times$ flatness + asymmetry + notch', comparing 108 congenital LQTS patients and 45 healthy subjects [PSSH$^+$17]. Taking one ECG for each patient at rest and automatically calculating the MSC score, they found that the MCS results differed significantly between LQTS patients (Mean = 117.8, *SD* = 57.4) and healthy control subjects (Mean = 71.9, *SD* = 16.2)(p-value $<$ 0.001), and showed that using the MCS score therefore has the potential to improve LQTS detection. They showed that using an MCS threshold of 85 can distinguish LQTS patients from healthy control subjects with 79% sensitivity and 82.6% specificity (ROC AUC = 0.878) [PSSH$^+$17].

Although all these approaches have shown potential for improving LQTS detection, they still rely on automated signal processing methods, the limitations of which were previously discussed in detail in Section 2.2.4. In a sole departure from this, Hajimolahoseini et al. (2019) extracted signal features from ECG images using a convolutional neural network (CNN) deep learning model [HRK19]. As deep learning model usually requires large datasets to perform well [ERR$^+$19, Top19a], they needed to segment the

12-lead ECG image into individual heartbeats. This increased the number of images used to train the CNN model, where each ECG image represents all 12-lead signals of a single heartbeat. We hypothesise here that the reason behind including all 12 leads is to increase the number of features within a single ECG image. Their approach provided high average accuracy results during cross-validation testing (Mean F1-score = 93.26%), showing a new promising route to extracting ECG features automatically without using signal processing measurement approaches [HRK19]. However, given that this approach applied only on single heartbeats, it did not consider detecting LQTS according to heart rate. As such, future work should investigate the potential of this approach in detecting LQTS across different heart rates. Another limitation of this study is that they labelled the data based on genetic testing results (not QTc values), and as such the primary purpose of their approach was to distinguish between genotype-positive LQTS patients (congenital LQTS patients) and genotype-negative healthy people based on ECG morphology only, without considering the QT-interval. Thus, it is not clear whether this approach would generalise to the acquired, drug-induced LQTS, where a genotype-negative healthy person can still develop LQTS as a result of taking medications.

LQTS risk assessment: The risk of LQTS can be assessed in many ways, and studies have thus differed in their methodology for determining it. A study by Couderc et al. (2007) aimed to determine the propensity of an individual with congenital LQTS to develop drug-induced TdP based on ECG baseline characteristics, before taking the QT-prolonging medication (Sotalol) [CKH$^+$07]. They recorded a single drug-free ECG baseline and a single follow-up drug-induced ECG for 34 patients with congenital LQTS. Seventeen of these patients developed one or more drug-induced TdP episodes as a result of taking the medication. The baseline ECGs were then divided into two groups: patients with a history of drug-induced TdP and patients without. The ECG features, including the QT-interval, RR-interval and T-wave morphology, were measured using a fully automated signal processing approach on all heartbeats to provide overall averaged values, and a semi-automated signal processing approach where three representative heartbeats were selected and a clinician manually adjusted the end of the T-wave if needed. A logistic regression model was then used to determine the baseline ECG features that identify individuals with a significant tendency to develop drug-induced TdP. The results showed that the T-wave morphology—particularly the early repolarisation duration (ERD) shown by the J–T^{peak} interval on the ECG—is the

strongest predictor of developing drug-induced TdP. They found that for each one millisecond increase in its duration, there was an 8.4% increased probability of developing TdP [CKH+07].

Two studies have investigated the use of data visualisation techniques to assess the risk of LQTS in different ways [PAS+16, PMX+17]. The first study by Page et al. (2016) examined the use of a 'QT clock' visualisation technique to monitor dynamic changes in heart rate corrected QT-interval (QTc) values automatically generated by a 24-hour Holter ECG monitor on a circular plot representing a 24-hour clock [PAS+16]. The clock's radius represents the QTc-interval values varying from 300 to 600ms from the centre to the perimeter. The Holter monitor recorded continuous ECGs over 24 hours, and the QTc-interval of each heartbeat was automatically measured using an open source ECG analysis software developed by Chesnokov et al. [CNG06], and then the percentage of prolonged QTc intervals within an hour was calculated. The QT clock visualises the percentage of beats with QTc prolongation across 24 hours. They examined the effectiveness of this visualisation for presenting periods of prolonged QTc interval over 24 hours for a single patient, as well as in showing QTc-prolongation period differences between patients across different LQTS types (202 LQTS type1, 89 type 2, and 14 type 3) in comparison with 200 healthy control subjects. The QT clock helped uncover dynamic QT-interval changes over 24 hours between LQTS patients type 1 and type 2; LQTS type 1 patients showed more frequent QTc prolongation during the day than at night, while LQTS type 2 patients showed the opposed pattern (p-value < 0.05) [PAS+16]. This provides evidence that the QT clock can be used to assess the risk of QT-prolongation over 24 hours for an individual patient and a multiple patient group.

The second study by Page et al. (2017) also used a 24-hour Holter ECG monitor, this time with both fully-automated and semi-automated approaches to measuring the QT-interval [PMX+17]. Automated QT measurements were taken for 192 healthy subjects and manual QT measurements were taken by a clinician for LQTS patients (138 LQTS type 1 and 97 LQTS type 2), who then calculated changes in QT and RR intervals. Because the QT-interval changes inversely with heart rate (see Section 2.2.4), their study examined an experimental web-based tool that visualises QT/RR interval changes using a heat map. The results showed that a simple heat map visualisation technique showing the correlation between QT and RR intervals changes

helps detect whether the QT is adapting to sudden heart rate changes, thus supporting the detection of specific QT adaptation patterns indicating a higher risk for cardiac events. They found that patients with cardiac events have large Δ QT and Δ RR interval variability, particularly when the heart rate is accelerating and the QT-interval getting longer [PMX$^+$17], although most clinical research in this area has shown that cardiac events in LQTS patients are often correlated with a slow heart rate [CNK$^+$15, Nam10, TRR$^+$07, CIKD07].

Morettini et al. (2019) aimed to assess the risk of drug-induced LQTS by using an artificial neural network (ANN) to classify the degree of hERG potassium-channel blocking as 'high block' or 'low block' based on T-wave morphology, after a QT-prolonging (dofetilide) administration [MPS$^+$19]. They used the manually measured T-wave morphology characteristics of 22 healthy subjects receiving a single 500 μg dose of dofetilide. Namely, they calculated the difference in the J–T^{peak} and T^{peak}-T^{end} intervals (the early and the late phases of repolarisation) between each follow-up ECG and a drug-free ECG baseline. To train the ANN classifier, they used these Δ JT and Δ TP values as inputs and clinical blood test data showing the amount of dofetilide plasma concentration indicating the hERG potassium-channel block level as outputs. The dataset was split into training, validation, and test sets (114, 29, and 143, respectively). The results showed that an ANN classifier, which uses changes in the T-wave duration (*i.e.* the J–T^{peak} and T^{peak}-T^{end} intervals) from a drug-free baseline as features, accurately determines the risk of hERG potassium-channel block levels caused by a QT-prolonging drug (ROC AUC = 0.91) [MPS$^+$19].

Given that many cancer treatments including chemotherapy are known to cause an acquired, drug-induced LQTS [KBS$^+$13, TBM$^+$04], a study by Chen et al. (2020) sought to identify significant risk factors for all-cause mortality of acquired LQTS patients using statistical analysis methods including random survival forests (RSF) and non-negative matrix factorisation (NMF) [CZY$^+$20]. They used clinical characteristics (including serum potassium and calcium levels and patient history of syncope and severe arrhythmia) and manually measured ECG features (including the PR-interval, QRS complex, QT-interval, J–T^{peak} and T^{peak}-T^{end} intervals) for patients who died with acquired LQTS, with or without cancer. Their study results showed that a combined approach of RSF-NMF statistical models can predict all-cause mortality from drug-induced LQTS including a prolongation of J–T^{peak} interval corrected to heart rate and QRS complex on the ECG, along with cancer, serum potassium and calcium levels [CZY$^+$20].

RQ2: Do the approaches focus on congenital or drug-induced LQTS? or both? Studies have focused either on congenital or drug-induced, but not both. Most of the studies have focused primarily on congenital LQTS ($n = 8$). Only two studies focused on acquired, drug-induced LQTS: one study has focused on patients receiving anti-cancer medications, while the other study focused on healthy people at risk of developing drug-induced LQTS [MPS+19].

RQ3: Do the approaches support risk assessment of potential TdP? Only two studies have assessed the risk of potential TdP based on the T-wave morphology. One study aimed to determine the propensity of an individual with congenital LQTS to develop drug-induced TdP using a logistic regression prediction model [CKH+07]. This found that the presence of a prolonged early repolarisation duration (represented on the ECG by the J–T^{peak} interval) in the ECG baseline can increase the risk of developing drug-induced TdP if a patient used a QT-prolonging drug [CKH+07]. The other study sought to determine the risk of TdP based on the hERG potassium-channel block level (*i.e.* they hypothesised that high hERG block level = high TdP risk) by using an artificial neural network (ANN) classifier [MPS+19]. They showed that increases in both the early and late repolarisation duration (represented by the J–T^{peak} and T^{peak}-T^{end} intervals of the T-wave) from a drug-free ECG baseline can increase the risk of developing TdP [MPS+19].

3.5 Identifying knowledge gaps and formulating our research hypothesis

Although computer-based ECG interpretation methods were introduced in the 1950s [Rau07]–where numerous efforts have been made to to improve the accuracy of detection of many cardiac conditions (notably arrhythmias)[LL20, SW17]—long QT syndrome has received scant attention in this area of research. This literature review shows that researchers recognised the difficulty of detecting LQTS based solely on measuring the QT-interval, and most studies have attempted to explore alternative ECG features to identify this critical condition. A considerable amount of research has focused on extracting the T-wave morphology features from the ECG using either manual measurements or automated signal processing measurements, then using these numerical feature values as inputs to various computerised statistical analysis approaches to predict the ECG features of LQTS patients that differ from healthy subjects. To date, no

study has investigated the impact of using computer-based methods to support visual LQTS detection on the standard ECG directly.

Although an acquired, drug-induced LQTS caused by many commonly prescribed medications is by far the most common cause of TdP arrhythmia attacks [NAS12, YC03, CMY08, ESTB20, WS20], most studies focused on congenital LQTS and remarkably few studies have investigated the impact of computer-based ECG interpretation methods in supporting the assessment of drug-induced LQTS at risk of TdP. This thesis seeks to bridge these gaps by exploring the potential impact that computer-based ECG interpretation methods could have on enabling non-experts to interpret the ECG. This could ultimately be used for improving primary and secondary prevention of drug-induced LQTS, and ultimately save many lives. Such methods could help less-experienced clinicians with ECG interpretation, particularly within emergency departments, to assess and monitor patients' QT-intervals before or during the provision of a QT-prolonging medication, and also support outpatients on a known QT-prolonging drug to self-monitor their own ECGs. To the best of our knowledge, no study has focused on supporting non-experts in ECG interpretation to detect or assess drug-induced LQTS and TdP by leveraging computer-based ECG interpretation approaches.

Research has shown that accurate human ECG interpretation depends on the perceptual skill of pattern recognition [WBA$^+$14, BZF$^+$14, DBV$^+$16], with experts using this first, and resorting to a more systematic measurement approach of ECG features if they run into barriers [DMH$^+$19]. To date, no study has used knowledge of human visual perception to enhance human-machine ECG interpretation. This thesis therefore hypothesises that using a science-of-perception-based approach to optimise the recognition of important signal patterns associated with LQTS could improve non-expert ECG interpretation accuracy, and endow machines with a 'human-like' perceptual ability to recognise LQTS signal patterns, with the potential to facilitate an explainable, shared human-machine ECG interpretation.

3.6 Chapter summary

Critical analysis of the current literature on computer-based ECG interpretation methods proposed for detecting or assessing the risk of LQTS shows that researchers have recognised the difficulty of detecting LQTS based solely on measuring the QT-interval, and as such most studies have attempted to explore alternative ECG features to identify this critical condition automatically. The vast majority of the work in this area has focused on extracting T-wave morphology features from the ECG using either manual or automated signal processing measurements and using the resulting numerical feature values as inputs to computerised statistical analysis approaches to predict the ECG features that distinguish LQTS patients from healthy subjects. To date, no study has investigated the impact of using computer-based methods to support visual LQTS detection on the standard ECG directly. Also, although the drug-induced LQTS caused by many commonly prescribed medications is by far the most common cause of TdP arrhythmia and sudden cardiac death, most studies have focused on congenital LQTS, and remarkably few studies have investigated the impact of using computer-based ECG interpretation methods for supporting the risk assessment of potential TdP. Moreover, although many QT-prolonging drugs are commonly prescribed (often for self-limited diseases), no study has focused on using computer-based ECG interpretation approaches to support non-experts in ECG interpretation to detect or assess drug-induced LQTS at risk of TdP.

Chapter 4

Can Laypeople Identify a Drug-Induced QT-Interval Prolongation? A Psychophysical and Eye-Tracking Experiment Examining the Ability of Non-Experts to Interpret an ECG

4.0 Chapter overview

4.0.1 Thesis context

The research in this thesis is motivated by the idea that laypeople will ultimately be able to self-monitor for LQTS at home. To date, no study has specifically investigated the ability of laypeople to interpret ECGs in general or to determine QT-interval prolongation in particular, discussed in Chapter 1 as the first research challenge. This chapter attempts to fill this gap by introducing a psychophysical and eye-tracking experiment that quantifies laypeople's ability to perceive drug-induced increases in the QT-interval from a drug-free 'normal' QT-interval (baseline) on the standard ECG. As a starting point, the investigation is limited to ECGs with a regular heart rate of 60 bpm. The reason behind this—given the fact that the QT-interval (QT) is the same as the corrected QT-interval (QTc) at 60 bpm [LMJM04]—is to examine their ability

to detect differences in the QT-interval length (between a drug-free ECG baseline and a drug-induced comparison ECG) with no QT measurement or heart rate correction required.

The results of this study show the potential for ECG self-monitoring, as the majority of laypeople were able to detect a clinically significant QT-prolongation (drug-induced QT $>$ 500ms, the difference from the baseline $\approx$ 88ms) at a low normal heart rate. A number of research questions emerged from this work, including: (1) how to support laypeople in detecting smaller increases in the QT-interval that may represent a clinically significant side effect of a QT-prolonging drug, which was explored and empirically studied in Chapter 5; and (2) how to further support them in perceiving increases in the QT-interval at higher or lower heart rates, which has been investigated in depth in Chapter 6.

The main content of this chapter is adapted from: Alaa Alahmadi, Alan Davies, Markel Vigo, and Caroline Jay. Can lay people identify a drug-induced QT-interval prolongation? A psychophysical and eye-tracking experiment examining the ability of non-experts to interpret an ECG. *Journal of the American Medical Informatics Association*, 2018.

4.0.2 Author's contributions

Alaa Alahmadi and Caroline Jay devised the idea for the work. Alaa Alahmadi designed the study, carried out the data collection, analysed the results and wrote the paper, with Alan Davies, Markel Vigo and Caroline Jay contributing significant edits. Caroline Jay assisted with study design and analysis. Alan Davies acted as the electrocardiogram domain expert throughout. Caroline Jay and Markel Vigo provided continuous guidance and discussion.

4.0.3 Published abstract

Objective: The study sought to quantify a layperson's ability to detect drug-induced QT-interval prolongation on an electrocardiogram (ECG) and determine whether the presentation of the trace affects such detection.
Materials and methods: Thirty layperson participants took part in a psychophysical and eye-tracking experiment. Following training, participants completed 21 experimental trials, in which each trial consisted of 2 ECGs (a baseline and a comparison stimulus, both with a heart rate of 60 bpm). The experiment used a 1 alternative forced-choice paradigm, in which participants indicated whether or not they perceived a difference in the QT-interval length between the 2 ECGs. The ECG trace was presented in 3 ways: a single complex with the signals aligned by the R wave, a single complex without alignment, and a 10-second rhythm strip. Performance was analyzed using the psychometric function to estimate the just noticeable difference threshold, along with eye-tracking metrics.
Results: The just noticeable difference 50% and 75% thresholds were 30 and 88ms, respectively, showing that the majority of laypeople were able to detect a clinically significant QT-prolongation at a low normal heart rate. Eye movement data indicated that people were more likely to appraise the rhythm strip stimulus systematically and accurately.
Conclusions: People can quickly be trained to self-monitor, which may help with more rapid identification of drug-induced long QT syndrome and prevent the development of life-threatening complications. The rhythm strip is a better form of presentation than a single complex, as it is less likely to be misinterpreted due to artefacts in the signal.

4.1 Background and significance

Drug-induced long QT syndrome (LQTS) is a cardiac abnormality that can increase the risk of a life-threatening arrhythmia, known as torsades de pointes (TdP), which may lead to syncope, drowning, and sudden cardiac death [GM08, AAKR$^+$02, MSC$^+$85]. LQTS is a side effect of more than 100 commonly prescribed QT-prolonging medications including antiarrhythmic drugs, antihistamines, and antidepressants [CMY08, YC03]. People taking these medications may not experience any symptoms, and sometimes a prolonged QT-interval can only be detected by examining an electrocardiogram (ECG) [Kha02, RV05, SC08].

An ECG is a graphical representation of the electrical activity of the heart and is widely applied in clinical practice to assess heart function and detect cardiac pathologies [ATOP14]. The QT-interval represents the duration of time to complete the ventricular depolarization and repolarization cycle and is measured in the ECG from the beginning of the QRS complex to the end of the T wave [AAKR$^+$02, GMZ06]. LQTS occurs when the repolarization of the heart following a heartbeat is delayed and appears as an elongated QT-interval on the ECG [AAKR$^+$02, GMZ06]. There is also a congenital LQTS caused by mutations in certain genes. People with this disorder might be excluded from using QT-prolonging drugs [SC08, CCDS08].

Frequent monitoring is advisable for people who are at high risk of acquiring LQTS including patients who take prescribed QT-prolonging medications [KW12] or patients participating in a clinical trial for a new drug [CSDB$^+$09, Sha02]. Several studies have investigated the effectiveness of utilizing ambulatory ECG devices to monitor patients' ECG remotely [Loc17, RKE$^+$89], but this approach still relies on clinicians being able to access and interpret the ECG. An additional complication is that health status, age, sex, and ethnicity all influence a patient's ECG in general and the QT-interval specifically [GMZ06, MMDY94, HST$^+$16]. It has been shown that there is no 'cutoff' value for deciding whether, in isolation, the QT-interval is normal, short, or prolonged [GMZ06]. A personalized monitoring solution that considers a patient's reading against their 'normal' baseline ECG has the potential to address some of these issues.

While there are computerized methods for measuring QT-interval, the reliability of these methods is limited [KBR$^+$14, GL13, MCbA01, RSG09, EI13, HG06, TABW11, TAS$^+$15], and human visual validation is strongly recommended [MCbA01, EI13, SW17]. In addition to this, the accuracy of automated ECG interpretation methods is affected by several factors including the presence of abnormal sinus rhythm such as atrial arrhythmias [HG06] or a poor-quality ECG signal [EI13, HG06, SW17]. Moreover, abstracting the ECG data purely into numbers also risks masking other potential abnormal clinically significant changes in the ECG morphology. For instance, specific T-wave patterns can aid detection of LQTS [CHA$^+$15], and large T-U waves are known to precede TdP [KFBW09]. As such the ECG morphology still provides the richest information for recognizing LQTS.

Studies have shown that clinicians find QT-prolongation detection difficult [BY05]. While QT experts achieve a high level of accuracy (96%), other clinicians, even those

who routinely read ECGs, can perform poorly ($< 25\%$) [VRS$^+$05]. Training is important; in a study in which students were taught to use the tangent method, they performed significantly better than arrhythmia experts and cardiologists [PDJVdBW08].

If patients or their carers or family members can use a clinically reliable ECG monitoring device at home and receive the right training to detect specific types of abnormality, this raises the possibility of self-monitoring outside of the clinical environment. Self-care and self-monitoring have been shown to empower patients with knowledge about their condition, which can reduce anxiety [LMG$^+$16].

Psychophysical experiments are used to model a human's ability to distinguish a difference in physical stimuli [Ste57, P$^+$16]. In a classical psychophysical experiment, the parameter of interest is typically the difference threshold, which estimates the smallest unit or change of a stimulus a person can detect [P$^+$16].

In cardiology, eye-tracking research has been used to study the visual behavior of medical practitioners reading an ECG [AT06, DVHJ16, DBV$^+$16]. To date, studies have neither applied psychophysical methods to understand ECG interpretation nor investigated the ability of laypeople to perceive differences in ECG morphology.

4.2 Objective

The primary objective was to quantify a layperson's ability to detect a clinically significant drug-induced QT-interval prolongation when compared to a 'normal' ECG (baseline). The secondary objective was to determine whether the presentation of the ECG (as a single complex or a 10-second rhythm strip) affects this ability.

4.3 Materials and methods

4.3.1 Participants

Thirty participants (15 men and 15 women) with no experience in ECG interpretation were recruited from a university campus (26 students and 4 staff). The mean age was 26 ∓ 6 years. Participants were asked to rate their knowledge of ECGs or ECG interpretation; only people who identified as having no knowledge were included. Participants' sight was normal or corrected-to-normal and they reported no motor or neurological disorders.

4.3.2 Stimuli design

The ECG stimuli were taken from a clinical study conducted to assess QT-interval changes in healthy subjects receiving medication known to cause QT prolongation [JVM+14]. As the study is motivated by the potential for self-monitoring, we selected data from a single participant, whose QT-interval was seen to rise to prognostically dangerous levels. The subject (a 35-year-old man) had normal QT-intervals (QT-interval $<$ 430ms) prior to taking the medication dofetilide (a class III antiarrhythmic); he subsequently experienced a gradual increase in the QT-interval, and eventually reached very high QT prolongation (QT-interval= 579ms). The ECGs sampled all had a heart rate of 60 bpm to ensure it was possible to compare QT-intervals without having to apply a heart rate correction formula (QTc). The QT values used were 417, 421, 430, 441, 485, 537, and 579ms. It was not possible to select a fixed increase of QT-interval for 2 reasons. First, the subject experienced a variable increase in the QT-interval over 24 hours, after receiving a single dose of the medication. Second, as we limited our selection to ECGs that have a heart rate of 60 bpm, only 7 ECGs were available for this representative case. The dataset and its sources can be found in the PhysioNet database [GAG+00], and the clinical trial study can be found in Johannesen et al [JVM+14].

4.3.3 Study design

The experiment used a counterbalanced within-subjects design with 2 independent variables:

- QT-interval difference (see Table 4.1), with 7 levels ranging from 0 (no difference) to 6 (highest difference);

- ECG signal presentation format (see Figure 4.1), with 3 versions, in which each consisted of a baseline complex with a normal QT-interval and a comparison complex with either a normal or prolonged QT-interval:

    1. Two single ECG complexes without R-wave alignment.
    2. Two single ECG complexes aligned on the R wave.
    3. Two 10-second rhythm strips showing 10 complexes.

Table 4.1: QT values acquired from the clinical trial between the baseline and the comparison stimuli.

Trial	Level of Difference	QT Value of the Baseline ECG (ms)	QT Value of the Comparison ECG (ms)	Value of QT Increase	Clinical Rating
1	0 (no difference)	417	417	0	Normal
2	1 (smallest difference)	417	421	4	Normal
3	2	417	430	13	Borderline
4	3	417	441	24	Borderline
5	4	417	485	68	Prolonged
6	5	417	537	120	Very prolonged
7	6 (highest difference)	417	579	162	Very prolonged

Note: As the heart rate was 60 bpm, the QT is the same as the corrected QT-interval using Bazett's formula. The clinical rating was determined based on the suggested Bazett-corrected QT-interval values for diagnosing QT prolongation in adult men [GMZ06, JA09].
ECG: electrocardiogram.

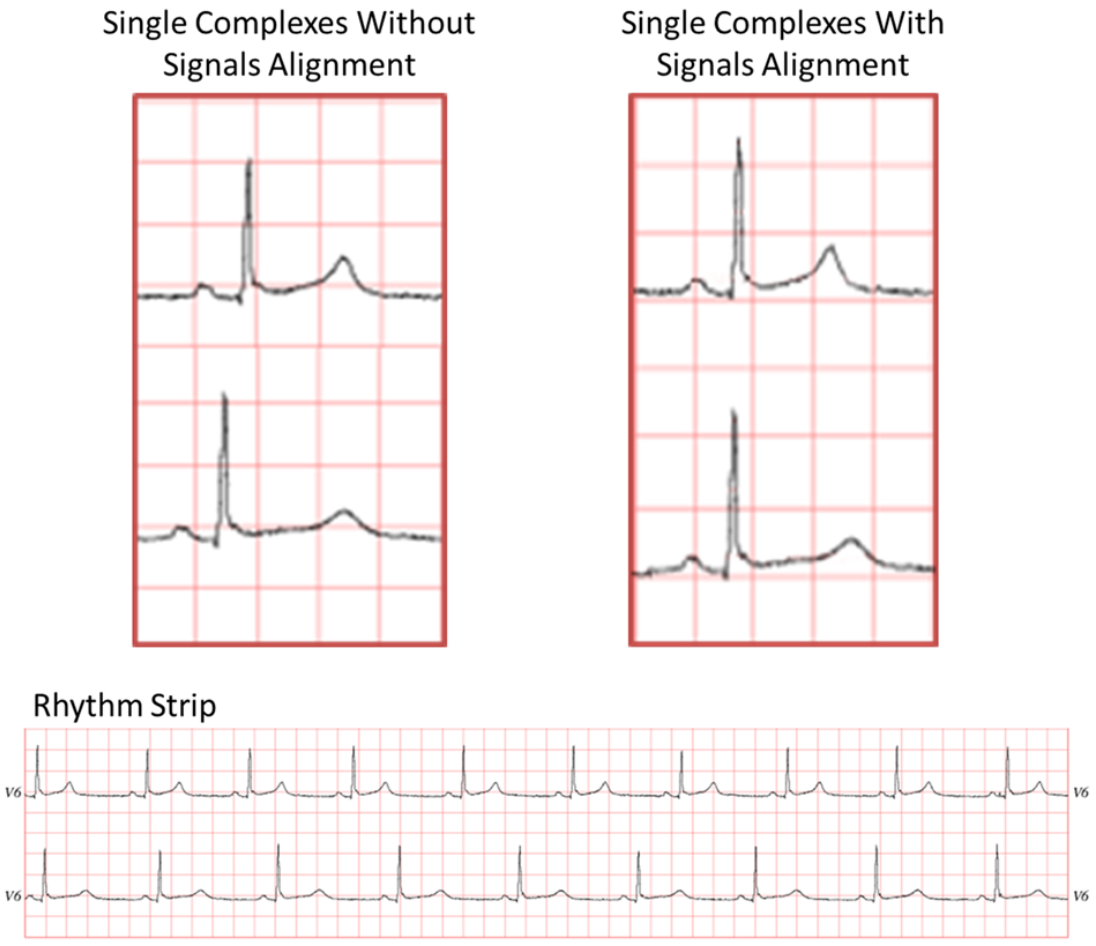


Figure 4.1: Example of the 3 presentation formats showing a baseline stimulus of a normal QT-interval (QT-interval=417ms, heart rate=60 bpm) above a comparison stimulus of a prolonged QT-interval (QT-interval=537ms, heart rate=60 bpm).

Each participant completed 21 trials (7 for each presentation format). We used the method of constant stimuli, in which the levels of QT-interval change in the comparison stimulus are presented randomly and are not related from one trial to the next. This reduces errors of habituation and expectation as the participant cannot predict the level of the next stimulus [BHW09]. Participants completed all trials for one format before moving to the next.

The order of presentation formats was counterbalanced using a balanced Latin square by dividing the thirty participants into six groups of five. Considering (A) to be the 'single ECG complexes without R-wave alignment', (B) to be the 'single ECG complexes aligned on the R-wave' and (C) to be the 'rhythm strip', the counterbalanced order of presentation format for groups 1 to 6 respectively was as follows: A, B, C; A, C, B; C, A, B; C, B, A; B, C, A; and B, A, C.

4.3.4 Apparatus

A Tobii X2–60 eye tracker and Tobii Studio 3.2 software were used to record eye gaze with a sampling rate of 60Hz. Gaze coordinates were recorded every 16.7ms. Audio was recorded to collect participants' verbal answers.

4.3.5 Task and procedure

All participants were informed of the motivation for the work (the potential for laypeople to self-monitor LQTS at home), and were presented with a sample ECG and shown how to identify the location of the QT-interval. Each participant then completed an assessment task, in which they were asked to highlight the location of QT-intervals visually on 3 different ECGs showing normal, borderline and prolonged QT-intervals. Participants were also shown how to determine the interval length by counting the grid squares between the beginning of the Q-wave and the end of the T-wave. People were not asked to determine what a normal QT was, but rather to look for a change in its length. This preparation session thus did not involve any medical terms, clinical methods or high-level training techniques typically associated with ECG interpretation. All participants passed the assessment task, correctly identifing QT-intervals for all ECGs.

The experiment used a classical psychophysical discrimination task known as 1-alternative forced-choice same-different task, also occasionally known as 2IAX or AX [P$^+$16]. Participants were presented with 2 ECGs—a baseline stimulus in which the QT-interval is normal (no QT-interval prolongation) above a comparison stimulus that represents a change in the QT-interval—and they had to decide whether the QT-intervals of the 2 stimuli were the same or different.

We presented the baseline stimulus above the comparison stimulus in all trials, and the participants were aware that the 'normal' baseline was always positioned at the top. One trial shows the same ECG for the baseline and the comparison stimuli. Another 6 trials present the baseline as the 'normal' QT-interval (417ms), and the comparison as 'longer' QT-interval of 421, 430, 441, 485, 537, or 579ms. Table 4.1 shows the difference between the 2 ECGs in each trial. The participants indicated verbally whether there was a difference in the QT-intervals. There was no time limit imposed. The answers were recorded on a paper sheet during the experiment by the researcher and reviewed via the audio recording after the experiment.

4.3.6 Analysis

Two types of assessment were used to analyze participants' responses.

Assessment 1

For the trial in which the QT-interval was the same for the baseline and comparison stimuli (*i.e.* level 0) (Table 4.1), participants' responses were assessed for detection of negative findings measured as true negatives (*i.e.* correct reject) and false positives (*i.e.* false alarm). A false alarm response is registered when there is no QT-interval difference but participants report that there is, and a correct reject response is recorded when they correctly identify the QT-intervals as the same.

Assessment 2

For the 6 trials that showed increases in the QT-interval (*i.e.* levels 1–6) (Table 1), participants' responses were assessed for detection of positive findings as true positives (*i.e.* when participants correctly perceived a difference in the QT-intervals) and false negatives (*i.e.* when they did not perceive a difference in the QT-intervals when a difference was present). This assessment was carried out using the psychometric function, an inferential model applied in psychophysical detection and discrimination tasks. It was used to model the relationship between the gradual increase in the QT-interval and the forced-choice responses of the participants. The psychometric function was plotted as the proportion of correct responses as a function of QT-interval, and the just noticeable difference (JND) threshold was estimated. In psychophysics, the JND is defined as the minimum amount of change necessary in a stimulus to be just noticeable and detectable [P$^+$16]. In this study, we defined it as the minimum amount of QT-interval change required to be just discriminable. We estimated the 50% and 75% JND thresholds as the value of QT-interval in the comparison stimulus at which the proportion of correct responses is equal to 0.5 and 0.75, respectively. These JND thresholds were then used to determine the point at which participants were able to detect a clinically relevant difference. The equations used for estimating the JND thresholds were defined as follows:

$$JND(in\ ms) = QT\ value\ of\ the\ comparison\ stimulus\ at\ 50\%\ correct\ answers - QT\ value\ of\ the\ baseline\ stimulus$$

Equation 4.1: The just noticeable difference (50%) threshold estimation formula.

$$JND(in\ ms) = QT\ value\ of\ the\ comparison\ stimulus\ at\ 75\%\ correct\ answers - QT\ value\ of\ the\ baseline\ stimulus$$

Equation 4.2: The just noticeable difference (75%) threshold estimation formula.

To facilitate the calculation of eye movement metrics, areas of interest (AOIs) were created on the stimuli using Tobii studio software. For the single ECG complex presentation format—with or without signals alignment—2 areas of interest were created: 1 for the baseline stimulus and 1 for the comparison stimulus. For the rhythm strip presentation format, an AOI was created for each ECG complex, resulting in 10 AOIs for the baseline stimulus and 10 AOIs for the comparison stimulus. Figure 4.2 illustrates these areas of interest for the rhythm strip presentation format.

The eye-tracking metric total fixation duration, which indicates the total length of time participants fixated on a given AOI, was calculated for the 3 presentation formats (in the case of the rhythm strip stimulus, this was cumulative across all AOIs). Additionally, the percentage fixated metric, which is the percentage of participants who fixated at least once within an AOI, was calculated for each ECG complex in the rhythm strip presentation format.

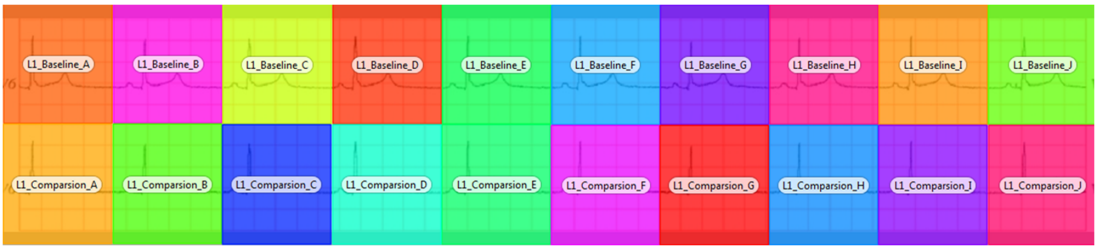


Figure 4.2: The areas of interest for the rhythm strip presentation format. Each area of interest represents 1 electrocardiogram complex.

4.4 Results

4.4.1 Detection accuracy

Assessment 1: correct reject and false alarm For the trial which showed no prolongation of QT-interval (*i.e.* the baseline and comparison were the same), the percentage of correct reject responses was 93.33% and false alarm rate was 6.66% in the rhythm strip presentation, demonstrating that only 2 participants of 30 incorrectly perceived a difference in QT-interval in which no difference exists. In the case of the single complex without signals alignment, correct reject rate was 90% and false alarm rate was 10%. In the condition with signals alignment, the correct reject rate was 100%.

Assessment 2: the psychometric function The psychometric function modeling shows an incremental cumulative distribution curve in the rhythm strip presentation, indicating that the proportion of people able to perceive the difference in the QT-interval grew as the QT-interval increased. Data from the single ECG complex presentations, both with and without signals alignment, showed a different pattern, as a large number of people appeared able to detect the smallest possible difference. As it is unlikely that a person can perceive a small increase in a stimulus level, but not perceive a higher increase, this is likely to be due to an artifact in the particular complex used as a stimulus. Figure 4.3 illustrates the psychometric function model for the 3 presentation formats.

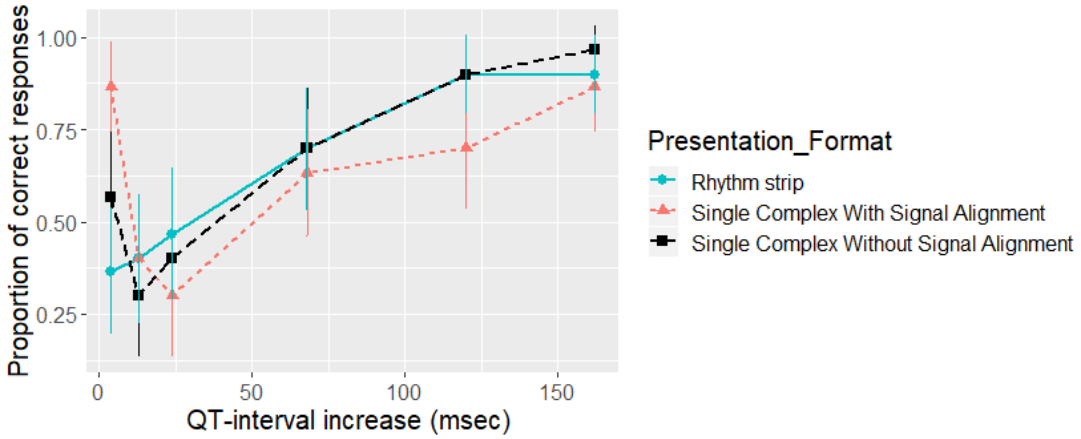


Figure 4.3: The psychometric function plot showing the proportion of correct responses on the Y-axis as a function of QT-interval on the X-axis for the 3 presentation formats. The error bars represent 95% confidence intervals.

JND threshold

The JND was estimated only from the rhythm strip format as it showed the most reliable results. The 50% and 75% JND thresholds were 30ms (QT-interval=447ms) and 88ms (QT-interval=505ms), respectively, and were determined from fitting the psychometric function using a logistic function with maximum likelihood estimation.

4.4.2 Total fixation duration

The mean of total fixation durations were 3.85 ∓ 5.21 seconds for the rhythm strip presentation, 1.82 ∓ 2.21 seconds for the single complex with signal alignment and 1.62 ∓ 2.75 seconds for the single complex without signal alignment across all trials. The mean of total fixation duration differs significantly between the 3 presentation formats for all trials when compared with a Friedman test, $\chi^2(2) = 0.20, p < 0.05$, as seen in Figure 4.4 and Table 4.2. This shows that people fixated significantly longer in the rhythm strip condition than either of the single complex conditions.

Table 4.2: Results of the Friedman test comparing the mean of total fixation duration across the 3 presentation formats.

Trial	Level of Difference	χ^2	P-value
1	0 (no difference)	7.008	0.030
2	1 (smallest difference)	7.681	0.021
3	2	7.267	0.026
4	3	7.681	0.021
5	4	12.067	0.002
6	5	7.800	0.020
7	6 (highest difference)	14.467	0.001

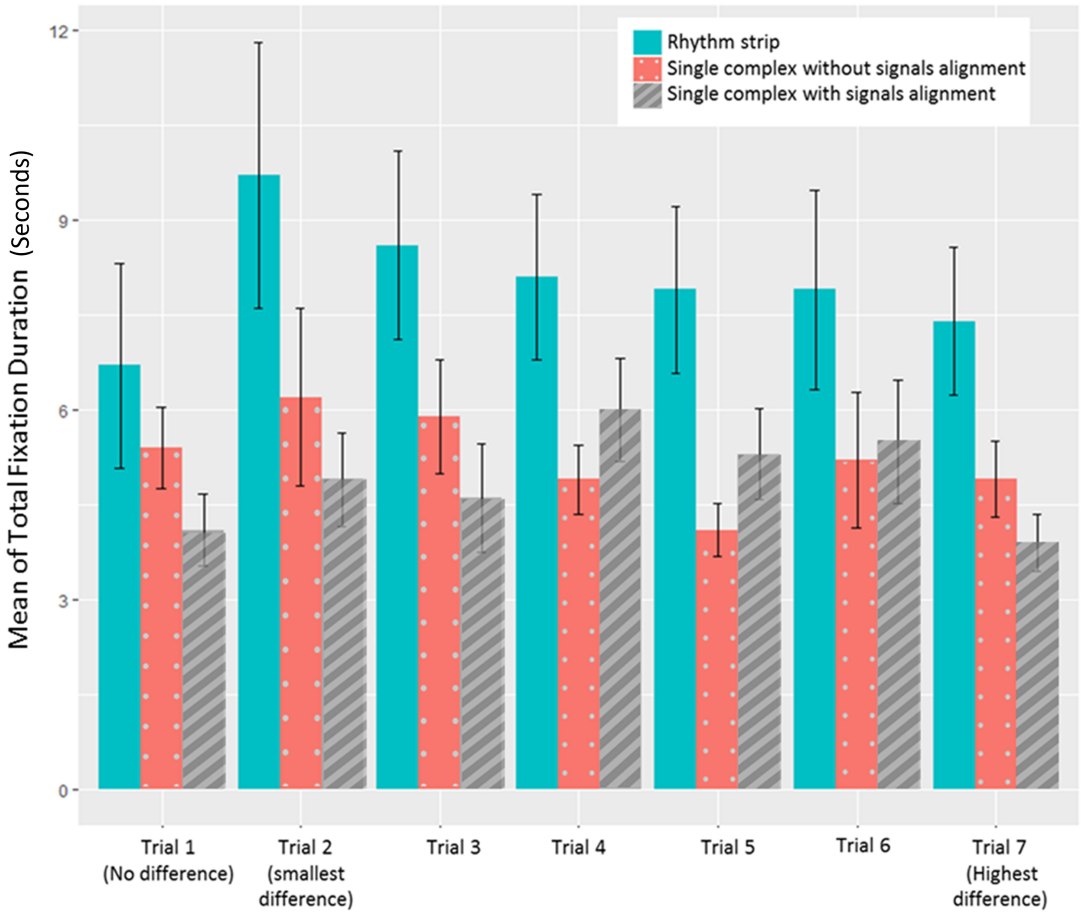


Figure 4.4: Mean of total fixation duration in seconds for the 3 presentation formats as a function of QT-interval difference across all trials. The error bars represent SEM.

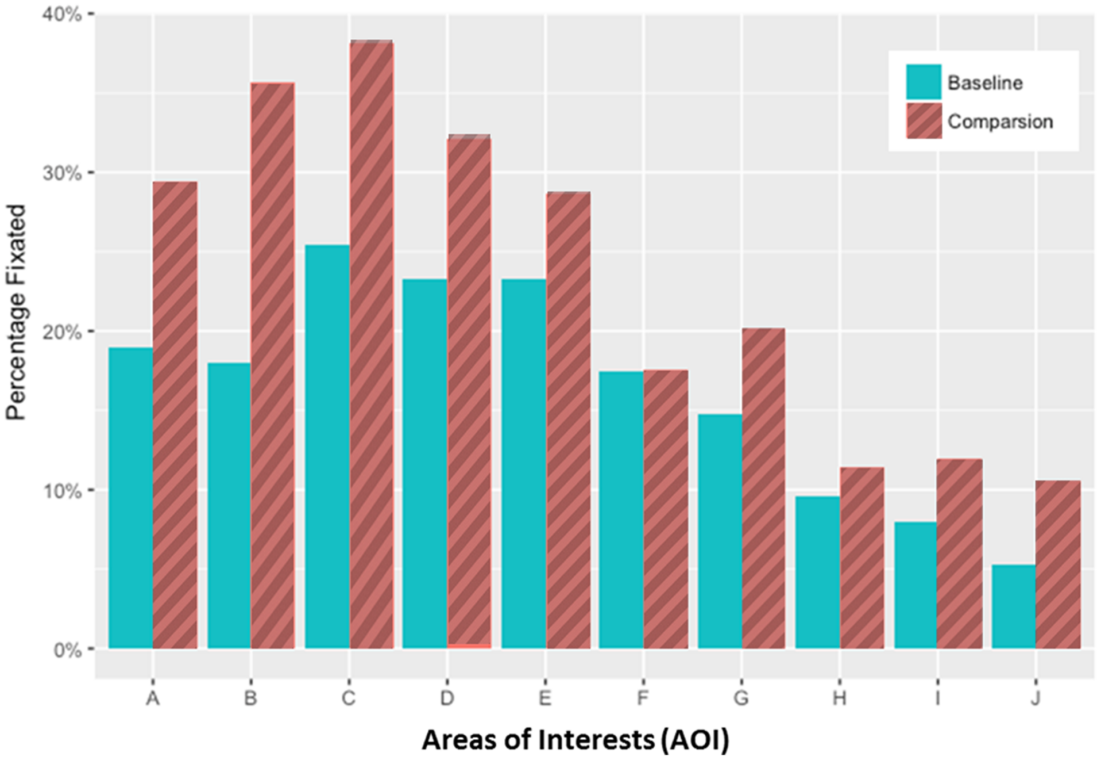


Figure 4.5: Percentage of people fixating on the areas of interest (AOIs) in the rhythm strip presentation, averaged across all trials. Each AOI represents a single electrocardiogram complex.

4.4.3 Percentage fixated in the rhythm strip AOIs

The percentage of rhythm strip AOIs fixated was calculated to determine whether people looked at more than 1 ECG complex before making their decision (see Figure 4.5). In any given trial, participants fixated on average at least 4 ECG complexes for either the baseline or the comparison stimulus before making their decision. Participants looked at the first 5 ECG complexes (from left to right, *i.e.* the AOIs A–E in Figure 4.5) more than the other complexes. Heat maps of mean fixation frequency also show this result (Figure 4.6C).

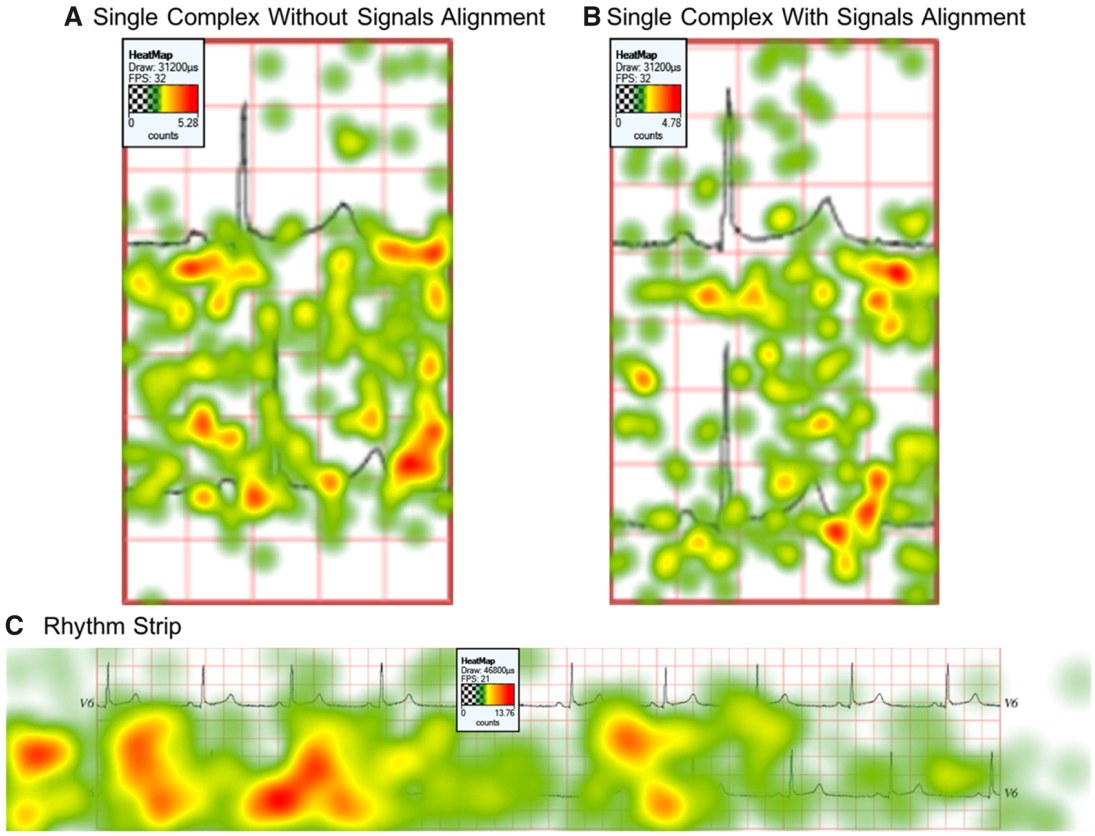


Figure 4.6: A heatmap of mean fixation count for the 3 presentation formats of trial number 1, showing the smallest difference in QT-interval. The presentation formats are (A) single complex without signals alignment, (B) single complex with signals alignment and (C) rhythm strip.

4.5 Discussion

The study showed that laypeople can perceive a clinically significant prolongation of the QT-interval at a low normal heart rate (60 bpm) with minimal training. The estimated JND thresholds indicate that 50% of people perceived the difference when the QT-interval was borderline (QT-interval=447ms and JND=30ms) and 75% of people perceive an even longer difference (QT-interval=505ms, JND=88ms). This provides evidence that people could be trained to self-monitor for LQTS. Although the QT-prolongation above 500ms is considered a risk factor for TdP [Tis16], clinical research has shown that even a small ($\approx$ 10ms) QT-interval increase from the baseline is considered a significant side effect of a QT-prolonging drug [RFF$^+$09, DNS06].

The analysis from both the psychometric function and eye-tracking data show that the rhythm strip presentation is preferable to the single complex presentation, as it is less susceptible to artefacts in the ECG morphology. The psychometric function model showed that participants' responses in the rhythm strip condition formed a linear curve showing a proportional relationship between the perceived difference and the gradual increase of QT-interval. This is in contrast with the single complex presentation, which appeared to show that people were able to detect a very small difference more easily than a longer one. This suggests that people need to view more than 1 ECG complex to come to an accurate decision. The eye-tracking data supports this argument. People looked on average at least 4 ECG complexes before making a decision (Figures 4.5 and 4.6C). Figure 4.6A and 4.6B show a heatmap of fixations in the single complex presentation, in which the majority occurred on the end of the T-wave.

4.5.1 Study limitations and future work

This study only examined the perception of QT-interval prolongation and it is not clear whether laypeople could identify other abnormalities, such as changes in ST-segment elevation. The ECGs had a single, normal heart rate of 60 bpm. Detecting a difference could be more difficult at higher or lower heart rates, and future work should investigate this. Although detection rates in this study compared favorably to those of some clinicians [VRS$^+$05], it should be noted that a different paradigm was used in the current study (forced choice rather than classification), and as such the results are not directly comparable. The study examined people's ability to detect a QT prolongation in terms of sensitivity (identifying true positives). Future work should also examine specificity (identifying true negatives) as well as measuring the predictive positive value, which

is the proportion of positive results reported by the participant that are truly positive. This is important for understanding the practical aspects of self-monitoring.

The data used to design the stimuli were acquired from a 12-lead ECG, and not a mobile monitoring device, in which the signal is likely to be less reliable and affected by noise. The psychophysical task employed in this study can yield a biased response, as people may be more inclined to respond by saying 'different' or 'the same.' A 2-alternative forced-choice task can guard against this, as it forces the participant to choose the stimulus that has the longer QT-interval.

4.6 Conclusions

Laypeople can detect a clinically significant QT-interval prolongation in a standard ECG signal presentation, when compared with a 'normal' ECG baseline. A rhythm strip, which shows more than 1 ECG complex, is less likely to cause misperception of the QT-interval. The results show the potential for training laypeople to self-monitor their ECG outside of the clinical environment, which may help with more rapid identification of drug-induced LQTS, and enable treatment to be altered to prevent the development of life threatening complications.

Chapter 5

Evaluating the Impact of Pseudo-Colour and Coordinate System on the Detection of Medication-Induced ECG Changes

5.0 Chapter overview

5.0.1 Thesis context

The work presented in Chapter 4 shows the need to explore other approaches to ECG interpretation that can support laypeople in detecting smaller, but clinically significant, QT-interval increases from a drug-free QT baseline. These can be particularly difficult to perceive visually without precise QT-interval measurement, as discussed in Chapter 1 as the second research challenge. This chapter therefore presents a study evaluating a novel ECG visualisation technique that uses pseudo-colour to expose QT-interval duration displayed on two popular coordinate systems (Cartesian *vs.* Polar), eliminating the need to measure the QT-interval. The technique draws from the field of pre-attentive processing theory in human vision, which is introduced in Chapter 2. The Cartesian coordinate system is the standard means of presenting the ECG signal in clinical practice, and displaying pseudo-colour on Polar coordinates—which are commonly used in time-series data visualisations to represent changes over time [AJB16]—could improve people's ability to detect small changes. The study used a

'two alternative forced choice' (2AFC) psychophysical discrimination task to systematically vary the QT-interval increase at a regular heart rate of 60 bpm. Eye-tracking was used to understand the locus of attention.

The results of this study show that using pseudo-colouring on the ECG significantly improves laypeople's visual sensitivity to drug-induced QT-prolongation from a drug-free QT baseline. People can perceive increases that are much smaller than a 1mm square on the standard ECG grid (which represents 40ms), with the effect being strongest for Polar coordinates, even when T-wave morphology is abnormal. These promising results sparked our curiosity to further investigate the effectiveness of the pseudo-colouring technique with the presence of other issues known to hinder QT-prolongation detection, as extensively discussed in Chapter 2, particularly changing heart rate. In addition, given the fact that the drug-free ECG baseline may not always be available, not to mention that it would have to be adjusted according to heart rate; thus, an additional research question has to do with whether laypeople can interpret a pseudo-coloured ECG without comparing it to a drug-free ECG baseline. Therefore, we designed and developed an enhanced version of the visualisation technique that automatically adjusts the pseudo-colour according to heart rate based on a clinically reliable TdP risk assessment method known as 'QT-nomogram' [CIKD07], enabling laypeople to assess the QT-interval without having to compare it with a drug-free baseline ECG. This was empirically evaluated in Chapter 6.

The main content of this chapter is adapted from: Alaa Alahmadi, Alan Davies, Jennifer Royle, Markel Vigo, and Caroline Jay. Evaluating the impact of pseudo-colour and coordinate system on the detection of medication-induced ECG changes. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2019.

5.0.2 Author's contributions

Alaa Alahmadi developed and evaluated the pseudo-colouring visualisation, designed the study, carried out the data collection, analysed the results and wrote the paper, with Alan Davies, Jennifer Royle, Markel Vigo and Caroline Jay contributing significant edits. Jennifer Royle (patient-centricity senior program leader in the digitalECMT) provided feedback about the initial idea of patient's self-monitoring for QT-interval changes, concept and direction of the research. Alan Davies acted as the electrocardiogram domain expert throughout. Caroline Jay and Markel Vigo provided continuous guidance and discussion.

5.0.3 Published abstract

The electrocardiogram (ECG), a graphical representation of the heart's electrical activity, is used for detecting cardiac pathologies. Certain medications can produce a complication known as 'long QT syndrome', shown on the ECG as an increased gap between two parts of the waveform. Self-monitoring for this could be lifesaving, as the syndrome can result in sudden death, but detecting it on the ECG is difficult. Here we evaluate whether using pseudo-colour to highlight wave length and changing the coordinate system can support lay people in identifying increases in the QT-interval. The results show that introducing colour significantly improves accuracy, and that whilst it is easier to detect a difference without colour with Cartesian coordinates, the greatest accuracy is achieved when Polar coordinates are combined with colour. The results show that applying simple visualisation techniques has the potential to improve ECG interpretation accuracy, and support people in monitoring their own ECG.

5.1 Introduction

A side effect of commonly prescribed medications including antihistamines, antibiotics and antidepressants is prolongation of the QT-interval, or drug-induced Long QT Syndrome (LQTS) [YC03, CMY08]. LQTS is a cardiac abnormality that can increase the risk of the life-threatening arrhythmia *torsades de pointes (TdP)*, which can lead to loss of consciousness or sudden death in young, otherwise healthy people [AAKR$^+$02, MSC$^+$91, GM08, YC03]. People may not experience symptoms and an electrocardiogram (ECG) is often the only way to identify LQTS [Kha02, SC08, RV05].

ECGs are a graphical representation of the electrical activity of the heart, widely used in clinical practice to assess heart function [SSS66]. ECG results are displayed as a line on a graph-like trace, where the 'waves' (peaks and troughs) are labelled with letters and represent different stages of the heartbeat. The duration of the QT-interval (the time period between the 'Q' and 'T' waves) represents the activity of the heart ventricles (Figure 5.1).

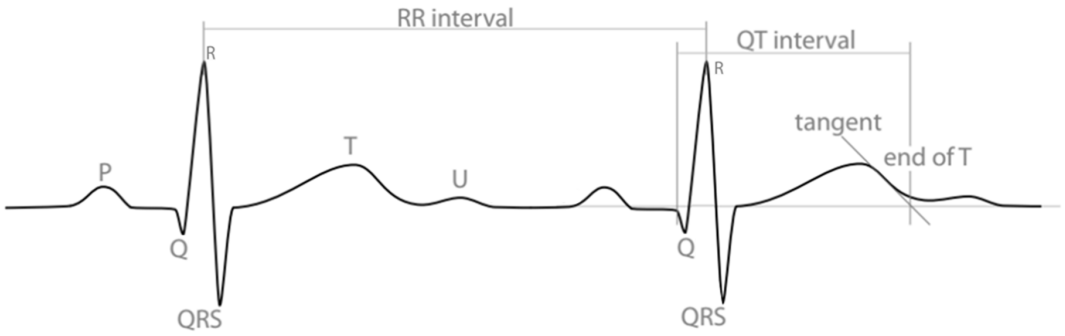


Figure 5.1: Measurement of the QT-interval on the ECG from the start of the Q-wave to the end of the T-wave.

Frequent ECG monitoring is advised for people at high risk of acquiring medication-induced LQTS [Sha02, KW12]. Recent technology innovations have made it possible to monitor ECGs outside of the clinical environment [Pod17] but this approach still relies on clinician interpretation [Loc17, AORD18]. This not only increases cost, but also makes it difficult (and sometimes impossible) to manage everyone who is at high risk. If lay people can interpret their own results, this may lead to a step-change in the detection and management of this potentially critical condition. However, ECG interpretation is known to be complex, even for clinicians [VRS$^+$05, STN08], and as such little work has examined self-monitoring.

Assessing the QT-interval, in particular, is known to be difficult [VRS$^+$05, STN08], and in a prior study the majority of clinicians were not able to recognise it [VRS$^+$05]. This may be due to the fact that whilst people find it easy to perceive quantity on a vertical scale, they are poor at judging it on a horizontal scale (see *e.g.* [LSZ$^+$09, WA84, Lib91, PK11]). Artefacts in the ECG signal can also cause misinterpretation of QT-interval length [ADVJ18].

Visualisation techniques have the potential to help highlight abnormalities within the ECG. Here we examine whether pseudo-colouring—representing continuously varying values using a sequence of colors [War12, SMY$^+$05]—and changing the coordinate

system can support lay people in identifying increases in the QT-interval. Using a psychophysical paradigm and eye tracking to systematically examine the issue, we find that:

1. Pseudo-colouring significantly increases lay individuals' ability to identify increases in the QT-interval, even when T-wave morphology is abnormal.
2. Coordinate system interacts with colour, such that people are most accurate in the condition where the ECG is presented using polar coordinates and pseudo-colour, and least accurate when presentation occurs with polar coordinates and no colour.
3. According to eye tracking data, pseudo-colour helps to focus visual attention, and people are most accurate when using the polar coordinates as this concentrates colour in the center of the screen.
4. People are significantly more satisfied when pseudo-colour is used.

5.2 Background

Previous research providing a foundation for the current study is described in this section. This covers: (1) ECG interpretation; (2) ECG visualisation methods; and (3) human perception of visualised data.

5.2.1 ECG interpretation

An ECG trace is a cyclic time series with each cycle representing a new heartbeat. The electrical activity is detected via leads placed on the body, where each lead produces a different electrical 'view' of heart activity. In hospital, clinicians commonly interpret short (10 second) ECGs via 12-leads [NKM+04]. This is the most comprehensive view, but useful ECG information can be gathered from a single lead 'view' (often used by mobile technologies). LQTS is detected by measuring from the beginning of the Q-wave to the end of the T-wave (identified using the tangent drawn at the maximum downslope of the T-wave) [AAKR+02, GMZ06] as shown in Figure 5.1.

The standard method for visualising ECG data is a Cartesian line graph showing the voltage of the heart on the Y-axis, and time in milliseconds (msec) on the X-axis [BPW+38]. A background grid supports the reader in measuring duration. To measure the QT-interval, the interpreter counts the small squares (each representing 40

msec) from the beginning of the Q-wave to the end of the T-wave [Pri10, AAKR+02, GMZ06].

Automated ECG interpretation was introduced in the 1950s to assist clinicians who had less training in ECG interpretation [Rau07]. It remains far from perfect, and even the best computational methods can produce significant errors [HG06, SW17]. Research has shown that QT-interval is underestimated or unreported by computational methods [TAS+15, KBD+18, GL13, MCbA01, TABW11, RSG09]. The main challenge lies in identifying the end of the T-wave, especially when the morphology (shape) of the T-wave is non-standard. [GPAW14, HC94, GMZ06, Mor01]. This is particularly problematic, as QT-prolonging drugs often affect the morphology of the T-wave, with some drugs (*e.g.* quinidine and ranolazine) causing large T-wave morphology changes [VJM+15].

Each person has a unique baseline ECG that reflects their individual heart function: health status, age, gender and ethnicity all influence the ECG in general, and the QT-interval in particular [GMZ06, MMDY94, HST+16]. This complicates population-level computer-derived QT calculations. Abstracting the QT-interval numerically also risks masking other potentially abnormal clinically significant changes in the ECG. For instance, specific T-wave patterns can aid detection of drug-induced LQTS [CHA+15], and large T-U waves are known to precede the life-threatening arrhythmia *Torsades de Pointes* [KFBW09]. As such the ECG morphology continues to provide the richest information for recognising LQTS. Current automated methods are thus a supplement to, rather than a substitute for, the human eye, and a combination of computer-visualisation methods with the gold-standard human interpretation remains the most accurate and reliable method [EI13, KBD+18, GL13, MCbA01, RSG09, TAS+15, SW17].

5.2.2 ECG visualisation

A number of studies have examined the effectiveness of visualisation techniques in supporting clinician-interpretation of ECGs. Chiang et al. [CCCK01] integrated ECG signals from the periodical and limb leads into two images of electrical heart function. This enabled clinicians to observe an overall integral heart view, which aided interpretation when viewed alongside the 12-lead ECG. Kors et al. [KvH08] presented a mirror image that converted the 12 leads to 24, and was shown to be effective in improving the detection-rate of heart attacks. Madias et al. [Mad04] used a 13th, multi-use lead, which provided a further 'view' of the heart. When exploring large

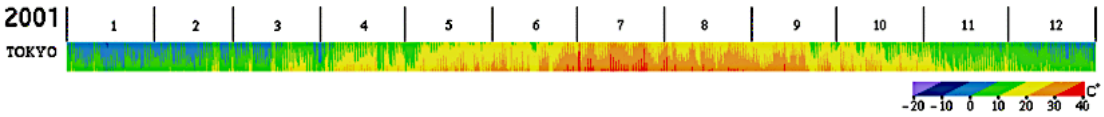


Figure 5.2: Using pseudo-colouring to represent temperature in time series data over the 12 months of the year [SMY+05].

scale ECG data, a glyph-based interactive system has been shown to be effective in detecting arrhythmia [XGC+18]. Vectorial methods have been used to represent direction and magnitude data [NWM+11, Fra56] and spatial visualisations have presented the ECG on a body surface potential map [Tac63, LSWA78]. As these methods provide data with respect to further dimensions of the heart, they are useful as a supplement, but do not replace the standard method [BFN+13]. Furthermore, previous work in this area focused on aiding clinicians. Here, the aim is to help lay-people identify when their ECG is different (*i.e.* has deviated) from their normal baseline, so they know when to seek help.

5.2.3 Human perception of visualised data

Here, we consider the problem from the perspective of the lay interpreter, rather than the data, using knowledge of visual perception to enhance the way the ECG is presented. In particular, we draw from the field of pre-attentive processing, which outlines a set of visual properties known to be detected rapidly and accurately by the human eye [Not93]. Examples of pre-attentive properties include colour, form, and spatial positioning. Using these properties in design can improve both the effectiveness and the efficiency of a visualisation [War12, HBE95, HBE96].

Colour is a pre-attentive attribute that is noticed without conscious effort [Not93, GCC17]. Many studies have shown the effectiveness of using colour to separate visual elements from their surroundings, saving the user from having to carry out a linear visual search [Hea96, War12, PGLS95]. A useful technique is pseudo-colouring, which represents continuously varying values using a sequence of colors [War12]. Pseudo-colouring is commonly used in geo- and time-series visualisations [War12, SMY+05]. Figure 5.2 shows an example of using pseudo-colouring to show changes in temperature over time.

Adnan et al. [AJB16] have examined perception of time-series visualisations. They showed Cartesian coordinates to be most effective for detecting trends and identifying

maximum and minimum values when used with positional and colour visual encodings, and Polar coordinates to be most effective for finding minimum values when using area visual encoding.

The circular layout used in the Polar coordinate system has also been employed to perceive changes in data over time. Page et al. [PSCA15, PAS+16] proposed an 'ECG Clock' generator, to visualise the changes in QT-interval values automatically generated by a 24-hour Holter ECG monitor. Circular layouts have been also used to detect symmetrical patterns in data [HGM+97] and to measure symmetry in graphs [WK17].

5.3 Method

5.3.1 Measuring visual perception

To systematically evaluate the effectiveness of pseudo-colour and coordinate system in supporting lay people's assessment of the QT-interval, we use methods from psychophysics and eye-tracking research. Psychophysical experiments investigate the relationship between physical stimuli and human perception, by varying the properties of a stimulus along one or more physical dimensions [Ste17]. Eye-tracking is used to quantify visual behavior when performing a given task, to understand differences in locus and level of attention [Duc07].

5.3.2 ECG data acquisition

The ECG datasets were taken from a clinical trial that assessed the effect of known QT-prolonging medication on healthy subjects [JVM+14]. As our work is motivated by supporting self-monitoring, we selected data from a single participant, whose QT-interval was seen to rise to clinically dangerous levels. The subject was a 35-year-old male who had normal QT-intervals (<430 milliseconds) prior to taking the medication 'Dofetilide' (an antiarrhythmic drug); he subsequently experienced a gradual increase in the QT-interval, which eventually reached very high levels (QT=579 milliseconds). The ECGs sampled all have a regular heart rate (HR=60 bpm) and are from lead-II, which is typically used to measure the QT-interval. The QT-values of the selected ECGs were 417, 421, 441, 447, 455, 468, 485, 537, 565 and 579 milliseconds. We categorised these values based on their clinical significance: normal (QT <430); borderline (QT >430 and <470); prolonged (QT >470 and <500); very prolonged levels

(QT >500) [JA09]. The open ECG dataset is available from the PhysioNet database [GAG+00].

5.3.3 Visualisation design

We used a co-design approach, creating the visualisation techniques with an expert in ECG interpretation (to ensure accuracy), and refining them with input from lay people. As a first step, R-peaks were detected in the raw ECG datasets, and a dashed vertical line used to show the halfway point of the R-R interval (Figure 5.1). This helps to identify the area of interest containing the QT-interval. Note that it is easy to detect the R wave in the vast majority of ECGs, as (unlike the other waveforms which vary considerably) it consistently has the greatest amplitude. We then applied pseudo-colouring, to shift the 'work' of QT-interval visual encoding from perceiving distance between two waves, to perceiving colour in terms of hue and intensity.

As spectrum-approximation sequences in particular help with reading values [War88], we used these as a foundation for the pseudo-colouring technique. Cool spectral colour codes (purple to blue to green) were used to indicate normal QT-interval ranges, and warm colours (yellow to orange to red) to show abnormal QT-interval ranges. We applied the pseudo-colouring sequence in the area between the 0 voltage baseline up to (or down to) the signal, from the beginning of the R-wave to the R-R interval halfway point (see *e.g.* Figures 5.3 and 5.4). The pseudo-colouring sequence was mapped to the ECG signal such that the colour code changed every 40 milliseconds, which is equal to a small square on the standard ECG background grid.

To understand the impact of coordinate system on ECG data interpretation, we displayed the ECG signals on Cartesian and Polar coordinates with and without pseudo-colouring. We used R [IG96] with RStudio software version 1.1.447 to create the visualisations. Figures 5.3 and 5.4 show ECGs with very prolonged and normal QT-intervals with and without the pseudo-colouring sequence on Cartesian coordinates. Figures 5.5 and 5.6 show the same ECGs, but on Polar coordinates. The ECGs are reduced in size for inclusion in the paper. The full size images, along with the scripts used to create them, can be found in the supplementary materials, and in our repository[1].

[1] `https://github.com/mbchxaa6/ECG_QT_Visualisation`.

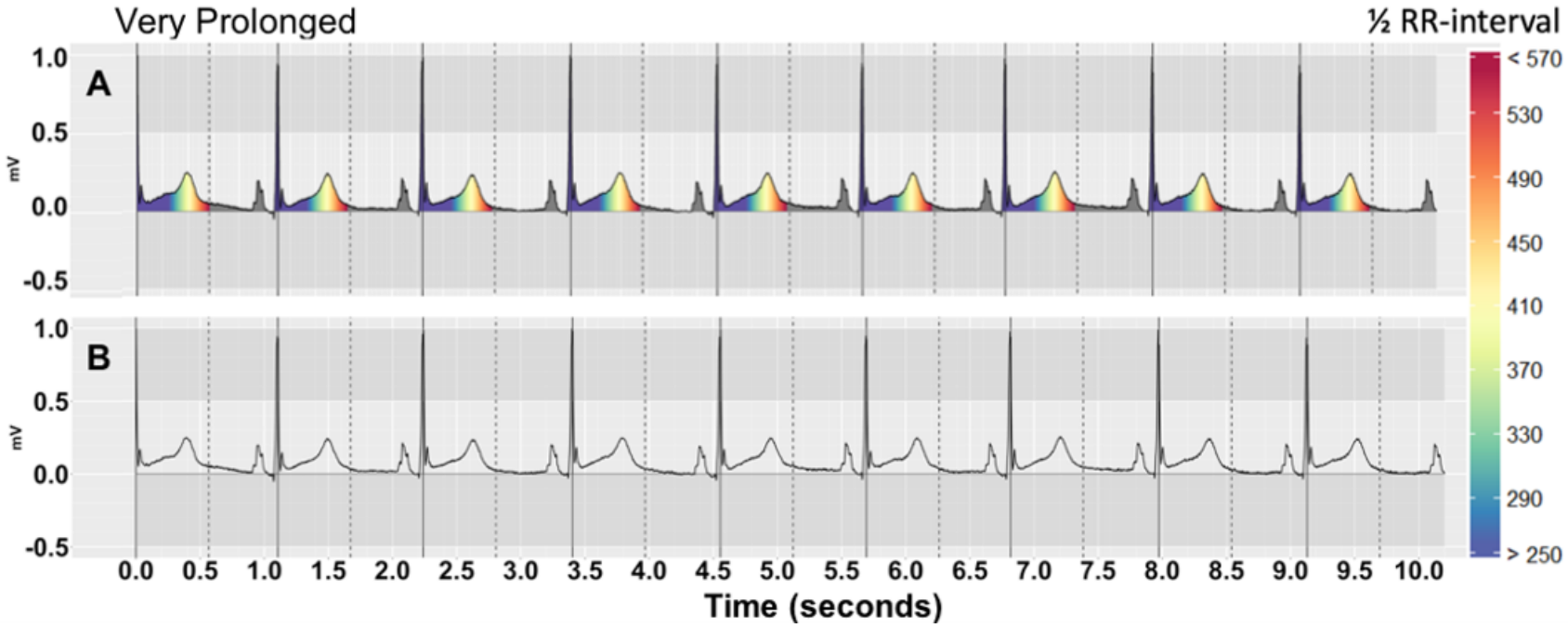


Figure 5.3: ECGs with very prolonged QT-intervals (QT= 579ms) on Cartesian coordinates with (A) and without (B) pseudo-colour.

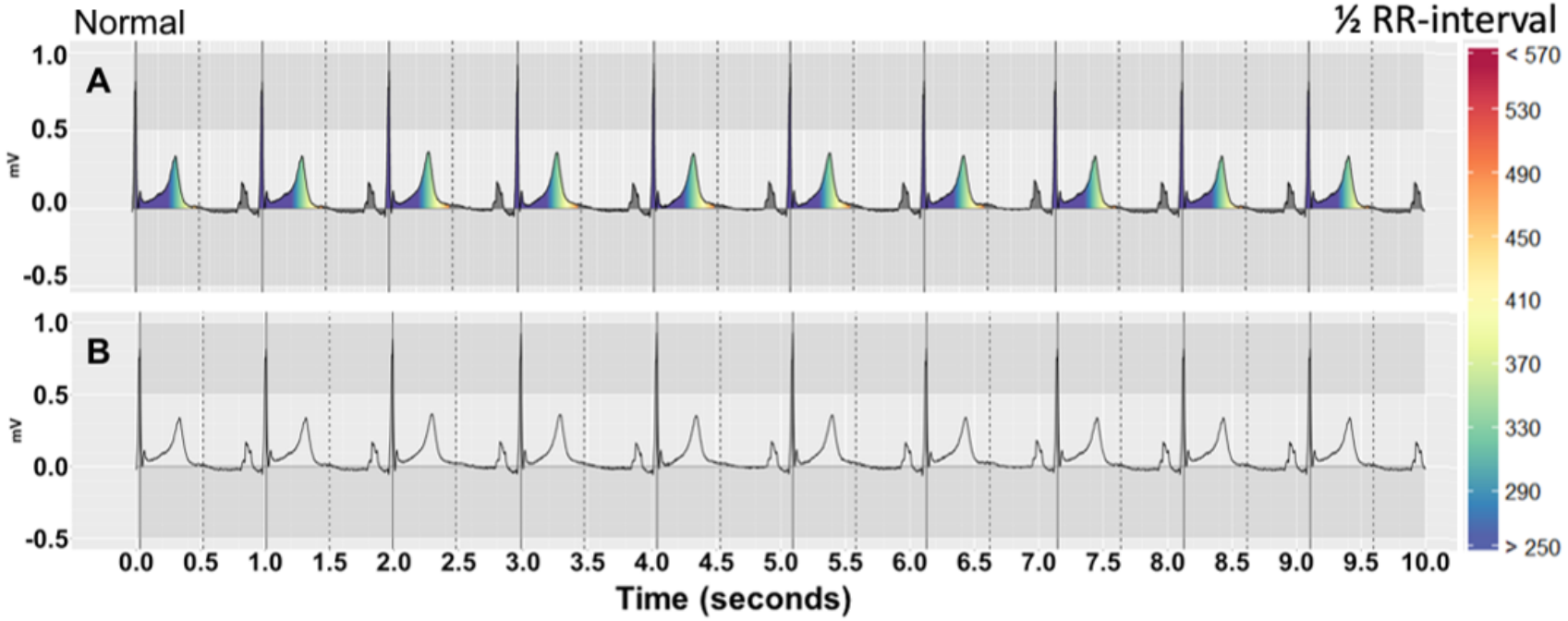


Figure 5.4: ECGs with normal QT-intervals (QT= 417ms) on Cartesian coordinates with (A) and without (B) pseudo-colour.

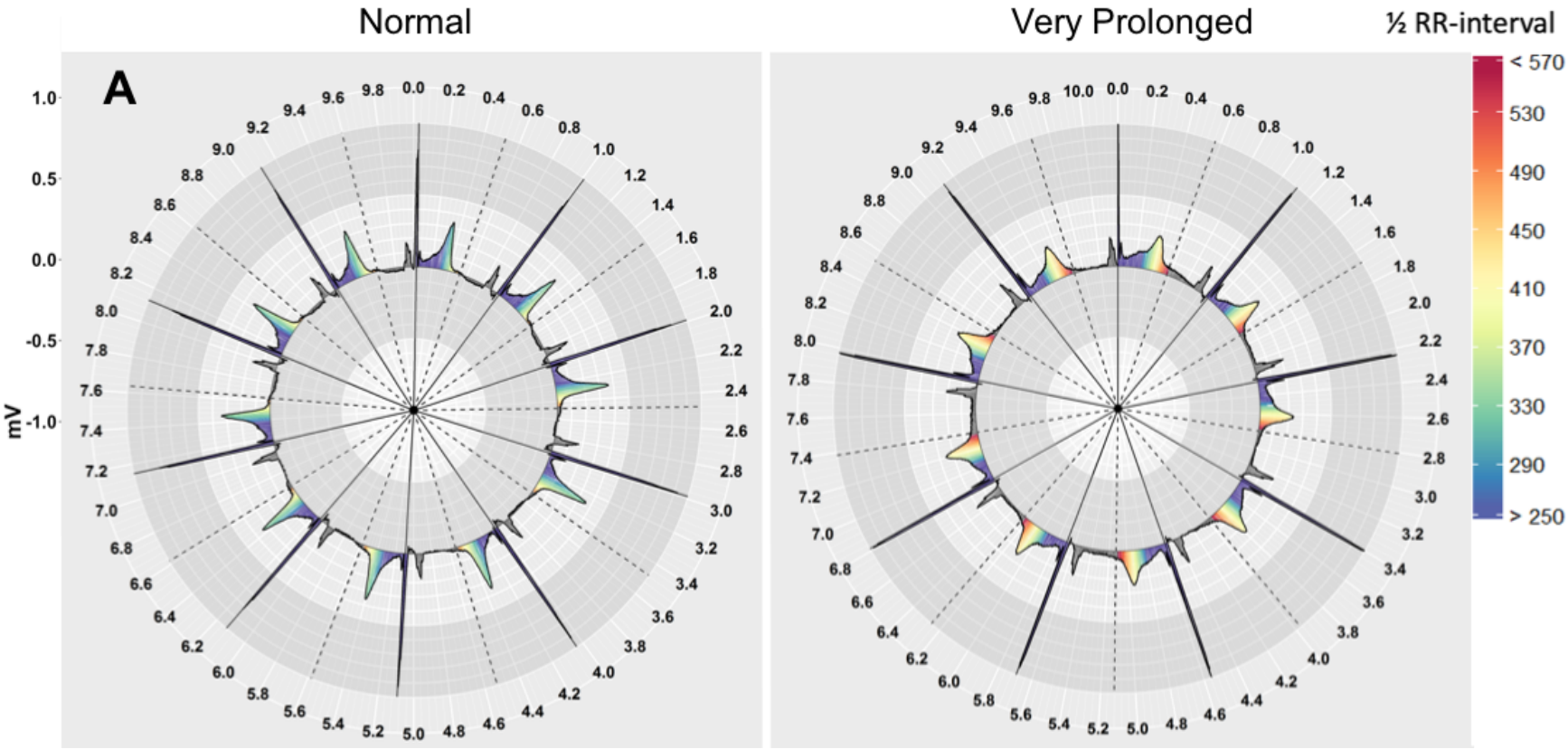


Figure 5.5: ECGs with normal (QT= 417ms) and very prolonged (QT= 579ms) QT-intervals on Polar coordinates with pseudo-colour.

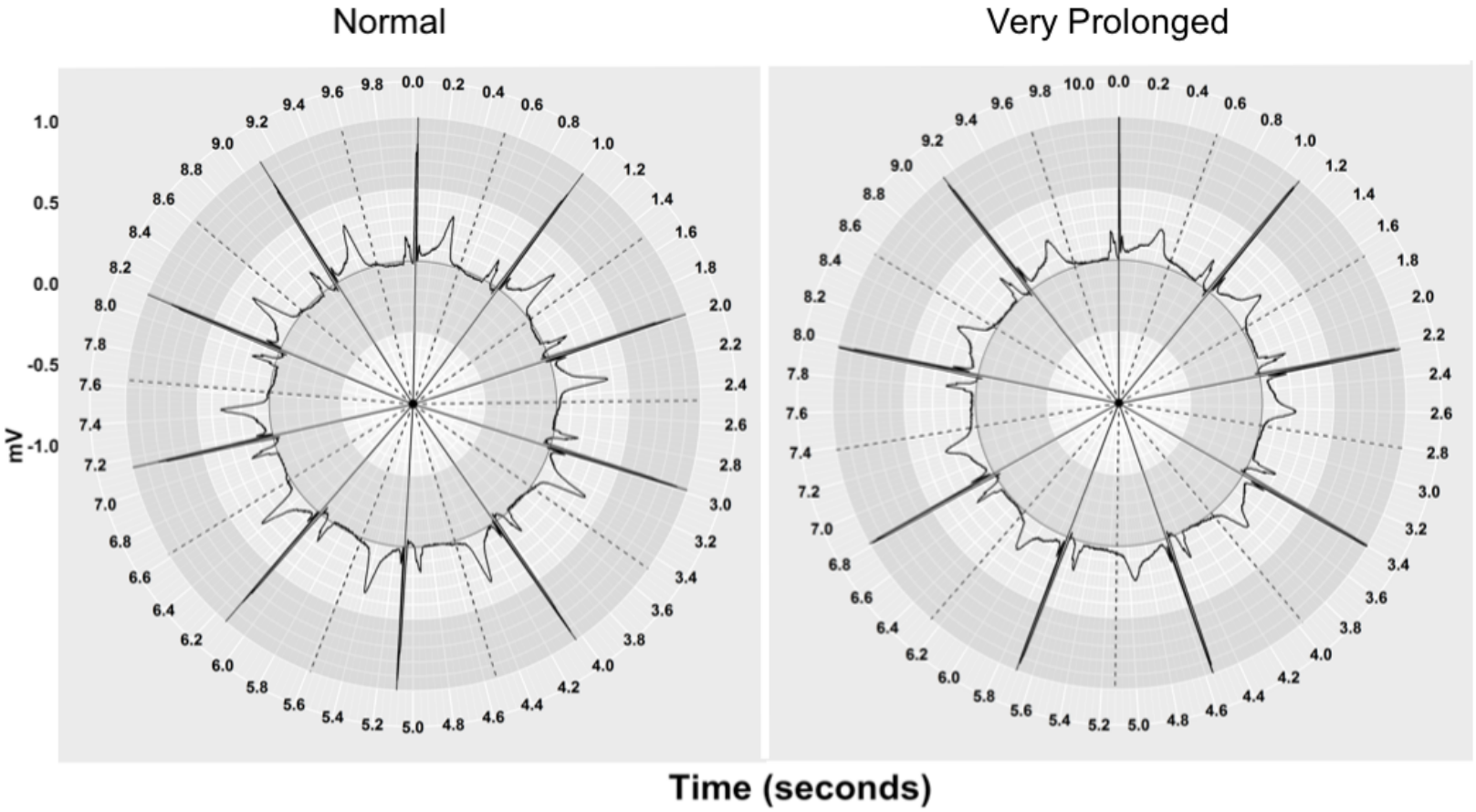


Figure 5.6: ECGs with normal (QT= 417ms) and very prolonged (QT= 579ms) QT-intervals on Polar coordinates without pseudo-colour.

5.3.4 Experiment design

We hypothesized that changes in the T-wave morphology (*e.g.* flattening of the wave, which can be caused by QT-prolonging medication [VJM+15]) might cause misperception of the QT-interval, and included this as a factor. The study thus used a counterbalanced within-subjects design with three independent variables, each with two levels:

1. Colour-coding: no colouring; pseudo-colouring.
2. Coordinate system: Cartesian; Polar.
3. The T-wave morphology: normal; abnormal.

The within-subjects factorial design yielded a total of 8 (2x2x2) experimental conditions for each participant. We counterbalanced the order of visualisation presentation using a balanced Latin square to minimize practice effects. The first forty participants were divided into four groups of ten participants. The two remaining participants were added to different groups at random. Considering (A) to be 'Cartesian no-colouring', (B) to be 'Cartesian pseudo-colouring', (C) to be 'Polar no-colouring' and (D) to be 'Polar pseudo-colouring', the counterbalanced order of visualisation presentation for group 1 to 4 respectively was as follows: A, B ,D, C; B, C, A, D; C, D, B, A; and D, A, C, B.

We assessed the effects of T-wave morphology on two separate tasks (described below). The order of the T-wave morphology condition (normal or abnormal) was counterbalanced across participants by dividing them into two groups: one group was presented with the normal T-wave condition first and the abnormal condition second; this order was switched for the second group. The dependent variables were response correctness, reaction time, fixation location, mean fixation duration, and satisfaction.

5.3.5 Participants

Forty two participants (22 males and 20 females) were recruited from a university campus. Eligibility for the study was determined by asking participants to rate their knowledge of ECGs/ECG interpretation, and including only people who reported no knowledge at all. Participants consisted of 34 students and 8 staff. The mean age was 30 (SD=7). The backgrounds of the participants were Computer Science (n=27), Education (n=3), Chemical Engineering (n=3), Electrical Engineering (n=3), Mathematics (n=4), History and Sociology (n=1) and Music/Violin Performance (n=1). Their sight was normal or corrected-to-normal and they reported no motor or neurological disorders.

5.3.6 Task and procedure

Participants completed a 10-minute training session where they were told about the motivation for the research (the potential for self-monitoring ECGs for drug-induced QT-prolongation), and shown how to identify the QT-interval using the standard ECG signal representation and the different visualisation techniques. Each participant then completed an assessment task to check that they understood how to perform the measurement, where they were asked to highlight the start and end point of the QT-intervals on two different ECGs—one with a normal QT-interval and the other with prolonged QT-interval—using the four visualisation techniques (*i.e.* 2 ECGs $\times$ 4 visualisation techniques = 8 stimuli). Participants were informed that the greater the amount of warm colors within the T-wave area, the longer the QT-interval. All participants passed the assessment task, correctly assessing the QT-intervals for all stimuli.

As the task requires differentiating between spectrum-approximation pseudo-colouring sequences, only participants with normal vision were included in the experiment. Whilst spectrum-approximation pseudo-colouring sequences presented in greyscale are potentially readable by people with colour blindness, as long as they vary by lightness or contrast, this would need further testing to ensure its efficacy [War21]. Another solution is to use a pseudo-colouring sequence that can be distinguished by most colour-blind individuals such blue-to-yellow and blue-to-red variations [War21].

The experiment used a 'two alternative forced choice' (2AFC) psychophysical discrimination task [Ste17]. Within a trial, the participant was presented with two ECG stimuli; a baseline showing no QT-prolongation and a comparator showing an increased (or the same) QT-interval; the participant had to select the ECG that they perceived to have the longer QT-interval using the left/right arrows in the Polar condition, and up/down arrows in the Cartesian condition, according to the stimulus' position on the screen. Participants completed all trials for one visualisation technique before moving to the next. The location of the ECG with the longer QT-interval (*i.e.* top/bottom or left/right) was counterbalanced in the design; stimuli were then presented at random. To determine the effects of T-wave morphology, we split the experiment into two separate 2AFC tasks, as follows.

Normal-T-wave In this condition, participants completed a total of 20 experimental trials. In each trial, two ECGs were presented; a baseline ECG showing a normal QT-interval with a normal T-wave morphology and a comparator ECG. Two trials showed exactly the same ECG for the baseline and the comparator stimuli, in order to test the validity of our method (the probability of choosing each alternative should be equal to 0.5). The other 18 trials presented the same ECG baseline (QT = 417 msec), and an ECG with a longer QT-interval that was selected from the following set of QT values, where each value was presented twice: 421, 441, 447, 455, 468, 485, 537, 565, 579 milliseconds. Figures 5.5, 5.6, 5.3 and 5.4 show examples of the ECGs used in the normal T-wave morphology condition.

Abnormal T-wave This condition was used to evaluate whether the visualisation techniques can help people to perceive QT-prolongation regardless of the T-wave morphology. To reduce potential fatigue, this condition contained only 8 trials. In each trial, two ECGs were presented: a baseline ECG showing a borderline QT-interval with an abnormal (flat) T-wave morphology and an ECG with an increased QT-interval. Participants had to choose the ECG with the longer QT-interval. The borderline ECG had a QT value of 447ms and the comparator ECG was selected from the following set of QT values, which had either a normal or abnormal T-wave as indicated: 468 (abnormal), 485 (normal), 565 (abnormal) and 579 (normal) milliseconds. Figures 5.7 and 5.8 show examples of the ECGs used in the abnormal T-wave morphology condition.

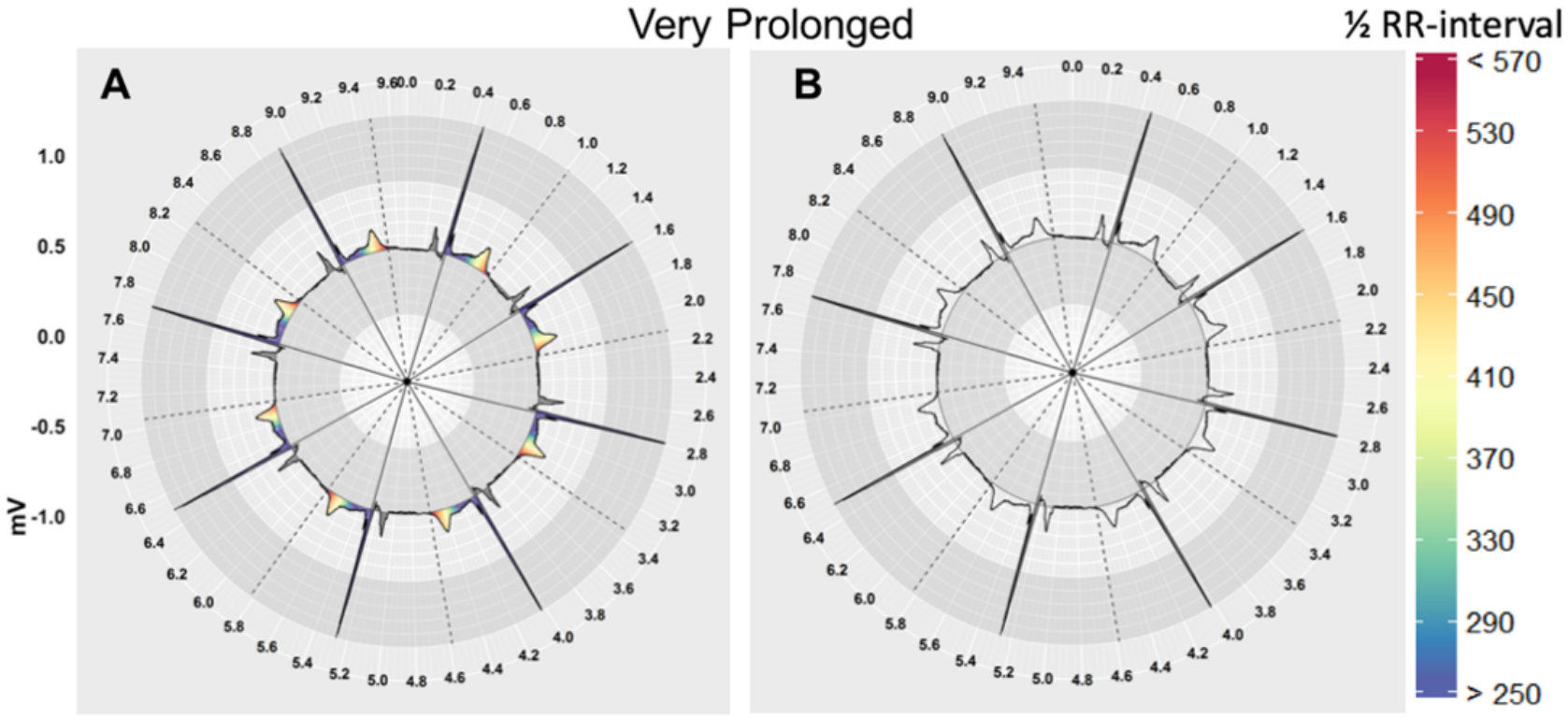


Figure 5.7: ECGs with very prolonged QT-interval (QT= 565ms) and abnormal T-wave morphology on Polar coordinates with (A) and without (B) pseudo-colour.

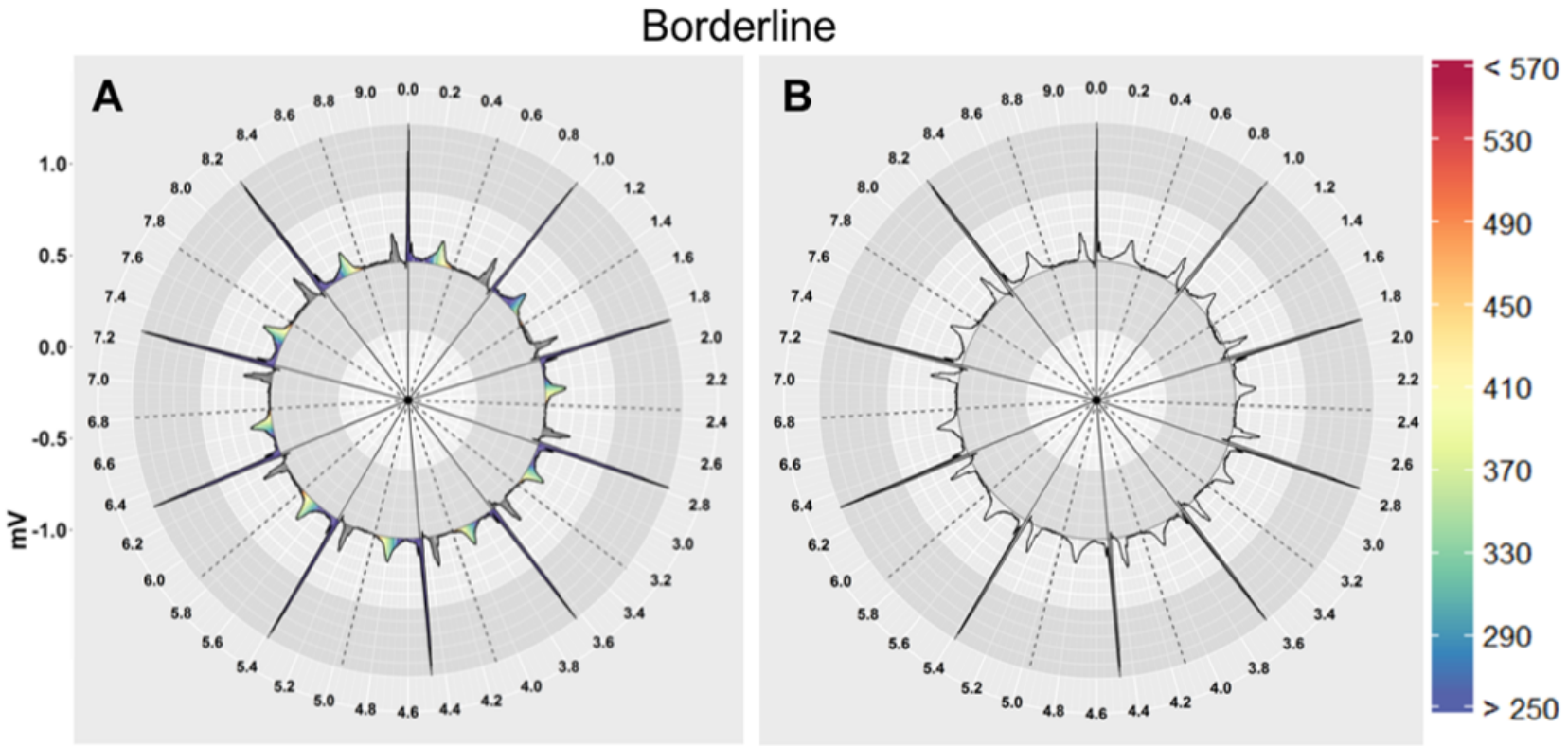


Figure 5.8: ECGs with borderline QT-interval (QT= 447ms) and abnormal T-wave morphology on Polar coordinates with (A) and without (B) pseudo-colour.

5.3.7 Apparatus

A Tobii Pro Spectrum eye-tracker and Tobii Pro lab 1.95 software were used to record eye gaze with a sampling rate of 600 HZ. Key press events were recorded to collect participants responses. The study was performed on a 23.8 inch (diagonal) Tobii Pro Spectrum eye-tracking monitor, with a resolution of 1920 x 1080 pixels. Each Cartesian coordinate ECG stimulus was 32.31cm x 6.14cm, and each Polar coordinate stimulus was 15.61cm x 12.93cm.

5.4 Results

All anonymised raw data, along with relevant R-scripts and SPSS outputs are available in our Github repository[2].

5.4.1 Accuracy

Psychometric function

We used a psychometric function, which is an inferential model employed in psychophysical detection and discrimination tasks, to model the relationship between the gradual increase in the QT-interval and the correctness of participants' responses across the four visualisation techniques. The psychometric function was plotted as the percentage of correct responses (trials where the longer QT-interval stimulus was correctly identified) as a function of the QT-interval increase (Figure 5.9). The results show that pseudo-colour significantly improves perception of QT-interval increases regardless of the T-wave morphology, with people able to detect smaller increases with the Polar coordinates than the Cartesian coordinates. This is important, as even these small increases are clinically significant.

When pseudo-colour is not used, T-wave morphology interacts with coordinate system in the detection of the QT-interval increases. When the T-wave morphology is normal, people perform better with Cartesian coordinates (Figure 5.9 (A)). However, when the T-wave morphology is abnormal (increased flattening of the T-wave), people perform better with Polar coordinates (Figure 5.9 (B)).

[2] `https://github.com/mbchxaa6/Data_Analysis`.

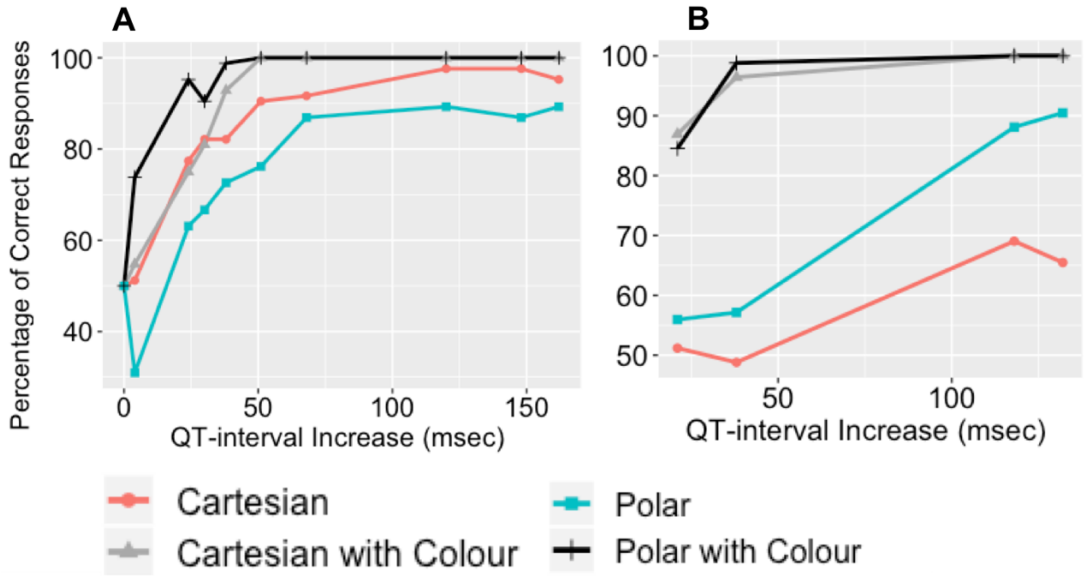


Figure 5.9: The psychometric function plot shows the percentage of correct responses as a function of the QT-interval increase from the baseline with (A) Normal T-wave morphology and (B) Abnormal T-wave morphology.

Just noticeable difference (JND) threshold

In psychophysics, the JND threshold is defined as the minimum amount of change in a stimulus necessary for it to be 'just noticeable'. In this study, we defined it as the minimum increase in the QT-interval required for it to be detectable. We estimated the 75% JND threshold as the value of the QT-interval increase from the normal baseline at which the percentage of correct responses is equal to 75%. Only the normal T-wave morphology condition was used for estimating the JND, as the abnormal condition contained insufficient trials for it to be fitted with this statistical model. The JND thresholds, determined by fitting the psychometric function using a logistic function with maximum likelihood estimation (MLE) (Figure 5.10), were 29, 19, 65 and 9 milliseconds for *Cartesian*, *Cartesian with pseudo-colour*, *Polar* and *Polar with pseudo-colour* respectively. Pseudo-colour thus reduces the JND in both co-ordinate systems, with the effect being strongest for Polar co-ordinates.

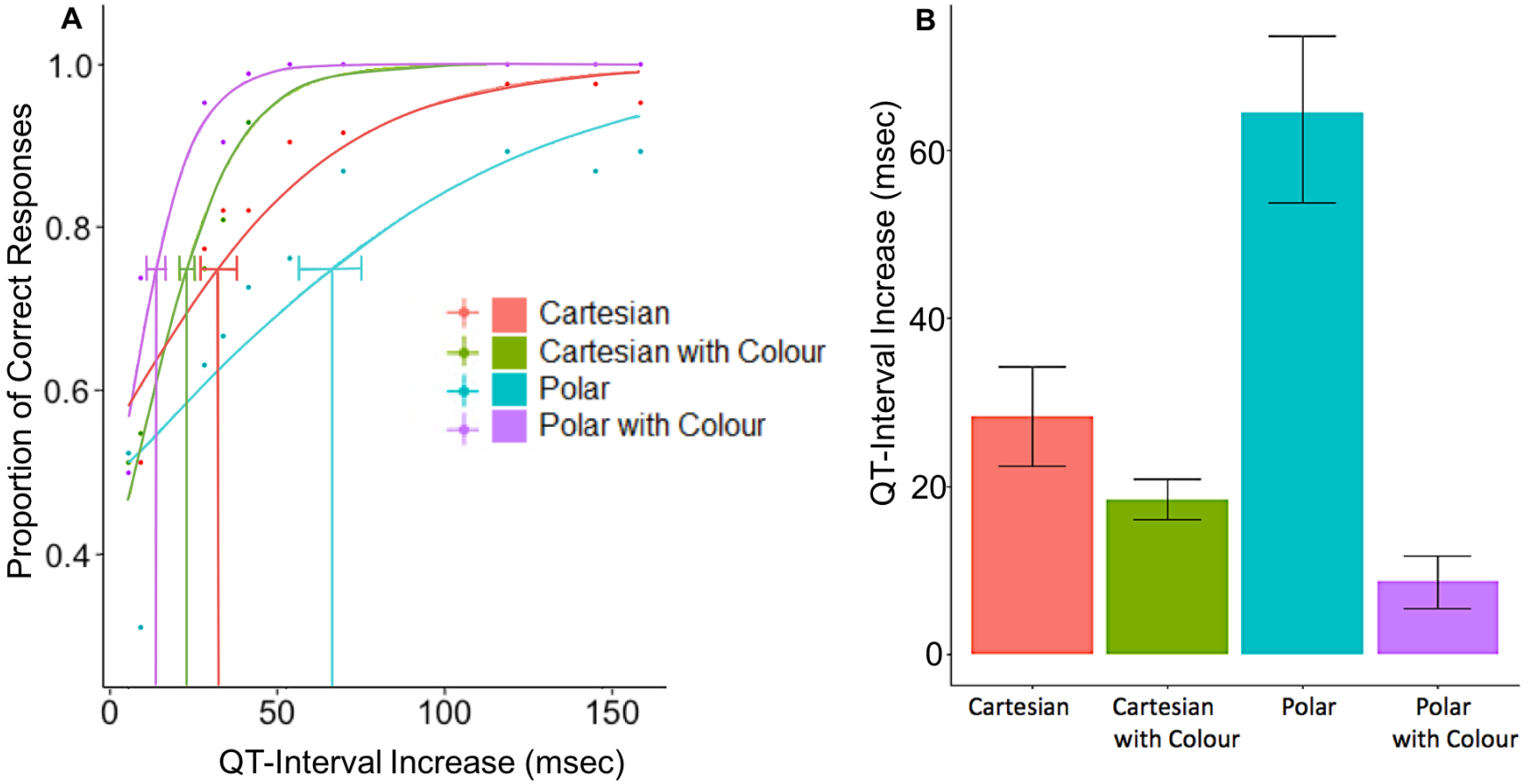


Figure 5.10: (A) The fitted psychometric function plot shows the proportion of correct responses as a function of the QT-interval increase from the normal baseline. (B) The just noticeable difference (JND) thresholds plot. The error bars represent bootstrap confidence intervals.

5.4.2 Reaction time

We measured reaction time as the period between the appearance of the stimuli on the screen and the key press event when people made their decision. As shown in Figure 5.11, pseudo-colour reduced the reaction time as the QT-interval increased in all conditions.

A Shapiro-Wilks test showed the reaction time data was not normally distributed ($p < 0.05$). We thus used a non-parametric Friedman test to compare reaction times across the four conditions. The test was conducted for each QT-interval increase and under each condition of the T-wave morphology. For all QT-interval increases, there was a statistically significant difference in reaction time according to visualisation technique, under both conditions of the T-wave morphology ($p < 0.05$) (Table 5.1).

To examine where the differences actually occur, *post hoc* pairwise comparisons were performed using a Wilcoxon signed-rank test with Bonferroni correction ($\alpha = 0.008$). This showed that when the QT-interval was clinically prolonged (equal to 485 msec and increased from the baseline by 68 msec) or very prolonged (greater than 500 msec and increased from the baseline by over 100 msec), reaction time was significantly faster when pseudo-colour was used, for both types of coordinate system ($p < 0.008$), regardless of T-wave morphology.

When the QT-interval was in the borderline range, the T-wave morphology and co-ordinate system interacted with pseudo-colour. In the trial that shows a borderline QT-interval (increased by 38 msec) with a normal T-wave morphology, reaction time when pseudo-colour was used was significantly faster for Polar coordinates than Cartesian coordinates ($Z = -3.806, p < 0.008$, Figure 5.11 (A)). However, in the trial showing a borderline QT-interval (increased by 21 msec), but with an abnormal T-wave morphology, there was the opposite effect, with people responding faster in the Cartesian with pseudo-colouring condition than in the Polar with pseudo-colouring condition ($Z = -2.870, p < 0.008$, Figure 5.11 (B)).

5.4.3 Eye-tracking metrics

To calculate eye movement metrics, Tobii Pro lab software was used to create two areas of interest (AOIs) for each experimental trial: one for the baseline ECG stimulus, and one for the comparator ECG stimulus.

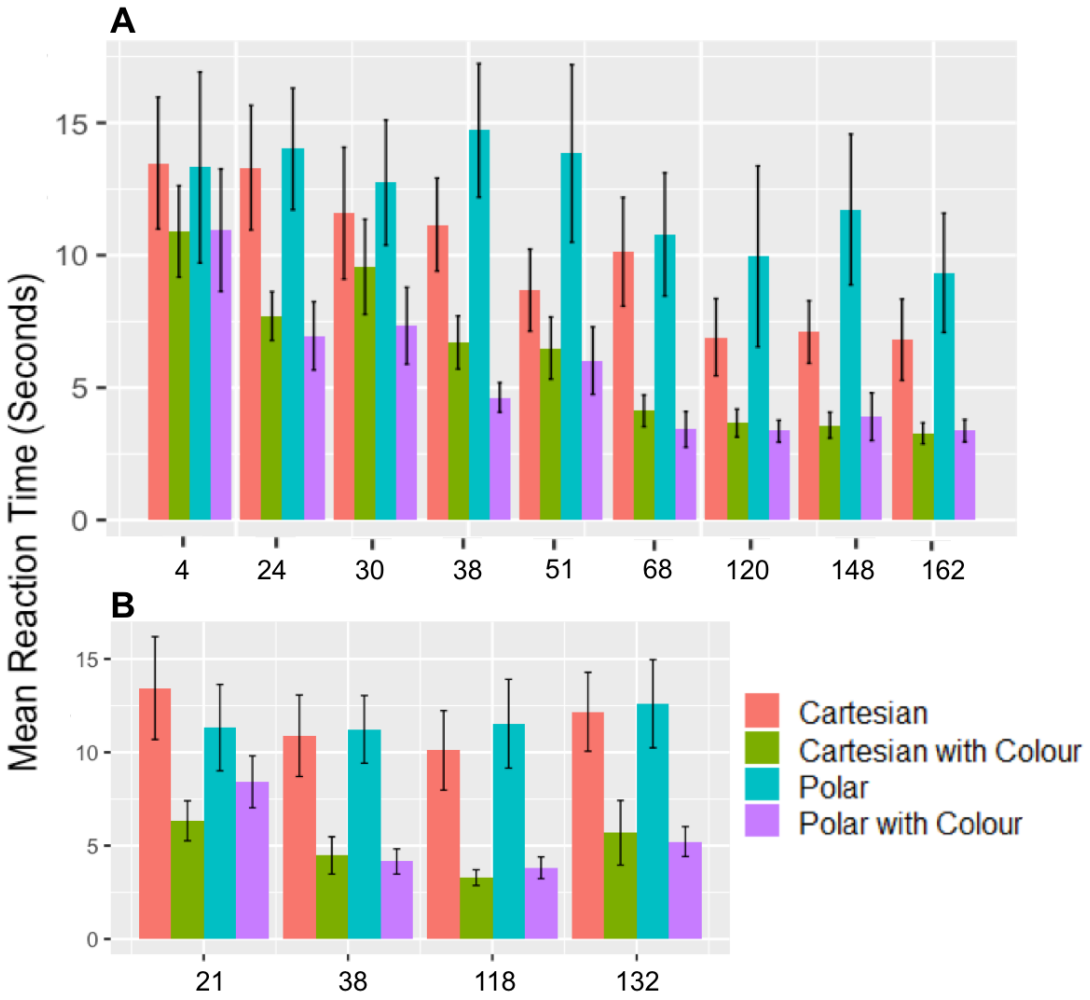


Figure 5.11: Mean reaction time in seconds over the QT-interval increases (msec) from the baseline with (A) Normal T-wave morphology (B) Abnormal T-wave morphology. Error bars represent 95% confidence intervals.

Table 5.1: Results of the Friedman test comparing reaction times in the four visualisation conditions for all QT-interval increases, and in each condition of the T-wave morphology. QT represents the value of the longer QT-interval in milliseconds. Δ QT represents the difference in milliseconds between the value of the longer QT-interval and the baseline QT-interval (*i.e.* the amount of QT-interval increase).

T-wave	QT	Δ QT	Range	$\chi^2(3)$	P-value
Normal	421	4	Normal	9.543	$p < 0.05$
morphology	441	24	Borderline	68.100	$p < 0.05$
	447	30	Borderline	13.443	$p < 0.05$
	455	38	Borderline	101.954	$p < 0.05$
	468	51	Borderline	41.471	$p < 0.05$
	485	68	Prolonged	101.886	$p < 0.05$
	537	120	Very prolonged	79.225	$p < 0.05$
	565	148	Very prolonged	90.286	$p < 0.05$
	579	162	Very prolonged	72.265	$p < 0.05$
Abnormal	468	21	Borderline	50.671	$p < 0.05$
morphology	485	38	Prolonged	114.783	$p < 0.05$
	565	118	Very prolonged	138.529	$p < 0.05$
	579	132	Very prolonged	94.409	$p < 0.05$

Mean fixation duration

The mean fixation duration metric, which is an indicator of cognitive load [HNA+11, DVDL14], was used to understand whether the visualisation techniques helped people to focus on the target ECG stimulus that had the longer QT-interval. As shown in Figure 5.12, regardless of T-wave morphology and coordinate system, using pseudo-colour results in longer fixations on the stimulus with the longer QT-interval, compared with the baseline stimulus, and this effect becomes more pronounced as the QT-interval increases.

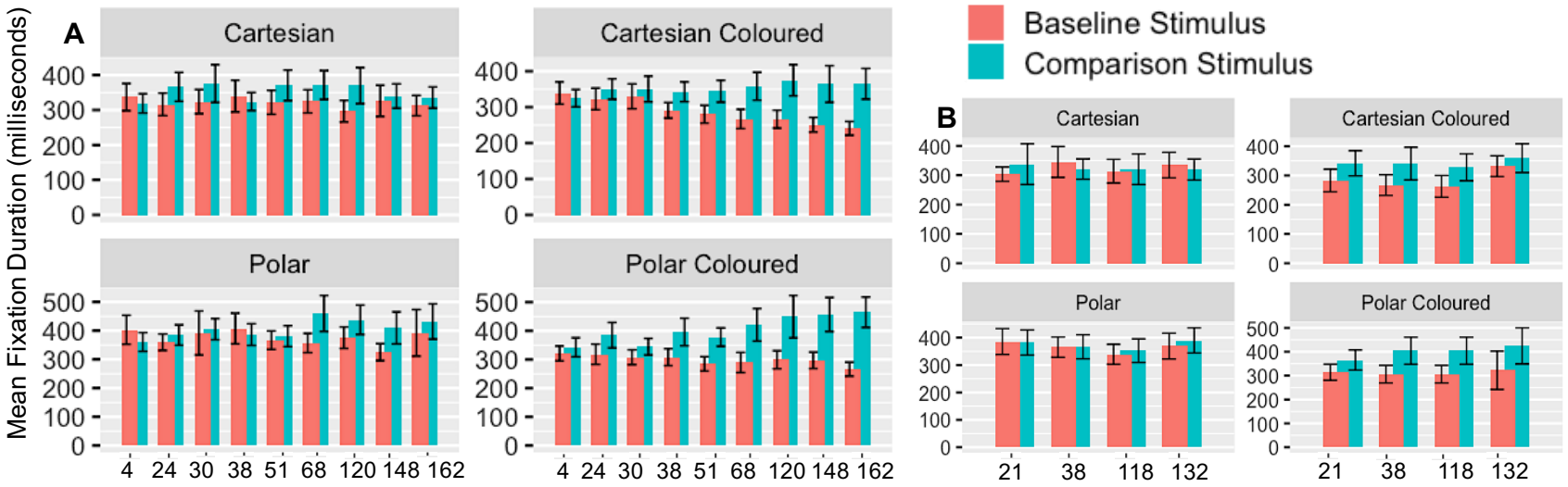


Figure 5.12: Mean fixation duration of the baseline and the comparator stimuli over the QT-interval increases (msec) with (A) Normal T-wave morphology and (B) Abnormal T-wave morphology. The error bars represent 95% confidence intervals.

5.4.4 Satisfaction

Following the experiment, participants completed a five point Likert-type scale ranging from 'bad' (1) to 'good' (5) to rate the effectiveness of each visualisation technique in supporting the detection of increases in the QT-interval. A Friedman test showed there to be a statistically significant difference in satisfaction depending on which visualisation technique was used ($\chi^2(3) = 90.860, p < 0.05$). A *post-hoc* analysis with a Wilcoxon signed-rank test utilising a Bonferroni correction ($\alpha = 0.008$) showed a significant preference for pseudo-colour ($p < 0.008$). However, although people were faster and more accurate in the Polar coordinate condition when pseudo-colour was used, there was no difference in people's satisfaction for either coordinate system ($Z = -0.435, p = 0.664$).

5.5 Discussion

Recognizing QT-interval prolongation on the standard ECG is difficult. Previous research assessed the ability of medical professionals to recognise LQTS when presented with four ECGs (two ECGs with LQTS patients and two ECGs of healthy subjects). The results showed that accurate classification of all QT-intervals as either 'prolonged' or 'normal' was achieved by 96% of QT experts and 62% of arrhythmia experts, but by less than 25% of cardiologists and noncardiologists [VRS+05]. The QT-interval is also underestimated or unreported by computerised methods [TAS+15, KBD+18, GL13, MCbA01, TABW11, RSG09], and as such, human visual validation is strongly recommended [EI13, SW17]. To support lay people, who have no experience in ECG interpretation, in detecting life-threatening changes in the ECG, we need to understand how people perceive ECG data, and the extent to which visualisation techniques can aid the interpretation process.

We used psychophysical methods to model lay people's detection of QT-interval increases when using four visualisation techniques. The results show that using pseudo-colour to represent time significantly improves accuracy in detecting increases in the QT-interval, for both coordinate systems. People are most accurate in detecting small, but clinically significant increases in the QT-interval with Polar coordinates, regardless of whether the T-wave morphology is normal or abnormal (Figure 5.9).

Clinical research has shown that even a small ($\approx$ 10 msec) QT-interval increase from the baseline is considered a significant side effect of a QT-prolonging drug [RFF+09, Bre10, FDA+05]. When the T-wave morphology is normal, the 75% just noticeable

difference (JND) thresholds were 29, 19, 65 and 9 milliseconds for *Cartesian*, *Cartesian with pseudo-colour*, *Polar* and *Polar with pseudo-colour* respectively. This shows that using a combination of Polar coordinates and pseudo-colour has the potential to support lay people in detecting the smallest clinically significant change. It also shows that colour can improve sensitivity to changes such that people can perceive increases that are much smaller than a 1mm square on the standard ECG grid (which represents 40 msec). As well as improving accuracy, using pseudo-colour reduced reaction times and increased attention to the longer QT-interval stimulus. Eye-tracking data showed that the average fixation duration increased more on the comparator stimulus, which has the longer QT-interval, than the baseline stimulus, as the interval length increased (Figure 5.12). Figures 5.13 and 5.14 show a heat map of absolute fixation duration across all participants on Cartesian and Polar coordinates respectively, demonstrating that even when the interval is borderline, rather than prolonged (QT = 455 msec, increased by 38 msec), people still fixate longer on the comparator stimulus. This shows the power of the colour codes used with the spectrum-approximation pseudo-colouring sequence, where warmer colours including orange and red help to attract attention to abnormal QT-interval levels.

5.5.1 Visualisation design implications

This study shows that colour as a pre-attentive attribute can support the detection of small differences in time-series data represented along a continuous scale, that are otherwise difficult to perceive. While people find it relatively easy to perceive quantity along a vertical scale, they are known to be poor at judging size or quantity displayed along a horizontal scale (see *e.g.* [LSZ+09, WA84, Lib91, PK11]). Time-series data are conventionally displayed horizontally. Although this study focused on a specific problem within ECG interpretation, the results may have a wider application to other forms of time-series data, for example, in supporting detection of change in seasonality in financial data, or follow-up months of survival rate among cancer treatments (*e.g.* in Kaplan-Meier curves). People are able to detect the smallest differences when the ECG is presented using Polar coordinates and pseudo-colour. Eye-tracking research has shown that people's initial eye movements are more commonly located in the center of the screen [Bin10]. According to the study's eye-tracking data, the warmer hues of the pseudo-colour helped to focus visual attention; as Polar coordinates concentrate more colour in the center of the screen than Cartesian coordinates, the increased salience may be easier to perceive in foveal vision.

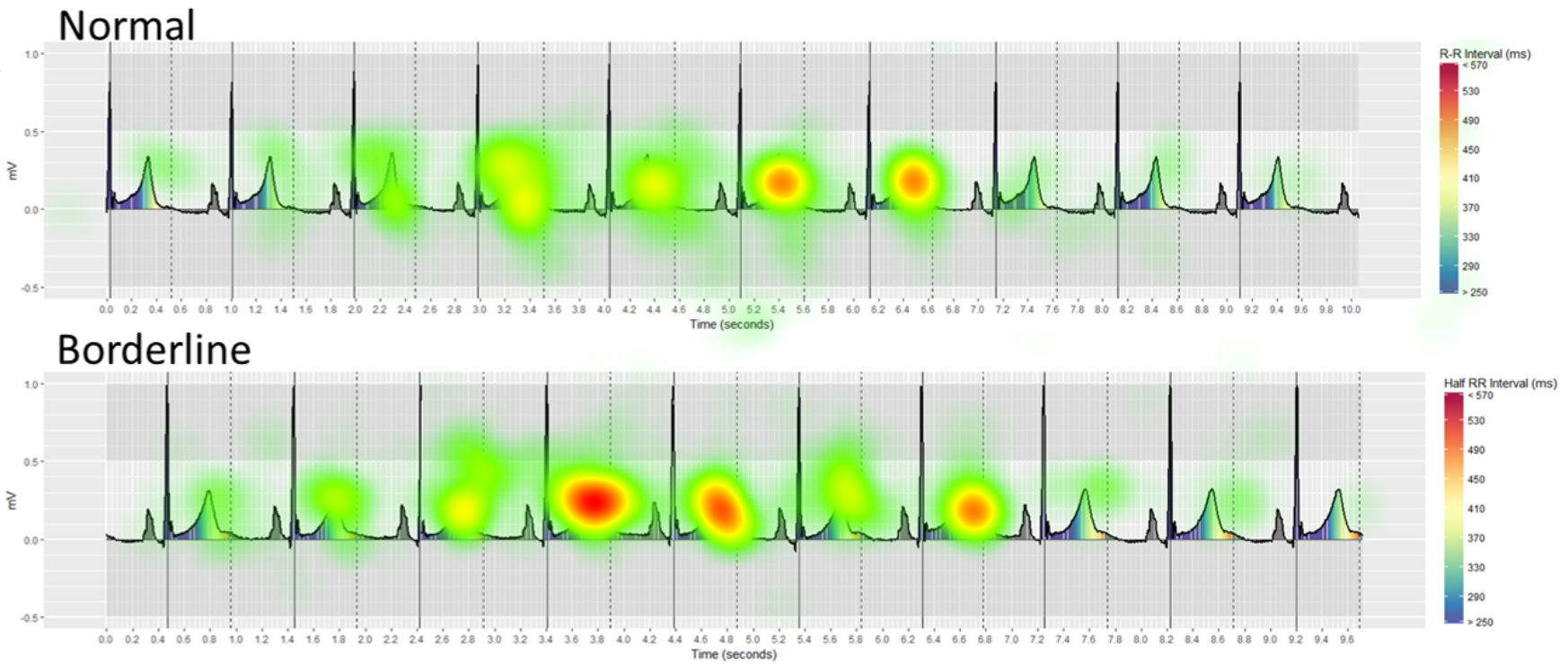


Figure 5.13: Heatmap of absolute fixation duration for all participants on Cartesian coordinates. Fixation is longer on the borderline QT-interval.

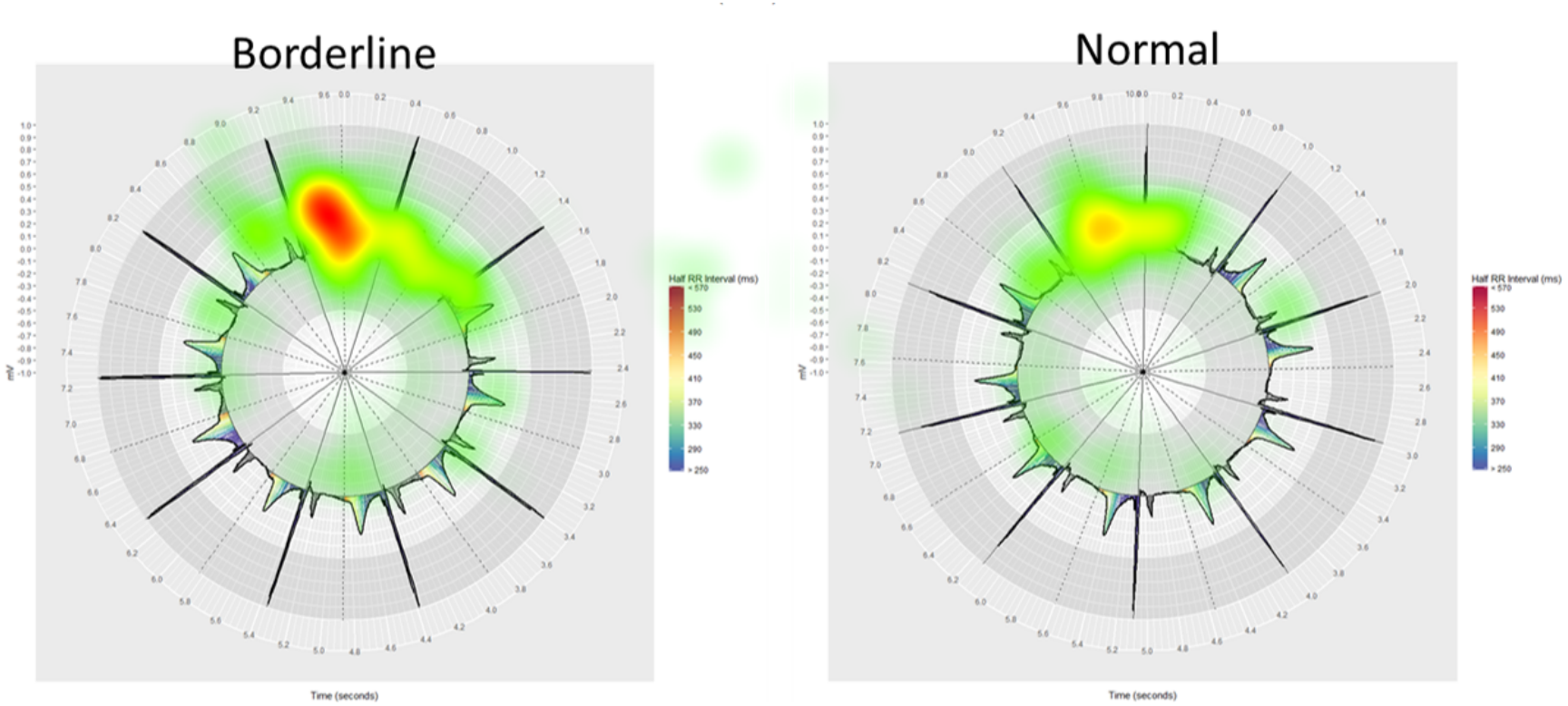


Figure 5.14: Heatmap of absolute fixation duration for all participants on Polar coordinates. Fixation is longer on the borderline QT-interval.

5.5.2 Limitations and future work

Limitations of this study include: (1) we investigated detection of QT-interval prolongation, and it is not clear whether these techniques would generalise to interpretation of other ECG abnormalities, such as changes in ST-segment elevation, or to other signal/time-series data; (2) the data used to design the stimuli were from a high quality signal with little noise; they were acquired from a 12-lead ECG, not a mobile monitoring device, where the signal is much more likely to be affected by noise; (3) we assessed an irregularity from the perspective of abnormal T-wave morphology, but this, of course, is one of many; fast or slow heart rates, abnormalities in ST-T changes and the presence of some common types of arrhythmia such as atrial fibrillation (AF) can all affect QT-interval calculation; (4) in a self-monitoring situation people may be using tablets or phones. We hypothesise that the visualisation techniques will still be beneficial, but this will need to be confirmed in a further study examining the effects of screen size and lighting setting on the visualisation techniques; and (5) our participants were highly educated, and we do not know whether the results would generalise to other demographics. Future work will include evaluating the visualisation techniques with more diverse clinical populations, particularly with low-literacy and low-income minority populations, who are taking medication that can lead to LQTS, and are using a mobile device with a wearable ECG monitor.

5.6 Conclusion

This study shows that using simple visualisation techniques significantly improves lay people's ability to accurately measure the QT-interval. This may help with self-monitoring drug-induced LQTS and enable treatment to be altered to prevent the development of life threatening complications. Whilst using a pseudo-colour sequence significantly improves people's ability to detect increases in the QT-interval when the ECG is displayed on a standard Cartesian coordinate system, the greatest accuracy is achieved when pseudo-colour is combined with Polar coordinates.

Chapter 6

Pseudo-Colouring an ECG Enables Laypeople to Detect QT-Interval Prolongation Regardless of Heart Rate

6.0 Chapter overview

6.0.1 Thesis context

The study reported in Chapter 5 demonstrated the effectiveness of using pseudo-colouring to support the detection of small increases in QT-interval at a regular heart rate of 60 bpm that are difficult to perceive without colour. Drug-induced changes in T-wave morphology and heart rate are known to complicate visual LQTS detection and TdP risk assessment. At present, the QT correction formulae (QTc) used in clinical practice to correct QT-interval to heart rate are known to be poor at identifying patients at risk of drug-induced TdP, particularly for fast and slow heart rates [DLD$^+$03, LMJM04] (discussed in Chapter 1 as the third research challenge). This chapter addresses this issue with an empirical study that evaluates whether applying the pseudo-colouring technique according to the QT-nomogram [CIKD07] and changing the coordinate system (Cartesian *vs.* Polar) can support laypeople in identifying QT-prolongation at risk of TdP regardless of T-wave morphology and heart rate visually, without the need for explicit QT measurement and heart rate correction. Participants assessed the ECGs of multiple patients on different QT-prolonging drugs for risk of TdP. The ECGs were

presented alone (without a drug-free ECG baseline comparison). Accuracy was quantified in terms of sensitivity and specificity using a multi-reader, multi-case (MRMC) receiver operating characteristic (ROC) study design within a psychophysical paradigm. Eye-tracking was used to determine the locus of visual attention.

The results of this study show that pseudo-colouring significantly improves laypeople's ability to detect QT-prolongation at risk of TdP visually and identify 'normal' QT-intervals showing no risk of TdP, regardless of heart rate, T-wave morphology and coordinate system. These results have motivated us to investigate whether modelling the visual perceptual process used by humans when interpreting the pseudo-coloured ECG can be used as a basis for an automated QT-prolongation detection algorithm, which was initially explored in Chapter 7 and comprehensively studied in Chapter 8.

The main content of this chapter is adapted from: Alaa Alahmadi, Alan Davies, Markel Vigo, and Caroline Jay. Pseudo-colouring an ECG enables lay people to detect QT-interval prolongation regardless of heart rate. *PLoS One*, 15(8):e0237854, 2020.

6.0.2 Author's contributions

Alaa Alahmadi expanded the development of the pseudo-colouring visualisation such that it automatically adjusted according to heart rate, designed the study, carried out the data collection, analysed the results and wrote the paper, with Alan Davies, Markel Vigo and Caroline Jay contributing significant edits. Alan Davies acted as the electrocardiogram domain expert throughout. Caroline Jay and Markel Vigo provided continuous guidance and discussion.

6.0.3 Published abstract

Drug-induced long QT syndrome (diLQTS), characterized by a prolongation of the QT-interval on the electrocardiogram (ECG), is a serious adverse drug reaction that can cause the life-threatening arrhythmia Torsade de Points (TdP). Self-monitoring for diLQTS could therefore save lives, but detecting it on the ECG is difficult, particularly at high and low heart rates. In this paper, we evaluate whether using a pseudo-colouring visualisation technique and changing the coordinate system (Cartesian *vs.* Polar) can support lay people in identifying QT-prolongation at varying heart rates. Four visualisation techniques were evaluated using a counterbalanced repeated measures design including Cartesian no-colouring, Cartesian pseudo-colouring, Polar no-colouring and

Polar pseudo-colouring. We used a multi-reader, multi-case (MRMC) receiver operating characteristic (ROC) study design within a psychophysical paradigm, along with eye-tracking technology. Forty-three lay participants read forty ECGs (TdP risk $n = 20$, no risk $n = 20$), classifying each QT-interval as normal/abnormal, and rating their confidence on a 6-point scale. The results show that introducing pseudo-colouring to the ECG significantly increased accurate detection of QT-interval prolongation regardless of heart rate, T-wave morphology and coordinate system. Pseudo-colour also helped to reduce reaction times and increased satisfaction when reading the ECGs. Eye movement analysis indicated that pseudo-colour helped to focus visual attention on the areas of the ECG crucial to detecting QT-prolongation. The study indicates that pseudo-colouring enables lay people to visually identify drug-induced QT-prolongation regardless of heart rate, with implications for the more rapid identification and management of diLQTS.

6.1 Introduction

6.1.1 Background and significance

An electrocardiogram (ECG or EKG) is a graphical representation of the electrical activity of the heart, widely used as a clinical tool for monitoring heart function and detecting cardiac pathologies [Sch71]. An important measurement on the ECG is the QT-interval, which represents the duration of the ventricular depolarization and repolarization cycle. It is measured from the beginning of the QRS complex (illustrating ventricular depolarization) to the end of the T-wave (showing subsequent repolarization) [GMZ06, Sch71], as shown in Figure 6.1.

The QT-interval is of considerable clinical importance, primarily because its prolongation can increase the risk of a life-threatening arrhythmia known as *Torsades de Pointes* (TdP), a form of polymorphic ventricular tachycardia that is the leading cause of sudden cardiac death in young, otherwise healthy people [AAKR$^+$02, MSC$^+$91, GM08, YC03]. Prolongation of the QT-interval indicates a cardiac disorder known as 'long QT syndrome'(LQTS), which is caused by the malfunction of cardiac ion channels impairing ventricular repolarization; it manifests as a longer QT-interval than normal on the ECG [MK05, DSR08]. The TdP arrhythmia is often precipitated by triggers such as physical activity (notably swimming) [BSW14], and stress-related emotion [VTVAV15]. Athletes with LQTS are thus particularly at risk of TdP

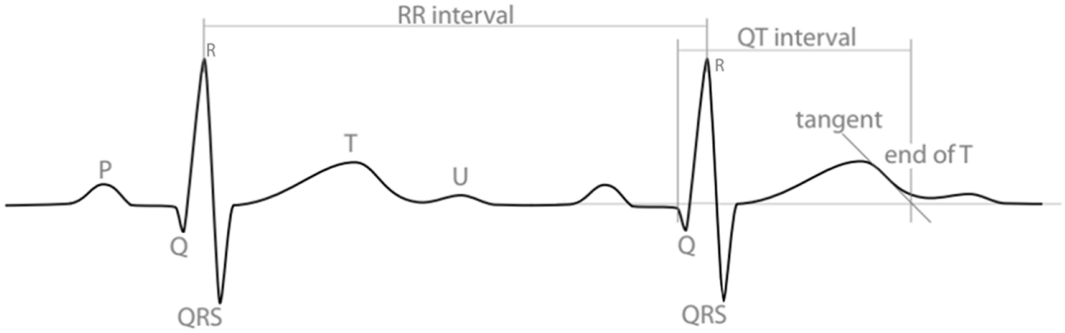


Figure 6.1: An illustration of electrocardiogram (ECG or EKG) waveforms and the QT-interval. Different 'waves' (peaks and troughs) are labelled with letters and represent different stages of a heartbeat. The QT-interval is measured from the beginning of the Q-wave to the end of the T-wave, identified here using the tangent method drawn at the maximum down-slope of the T-wave.

episodes [JA13, BSW14, GPR16]. A major difficulty with identifying LQTS is that it is often asymptomatic; sudden cardiac death can be the first clinical manifestation, and therefore it may go undiagnosed, or underdiagnosed, without ECG assessment [Kha02, SC08, RV05].

LQTS can be congenital—a result of genetic mutations in cardiac channelopathies, as seen in Romano–Ward syndrome [PSN$^+$98]—or acquired, resulting from the clinical administration of certain drugs [YC03, CMY08]. There are many commonly prescribed medications that can prolong the QT-interval, including antihistamines, antibiotics, antidepressants and antiarrhythmic drugs [YC03, CMY08]. Drug interactions that block the human ether-a-go-go-related gene (hERG) potassium channel are the most common cause of drug-induced long QT syndrome (diLQTS) [TKK06, RPM$^+$05]. These drugs are associated with large and increasing numbers of sudden deaths, and preventive strategies could therefore save many lives [Wys07, YC03, Bro04].

Frequent ECG monitoring is strongly recommended for people at risk of acquiring diLQTS, including patients on a known QT-prolonging drug and patients participating in a clinical trial testing a new drug [Sha02, KW12]. A baseline ECG, before taking a QT-prolonging drug, and follow-up ECGs to monitor the QT-interval are advisable [NAS12, DAF$^+$10]. Recent innovations in home healthcare technologies have made it possible to record high-quality, clinically reliable ECG data outside of the clinical environment [SGH$^+$17, SPS$^+$19, WRY19, RFP19], but interpreting such data correctly in a timely manner still remains a major challenge [SW17, HG06]. One approach

involves monitoring patients' ECGs remotely, where the ECG report is referred to a clinician who specialises in ECG interpretation [SGH+17, SPS+19, WRY19, RFP19]. However, a single dose of a QT-prolonging drug could dramatically prolong the QT-interval within 24 hours for some patients, with the risk of TdP increasing with continued use [LVL+18]. At present, remote expert monitoring is too infrequent and costly to be an effective way of managing this issue for everyone at risk of acquiring diLQTS.

While there are computerised methods for interpreting the ECG, the accuracy of these methods remains limited [HG06, SW17]. Challenges of automating ECG interpretation include the correct recognition of the ECG waveforms, in particular the amplitude and duration characteristics (which differ substantially across individuals), and the precise determination of the onset and offset of the different waves and complexes (P-wave, QRS complex, T-wave) [SW17].

Automated QT measurement algorithms have proved unsatisfactory for detecting LQTS in particular [TAS+15, KBD+18, EI13, GL13, MCbA01, TABW11, RSG09]. Garg and Lehmann [GL13] found that even a widely used computerized ECG analysis system was not able to detect QT-interval prolongation in 52.5% of patients affected. Research has also shown that drug-induced QT-prolongation can be underestimated and under-reported by computerised methods in patients on Methadone, a drug that is infamous for prolonging the QT-interval and increasing the risk of TdP [TAS+15]. A major challenge for automated QT algorithms is identifying the precise end of the T-wave (the terminal point), especially when the T-wave's morphology is abnormal [GPAW14, GMZ06, Mor01]. This is particularly problematic, as QT-prolonging drugs also often affect the morphology of the T-wave; for example, patients on pure hERG-blocking drugs can develop flat, asymmetric, and notched T waves, whilst patients on multi-channel blocking drugs (*e.g.* hERG-blocking with lesser calcium and late sodium block) can develop more distorted, bizarre, T-waves [VJM+15]. In fact, specific T-wave patterns can aid detection of LQTS [CHA+15], and large T-U waves are known to precede TdP [KFBW09]. These indicators are lost by automated QT algorithms, however, as they work by abstracting the ECG data into a number (the calculated length of the QT-interval). As such, the human visual validation of QT-prolongation on the ECG remains mandatory in clinical practice, and provides the richest information for recognizing LQTS [EI13, KBD+18, TABW11, GL13, MCbA01, RSG09, TAS+15, SW17].

Giving laypeople the means to detect when the QT-interval is prolonged without having to rely on an external interpreter could lead to a step-change in the detection

and management of diLQTS, as well as helping to reduce the development of life-threatening complications in situations that place individuals at risk of QT-interval prolongation. ECG interpretation is complex, however, and detecting QT-prolongation on the ECG is particularly difficult, even for clinicians [VRS+05, STN08]. From a perceptual-cognitive perspective, this may be related to the fact that people are poor at perceiving quantity represented along a horizontal scale [LSZ+09, WA84, PK11]. Research has also shown that changes in the T-wave morphology and/or artefacts in the ECG signal can cause misinterpretation of the QT-interval length [ADVJ18, ADR+19, GPAW14, Mor01]. The effect of heart rate on the QT-interval is another challenge, as it is the proportionate rather than absolute length that is important, and it is common to misinterpret the QT-interval at heart rates that differ from the 'standard' 60 bpm [LMJM04]. Visualisation techniques have the potential to highlight abnormal changes within the ECG, by supporting intuitive visual perception of the issues. Pre-attentive processing theory, which outlines a set of visual properties known to be detected rapidly and accurately by the human eye, is considered to be especially powerful for designing effective visualisations [War12, HBE95, HBE96]. Colour is one of the most effective pre-attentive attributes that is noticed without conscious effort [Not93, GCC17]. A useful technique is pseudo-colouring, which encodes continuously varying values using a sequence of colours [War12]. It has been widely used to support diagnosis from medical images, including breast disease [ZSST11], and for highlighting details in organs and bones structures that would otherwise be difficult to perceive [Quw09, Sem18]. It is also used extensively in geographic and time series visualisations, where applications include encoding elevation in the data or showing changes over time [War12, SMY+05]. To our knowledge, no prior work has used pseudo-colouring to support ECG interpretation.

In this paper, we investigate whether a pseudo-colouring technique can reliably show prolongation of the QT-interval on an ECG, such that it be identified by a lay person. In an earlier feasibility study, we found that superimposing pseudo-colouring on the ECG using a spectrum-approximation colour sequence significantly improved people's ability to detect increases in the QT-interval at a low normal heart rate, when compared with a reference ECG stimulus showing a normal QT baseline [ADR+19]. This initial investigation had several limitations that affected the generalisability of the results. Firstly, the ECGs all had a single, regular heart rate of 60 bpm, and it is unknown how the technique would perform at higher or lower heart rates. The ECGs belonged to the same patient, who was taking the QT-prolonging drug ('Dofetilide',

an antiarrhythmic drug with a pure hERG blocker) and subsequently experienced a gradual increase in the QT-interval from the normal QT baseline. It is well-known that the ECG differs from one individual to another, and that different types of drug can affect the ECG in different ways. The study examined only sensitivity (identifying true positives); for self-monitoring, it is important also to examine specificity (identifying true negatives). Finally, and most importantly, the study design allowed people to compare the ECG with a normal baseline. Such a baseline would be hard to produce in self-monitoring situations, not least because it would have to be adjusted according to heart rate. Therefore, this study tests the technique by presenting the ECGs one by one, asking participants to judge whether QT-prolongation has occurred when viewing just a single image, rather than comparing it with a baseline.

6.1.2 Objective

The objective of the study was to evaluate the effectiveness of the pseudo-colouring technique in displaying QT-prolongation risk threshold on the ECG, when the data is displayed on two coordinate systems: Cartesian and Polar. Accuracy was quantified in terms of both sensitivity and specificity. The technique was tested with QT-intervals across several heart rates for multiple patients on different QT-prolonging drugs. No comparison baseline ECG was provided.

6.2 Materials and methods

In this section, we provide a detailed description of how the ECGs images were produced and displayed, and the experiment designed to evaluate the effectiveness of the technique. The ECGs shown in the figures are reduced in size for inclusion in the paper. All of the full size ECG images used in the experiment, along with the R scripts used to create them, can be found in [ADVJ20b].

6.2.1 ECG data acquisition

The ECG datasets were acquired from a clinical trial study that was conducted to assess the effects of different types of known QT-prolonging drugs, including a pure hERG potassium channel blocker and multichannel blockers, on the ECG of healthy subjects [JVM$^+$14]. The open ECG datasets are available online from the PhysioNet database [GAG$^+$00].

6.2.2 Visualisation design

The visualisation was created in two layers. The first layer plotted the ECG using the standard method, which is a two dimensional line graph displaying the amplitude, or voltage, of the electrical signal along the Y-axis and the time in milliseconds (ms) along the X-axis. We used the ECG data from lead II, as the QT-interval is conventionally measured in this lead [GMZ06, CMY08]. The pseudo-colouring was then applied in a second layer, following a series of steps: identifying individual heartbeats via R-peak detection; applying pseudo-colouring to each heartbeat; adjusting the colouring according according to heart rate; displaying the data on either a Cartesian (current standard) or Polar coordinate system.

Identifying individual heartbeats

The R-wave peaks (points with greatest amplitude) were detected in the raw ECG datasets using an automated numerical math function that finds the greatest peaks (maxima) in a time series [Bor]. A solid vertical line was superimposed on the image at this point perpendicular to the X-axis to show the location of the R-waves on the ECG, and delineate individual heartbeats. Detecting the R wave in the vast majority of ECGs is straightforward, as it consistently has the greatest amplitude. This is in contrast to the other waveforms which vary considerably across ECGs. As high quality ECG signals were used in the study, this method proved accurate and efficient. Where signal data is lower quality, it may be necessary to use alternative detection methods that can reliably distinguish R-peaks from high amplitude noises (see *e.g.* [FNY$^+$13, PHC05, LLL$^+$14, SMM14]).

Applying the pseudo-colouring

The traditional way to measure the QT-interval clinically is to count the small squares (each representing 40ms) on the standard ECG background grid, from the beginning of the Q-wave to the end of the T-wave. For the purposes of applying the pseudo-colouring, we added sequential time in ms as a third dimension to each heartbeat, from the R-peak minus 20ms (the estimated start of the Q-wave), to the clinically significant point of maximum risk for TdP—the method for estimating this is provided in the next section. We then applied pseudo-colouring to the area under the curve of the ECG signal by mapping every 40ms of this sequential time to a colour code or hue, with the intensity of the hue changing every millisecond. We used a spectrum-approximation

sequence as it is effective for supporting people in reading continuous values [War88]; it ranges from cool colours (purple to blue to green), which were used to indicate normal QT-interval levels, to warm colours (yellow to orange to red), used to show prolonged QT-interval levels.

Adjusting pseudo-colouring according to heart rate using the QT nomogram

To apply the pseudo-colouring accurately for different heart rates, it was necessary to identify the value of QT-prolongation that determines being 'at risk' for TdP. In the previous study [ADR+19], where all ECGs had a 60 bpm heart rate, we used the half R-R interval rule to identify at risk QT-prolongation. This states that the QT-interval is prolonged if it is equal to or longer than the half R-R interval (midpoint between two consecutive R-peaks) [LMJM04, Dub00]. This rule is accurate with a heart rate of 60 bpm, but lacks sensitivity in detecting at risk QT-prolongation with both lower and higher heart rates [BI15, LMJM04]. In clinical practice it is common to apply a QT correction formula (QTc), and then use a 'cut off' value to identify at risk QT-prolongation. Examples of such QTc formulae include Bazett's, Fredericia's, Framingham's and Hodge's [LMJM04]. However, recent research has shown that these correction formulae are inaccurate in identifying patients at risk of drug-induced TdP, particularly for fast and slow heart rates [DLD+03, LMJM04].

An alternative approach, known as the 'QT nomogram', is a risk assessment method designed specifically for identifying patients at risk of drug-induced TdP according to heart rate [FWM+05, CIKD07]. It was developed after screening 129 ECG cases that reported drug-induced TdP and comparing these with control cases (*i.e.* when no TdP was reported). The actual QT-interval value (not the QTc) and the heart rate for each case were plotted as coordinates on a graph; this produced a line showing the upper bound of the QT-interval value at risk for TdP as a function of heart rate. If the QT/HR value falls on or above this risk line, the patient is at risk of TdP; below the line the patient is not considered at risk of TdP [CIKD07]. The nomogram plot can be seen in Figure 6.2. Further evaluation of the QT nomogram showed that it had higher sensitivity and specificity than widely accepted QTc formulas [WGG+10] and the half R-R interval rule [BI15]. We therefore used it as the foundation for modifying the pseudo-colour as a function of heart rate.

For each ECG, the heart rate (HR) was calculated and then the corresponding QT value for the 'at risk' threshold was acquired from the QT nomogram and plotted on the ECG as a dashed vertical line. To apply the pseudo-colouring, we used nine

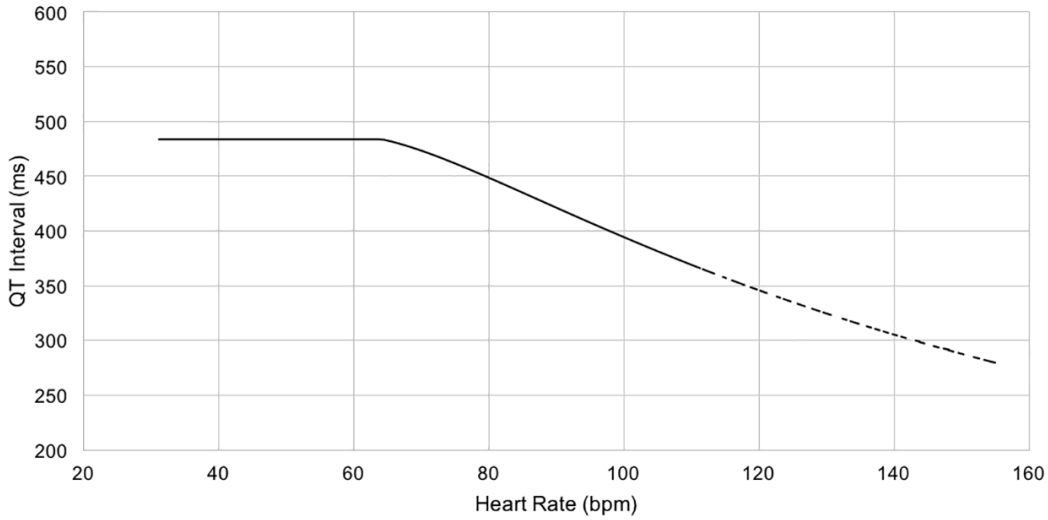


Figure 6.2: The QT-nomogram for identifying QT-prolongation at risk of TdP over heart rate. If the QT/HR value falls on or above the risk line, the patient is at risk of TdP; below the line the patient is not considered at risk of TdP [CIKD07].

indices, where each was mapped to a colour code and represented by a small square on the standard ECG background grid. A time value on the nomogram threshold line was mapped to dark orange, and values 40ms and 80ms above the nomogram line were mapped to red and dark red respectively, showing a higher risk for TdP. Time values below the nomogram line were mapped to progressively cooler colours, such that a time value below the line by five small squares (200ms), was mapped to blue. Figure 6.3 illustrates how the pseudo-colouring technique was applied according to the standard ECG background grid. The mappings between colours and time values according to the nomogram is shown in Table 6.1. Note that this approach does not provide a numerical value for the QT-interval, but rather colours the area under the curve of the signal over a specific time interval (40ms $\times$ 9 indices = 360ms) according to the nomogram. The T-wave area lies within this time interval: if the T-wave is in a cool colour location, then the QT-interval is in the normal range; if it is in a warm colour location, the interval is prolonged.

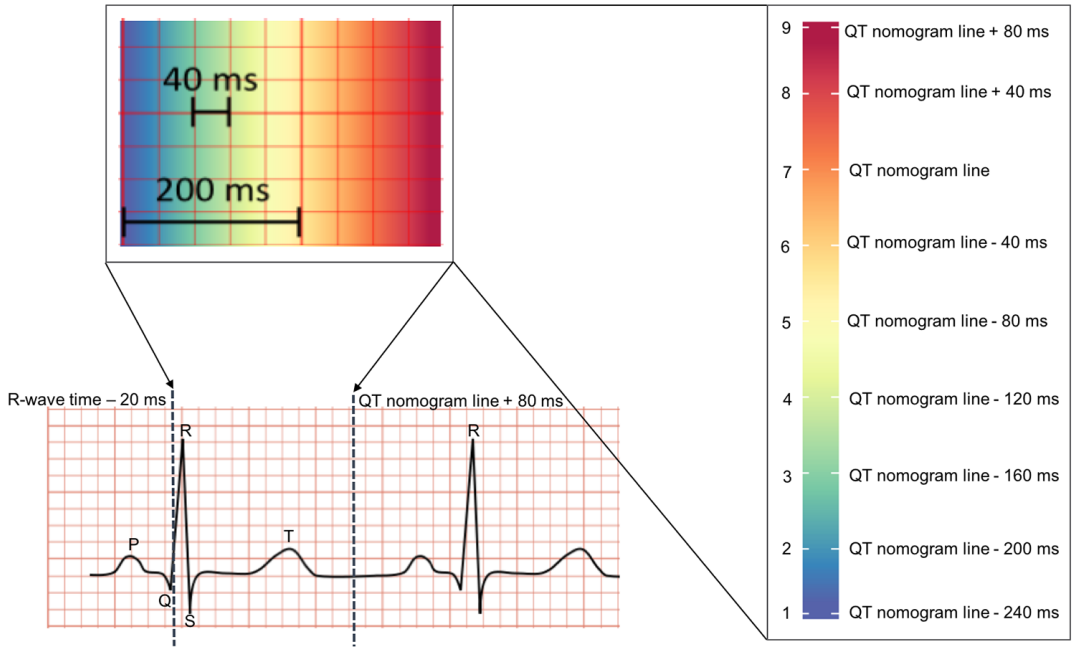


Figure 6.3: Mapping the pseudo-colouring to the ECG, according to the QT-nomogram. A small square on the standard ECG background grid is equal to 40ms.

Table 6.1: The nine indices on the pseudo-colouring scale with their corresponding time values and colour codes.

Index	The corresponding time value (ms)	Colour code
1	QT-value of the nomogram line - (40 × 6)	Purple
2	QT-value of the nomogram line - (40 × 5)	Blue
3	QT-value of the nomogram line - (40 × 4)	Green
4	QT-value of the nomogram line - (40 × 3)	Lime
5	QT-value of the nomogram line - (40 × 2)	Yellow
6	QT-value of the nomogram line - (40 × 1)	Orange
7	QT-value of the nomogram line	Dark orange
8	QT-value of the nomogram line + (40 × 1)	Red
9	QT-value of the nomogram line + (40 × 2)	Dark red

Coordinate system

Our previous work indicates that people may be able to detect QT-prolongation more accurately on a Polar coordinate system [ADR+19]. To see whether this held when ECGs with varying heart rates are presented in isolation, we compared two coordinate systems (Cartesian and Polar) with and without pseudo-colouring. We used the R programming language [IG96] using RStudio software version 1.1.447 to create the visualisations. Figures 6.4 and 6.5 show examples of ECGs with pseudo-colouring that have different heart rates, but similar QT-interval risk levels for TdP, on Cartesian coordinates. Figures 6.6 and 6.7 show the same ECGs on Polar coordinates.

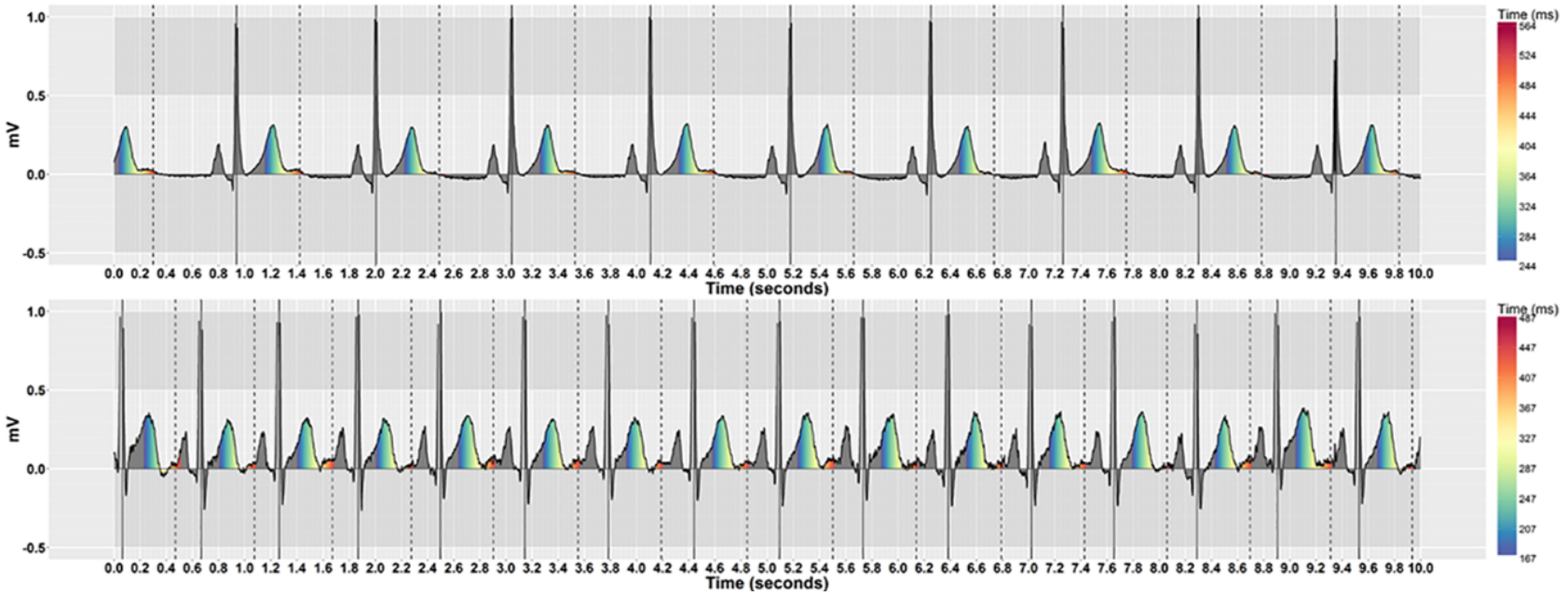


Figure 6.4: Pseudo-coloured ECGs that have different heart rates, but similar ‘normal’ QT-interval risk levels, on Cartesian coordinates. The QT-intervals are below the nomogram line by 80ms for both ECGs (no risk of TdP). The top stimulus has a low heart rate (HR = 57, QT = 401) and the bottom has a high heart rate (HR = 95, QT = 339).

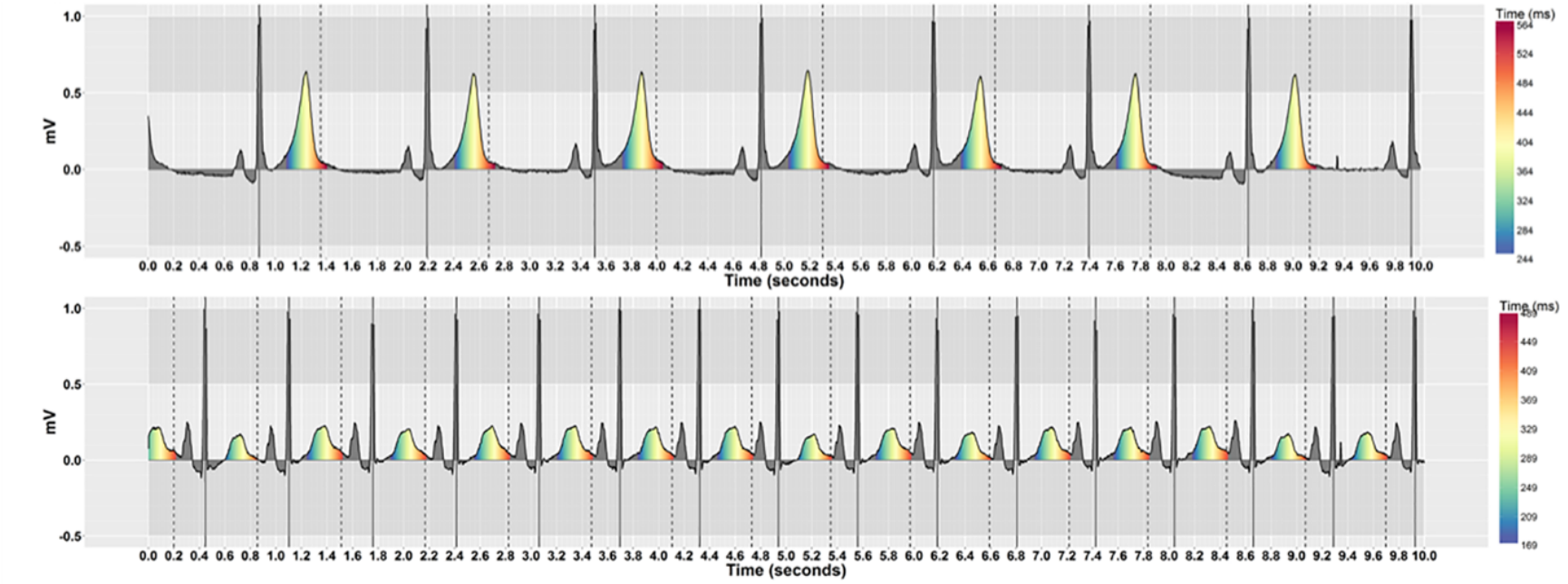


Figure 6.5: Pseudo-coloured ECGs that have different heart rates, but similar prolonged QT-interval risk levels, on Cartesian coordinates. The QT-intervals are on the nomogram line in both ECGs (risk of TdP). The top stimulus has a low heart rate (HR = 46, QT = 487) and the bottom has a high heart rate (HR = 94, QT = 410).

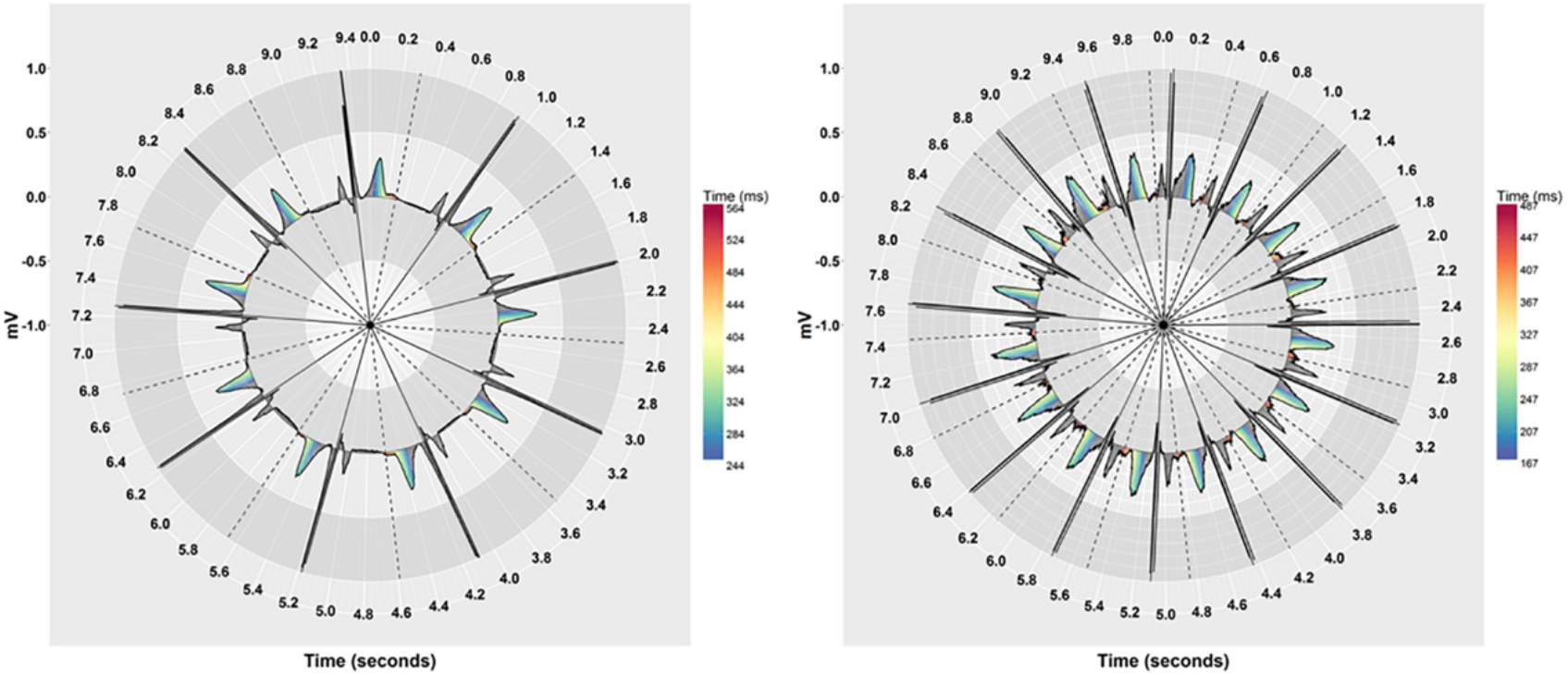


Figure 6.6: Pseudo-coloured ECGs that have different heart rates, but similar 'normal' QT-interval risk levels, on Polar coordinates. The QT-intervals are below the nomogram line by 80ms for both ECGs (no risk of TdP). The left stimulus has a low heart rate (HR = 57, QT = 401) and the right has a high heart rate (HR = 95, QT = 339).

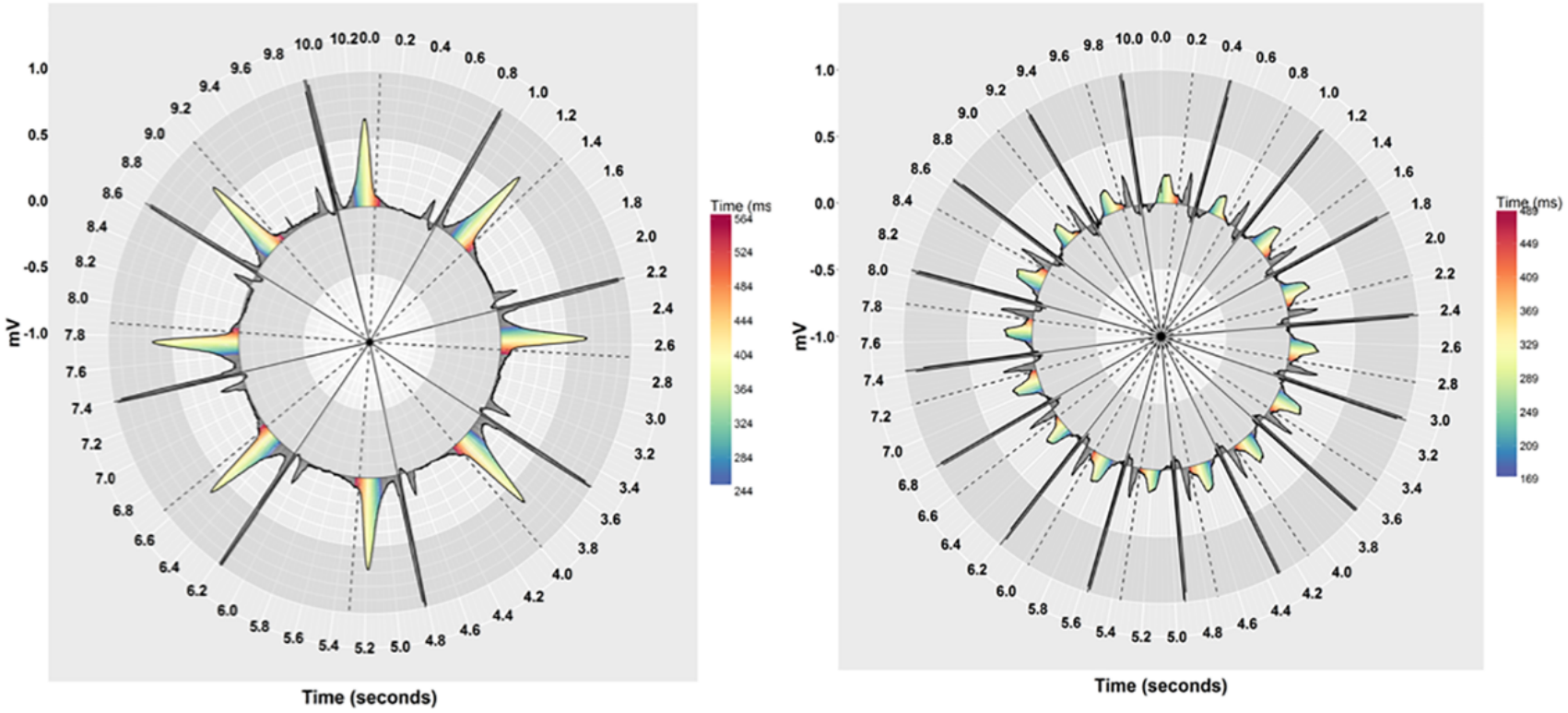


Figure 6.7: Pseudo-coloured ECGs that have different heart rates, but similar prolonged QT-interval risk levels, on Polar coordinates. The QT-intervals are on the nomogram line in both ECGs (risk of TdP). The left stimulus has a low heart rate (HR = 46, QT = 487) and the right has a high heart rate (HR = 94, QT = 410).

6.2.3 Experimental design

We used a multi-reader, multi-case (MRMC) receiver operating characteristic (ROC) study design within a psychophysical paradigm. The MRMC ROC design is an evaluation method commonly used to assess diagnostic performance in medical imaging studies, where multiple human observers (readers) interpret multiple patient images (cases) [HBM08]. The area under the ROC curve, a plot of sensitivity versus 1-specificity, is used to measure diagnostic accuracy. The MRMC design helps to increase the generalisability of study results and enhance statistical power, particularly when evaluating different computer-assisted detection (CAD) systems [HBM08]. Psychophysical experiments use detection and discrimination tasks to investigate the relationship between the intensity of a physical stimulus and human perception and sensation, by systematically varying the properties of the stimulus along one or more physical dimensions [P$^+$16]. Here the paradigm was used to systematically evaluate the impact of the visualisation technique on people's sensitivity to increases in the QT-interval. We also used eye-tracking, a non-invasive sensor technology that measures human eye movements, to help us interpret the visual behaviour underpinning the results on a *post hoc* basis [Duc07].

The study used a counterbalanced repeated measures design with two independent variables, each with two levels:

1. Colour-coding (no colouring and pseudo-colouring).

2. Coordinate system (Cartesian and Polar).

The within-subjects factorial design yielded a total of 4 (2×2) experimental conditions for each participant: Cartesian no-colouring; Cartesian pseudo-colouring; Polar no-colouring; Polar pseudo-colouring. We counterbalanced the order of visualisation presentation using a balanced Latin square to minimize practice effects. The participants were divided into four groups of eleven participants, except for the last group which included ten participants. Although an additional participant would have resulted in a more rigorous counterbalanced design, this small effect was mitigated by the systematic approach we adopted when analysing the results, *i.e.* by considering the comparison of sensitivity and specificity within-participant as well as across participants for each visualisation (discussed in detail in Section 6.2.9). Considering (A) to be 'Cartesian no-colouring', (B) to be 'Cartesian pseudo-colouring', (C) to be 'Polar no-colouring' and (D) to be 'Polar pseudo-colouring', the counterbalanced order of

Table 6.2: The six QT-levels with their corresponding QT-value relative to the nomogram and an estimated range.

QT-level	QT-value relative to the QT-nomogram line	Estimated range
1	QT-value of the nomogram line - (40 × 3)	Normal
2	QT-value of the nomogram line - (40 × 2)	Normal
3	QT-value of the nomogram line - (40 × 1)	Borderline
4	QT-value of the nomogram line	Prolonged
5	QT-value of the nomogram line + (40 × 1)	Very prolonged
6	QT-value of the nomogram line + (40 × 2)	Severely prolonged

visualisation presentation for groups 1 to 4 respectively was as follows: A, B, D, C; B, C, A, D; C, D, B, A; and D, A, C, B. The dependent variables were accuracy (broken down into sensitivity and specificity), reaction time in ms, satisfaction with the visualisation on a five point Likert-type scale from low (1) to high (5), mean fixation duration, total fixation duration and spatial fixation distribution.

6.2.4 Stimulus design

The heart rates of the ECGs acquired from the clinical trial ranged from 40 to 96 bpm, and the QT-interval values ranged from 300 to 579ms. As such, bradycardia (HR <60 bpm) was included in the range of heart rates, but not tachycardia (HR >100 bpm).

We followed psychophysical experimental design principles to systematically select the study's stimuli. We considered QT interval value, relative to the nomogram line, to be the physical dimension along which the stimulus was varied, in units of 40ms. Six levels of the QT-interval were identified: three levels below the nomogram line (no/low risk to borderline), one level on the nomogram line (moderate risk), and two levels above the nomogram line (high risk). The six QT levels were represented by indices 4 to 9 on the pseudo-colouring scale respectively (see Table 6.1). Figure 6.8 shows examples of ECGs with the first QT-level (normal), and the sixth QT-level (severely prolonged), while Figures 6.4, 6.5, 6.6 and 6.7 show examples of ECGs with the second QT-level (normal) and the fourth QT-level (prolonged). Table 6.2 shows the corresponding estimated QT-interval prolongation for the six levels.

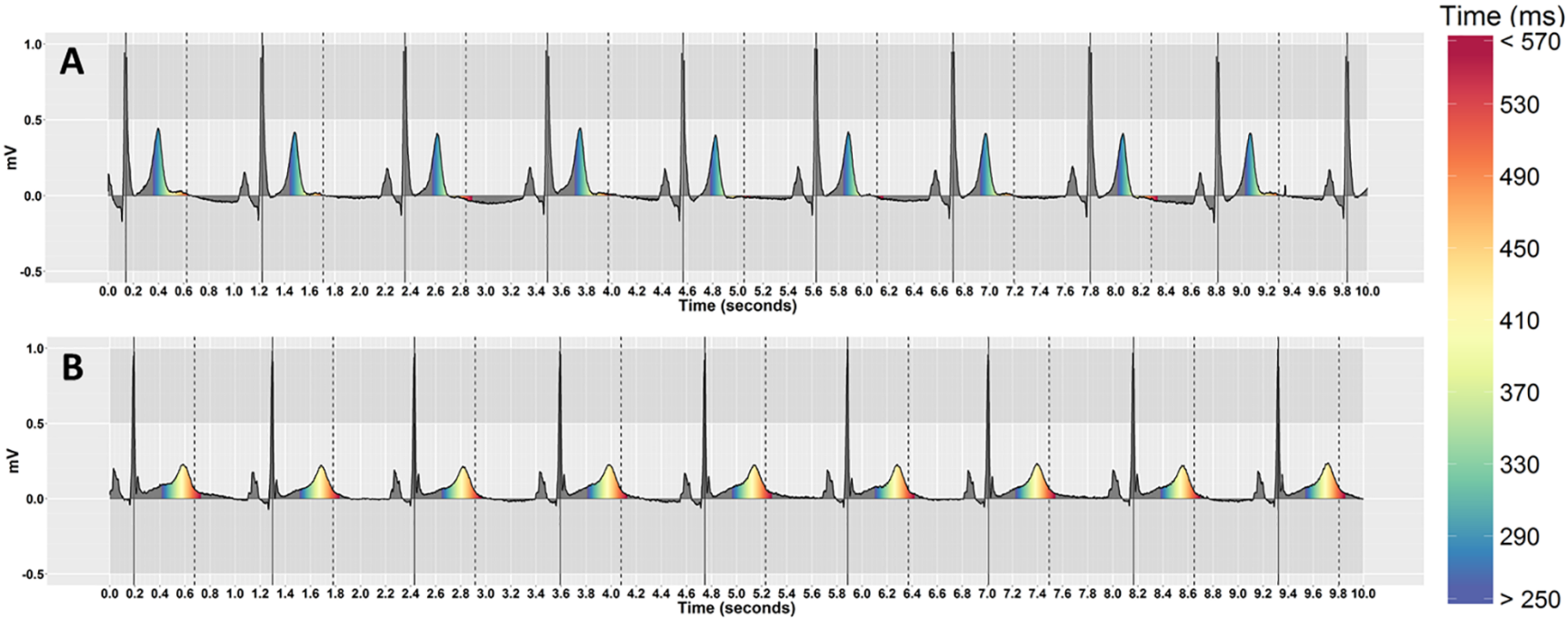


Figure 6.8: Pseudo-coloured ECGs that have similar heart rates, but different QT-interval risk levels, on Cartesian coordinates. (A) Normal QT-interval (HR = 55, QT = 361). (B) Severely prolonged QT-interval (HR = 52, QT = 579).

6.2.5 ECG case selection

Forty ECG cases were selected from multiple patients ($n = 17$) to match the six QT-levels. Twenty ECGs were below the nomogram line (*i.e.* no risk of TdP) and 20 ECGs were on or above the nomogram line (*i.e.* at risk of TdP). Six ECG cases were on a placebo, and had values corresponding to the first QT-level, across heart rates ranging from 48 to 90 bpm. The other 34 ECG cases (18 on Dofetilide, a known QT-prolonging drug with a pure hERG potassium channel blocker, and 16 on Quinidine, a multichannel blocker with a strong hERG block, and lesser calcium and late sodium blocks), represented the second to sixth QT-levels across heart rates from 42 to 96 bpm. Table 6.3 shows the metadata of each ECG case in terms of QT value, heart rate, drug type, patient ID, the corresponding QT-level relative to the nomogram line and whether the case was at risk for TdP or not based on the nomogram plot.

6.2.6 Participants

A total of forty three participants, with no experience in ECG interpretation, were recruited from a university campus (30 students and 13 staff). Eligibility for the study was determined by asking participants to rate their knowledge of ECG interpretation, and only people who identified as having no knowledge at all were included in the study. There were 18 male and 25 female participants between 20 and 56 years old (Mean = 31, $SD = 8$). Table 6.4 shows the participants' demographic data. Participants' sight was normal or corrected-to-normal and they reported no motor or neurological disorders. The research was approved by the University Research Ethics Committee). All participants provided written informed consent.

6.2.7 Apparatus

Stimuli were displayed on a 23.8 inch (diagonal) wide-screen Spectrum eye-tracking monitor, with a resolution of 1920 $\times$ 1080 pixels. Eye gaze was recorded using the Tobii Pro Spectrum eye-tracker and Tobii Pro Lab 1.95 software with a sampling rate of 600Hz. Key press events were recorded during the experiment to measure reaction times. Each Cartesian coordinate ECG stimulus measured 32.31cm $\times$ 6.14cm, and each Polar coordinate stimulus was 15.61cm $\times$ 12.93cm.

Table 6.3: Characteristics of the 40 selected ECG cases used in the study. The meta-data were acquired from the clinical trial study [JVM+14], published by PhysioNet database [GAG+00].

ECG ID	QT value	Heart rate	Drug type	Patient ID	QT-level	Risk for TdP on the nomogram
1	370	48	Placebo	11	1	Not at risk
2	361	55	Placebo	13	1	Not at risk
3	350	68	Placebo	16	1	Not at risk
4	343	72	Placebo	16	1	Not at risk
5	329	83	Placebo	18	1	Not at risk
6	335	90	Placebo	4	1	Not at risk
7	401	57	Dofetilide	12	2	Not at risk
8	389	75	Dofetilide	18	2	Not at risk
9	339	95	Dofetilide	20	2	Not at risk
10	419	47	Quinidine	8	2	Not at risk
11	396	68	Quinidine	16	2	Not at risk
12	355	82	Quinidine	18	2	Not at risk
13	445	46	Dofetilide	2	3	Not at risk
14	441	67	Dofetilide	7	3	Not at risk
15	431	75	Dofetilide	8	3	Not at risk
16	417	80	Dofetilide	11	3	Not at risk
17	371	94	Dofetilide	17	3	Not at risk
18	444	58	Quinidine	4	3	Not at risk
19	424	76	Quinidine	10	3	Not at risk
20	363	95	Quinidine	22	3	Not at risk
21	487	46	Dofetilide	4	4	At risk
22	468	72	Dofetilide	8	4	At risk
23	451	79	Dofetilide	9	4	At risk
24	445	81	Dofetilide	10	4	At risk
25	486	54	Quinidine	7	4	At risk
26	485	64	Quinidine	7	4	At risk
27	419	91	Quinidine	15	4	At risk
28	410	94	Quinidine	19	4	At risk
29	523	42	Dofetilide	4	5	At risk
30	494	71	Dofetilide	9	5	At risk
31	470	85	Dofetilide	18	5	At risk
32	565	49	Dofetilide	4	6	At risk
33	579	52	Dofetilide	18	6	At risk
34	547	64	Dofetilide	21	6	At risk
35	518	54	Quinidine	4	5	At risk
36	509	68	Quinidine	7	5	At risk
37	482	80	Quinidine	9	5	At risk
38	417	96	Quinidine	21	5	At risk
39	518	77	Quinidine	21	6	At risk
40	507	79	Quinidine	21	6	At risk

Table 6.4: The participants' demographic data.

Category	Count	Percentage %
Age		
20-30	25	58%
31-40	13	30%
41-50	3	7%
50+	2	5%
Sex		
Male	18	42%
Female	25	58%
Occupation		
University Staff	13	30%
Student	30	70%
Background		
Computer science	27	63%
Chemical engineering	1	2%
Psychology	3	7%
Biochemistry	4	9%
Interior design	1	2%
Mathematics	1	2%
Art, history and sociology	1	2%
Biomedical sciences	4	9%
Business	1	2%

6.2.8 Task and procedure

Participants completed a 15 minute training session prior to starting the experiment, where they were introduced to the ECG trace, and taught how to identify the QT-interval using the tangent method to identify the end of the T-wave (see Figure 6.1 and [PDJVdBW08]). Participants were instructed to use the tangent method visually on the screen (*i.e.* without any additional equipment) to roughly identify the end of the T-wave; the training session did not involve any medical terms or high-level training techniques typically associated with clinical ECG interpretation and focused solely on identification of the QT interval. Participants were then introduced to the four visualisation techniques, and shown how to use the pseudo-colouring scale and the risk threshold (represented by a dashed line) to assess the QT-interval. Participants were told that a QT-interval is considered prolonged if the T-wave ends on or exceeds the risk threshold dashed line, and when the pseudo-colouring within the QT-interval, and particularly in the Tpeak–Tend interval, contains warm colours (yellow/orange/red), with a greater visibility of warm colours indicating a higher level of QT-prolongation. Finally, participants completed an assessment task, in which they were asked to highlight the start and end points of the QT-intervals on two different ECGs (one with normal QT-interval and the other with a prolonged QT-interval) and decide whether it was normal or abnormal using the visualisation techniques (*i.e.* 2 ECGs $\times$ 4 visualisation techniques = 8 stimuli), to ensure that they understood how to perform the task. All participants passed the assessment task , correctly categorising the QT-intervals for all stimuli.

Participants were tested individually. They completed four tests, one for each visualisation technique, in which they read all ECG cases for one test before moving to the next. Participants read the same ECG cases using each visualisation technique. ECGs were presented at random within a test. Each test began with a five-point calibration of the eye-tracking apparatus. ECGs were presented one at a time, and participants completed two tasks: the first task used a psychophysical one alternative forced-choice paradigm (also known as a yes-no detection task), in which participants indicated verbally, as quickly as possible, whether they perceived the QT-interval as ('normal' or 'abnormal'), whilst pressing the space-bar on the keyboard to collect reaction time; the second task was to rate their confidence in their response using a 6-point scale of 'very likely normal' (1), 'probably normal' (2), 'possibly normal' (3), 'possibly abnormal' (4), 'probably abnormal' (5), and 'very likely abnormal' (6). There was no time limit

imposed, but participants were encouraged to do the tasks as quickly as possible. The binary responses and the confidence scores were recorded on a spreadsheet during the experiment by the researcher. At the end of the experiment, participants were asked to provide their satisfaction with the visualisations by rating how much each helped them to assess the QT-interval using a five point Likert-type scale ranging from 'not very much' (1) to 'a lot' (5).

6.2.9 Statistical analysis methods

Response accuracy was measured by calculating the average area under the receiver operating characteristic (ROC) curves (AUC) with 95% confidence intervals (CIs). QT-levels 1-3 inclusive were classified as negative (normal), and levels 4-6 as positive (prolonged). We used the Dorfman-Berbaum-Metz multi-reader multi-case (DBM MRMC) software from the University of Iowa [KMSPB], to calculate and compare the AUCs of the four visualisation techniques, based on the methods of Dorfman, Berbaum and Metz [DBM92] and Obuchowski and Rockette [OJRJ95] and later unified and improved by Hillis and colleagues [HOSB05, Hil07, HBM08]. The DBM MRMC method uses jackknifing and analysis of variance (ANOVA) methods and we considered both readers and cases as random variables; this allows the results to be generalised to the population of readers and cases. The ROC curve was fitted using the trapezoidal area and Wilcoxon method. The statistical power of ROC MRMC was greater than 90%, and the sample size estimation was performed using [HS18].

All statistical tests were performed at a 5% significance level ($\alpha = 0.05$); two-sided 95% confidence intervals (CIs) were used to quantify uncertainty. Within-participant comparisons of sensitivity and specificity were performed using McNemar's chi-squared test. We measured perceptual sensitivity to QT-levels using the psychometric function and just noticeable difference (JND) threshold, along with the signal detection analysis method R packages [LLM16, Mak18]. Differences in reaction time and satisfaction scores were determined via a Friedman test and *post hoc* pairwise comparisons performed with a Wilcoxon signed-rank test utilising a Bonferroni correction ($\alpha = 0.008$). Visual behaviour was analysed using eye tracking metrics including mean fixation duration, total fixation duration and fixation distribution using the nearest neighbour index (NNI) method, as implemented by Davies *et al.* [DVHJ16].

Table 6.5: The average area under the ROC curve across all participants and for each visualisation technique. SE = Standard error. CI = %95 Confidence intervals.

Visualisation technique	Average AUC	SE	%95 CI
Cartesian no-colouring	0.895	0.024	0.848 to 0.943
Cartesian pseudo-colouring	0.935	0.020	0.895 to 0.974
Polar no-colouring	0.878	0.029	0.821 to 0.935
Polar pseudo-colouring	0.934	0.023	0.888 to 0.980

6.3 Results

All anonymised data and related metadata underpinning the findings reported in this article can be found in [ADVJ20b].

6.3.1 Accuracy

Area under the ROC curve Pseudo-colouring significantly increased the average area under the ROC curve (AUC) for both coordinate systems (Tables 6.5 and 6.6). When Cartesian coordinates were used, the average AUC increased from 0.895 (standard error (SE) = 0.024) to 0.935 (SE = 0.020), while the increase was from 0.878 (SE = 0.029) to 0.934 (SE = 0.023) for Polar coordinates (Table 6.5). The average increase in AUC as a result of pseudo-colouring was therefore 0.04 for Cartesian and 0.056 for Polar coordinates, a statistically significant increase in both cases ($p = 0.014$, $p < 0.001$ for Cartesian and Polar coordinates respectively, see Table 6.6). There was no significant difference in average AUC as a function of coordinate system when pseudo-colouring was used ($p = 0.978$, see Table 6.6). Although the average AUC was higher for Cartesian coordinates than for Polar coordinates when no-colouring was used, the difference was not statistically significant ($p = 0.273$, see Table 6.6).

Table 6.6: Pairwise comparisons of the area under the ROC curve for the four visualisation techniques. Significant p-values are in bold. SE = Standard error. CI = %95 Confidence intervals.

Pairwise comparisons	AUC Difference	SE	P-value	95% CI
Cartesian no-colouring, Cartesian pseudo-colouring	- 0.039	0.015	**0.014**	-0.070 to -0.007
Cartesian no-colouring, Polar no-colouring	0.017	0.015	0.273	-0.013 to 0.048
Cartesian no-colouring, Polar pseudo-colouring	-0.038	0.015	**0.015**	-0.070 to -0.007
Cartesian pseudo-colouring, Polar no-colouring	0.056	0.015	**<0.001**	0.025 to 0.087
Cartesian pseudo-colouring, Polar pseudo-colouring	0.000	0.015	0.978	-0.030 to 0.031
Polar no-colouring, Polar pseudo-colouring	-0.056	0.015	**<0.001**	-0.087 to -0.024

Table 6.7: Comparisons of mean specificity between the visualisation techniques. The table shows the mean ± standard error, and 95% two-sided confidence intervals. Significant p-values are in bold and were calculated using a McNemar's chi-squared test.

Coordinate	Pseudo-colouring	No-colouring	Difference	P-value
Cartesian	0.90 ± 0.002, 0.870 to 0.929	0.94 ± 0.002, 0.913 to 0.966	-0.04 ± 0.002, -0.069 to -0.014	**<0.001**
Polar	0.90 ± 0.002, 0.874 to 0.925	0.88 ± 0.002, 0.849 to 0.910	0.02 ± 0.002, -0.024 to 0.044	0.460
Difference	0.00 ± 0.002, -0.030 to 0.030	0.06 ± 0.002, 0.0260 to 0.073		
P-value	1.000	**<0.001**		

Specificity Specificity was calculated as the proportion of true negative (*i.e.* 'normal') responses for QT-levels 1 to 3. The mean specificity across all participants was high and similar for both coordinate systems regardless of pseudo-colouring was used (Table 6.7). Although pseudo-colouring increased the mean specificity for Polar coordinates from 0.88 (SE=0.002) to 0.90 (SE=0.002), the increase was not statistically significant ($p = 0.460$). The highest mean specificity was 0.94 (SE=0.002) for Cartesian coordinates when pseudo-colouring was not used, which was significantly higher than when pseudo-colouring was used with this coordinate system ($p < 0.001$). There was no difference in the mean specificity between the two coordinate systems when pseudo-colouring was used ($p = 1.000$).

Sensitivity Sensitivity was measured using the ROC method as the proportion of true positive (*i.e.* 'abnormal') responses for QT-levels 4 to 6. The mean sensitivity increased with pseudo-colouring from 0.63 (SE=0.004) to 0.83 (SE=0.003) with Cartesian coordinates, and from 0.72 (SE=0.004) to 0.82 (SE=0.003) with Polar coordinates (Table 6.8). The increase in sensitivity of 0.2 (SE=0.004) with Cartesian and 0.1 (SE=0.003) with Polar coordinates was statistically significant ($p < 0.001$). Sensitivity was higher when Cartesian coordinates were used than when Polar coordinates were used by 0.01, a statistically significant difference ($p = 0.001$). When pseudo-colouring was not used, the mean sensitivity was 0.09 (SE=0.003) higher with Polar coordinates than with Cartesian coordinates, a difference that was statistically significant ($p < 0.001$).

Table 6.8: Comparisons of mean sensitivity between the visualisation techniques. The table shows the mean ± standard error, and 95% two-sided confidence intervals. Significant p-values are in bold and were calculated using a McNemar's chi-squared test.

Coordinate	Pseudo-colouring	No-colouring	Difference	P-value
Cartesian	0.83 ± 0.003, 0.773 to 0.877	0.63 ± 0.004, 0.570 to 0.689	0.2 ± 0.004, 0.138 to 0.261	**<0.001**
Polar	0.82 ± 0.003, 0.779 to 0.860	0.72 ± 0.004, 0.665 to 0.774	0.1 ± 0.003, 0.050 to 0.149	**<0.001**
Difference	0.01 ± 0.003, -0.030 to 0.050	-0.09 ± 0.003, -0.140 to -0.039		
P-value	**0.001**	**<0.001**		

We also measured perceptual sensitivity across QT-levels using the following psychophysical detection methods:

(1) Psychometric function and just noticeable difference (JND)

(2) Signal detection analysis The psychometric function and JND threshold quantify sensitivity in terms of a binary response (here, 'normal' or 'abnormal'). In psychophysics, signal detection theory recognises that there is uncertainty in the task, such that the decision may be affected by noise either externally, for example where a stimulus changes between presentations, or internally, as people may be biased towards saying 'normal' or 'abnormal' more frequently [P$^+$16]. We attempted to minimise the effect of decision bias by increasing sample size, counterbalancing experimental conditions, and including confidence scores, but some uncertainty is likely to remain. Signal detection theory measures the sensitivity, while taking response bias into account, using a sensitivity index d' [P$^+$16]. This was calculated for each participant and then averaged across all participants. Table 6.9 shows the mean sensitivity index (d') with the standard deviation and standard error for each visualisation technique. The results show that pseudo-colouring increased the mean sensitivity with both coordinate systems independent of decision bias.

Table 6.9: The mean sensitivity index for each visualisation technique. SD = Standard deviation. SE = Standard error.

Visualisation technique	Sensitivity index (d')	SD	SE
Cartesian no-colouring	1.87	0.55	0.01
Cartesian pseudo-colouring	2.22	0.49	0.01
Polar no-colouring	1.82	0.57	0.01
Polar pseudo-colouring	2.15	0.44	0.01

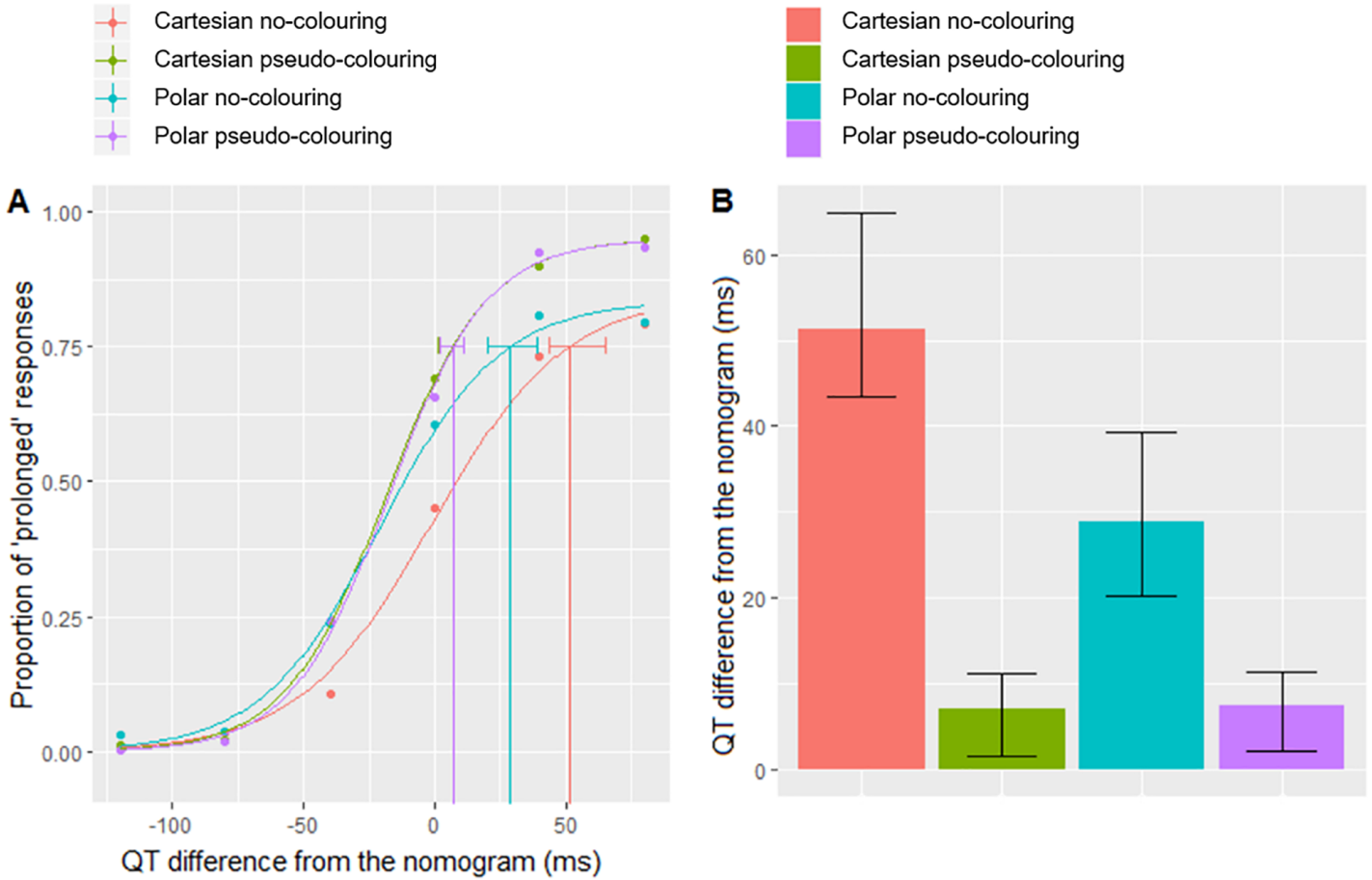


Figure 6.9: Psychophysical detection measures of sensitivity. (A) The fitted psychometric function plot shows the proportion of responses indicating QT-prolongation as a function of the QT-interval difference from the nomogram line, which corresponds to the six QT-levels. The QT value of the nomogram line (*i.e.* the fourth QT-level) is equal to 0 on the X-axis. Cartesian no-colouring is represented by a red line, Cartesian pseudo-colouring by a green line, Polar no-colouring by a turquoise line and Polar pseudo-colouring by a purple line. (B) The just noticeable difference (JND) thresholds plot. The error bars represent bootstrap confidence intervals.

Table 6.10: Results of the Friedman test comparing reaction times across the four visualisation techniques.

QT-level	Estimated Range	Risk for TdP	$\chi^2(3)$	p-value
1	Normal	No risk	35.251	**<0.001**
2	Normal	No risk	26.307	**<0.001**
3	Borderline	No risk	16.349	**0.001**
4	Prolonged	At risk	18.216	**<0.001**
5	Very Prolonged	At risk	32.223	**<0.001**
6	Severely prolonged	At risk	48.008	**<0.001**
Overall	All ECG stimuli		141.737	**<0.001**

6.3.2 Reaction time

We measured reaction time as the period from the appearance of the stimulus on the screen to the key press event when people made their binary decision. The mean reaction times across all stimuli were 7.9 and 8.6 seconds for Cartesian coordinates with and without pseudo-colouring respectively, and 8.1 and 8.9 seconds for Polar coordinates with and without pseudo-colouring respectively. A Friedman test was conducted for each QT-level stimulus and overall for each condition. People were significantly faster when pseudo-colouring was used for each QT-level and over all ECG stimuli cumulatively ($p \leq 0.001$)(Table 6.10, Figure 6.10). A *post hoc* pairwise comparison using a Wilcoxon signed-rank test with Bonferroni correction ($\alpha = 0.008$) showed that although participants responded faster when pseudo-colouring was used with Cartesian coordinates than with Polar coordinates, the difference was not significant ($p = 0.808$).

6.3.3 Satisfaction

A Friedman test showed there was a statistically significant difference in participants' satisfaction as a function of visualisation technique ($\chi^2(3) = 50.954$, $p < 0.001$). A *post hoc* pairwise comparison using a Wilcoxon signed-rank test with Bonferroni correction ($\alpha = 0.008$) showed that people preferred pseudo-colouring to no-colouring ($p < 0.008$); and Cartesian coordinates to Polar coordinates whether pseudo-colouring was used ($Z = -3.340^c$, $p = 0.001$), or not ($Z = -3.029^c$, $p = 0.002$). Although people were more sensitive to QT-prolongation when the ECG was presented on Polar coordinates with pseudo-colouring, people preferred Cartesian coordinates without pseudo-colouring to Polar coordinates with pseudo-colouring ($Z = -2.142^b$, $p = 0.032$).

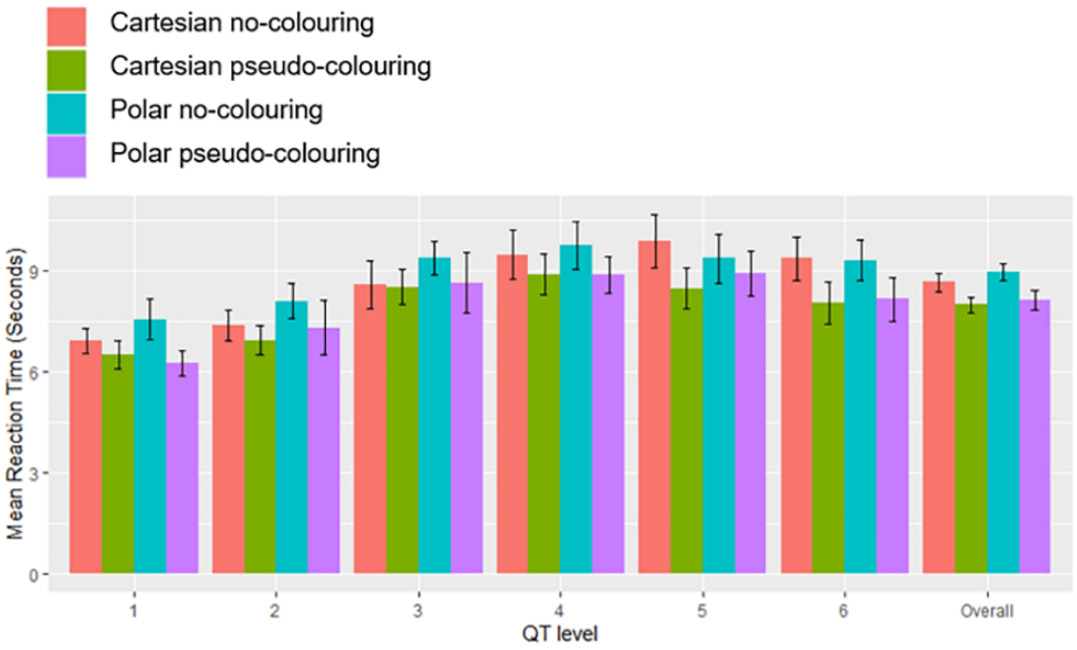


Figure 6.10: Mean reaction time in seconds for each QT-level and all stimuli combined. Error bars represent 95% confidence intervals.

6.3.4 Eye movement analysis

An area of interest (AOI) was created on the whole ECG for each stimulus. As such, 160 AOIs were created (40 ECGs $\times$ 4 visualisation techniques).

Mean fixation duration The mean fixation duration metric, which is an indicator of cognitive load [HNA$^+$11, DVDL14], was calculated across all participants for each ECG stimulus and then averaged over each QT-level. Mean fixation duration increased as the QT-level increased when pseudo-colouring was used (Figure 6.11A). When considered alongside the psychometric function (Figure 6.9A), this indicates that pseudo-colour helped people to focus visual attention. When pseudo-colouring was not used, mean fixation duration continued to increase at higher QT-levels for Cartesian coordinates, but this effect was not seen with Polar coordinates, indicating a difference in how people interpreted the ECG as a function of coordinate system.

Total fixation duration The total fixation duration (also known as the dwell time) quantifies the amount of time spent fixating on the stimulus [HNA$^+$11]. It was calculated across all participants for each stimulus and then averaged for each QT-level. The results show that total fixation duration was lower when pseudo-colouring was

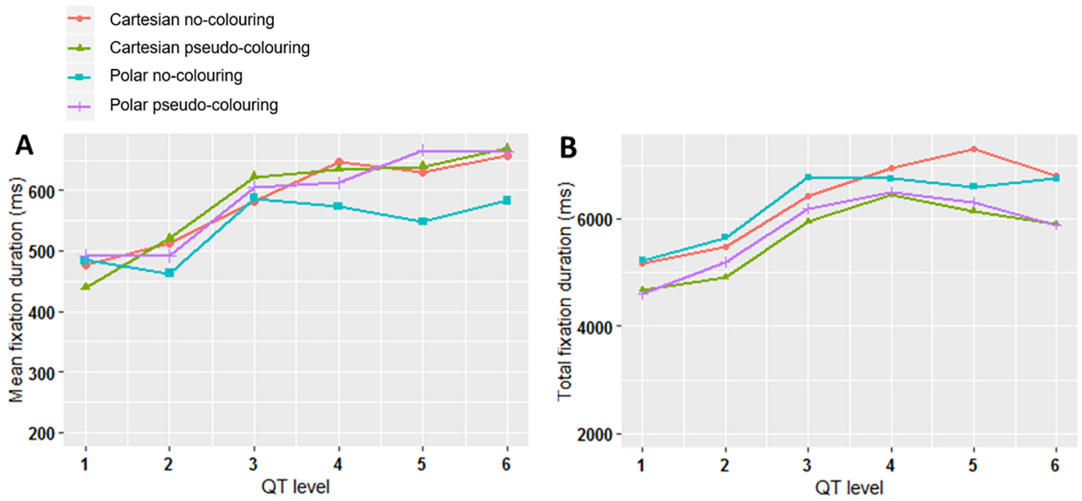


Figure 6.11: Eye movement analysis results. (A) Mean fixation duration in milliseconds on the ECG stimuli averaged for each QT-level. (B) Total fixation duration in milliseconds on the ECG stimuli averaged for each QT-level.

used for both coordinate systems (Figure 6.11B). People spent less cumulative time fixating on the ECG stimuli with pseudo-colouring, but individual fixations were more focused (Figures 6.11A and 6.11B). When pseudo-colouring was not used, there were more, shorter fixations, indicating a less-focused interpretation strategy.

Fixation distribution Figures 6.12 and 6.13 show heatmaps of fixations across an ECG stimulus with and without pseudo-colouring, for Cartesian and Polar coordinate systems respectively. We hypothesised that pseudo-colour helped to draw people's attention to the coloured T-wave area, and that fixations would therefore be more clustered when pseudo-colour was used. We used the Nearest Neighbour Index (NNI) to determine if the fixations were randomly spaced or clustered. The NNI provides a ratio of the distribution pattern of points in space from 0 (clustered) to 1 (dispersed) [DVHJ16]. We calculated the average Nearest Neighbour Index (NNI) for each ECG stimulus across all participants and for all visualisation techniques. Figure 6.14 shows a histogram of the distribution of the Nearest Neighbour Index (NNI), along with a box-plot for Cartesian and Polar coordinates, with and without pseudo-colouring. When Cartesian coordinates were used (Figure 6.14A), pseudo-colouring resulted in clustering in more stimuli than no-colouring. When Polar coordinates were used (Figure 6.14B), however, pseudo-colouring did not appear to affect the fixation distribution.

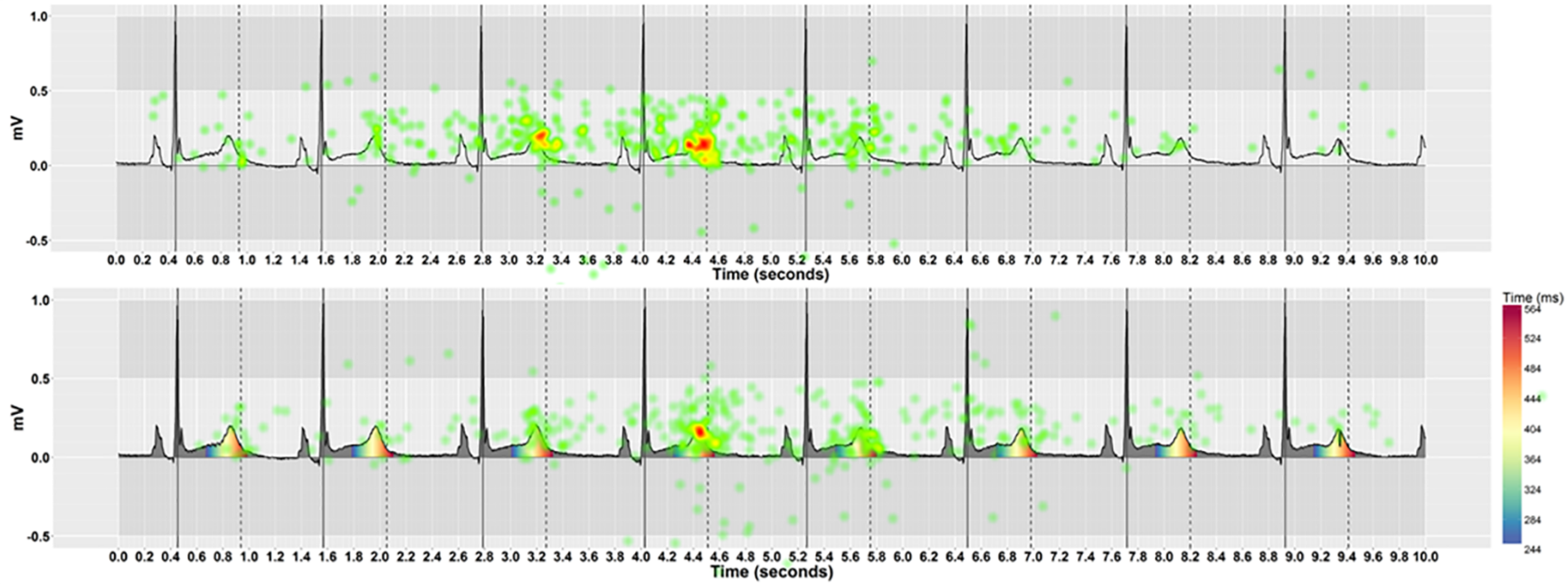


Figure 6.12: A heatmap of fixation count shows the cumulative number of fixations across all participants on Cartesian coordinates. With pseudo-colouring, people made fewer, more clustered fixations on the coloured T-wave area compared with no-colouring on the same ECG.

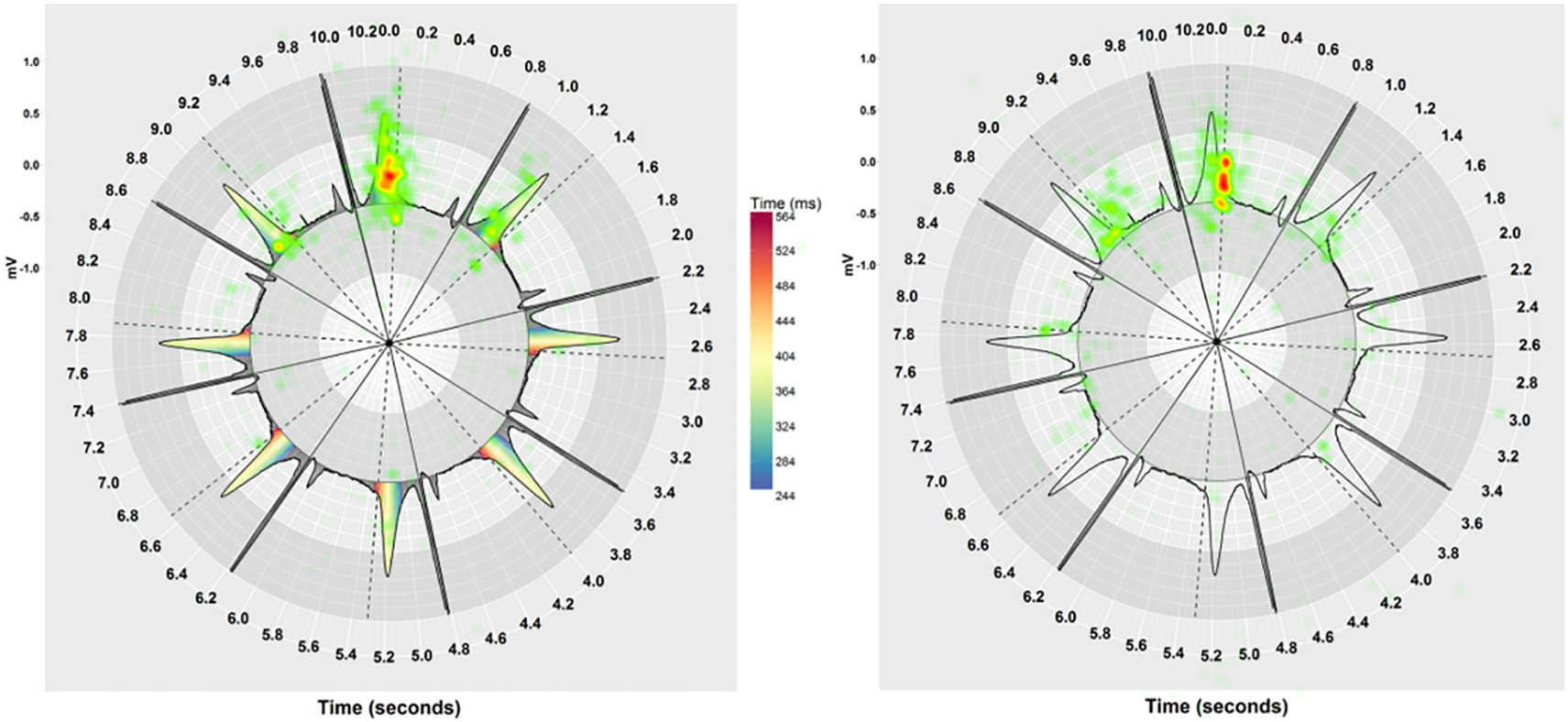


Figure 6.13: A heatmap of fixation duration shows the cumulative mean fixation duration across all participants on Polar coordinates. With pseudo-colouring, fixations were longer on the coloured T-wave area and the vertical dashed line that represents the QT risk threshold, while with no-colouring fixations were longer on the vertical dashed line only.

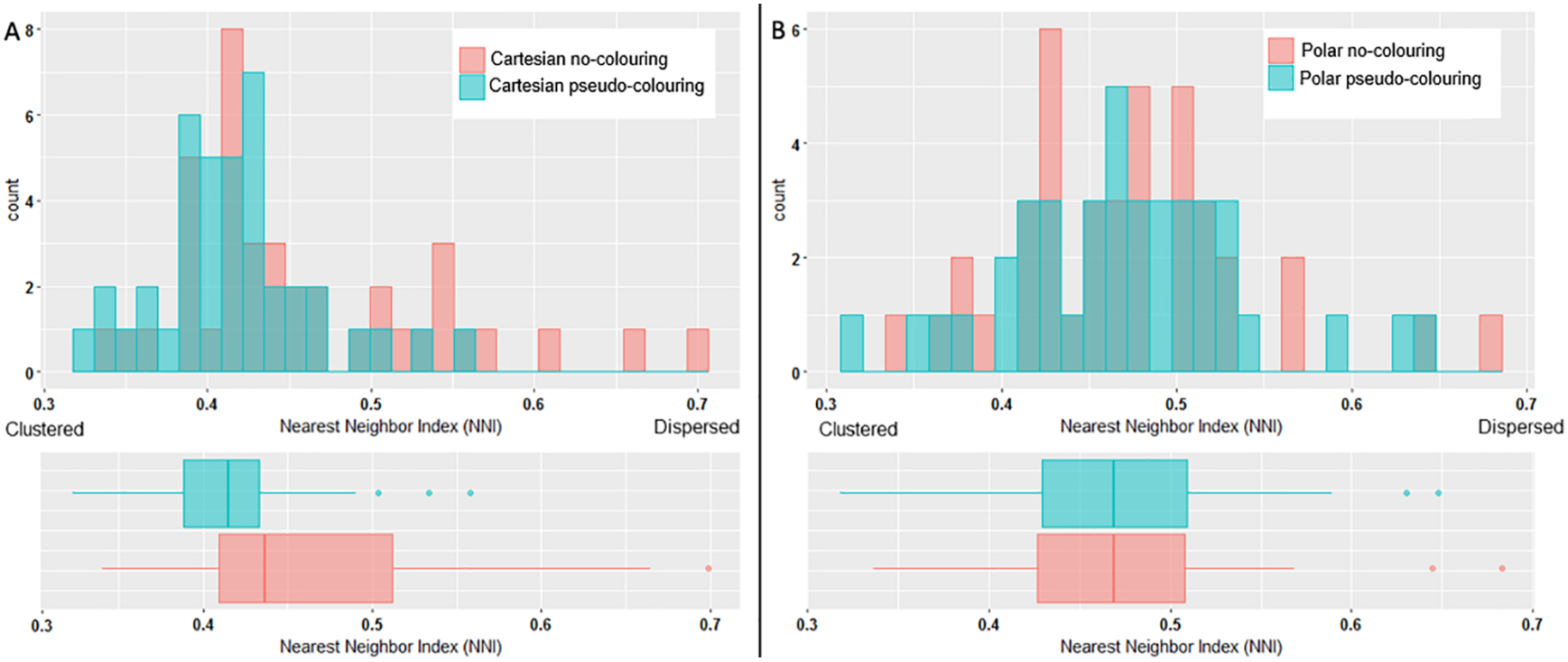


Figure 6.14: Nearest Neighbour Index (NNI) of fixations on ECGs with and without pseudo-colouring. (A) Cartesian coordinates. (B) Polar coordinates.

6.4 Discussion

This study demonstrates that applying pseudo-colouring to an ECG relative to the QT nomogram significantly increases lay people's accuracy in visually assessing QT-intervals regardless of heart rate (Tables 6.5 and 6.6). In particular, pseudo-colouring increased sensitivity to prolongation (true positive cases); whilst people were able to identify normal QT-intervals without colour, adding pseudo-colouring resulted in more accurate identification of at risk QT-prolongation cases, shown by both the ROC curve analysis and the psychophysical detection measures of sensitivity (Tables 6.8 and 6.9; Figure 6.9). The psychometric function shows that pseudo-colour helped increase sensitivity to QT-interval increases even before a patient is at risk of TdP (Figure 6.9A), specifically in the borderline range (*i.e.* the third QT-level). This level could be critical from a clinical perspective as just a 40ms increase from this point could put someone at risk of TdP. Clinical research has shown that even small ($\approx$ 10ms) QT-interval increases from the baseline should be regarded as a significant side effect of a QT-prolonging drug [RFF$^+$09, Bre10, FDA$^+$05]. From the perspective of our study design, this meant that pseudo-colouring with Cartesian coordinates reduced specificity (true negative cases) at this level, as we considered it to be within the normal range. From a practical perspective, however, considering this level to be abnormal may serve a useful purpose.

Pseudo-colour significantly reduced reaction times and helped to focus visual attention on the areas of the ECG crucial for detecting QT-prolongation. Eye-tracking data showed that pseudo-colour consistently increased mean fixation duration as the QT-level increased (Figure 6.11A); when the ECG was presented on Cartesian coordinates, fixations were more clustered when pseudo-colouring was used (Figure 6.14A).

Obtaining a precise measurement of the QT-interval is known to be challenging. An important QT-interval increase may be just a few milliseconds, less than one small square on the ECG. This would be particularly difficult to detect at high heart rates, as the QT-interval length shrinks with the R-R interval length [Mor01, GPAW14]. Here, we demonstrate that pseudo-colour can support detection of QT prolongation regardless of heart rate without needing to measure the QT-interval (Figures 6.4,6.5, 6.6 and 6.7). Figure 6.13 shows a heat map of absolute fixation duration across all participants on an ECG with a QT-interval 40ms above the nomogram line. When pseudo-colouring was not used, people's fixations appear to be more focused on measuring the gap between the end of the T-wave and the QT risk threshold dashed line. When pseudo-colouring is present on the same ECG, fixations appear more focused

on the T-wave. Focusing attention in this area is particularly useful, as long QT syndrome (LQTS) is associated with prolonging ventricular repolarization [MK05, DSR08], which is represented by the T-wave on the ECG [GMZ06, Sch71]. Research examining the effects of QT-prolonging drugs is investigating precisely how the T-wave responds to them to provide further insights of drug ion channel interactions and TdP risk [VJM$^+$15]. Pseudo-colour might ultimately help clinicians to determine which part of the T-wave—the initial half of the T-wave (J-T-peak) or the second half (T-peak T-end)—underpins QT-prolongation.

6.4.1 Limitations and future work

Whilst this work indicates the potential for pseudo-colouring to assist in the self-monitoring of QT-interval, the study took place in a controlled setting with a limited number of stimuli, and the transferability of the technique to practice remains an open question. Cartesian and Polar coordinates support the same level of accuracy, but people expressed a preference for pseudo-colour displayed on Cartesian coordinates. It may be, however, that the ECG trace would be better presented with Polar coordinates on smaller screens like smart watches, and future work should thus examine the effects of screen size and lighting setting on accuracy.

In this study, we only investigated the assessment of QT-interval, and it is not clear whether these visualisation techniques would generalise to interpretation of other ECG abnormalities, such as changes in ST-segment elevation. Research has shown that a large number of clinicians lack the skills to interpret QT-prolongation accurately [VRS$^+$05, STN08]; pseudo-colouring could also be used to help clinicians, especially within emergency departments, to visually assess and monitor patients' QT-intervals before or during the provision of a QT-prolonging medication. Future work should evaluate the visualisation techniques in clinical trials with more diverse clinical populations.

6.5 Conclusion

Applying pseudo-colouring to ECGs according to the QT-nomogram supports lay people in detecting QT-prolongation visually, regardless of heart rate. The results indicate that self-monitoring ECGs for drug-induced LQTS is feasible, with the potential to prevent the development of life threatening complications.

Chapter 7

Human-Machine Perception of Complex Signal Data

7.0 Chapter overview

7.0.1 Thesis context

At present, clinicians do not regard automated QT measurement algorithms as a reliable way of detecting QT-prolongation (discussed in Chapter 1 as the fourth research challenge). Current automated ECG interpretation approaches rely on recognizing different ECG waves, identifying the beginning of the Q-wave and the end of the T-wave to measure the QT-interval—a process that quickly becomes challenging in the presence of anomalies, artefacts or non-standard ECG waves, as discussed in detail in Chapters 2 and 3. Based on the evidence that pseudo-colouring significantly improves laypeople's ability to detect QT-prolongation by perceiving the amount of warm colour in the signal, alleviating the need to measure the QT-interval (Chapters 5 and 6), this chapter introduces an exploratory algorithm (Section 7.3) that uses an understanding of human visual perception of ECG data to improve automated QT-prolongation detection.

An empirical evaluation of the algorithm demonstrates how modelling the perceptual process that we hypothesise humans are using to interpret the pseudo-coloured ECGs can be used as a basis for a simple, explainable rule-based algorithm that is more accurate than current signal processing techniques (Section 7.3.2), and has the benefit of the human and machine sharing the same representation of the data (Section 7.3.1). We discuss the potential of this novel, shared human-machine interpretation

approach, in terms of its accuracy and acceptability in clinical practice (Section 7.4).

The promising results of the work reported in this chapter encouraged further evaluation of this new 'human-like' approach to the ECG interpretation, testing its effectiveness in detecting QT-prolongation at risk of TdP on a larger number of ECGs with a wide range of drug-induced ECG morphological changes (Chapter 8).

The main content of this chapter is adapted from: Alaa Alahmadi, Alan Davies, Katherine Dempsey, Markel Vigo, and Caroline Jay. Human-machine perception of complex signal data. In *Human-Like Machine Intelligence*. Oxford University Press, 2021.

7.0.2 Author's contributions

Alaa Alahmadi and Caroline Jay conceptualised and devised the idea for the work. Alaa Alahmadi developed and evaluated the human-like algorithm, carried out the ECG data acquisition, designed and analysed the study and wrote the paper, with Alan Davies, Markel Vigo and Caroline Jay contributing significant edits. Katherine Dempsey assisted with comparing the human-like algorithm with signal processing techniques. Alan Davies acted as the electrocardiogram domain expert throughout. Caroline Jay and Markel Vigo provided continuous guidance and discussion.

7.0.3 Published abstract

Electrocardiograms (ECGs), which capture the electrical activity of the human heart, are widely used in clinical practice, and notoriously difficult to interpret. Whilst there have been attempts to automate their interpretation for several decades, human reading of the data presented visually remains the 'gold standard'. We demonstrate how a visualisation technique that significantly improves human interpretation of ECG data can be used as a basis for an automated interpretation algorithm that is more accurate than current signal processing techniques, and has the benefit of the human and machine sharing the same representation of the data. We discuss the potential of the approach, in terms of its accuracy and acceptability in clinical practice.

7.1 Introduction

This research explores a new approach to the processing of complex signal data, which exploits an understanding of the human perceptual system to facilitate its interpretation by humans and machines simultaneously. We focus on the interpretation of electrocardiogram (ECG) data, and in particular on the heart condition known as 'long QT syndrome (LQTS)', which is associated with a life-threatening arrhythmia (*Torsades de pointes* (TdP)) that can lead to sudden cardiac death. The syndrome is characterized by a prolongation of the QT-interval on the ECG, which represents the duration of the ventricular depolarization and repolarization cycle, and is measured from the beginning of the Q-wave to the end of the T-wave [GM08], as shown in Figure 7.1. Whether or not the interval is considered to be prolonged depends on a number of factors, and particularly heart rate. Given a heart rate of 60bpm, a normal QT-interval would generally be 430ms or less [GM08, YC03].

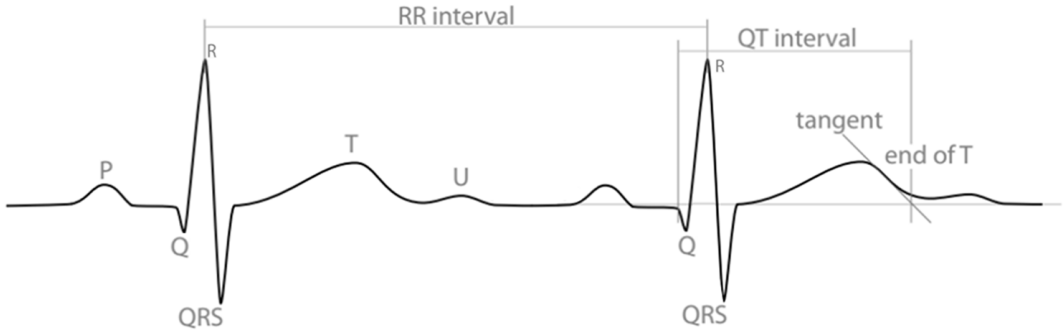


Figure 7.1: Measurement of the QT-interval on the ECG from the beginning of the Q-wave to the end of the T-wave. Figure taken from [ADR+19].

7.1.1 Interpreting the QT-interval on an ECG

The electrocardiogram (ECG), a recording of the complex signal data representing the heart's electrical activity, is widely used in clinical practice for assessing cardiac function and detecting pathologies. The standard method for visualising ECG data is with a two dimensional line graph showing the amplitude of the recorded electrical signal of the heart on the Y-axis and the time in milliseconds (ms) on the X-axis. The ECG 'wave forms' (peaks and troughs) are labelled with letters and represent different stages of the heartbeat, as shown in Figure 7.1.

Long QT syndrome is indicated by a prolongation of the QT-interval on the ECG, representing a delay in the ventricular repolarization activity of the heart [GM08]. Many commonly prescribed medications can prolong the QT-interval, leading to acquired drug-induced long QT syndrome (diLQTS).

Whilst there have been attempts to automate ECG interpretation for several decades, expert human reading of the data presented visually remains the 'gold standard' [WBA+14, SW17]. Nevertheless, ECG interpretation is complex and requires extensive training, with some conditions known to be very challenging to recognise, even for clinicians who routinely read ECGs [VRS+05]. To date, long QT syndrome has remained difficult to recognise on the ECG, from both a human and a machine perspective [VRS+05, TAS+15, KBD+18, GL13, MCbA01, TABW11, RSG09]. Here, we demonstrate how a visualisation technique that significantly improves human interpretation of ECG data—without the need for prior training—can be used as the basis for an automated human-like algorithm.

7.1.2 The broad notion of human-machine perception

Many approaches to computer vision, including deep learning, attempt to mimic or improve on human ability, but are often only loosely related to human visual processes [SANC14, SBOR06]. A growing area of research is investigating the role that models of human perception may play in improving computer vision. Perception, broadly speaking, is the process of recognising and interpreting sensory information [HW97]. The human visual system has a highly developed capability for perceiving different patterns and objects simultaneously, both whole and in parts, without prior training [SANC14, SBOR06, OAD+12]. This appears to be due to a primitive perceptual organization process that derives relevant groupings and patterns from an image or scene without prior knowledge of its contents [Low12, vdH17, Sta19], a phenomenon first noted in the Gestalt principles of visual perception, which articulate factors that regulate perceptual grouping, including proximity, similarity, closure, continuation and symmetry [Wer23]. Today, there is evidence of a multiplicity of perceptual grouping processes that vary in attentional demands [RYK17, BK03, KHBP05, KRA04], with many operating before the higher level cognitive system applies top-down knowledge to recognising a scene [vdH17, Sta19].

Research in computer vision has shown that the capability of a machine to organise and interpret sensory information in a 'human-like' way, termed *machine perception* [Vel10] can dramatically decrease the search space required for object recognition [Low12]. Combining a perceptual grouping approach with the principle of simplicity, which states that people tend to perceive the simplest possible interpretation of any given visualised information [Fel16, Hoc57, HE85], has been shown to enhance machine vision further [Fel16, vdH15, DSP90, Fel97].

Pre-attentive processing theory outlines a set of visual properties known to be detected rapidly and accurately by the human eye, which are important in the perceptual grouping process [WU19, The13, Not93]. Examples of pre-attentive properties include colour, shape, and size. Colour, in particular, is known to aid and influence the perceptual organization of the visual scene [KRA04]. Zavagno et al. (2014) have shown colour to be a relatively strong grouping factor that functions according to the principles of Gestalt theory, and can override other types of pre-attentive property including shape and size [ZD14]. In the technique we describe here, colour serves as the foundation for drawing the human reader's attention to the QT-interval in an ECG image, such that he or she can make a decision about whether it is dangerously prolonged. The salience information used by humans to make this judgement is then mapped to quantitative values, which can be used by an algorithm to automate the detection of prolongation.

7.1.3 Human-machine perception of ECG data

The majority of machine vision algorithms use machine learning, and in particular neural networks, as the basis for processing image data. Here, we take a different approach, in terms of both the data representation, and the role of human vision in the resulting algorithm. Our starting point is ECG signal data that is presented visually for human interpretation, and is typically analysed computationally using signal processing methods. As a first step, the signal data is visualised with colour, such that a human can easily draw the relevant information from it. Following this, the perceptual process we hypothesise the human is using to interpret the data is modelled, and forms the basis of a simple rule-based algorithm that accurately classifies whether the interval is prolonged.

7.2 Using pseudo-colour to support human ECG interpretation

Recognizing QT-interval prolongation on the standard ECG is notoriously difficult, even for trained medical professionals [VRS+05]. From a perceptual perspective, this is likely to be related to the fact that humans are poor at judging quantity on a horizontal scale [LSZ+09, WA84, Lib91, PK11]. Morphological diversity of the wave forms and artefacts in the ECG signal serve to exacerbate this issue [ADVJ18, THTA07, PDJVdBW08].

Our previous work—motivated by the potential benefits of self-monitoring for drug-induced long QT syndrome—considered the problem from the perspective of the lay person with no experience of ECG interpretation. To support this challenging target population we used knowledge of visual perception to enhance the way the ECG is visualised. Colour is a particularly powerful way of attracting visual attention in complex scenes, and aids visual recognition via perceptual grouping [Tre83, OS00, KRA04]. Based on this phenomenon, we produced a visualisation technique that highlights the duration of the QT-interval on the ECG using pseudo-colouring, a salient means of representing continuously varying values using a sequence of meaningful colours [War12]. This shifts the visual encoding process from perceiving a distance between two waves, to perceiving colour in terms of hue and intensity. Applying pseudo-colouring to the ECG significantly increased the speed and accuracy of human perception of QT-prolongation, at both a regular 60 beat-per-minute heart rate [ADR+19] and at varying heart rates, with diverse T-wave morphologies [ADVJ20a].

7.2.1 Pseudo-colouring method

The first step in the application of pseudo-colouring was to detect the R-peaks on the ECG, which identify heartbeats (see Figure 7.1). Note that it is trivial to detect the R-peaks in the vast majority of ECGs, as they consistently have the greatest amplitude (the amplitude of the other waves varies considerably). Pseudo-colour was then applied as follows:

- **Identifying the risk threshold for QT-prolongation:** The risk threshold for QT-prolongation changes according to heart rate. The QT-nomogram, a clinical assessment method that shows the risk of TdP by considering QT-interval as a

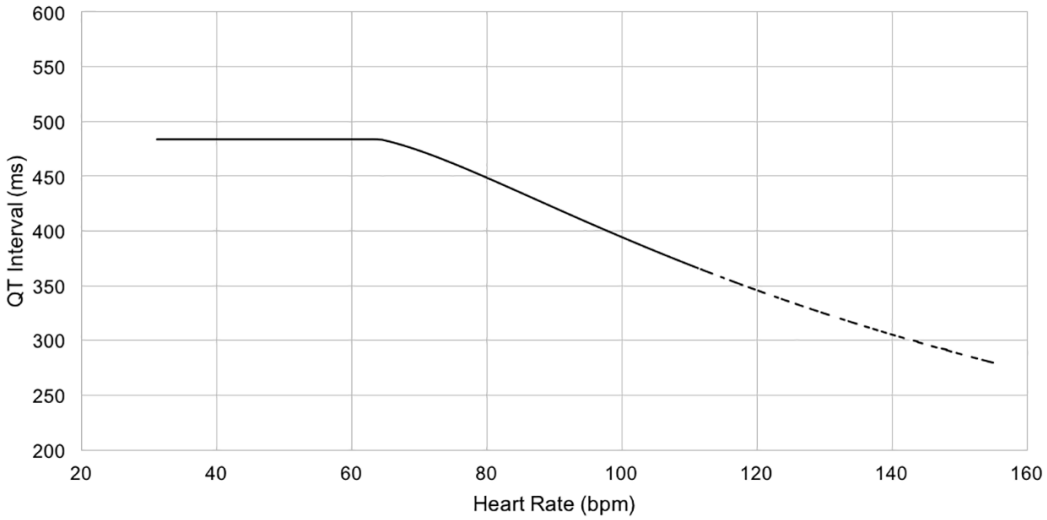


Figure 7.2: The QT-nomogram [CIKD07].

function of heart rate, was used to identify the threshold [CIKD07]. Figure 7.2 shows the QT nomogram plot. If the QT/HR pair plots on or above the risk line, the patient is at clinically significant risk of TdP; below the line the patient is not considered at risk. The risk threshold was calculated for each heart rate using the nomogram risk line. The risk threshold for a given heart rate was termed 'QT-value at risk'.

- **Applying pseudo-colouring to each heartbeat:** In clinical practice, the QT-interval is measured by counting the small squares (each representing 40ms) on the standard ECG background grid from the beginning of the Q-wave to the end of the T-wave. The time period of interest (*i.e.*, an approximation of the QT-interval) was calculated for each heartbeat from the R-peak minus 20ms (estimated as the start of the Q-wave) to the maximum potential QT-interval, which was estimated as the QT-value at risk plus two small squares (80ms). This formed an additional time dimension, to which the pseudo-colour could be mapped. Pseudo-colouring was then applied periodically over each heartbeat to the area between the isoelectric baseline (where amplitude is zero) and the ECG signal, by mapping the relevant area of the heartbeat time to a pseudo-colouring sequence.

We used a spectrum-approximation pseudo-colouring sequence, where cool spectral colour codes (purple to blue to green) were used to indicate normal QT-interval ranges, and warm colours (yellow to orange to red) to show abnormal QT-interval ranges. This produced nine indices on the pseudo-colouring scale, where each index was mapped to a colour code and represented a small square on the ECG, starting backwards from the nomogram line showing the QT-value at risk plus two small squares (80ms) to six small squares (240ms) below the nomogram line.

The QT-value at risk was mapped to dark orange. Forty ms and 80ms above this were mapped to red and dark red respectively, to indicate higher risk. Within each square, the intensity of the hue changed every millisecond, to show time progression. Figure 7.3 shows an illustration explaining how the pseudo-colouring technique was applied according to the standard ECG background grid. The way in which colours were mapped to the time period of interest according to the nomogram is shown in Table 7.1. Figure 7.4 shows examples of ECGs with normal and very prolonged QT-intervals visualised using the pseudo-colouring technique. Figures 7.5 and 7.6 show how the pseudo-colouring was adjusted according to heart rate (based on the QT-nomogram), where the pseudo-colouring shows the same level of risk, despite heart rate differences. The R script used to implement the visualisation technique can be found in [AVJ20].

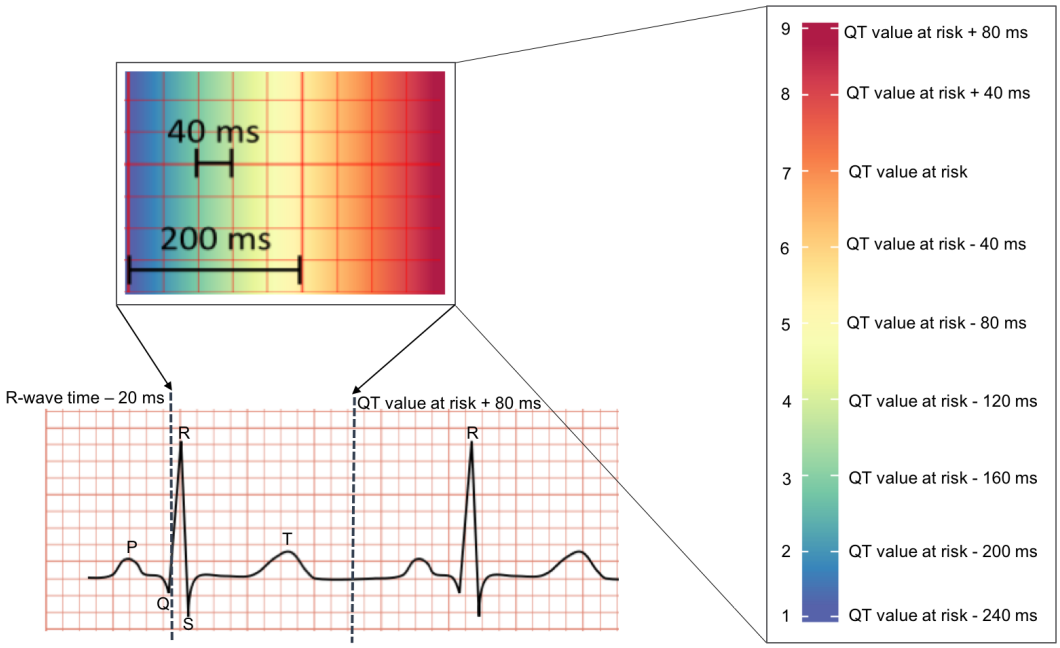


Figure 7.3: Mapping the pseudo-colouring to the ECG. A small square on the grid is equal to 40ms.

Table 7.1: The nine indices on the pseudo-colouring scale with their corresponding time value in milliseconds (ms) and colour code.

Index	Corresponding time value (ms)	Color code
1	QT-value at risk - (40 $\times$ 6)	Purple
2	QT-value at risk - (40 $\times$ 5)	Blue
3	QT-value at risk - (40 $\times$ 4)	Green
4	QT-value at risk - (40 $\times$ 3)	Lime
5	QT-value at risk - (40 $\times$ 2)	Yellow
6	QT-value at risk - (40 $\times$ 1)	Orange
7	QT-value at risk	Dark orange
8	QT-value at risk + (40 $\times$ 1)	Red
9	QT-value at risk + (40 $\times$ 2)	Dark red

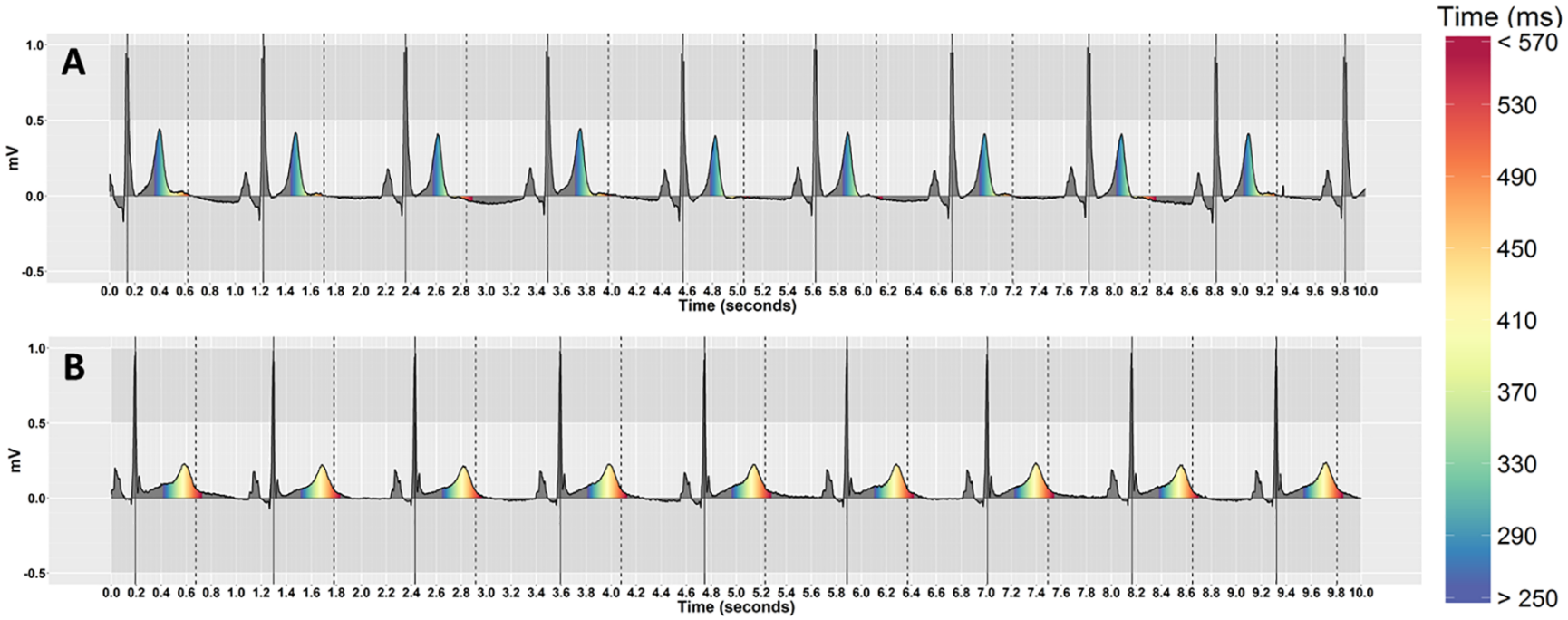


Figure 7.4: Examples of ECGs with pseudo-colouring, showing (A) a normal QT-interval (HR = 55, QT = 361ms) and (B) a dangerously prolonged QT-interval (HR = 52, QT = 579ms).

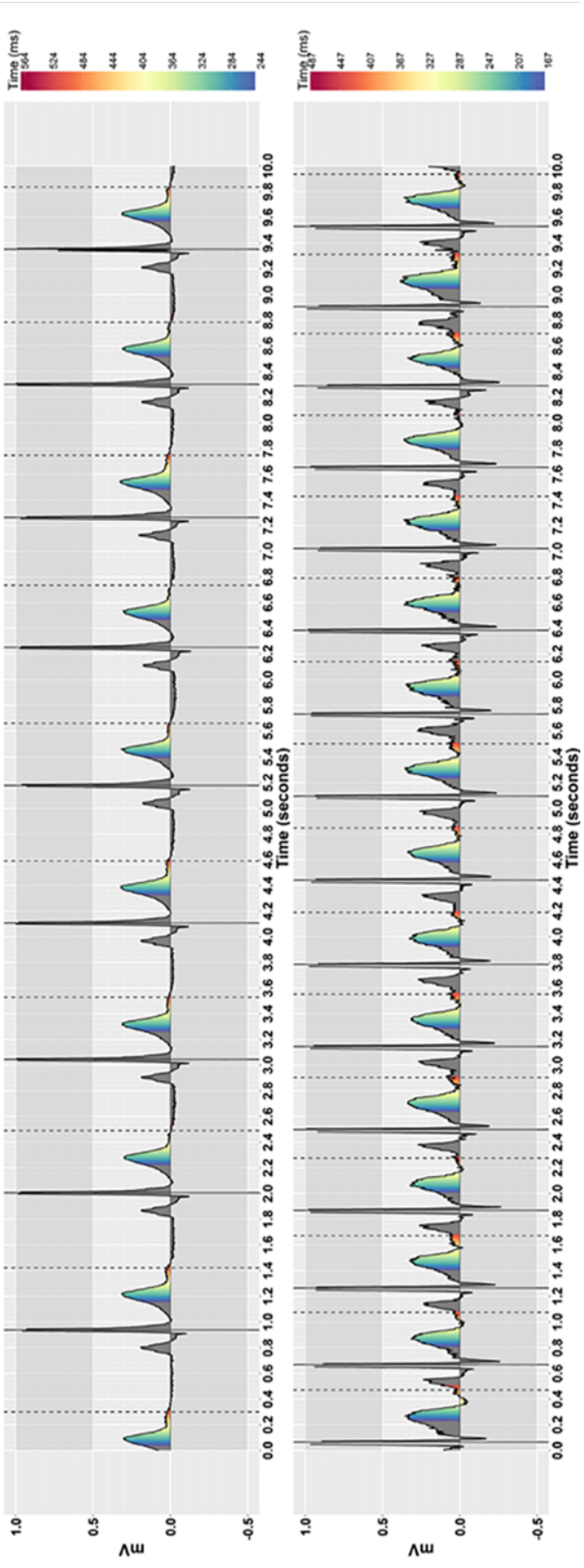


Figure 7.5: Examples of ECGs with pseudo-colouring that have the same normal QT-level, but different heart rates.

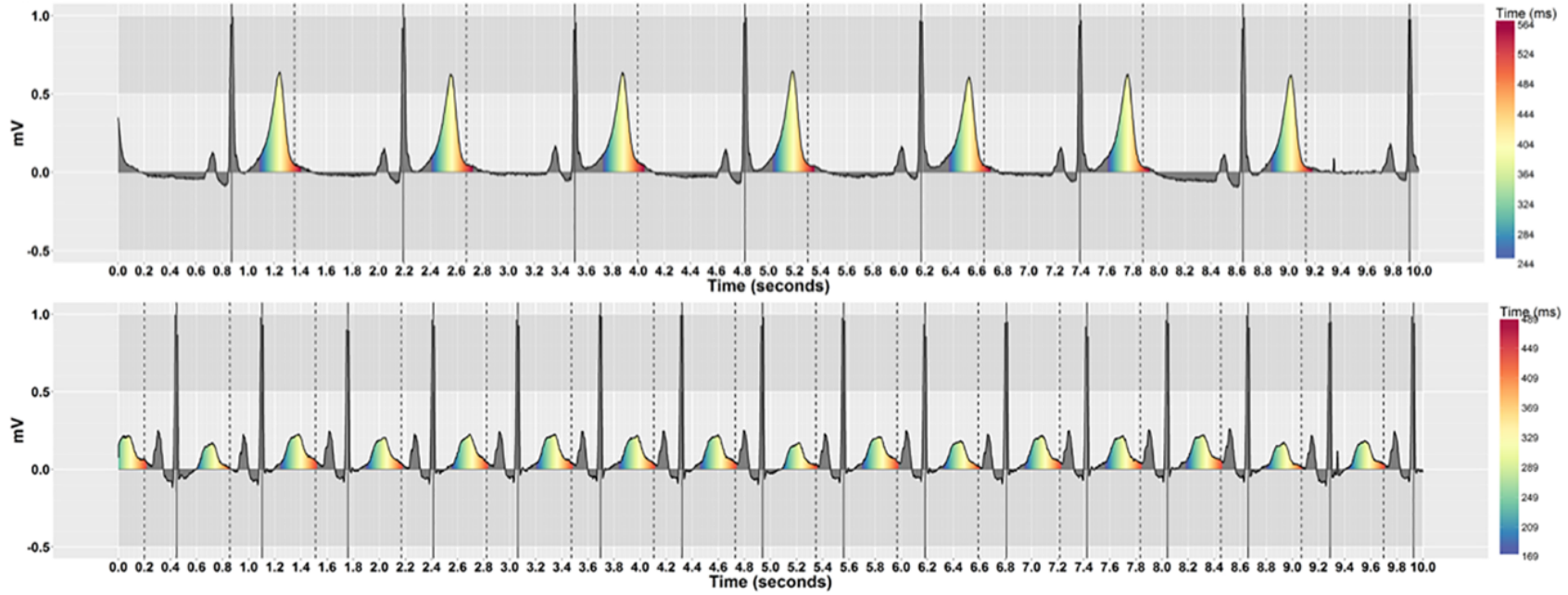


Figure 7.6: Examples of ECGs with pseudo-colouring that have the same abnormal QT-level, but different heart rates.

The pseudo-colouring technique was shown in an evaluation to be very effective, enabling lay people with no prior training in ECG interpretation to detect QT prolongation with 83% sensitivity, compared with 63% when pseudo-colour was not used. For more information about the visualisation technique and its evaluation with humans see [ADR+19] and [ADVJ20a].

7.3 Automated human-like QT-prolongation detection

Automated ECG interpretation systems have typically proved poor at detecting LQTS [TAS+15, KBD+18, EI13, GL13, MCbA01, TABW11, RSG09]. A major challenge for automated QT-detection algorithms is identifying the precise end of the T-wave (the terminal point), particularly when the T-wave's morphology is abnormal [GPAW14, HC94, GMZ06, Mor01]. This is particularly problematic, as medications that prolong the QT-interval often change levels of the blood's electrolytes including potassium, calcium, sodium, and magnesium, which can affect T-wave morphology [VJM+15]. Methadone, a drug that is infamous for prolonging the QT-interval and increasing the risk of TdP, also causes changes in the T-wave that cause the interval to be underestimated [TAS+15].

The pseudo-colouring technique was able to communicate QT-prolongation in such a way that humans were able to accurately perceive risk of TdP, as the signal from the colour reduces the need to identify the end of the T-wave. It thus follows that an algorithm using the same or an equivalent process to 'perceive' the information encoded in the colour should also be able to perform this task. We hypothesised that quantifying computationally the amount of warm colour displayed in the ECG signal could help a machine to detect LQTS, alleviating the need to measure the QT-interval directly.

The heuristic of quantifying area, rather than identifying the interval *per se*, is effective because the T-wave generally has the largest area under the curve of the ECG signal. The QT-interval is considered normal when the T-wave is located in the cool colour region, and prolonged when in the warm colour region. The pseudo-colour thus highlights the T-wave position in relation to the inter-heartbeat time dimension, without needing to identify either the peak or end of the wave. As such, the first step in the computational human-like algorithm is to calculate the area under the curve of the ECG signal using the trapezoidal rule. In mathematics, the trapezoidal rule is a method commonly used for approximating the definite integral that estimates the area under the curve of a linear function by dividing the area into a number of strips of equal width. That is, given a linear function $f(x)$ of a real variable x and an interval [a, b], the rule estimates the area under the graph of the function $f(x)$ as a trapezoid, calculating its area as follows:

$$\int_a^b f(x)dx \approx \frac{(f(a)+f(b))}{2}(b-a) \tag{7.1}$$

The raw ECG signal has an array of X and Y values, where X represents the time of the ECG signal in milliseconds, and Y represents the amplitude of the ECG signal. We considered [a, b] to be the interval of two successive timestamps in the X array of the ECG signal. If we consider the time interval of x_1 and x_2, the trapezoidal rule was applied by taking the average of amplitude of the ECG signal on the Y-axis of this interval as $f(x_1)$ and $f(x_2)$, and multiplying it by the difference in time between x_1 and x_2. As the time of the ECG is represented by integer numbers ranging from 1 to 10000 milliseconds for a 10-second ECG recording, the difference between any two successive times x_1 and x_2 is always equal to one. Using the finest possible level of granularity maximised the precision of the estimation. Then, the total area under the curve was calculated for every 40ms index on the pseudo-colouring scale. Cool spectral colours from indices 1 to 4 represented normal QT-intervals, and warm spectral colours from indices 5 to 9 represented prolonged QT-intervals. The percentage of the area under the curve was then calculated for warm and cool colours respectively. The QT-interval was considered 'prolonged' by the algorithm if the proportion of warm colours was greater than the that of cool colours; otherwise it was considered 'normal'. In summary, the QT-interval was considered 'prolonged' if the warm colours occupied more than 50% of the area under the curve of the ECG signal. The full human-like

algorithm is described in the pseudo-code below and the R script used to implement it can be found in [AVJ20].

Algorithm 1 Human-like algorithm

Data: ECG signal with X and Y

Input: X is time list $x[i]$, $i = 1,2,\cdots,10000$.

Y is amplitude list $y[i]$, $i = 1,2,\cdots,10000$.

Result: Normal or prolonged QT-interval

```
// Calculate area under the curve (AUC)
```

$i \leftarrow 1$

while $x[i] \leq 10000$ **do**

$delta\_x \leftarrow x[i+1] - x[i]$

$AUC[i] \leftarrow ((y[i] + y[i+1])/2)(delta\_x)$

$i \leftarrow i+1$

end

```
// Acquire the time points of the nine indices of the
   pseudo-colouring scale
```

Pseudo-colouring scale is time list $index[i]$, $i = 1,2,\cdots,9$.

for *each index i in the pseudo-colouring scale* **do**

Calculate the sum of AUCs of the 40ms interval:

$Sum\_index[i] \leftarrow \sum_{index[i]}^{index[i+40]} AUC[i]$

end

```
// Calculate percentage of AUC for warm colours
```

$Sum\_cool\_areas \leftarrow \sum_{index[1]}^{index[4]} AUC[index]$

$Sum\_warm\_areas \leftarrow \sum_{index[5]}^{index[9]} AUC[index]$

$Total \leftarrow Sum\_cool\_areas + Sum\_warm\_areas$

$Warm\_Colours\% \leftarrow (Sum\_warm\_areas/Total) \times 100$

```
// Inference rule
```

if $Warm\_Colours\% \geq 50$ **then**

QT is prolonged

else

QT is normal

end

Table 7.2: The sensitivity, specificity and overall accuracy of the human-like algorithm and human participants (mean values) when classifying the 40 ECGs.

	Human-like algorithm	Human	Difference
Sensitivity	0.85	0.83	0.02
Specificity	0.95	0.90	0.05
Overall accuracy	0.90	0.86	0.04

7.3.1 Comparison with human interpretation

We evaluated the accuracy of the human-like algorithm by first comparing it with the results of a study conducted with humans [ADVJ20a]. The ECGs ($n = 40$) were acquired from a clinical study conducted to assess QT-interval prolongation in healthy subjects receiving medication known to cause this issue [JVM$^+$14]. As part of the clinical study QT-intervals were calculated for all ECGs, and it is these values that were used as ground truth for our subsequent evaluation. ECGs were selected from multiple patients ($n = 17$), and represented different values of the QT-interval and heart rate, with 20 ECGs showing a normal QT-interval, and 20 ECGs showing clinically significant QT-prolongation. The ECGs had different heart rates, with some morphological T-wave changes caused by the QT-prolonging drugs. The ECG datasets can be found in the PhysioNet database [GAG$^+$00], and the clinical trial study is described in [JVM$^+$14].

We measured the sensitivity, specificity and overall accuracy of the classification. The sensitivity is the ability of the classifier (human/algorithm) to correctly identify those patients with the disease, and was calculated as the proportion of correctly classified 'prolonged' ECGs. The specificity is the ability of the classifier to correctly identify those patients without the disease, and was calculated as the proportion of correctly classified 'normal' ECGs. The overall accuracy was calculated as the proportion of correct classification of the 40 ECGs, *i.e.* the average of the sensitivity and specificity. The results of the human-like algorithm were very similar to those of the human participants, with both showing slightly higher specificity than sensitivity, and the algorithm being slightly more accurate overall (see Table 7.2 and Figure 7.7).

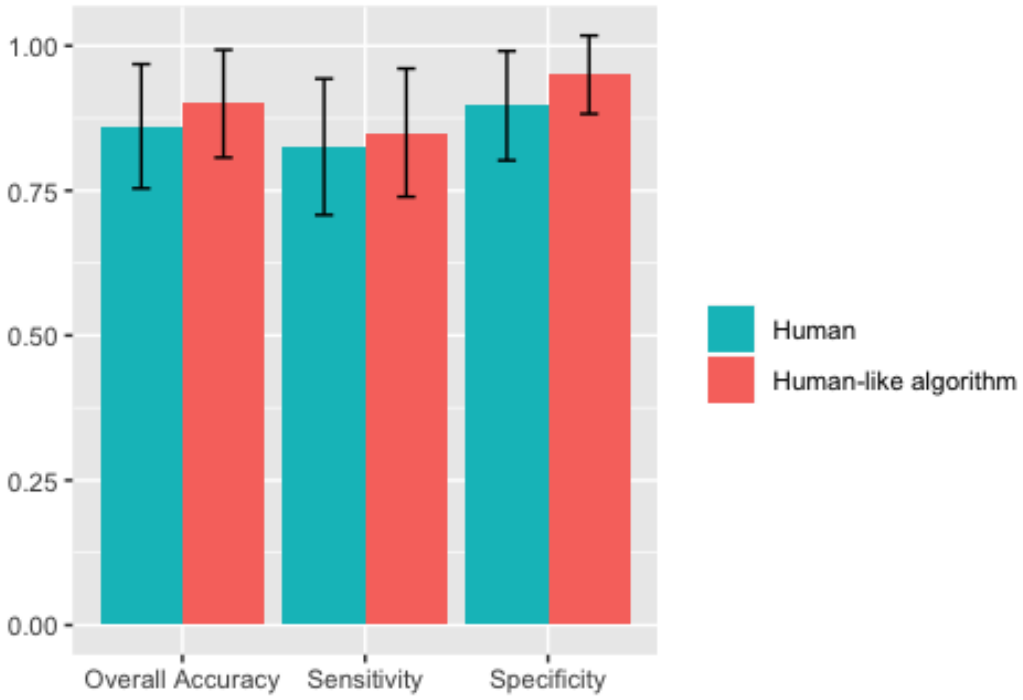


Figure 7.7: The sensitivity, specificity and overall accuracy of the human-like algorithm and human participants (mean values) when classifying the 40 ECGs. The error bars represent 95% confidence intervals.

7.3.2 Comparison with signal processing approaches

The majority of automated QT-interval analysis algorithms are proprietary or unavailable, and as such formally bench-marking the performance of our algorithm is not possible. Here, we compare it with signal processing approaches reported in the literature that have previously been applied to QT-interval measurement. The logic behind the human-like algorithm differs considerably from that used by standard signal processing methods, as it takes a naive perspective, calculating the percentage area of warm colours relative to cool colours, without prior identification and detection of the Q-wave or T-wave. By contrast, traditional signal processing approaches to ECG interpretation are based on the precise determination of the onset and offset of the different waves and complexes (P-wave, QRS complex, T-wave). This process is relatively straightforward if the ECG signal has a normal sinus rhythm, but it quickly becomes challenging in the presence of anomalies, artefacts or non-standard ECG waves [SW17].

ECG wave characteristics are also known to differ substantially across individuals, and are affected by factors including age, race, sex and health status [GMZ06, MMDY94, HST+16]. At present, there are no standard definitions for the ECG waves [Wil80, Par85, SW17]. As a result, differences in signal processing measurements persist. In addition to the challenge of correctly recognising the different ECG waves, accurate measurements of intervals (PR, QRS, QT) are particularly difficult to make, and thus the methods of determining onset and offset of waves vary among algorithms.

We implemented two common signal processing algorithms that use different methods to identify the end of the T-wave. The first algorithm, proposed by Hermans et al. (2017), uses an automated tangent method to identify the point of the maximum T-wave down-slope as the end of the T-wave [HVB+17]. The second algorithm uses the 15% threshold method which determines the end of the T-wave as the point in time when the ECG signal crosses the threshold at 15% of the amplitude of the T-wave peak [Hun05]. In both cases, the Q-wave onset was determined using the same method as the visualisation technique, which is the time of the R-peak - 20ms.

We compared the sensitivity, specificity and overall accuracy of the two signal processing methods with the human-like algorithm using the same 40 ECGs tested in the human interpretation study [ADVJ20a]. The signal processing methods measured the QT-interval of the ECGs as numbers and calculated the heart rate. As the pseudo-colouring technique was adjusted according to the nomogram to detect normal and prolonged QT-intervals across different heart rates, the signal processing methods were implemented to classify the QT-interval as 'normal' if the calculated QT/HR plot was below the nomogram line, and otherwise as 'prolonged'.

Table 7.4 shows the 40 ECGs ordered by QT-level from low-risk (1-3) to high-risk (4-6). The results show that both signal processing methods significantly underestimate the QT-interval, with most ECGs with a prolonged QT-interval being classified as 'normal' based on the nomogram (see Figure 7.8 and Table 7.3). The mean difference in milliseconds between the actual QT values and the calculated QT values were 61ms and 62ms for the automated tangent method and 15% threshold method respectively. The accuracy of the human-like algorithm was considerably better, due to its dramatically higher sensitivity, as shown in Figure 7.8. Table 7.4 illustrates how, as the QT-level relative to the nomogram increased, the percentage of warm colours calculated by the human-like algorithm also increased.

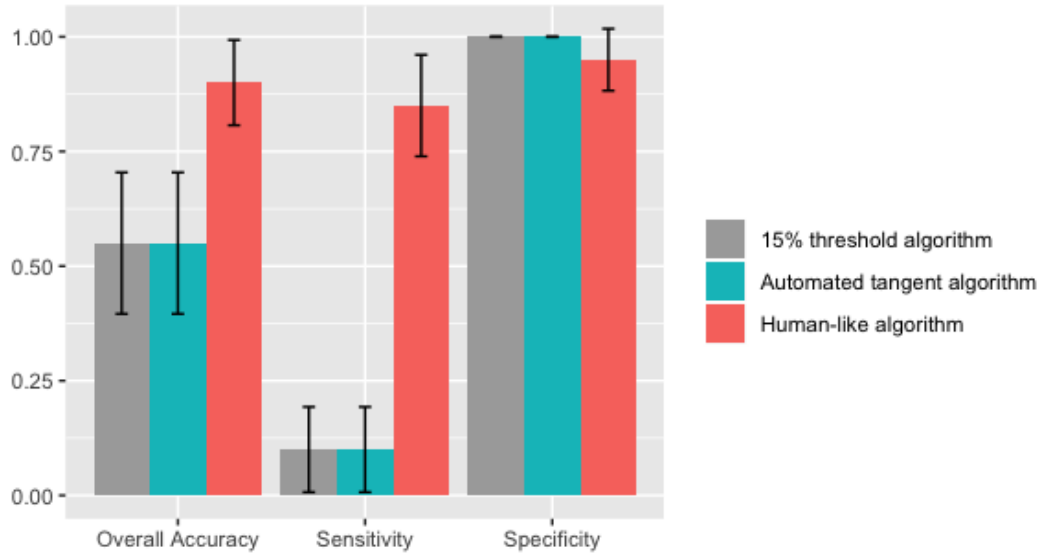


Figure 7.8: The sensitivity, specificity and overall accuracy of the human-like algorithm and two signal processing algorithms when classifying the 40 ECGs. The error bars represent 95% confidence intervals.

Table 7.3: The sensitivity, specificity and overall accuracy of the human-like algorithm and the two signal processing algorithms when classifying the 40 ECGs.

	Human-like algorithm	Automated tangent algorithm	15% threshold algorithm
Sensitivity	0.85	0.10	0.10
Specificity	0.95	1.00	1.00
Overall accuracy	0.90	0.55	0.55

Table 7.4: The 40 ECGs with, from left to right, the actual values as measured in the clinical study of the QT-interval in milliseconds, the heart rate (HR) in beats per minute, the QT-level at risk, the calculated QT values in milliseconds of the two signal processing methods, and the percentage of warm colours calculated by the human-like algorithm.

ECG	QT	HR	QT-level	Automated tangent algorithm (QT)	15% threshold algorithm (QT)	Human-like algorithm (warm colours%)
1	370	48	1	324	338	14%
2	361	55	1	319	332	1%
3	350	68	1	437	309	2%
4	343	72	1	417	306	5%
5	329	83	1	361	287	1%
6	335	90	1	332	290	20%
7	401	57	2	344	359	12%
8	389	75	2	306	324	19%
9	339	95	2	317	286	11%
10	419	47	2	354	367	28%
11	396	68	2	330	317	3%
12	355	82	2	364	327	2%
13	445	46	3	384	401	21%
14	441	67	3	378	390	49%
15	431	75	3	360	372	42%
16	417	80	3	343	367	32%
17	371	94	3	320	309	57%
18	444	58	3	390	403	29%
19	424	76	3	325	346	26%
20	363	95	3	315	310	39%
21	487	46	4	454	460	38%
22	468	72	4	417	409	69%
23	451	79	4	358	374	50%
24	445	81	4	367	396	54%
25	486	54	4	419	435	61%
26	485	64	4	405	422	48%
27	419	91	4	316	349	71%
28	410	94	4	312	326	74%
29	523	42	5	432	452	57%
30	494	71	5	427	403	69%
31	470	85	5	353	421	63%
32	518	54	5	413	431	48%
33	482	80	5	378	398	55%
34	417	96	5	310	344	51%
35	507	79	6	377	251	62%
36	565	49	6	478	528	65%
37	579	52	6	477	548	65%
38	547	64	6	451	448	68%
39	509	68	6	435	414	79%
40	518	77	6	388	405	87%

7.4 Human-machine perception: differences, benefits and opportunities

Perception, broadly speaking, involves two forms of processing: bottom-up processing, which is driven by incoming stimuli, and uses perceptual organisation to form a representation of an object; and top-down processing, which uses contextual information to aid the perception of patterns [EY97, CEY04, FAGB06]. ECG interpretation is thought to be dependent primarily on top-down processing [WBA+14, SW17]. From a human perspective, the visualisation technique described here works by harnessing bottom-up processing, drawing visual attention to the critical information contained within the ECG signal. Using a simple model of this process, the human-like algorithm was able not only to match human performance, but to exceed it. It is interesting to note that specificity was slightly higher for both the human participants and the algorithm. Whilst we cannot be sure that humans were internally employing a version of the algorithm to make decisions, this indicates that similar processes may be at work.

The shared representation of the data, and the shared model of how to interpret it, are important when considering the real world application of this approach. At present, clinicians do not regard automated ECG interpretation as reliable. In broader terms, the public also have concerns about the use of algorithms for medical decision support. A survey carried out for the Wellcome Trust highlighted transparency as an important factor in automation [FSB18]. A human can, in theory, explain an error, and it is therefore possible to establish negligence or malicious intent, an important consideration for the respondents. The majority of those surveyed also stated they would not like machines to suggest treatments or answer medical questions. This sits in contrast to the proposed digitisation of healthcare, which includes the potential use of AI and robotics in healthcare delivery [Top19b]. To achieve a shift in people's attitudes towards this technology, a number of challenges need to be addressed, including improved data protection and privacy standards, fairness (guarding against bias) and transparency in how technology works and decisions are made [VBC18]. Whilst we have not evaluated trust in the algorithm described here, we hypothesise that its inherent transparency will be of benefit in this regard. Its explainability goes beyond that offered by many rule-based algorithms, which are theoretically explainable, but may nevertheless be extremely complex and difficult for humans to understand. The human-like algorithm uses a representation of the data that matches that used by the person making the decision. We suggest that this operates not just at a conscious level,

using deliberate 'system 2' processes, but also at a sub-conscious, perceptual level, and is thus able to engage fast 'system 1' processes, such that people can understand the data quickly and relatively effortlessly [Kah11].

7.5 Future work

Biologically-inspired algorithms such as deep Convolutional Neural Networks (CNN), are now capable of image classification on a par with adult human capabilities [ZF19]. Whilst they have achieved impressive practical successes across a number of application domains including medicine, deep learning models must be trained on large datasets, and the way they represent and use data internally is often unclear [HBPK17]. They also exhibit weaknesses, as demonstrated by adversarial learning, a field of study that evaluates the safe use of machine learning techniques in adversarial settings such as spam filtering, cybersecurity, and biometric recognition, by attempting to fool the models through malicious input [LM05]. An example of this is provided by Goodfellow et al. (2015), who took an image of a panda that was that used in an image classification task and introduced a small perturbation to the image data [GSS15]. This changed the algorithm's classification from 57.7% confidence the image showed a panda, to 99.3% confidence that the image showed a gibbon [GSS15]. To a human observer, the image still clearly resembles that of a panda, but to a machine, the small change was catastrophic. Further exploration of human information processing may be the key to addressing these challenges.

Human visual perception is described as the construction of efficient representation formalisms of visual information [Can13]. The challenges of visual perception have drawn the curiosity of computer scientists for many years, particularly in terms of understanding representation formalisms that may be effective in machine perception and artificial intelligence [OPN08]. Human representation formalisms used in artificial intelligence include *relational* representations based on networks, graphs or frame schemes, *propositional* representations that use linguistics based on first order predicate logic, and *procedural* representations, also known as pattern-directed schemes [Can13]. A promising direction is using these representations to mimic the sophisticated and flexible perceptual capabilities of human perception to organise information in a preprocessing step. Merging deep learning models with symbolic *procedural* representations based on perceptual schemes has the potential to advance AI systems [TCX+17]. A good example of the potential of this approach is provided by Stettler et

al. (2018), who demonstrated that mimicking the process involved in perceiving visual obstruction can improve the performance of a CNN in a letter recognition task [SF18].

The human-like algorithm developed here also has the potential to aid CNN design. For example, using pseudo-colouring to improve information segmentation in a pre-processing step, may help to improve a CNN's accuracy in classifying ECGs with LQTS. Figures 7.9 and 7.10 show two ECGs, one with a normal and one with a prolonged QT-interval respectively. Below each ECG are the results of using an image pre-processing method, available online [Ana20], for detecting orange to red colours using RGB (red, green, blue) colour space and then masking the ECG image to show only the parts containing the warm colours. The information that is salient to a human can thus be prioritised as input to the CNN, reducing the search space.

The field of human perception has long been a great source of inspiration for developing and improving machine perception. A motivation for our approach was not only improving the accuracy of ECG interpretation, but also producing data representations that can be used to provide a transparent, understandable and explainable interpretation that keeps the human in the loop. Further exploration of the potential that human perceptual processes have for informing machine interpretation is a promising avenue for future research.

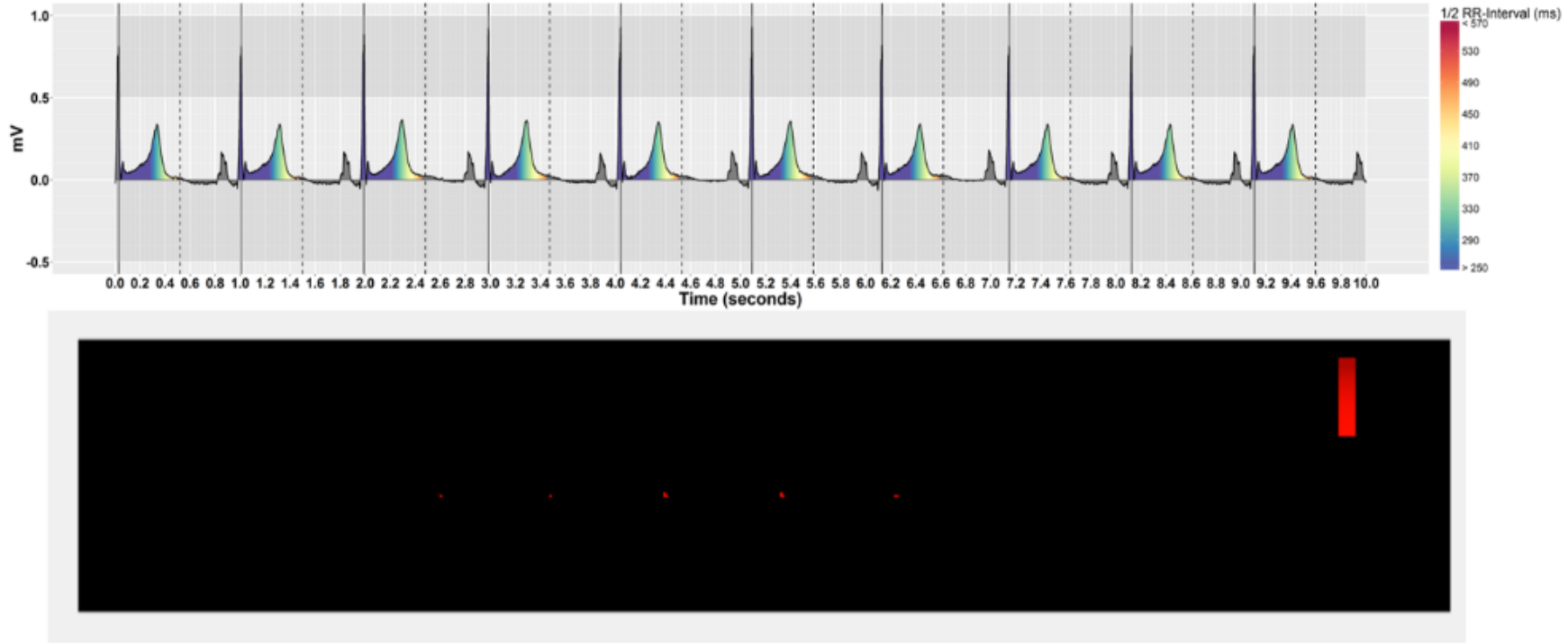


Figure 7.9: Examples of two ECGs and corresponding images pre-processed to display orange to red pixels, showing an ECG with a normal QT-interval.

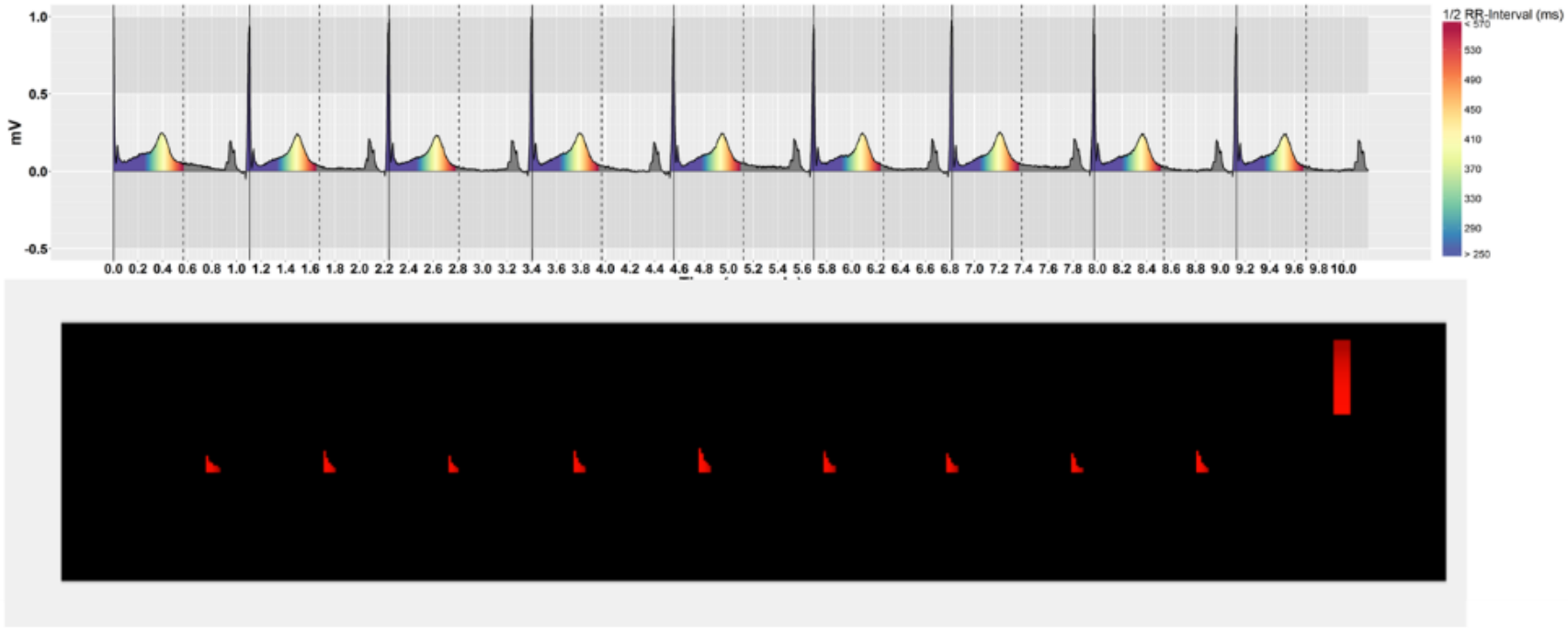


Figure 7.10: Examples of two ECGs and corresponding images pre-processed to display orange to red pixels, showing an ECG with a prolonged QT-interval.

Chapter 8

An Explainable Algorithm for Detecting Drug-Induced QT-Prolongation at Risk of Torsades de Pointes (TdP) Regardless of Heart Rate and T-Wave Morphology

8.0 Chapter overview

8.0.1 Thesis context

Drug-induced QT-prolongation is known to invalidate automated QT-interval measurement algorithms due to the substantial changes it causes to T-wave morphology and heart rate, thus causing underestimation of the risk of TdP [TAS$^+$15, CMY08] (discussed in Chapter 1 as the fifth research challenge). This chapter therefore introduces a study that extends the development of the 'human-like' algorithm reported in Chapter 7 and evaluates its effectiveness in determining drug-induced QT-prolongation at risk of TdP regardless of heart rate and T-wave morphology across a wide range of ECG cases. It empirically compares two approaches to building the enhanced algorithm: a manually-curated 'expert' algorithm that incorporates knowledge from the clinical literature; and a decision tree, which automates the generation of the rules from the data. The study also reports the results of two focus groups—one consisting of patients, the other of clinicians—which explored the relevance of our approach to clinical practice.

The results show that combining the 'human-like' perceptual heuristics provided by the pseudo-colour that serve as a proxy QT-interval length with additional clinical knowledge about T-wave morphology significantly improves machine detection of drug-induced QT-prolongation at risk of TdP regardless of heart rate and T-wave morphology (sensitivity = 0.94 and specificity = 0.99, when tested on 5050 ECGs). The study also shows that whilst the decision tree is highly accurate when tested using cross-validation, it performs less well on unseen data than the 'expert' algorithm, due to the generation of spurious rules that lack clinical validity. We conclude that the 'expert' algorithm significantly improves the detection of drug-induced QT-prolongation at risk of TdP, and hypothesise that a human-like approach to the development of diagnostic algorithms—incorporating both human pattern recognition ability and clinical expertise—may have much wider utility. These findings underpin several directions for future work, which are discussed in detail in Chapter 9.

The content of this chapter is adapted from: Alaa Alahmadi, Alan Davies, Jennifer Royle, Leanna Goodwin, Katharine Cresswell, Zahra Arain, Markel Vigo, and Caroline Jay. An explainable algorithm for detecting drug-induced QT-prolongation at risk of torsades de pointes (TdP) regardless of heart rate and T-wave morphology. *Computers in Biology and Medicine*, 2021.

8.0.2 Author's contributions

Alaa Alahmadi and Caroline Jay conceptualised and devised the idea for the work. Alaa Alahmadi developed and evaluated the algorithm, carried out the ECG data acquisition, designed and analysed the study and wrote the paper, with Alan Davies, Markel Vigo and Caroline Jay contributing significant edits. Caroline Jay assisted with study design and drew out the theoretical contribution. Alan Davies acted as the electrocardiogram domain expert throughout. Jennifer Royle, Leanna Goodwin, Katharine Cresswell and Zahra Arain coordinated and organised the focus group events.

8.0.3 Published abstract

Torsade de points (TdP), a life-threatening arrhythmia that can increase the risk of sudden cardiac death, is associated with drug-induced QT-interval prolongation on the electrocardiogram (ECG). While many modern ECG machines provide automated measurements of the QT-interval, these automated QT values are usually correct only for a noise-free normal sinus rhythm, in which the T-wave morphology is well defined. As QT-prolonging drugs often affect the morphology of the T-wave, automated QT measurements taken under these circumstances are easily invalidated. An additional challenge is that the QT-value at risk of TdP varies with heart rate, with the slower the heart rate, the greater the risk of TdP.

This paper presents an explainable algorithm that uses an understanding of human visual perception and expert ECG interpretation to automate the detection of QT-prolongation at risk of TdP regardless of heart rate and T-wave morphology. It was tested on a large number of ECGs ($n = 5050$) with variable QT-intervals at varying heart rates, acquired from a clinical trial that assessed the effect of four known QT-prolonging drugs versus placebo on healthy subjects. The algorithm yielded a balanced accuracy of 0.97, sensitivity of 0.94, specificity of 0.99, F1-score of 0.88, ROC (AUC) of 0.98, precision-recall (AUC) of 0.88, and Matthews correlation coefficient (MCC) of 0.88.

The results indicate that a prolonged ventricular repolarisation area can be a significant risk predictor of TdP, and detection of this is potentially easier and more reliable to automate than measuring the QT-interval distance directly. The proposed algorithm can be visualised using pseudo-colour on the ECG trace, thus intuitively 'explaining' how its decision was made, which results of a focus group show may help people to self-monitor QT-prolongation, as well as ensuring clinicians can validate its results.

Graphical abstract

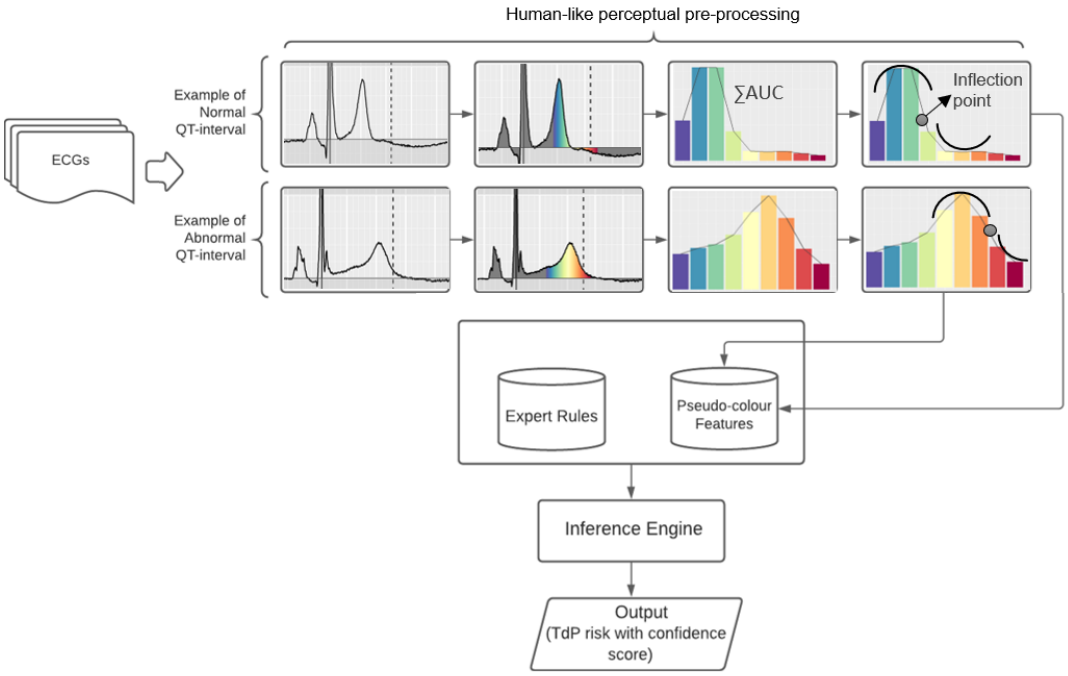


Figure 8.1: Graphical abstract of the ‘expert’ algorithm.

Highlights

- We develop an explainable algorithm to detect drug-induced QT-prolongation at risk of torsades de pointes (TdP) regardless of heart rate and robust to T-wave morphology changes using a 'human-like' approach, where human perception of the ECG signal is used to determine features, and diagnostic rules are determined according to the clinical literature.
- Testing the algorithm on a large number of ECGs ($n = 5050$) with variable QT-intervals at varying heart rates yields a Matthews correlation coefficient (MCC), balanced accuracy, sensitivity and specificity of 0.88, 0.97, 0.94, 0.99 respectively.
- We explore whether we can improve the algorithm by automating the rule-generation with a decision tree. We find that we can, but the rule-based decision tree algorithm overfits and generalises less well than the 'expert' evidence-based rule algorithm, and that whilst the decision tree rule algorithm is technically explainable, it is not clinically explainable.
- A focus group evaluation confirms the explainability of the algorithm is important, supporting an 'expert' rule-based approach, rather than a fully automated approach.
- Defining features according to human interpretation heuristics provides a promising route to automating ECG interpretation, and has the benefit of being explainable to patients and clinicians. Automating rule generation based on these features has potential, but here it is not sufficient to produce a reliable, trustable algorithm. A human-in-the-loop approach, where machine learning is a tool used to surface potential rules, but these are validated empirically before being implemented, may be the best approach to use in clinical practice.

8.1 Introduction

This paper provides a new perspective on the challenging topic of electrocardiogram (ECG) interpretation. ECGs, which represent the electrical activity of the human heart, are powerful and widely used diagnostic instruments. The data they contain is rich and extremely complex, however, and learning how to accurately interpret them can take many years. Automated systems can identify a normal sinus rhythm with reasonable accuracy, but are much poorer at reliably detecting abnormalities [SW17, EI13, RSG09]. A clinically significant cardiac abnormality that automated methods have been shown to be particularly unreliable at detecting is QT-interval prolongation [TAS$^+$15, KBD$^+$18, EI13, GL13, TABW11, RSG09]. This can be congenital, or acquired, resulting from the clinical administration of certain pharmacological drugs, and is associated with a life-threatening arrhythmia known as Torsade de Pointes (TdP) [YC03, CMY08].

In this work, we combine knowledge of human perception with clinical expertise to develop an automated algorithm that can reliably detect patients with QT-prolongation at risk of Torsade de Points (TdP). We take as our starting point a visualisation technique that displays the ECG signal such that a lay person can detect QT-interval prolongation quickly and accurately (Section 8.2.2). The information encoded in the visualisation is mapped to a set of features, which form the basis of two interpretation algorithms: an 'expert' set of rules, which are formed according to clinical practice (Section 8.3.2); and a decision tree, which automates the generation of rules from the same set of features (Section 8.4.3). Whilst the decision tree appears more accurate when trained and tested with cross-validation, the expert algorithm is more accurate when tested on unseen data.

8.2 Background and significance

8.2.1 Identifying patients at risk of TdP

Torsades de pointes (TdP), or 'twisting of the points', was a term first used in 1966 by Francois Dessertenne to describe a form of polymorphic ventricular tachycardia, in which the continuously changing polarity and amplitude of the QRS complexes appear to twist around the isoelectric line of the electrocardiogram (ECG) [Des66]. It is a potentially lethal arrhythmia that can degenerate into ventricular fibrillation, the

leading cause of sudden cardiac death in young individuals with structurally normal hearts [YC03, MK05, ESTB20, WS20]. It is often precipitated by triggers such as emotional stress or exercise, especially swimming [BSW14, VTVAV15]. Since the time of its original description by Dessertenne, it has been well-established that TdP is frequently associated with QT-interval prolongation on the ECG, and is caused by a cardiac ion channelopathy known as 'long QT syndrome' (LQTS) [Des66, YC03, ESTB20, WS20].

Whilst TdP can result from both congenital and acquired long QT syndrome, acquired LQTS caused by pharmacological drugs is by far the most common cause of TdP [YC03, CMY08, ESTB20, WS20]. A steadily increasing number of medications have been reported to cause drug-induced QT-prolongation, TdP and sudden cardiac death [Kha02, CMY08, ESTB20]. This, in turn, has troubled clinicians, the pharmaceutical industry and regulatory authorities, particularly because many of these QT-prolonging drugs, including antihistamines, psychotropics, antibiotics and antiarrhythmic agents, are widely prescribed, often for self-limited diseases [YC03, WS20].

The QT-interval represents the duration of the ventricular depolarisation and repolarisation cycle; it is measured on the ECG from the beginning of the QRS complex (reflecting ventricular depolarisation) to the end of the T-wave (representing subsequent repolarisation) [GMZ06]. In LQTS, a delay occurs in ventricular repolarisation, which increases the risk of premature ventricular contraction (PVC) occurring during the relative refractory period of repolarisation, reflected on the ECG by a unique pattern known as the R-on-T phenomenon that initiates TdP [YC03, AAKR$^+$02] (see Figure 8.2). Dividing the ventricular repolarisation period into early and late repolarisation, shown on the ECG as the J–T^{peak} and T^{peak}-T^{end} intervals respectively (Figure 8.2), is recommended for assessing the risk of drug-induced TdP; many TdP episodes are reported following a prolonged T^{peak}-T^{end} interval in particular [CZSZ08, JVM$^+$14, KGS$^+$12, BSR$^+$16].

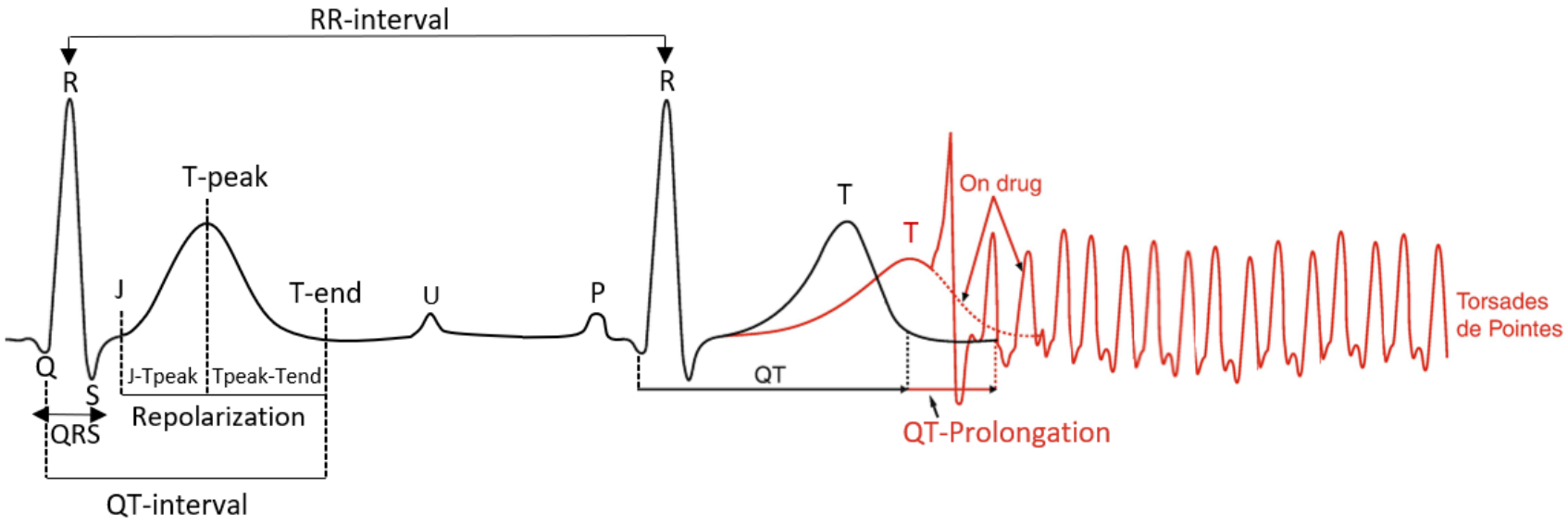


Figure 8.2: An illustration of drug-induced QT-interval prolongation and the R-on-T phenomenon that initiates torsades de pointes (TdP) arrhythmia on the electrocardiogram (ECG).

Measuring the QT-interval, however, is known to be a challenging task [MS99, Mor01, GPAW14, CMY08], making drug-induced QT-prolongation detection difficult, even for clinicians who routinely read ECGs [VRS+05, STN08]. Whilst it is relatively easy to determine the beginning of the QRS complex (or RS complex if there is no Q-wave), identifying the end of the T-wave is recognized as being the most difficult aspect of measuring the QT-interval [CMY08, GPAW14, GMZ06, Mor01]. Numerous methods have been proposed [MCM96, MCM95b, KGK+96, GMZ06, PDJVdBW08, PKN+09, SKP+11, IP13, GPAW14], among which the tangent method is the most popular [PDJVdBW08], but all techniques have been shown to be associated with potential inaccuracies [CMY08, Mor01]. A major challenge lies in the fact that the morphology (shape) of the T-wave itself can be very variable, and QT-prolonging drugs may cause abnormal changes in it [KGK+96, Mor01, CMY08, VJM+15]. Variability in T-wave amplitude [MCM95a], a prolonged, flat T-wave, and fused T-U waves (also known as T-U complexes) can easily invalidate QT-interval measurement [CMY08, IRHH98, KNK+10, KGK+96]. This issue complicates both manual and automated QT-interval measurement, and the best method to use remains a subject of debate [CMY08, KOO+11, SW17]. At present, despite numerous efforts [KOO+11, PKN+09, HVB+17, Hun05, MCM95b, MCM96], no automated QT-interval measurement method is considered reliable enough to be used in the clinical assessment of the cardiac safety of a QT-prolonging drug [CMY08, SW17, TAS+15, KBD+18, EI13, GL13, MCbA01, TABW11, RSG09].

Automated methods in the literature fall into two categories according to how the end of the T-wave is defined: threshold-based methods and slope-based methods [PKN+09, MCM96, MCM95b, CMY08]. Threshold-based methods identify the end of the T-wave based on threshold levels, defined as a fraction ranging from 5% to 15% of the amplitude or differential of the T-wave [Hun05, CMY08]. In slope-based methods, the end of the T-wave is defined as the interception of the maximum T-wave slope with the isoelectric line [Hun05, CMY08]. As the steepest part of the descending portion of the T-wave is affected by T-wave morphology, different methods have been proposed to identify the maximum T-wave slope, including the tangent method, peak slope method, and least-square fitting method [MCM96, MCM95b, Hun05, CMY08, WLX+16]. Generally, slope-based methods provide a better estimation of the QT-interval than threshold-based methods, as the results of the latter vary according to the threshold level and the T-wave morphology [XR98, CMY08, HVB+17]. However, the accuracy of slope-based methods can be significantly reduced when the T-wave has a

slow-moving deflection that fails to return to the ECG baseline, making the steepest slope hard to define [CMY08]. In addition, these methods depend on an accurate identification of the isoelectric line, which is known to be influenced by noise and ECG baseline wander (a type of artifact) [SW17, CMY08]. Bizarre automated QT measurements are also sometimes reported, even for normal, noise-free ECGs [CMY08].

Assessing the risk of drug-induced TdP also benefits from consideration of the ventricular repolarisation morphology more generally [TRR$^+$07, HES92]. Augmented U-waves, or partial/complete T-U wave fusion, where the T and U waves cannot be distinguished, are significant risk predictors of TdP [KNK$^+$10, KFBW09]. In most cases, this is shown to be, in fact, a prolonged biphasic, or notched T-wave [YA98, VJM$^+$15]. For certain T-U morphologies commonly reported in drug-induced long QT syndrome, measuring the QT-interval as a simple time period may lack diagnostic validity, and the presence of prolonged fused T-U waves should be viewed as a more accurate predictor of TdP [CMY08, HES92].

8.2.2 A human-like approach to automated ECG interpretation

We have shown in previous work that superimposing pseudo-colouring, a technique that represents continuously varying values using a sequence of colours [War12], on the ECG significantly improves people's ability to detect QT-prolongation at risk of TdP, regardless of heart rate [ADVJ20a]. Our technique draws from the field of pre-attentive processing theory in human vision, which outlines a set of visual properties including colour that can be detected rapidly and accurately by the human eye [War21, HBE96, HBE95]. This form of data presentation helps to overcome the problem of identifying the end of the T-wave, directing the observer instead to inspect repolarisation morphology. The technique was designed to support 'intuitive' visual perception of drug-induced LQTS, both for patients on a QT-prolonging drug, and clinicians with less training in ECG interpretation. The technique colours the area under the curve of the ECG signal within the ventricular repolarisation period relative to the R-peak, with a gradient of cool colours (purple to blue to green) indicating normal QT-interval ranges, and warm colours (yellow to orange to red) showing abnormal QT-interval ranges. As the ventricular repolarisation period is represented by the T-wave on the ECG [CMY08], the pseudo-colour highlights its position in relation to the inter-heartbeat time dimension, without needing to identify either the peak or end of

the T-wave. The pseudo-colouring was adjusted for heart rate using the 'QT nomogram', which is a risk assessment method designed specifically for identifying patients at risk of drug-induced TdP according to heart rate [FWM+05, CIKD07]. Figure 8.3 shows examples of ECGs with pseudo-colouring that have different heart rates, but similar levels of TdP risk according to QT-interval; the dashed lines represent the QT nomogram line value at risk of TdP for a given heart rate. More details about the pseudo-colouring technique and its evaluation can be found in [ADR+19, ADVJ20a].

Here, we exploit an understanding of how humans use pseudo-colouring to determine QT-interval length to inform a new approach to the automated detection of QT-prolongation. We term this approach *human-like*, as the knowledge representation and reasoning processes used within the algorithm are inspired by the way in which humans interpret ECGs.

Long QT syndrome refers to a prolongation of the ventricular repolarisation period, represented by the T-wave on the ECG [CMY08]. Although the interval of interest starts at the Q-wave, it is the prolongation of the T-wave in particular that is the ECG marker for LQTS, rather than the QRS complex that represents the ventricular depolarisation period [CMY08]. As the T-wave generally has the largest area under the curve (AUC) of the ECG signal, the pseudo-colouring provides a way to approximate QT prolongation, alleviating the need to measure the QT-interval distance precisely.

Although using AUC to assess the risk of drug-induced LQTS has yet to appear in the automated ECG interpretation literature, it has been theoretically suggested by Bonate *et al.* (1999), who proposed a univariate summary measure that calculated the total area under the QT interval for both the ECG baseline and the post-dose ECG [BR99]. However, this method still required identification of the beginning of the Q-wave and the end of the T-wave to locate the QT-interval for comparison, and lacked clarity on whether both positive and negative values from the ECG isoelectric line be used in the calculation, or only positive values.

Our proposed algorithm avoids the necessity of locating the start and end of the QT-interval by considering the AUC in reference to the easily detectable R-peak, as described in Section 8.3.2. In a pilot study, we took the ECG signal with superimposed pseudo-colour as a starting point, and calculated the AUC for each colour in the sequence using the trapezoidal rule [ADD+21]. The QT-interval was considered 'prolonged' if the warm pseudo-colours (yellow to orange to red) occupied more than 50% of the AUC of the ECG signal, *i.e.* when the T-wave contained more warm than cool colours; otherwise it was considered 'normal'. When tested on a clinical dataset of 40

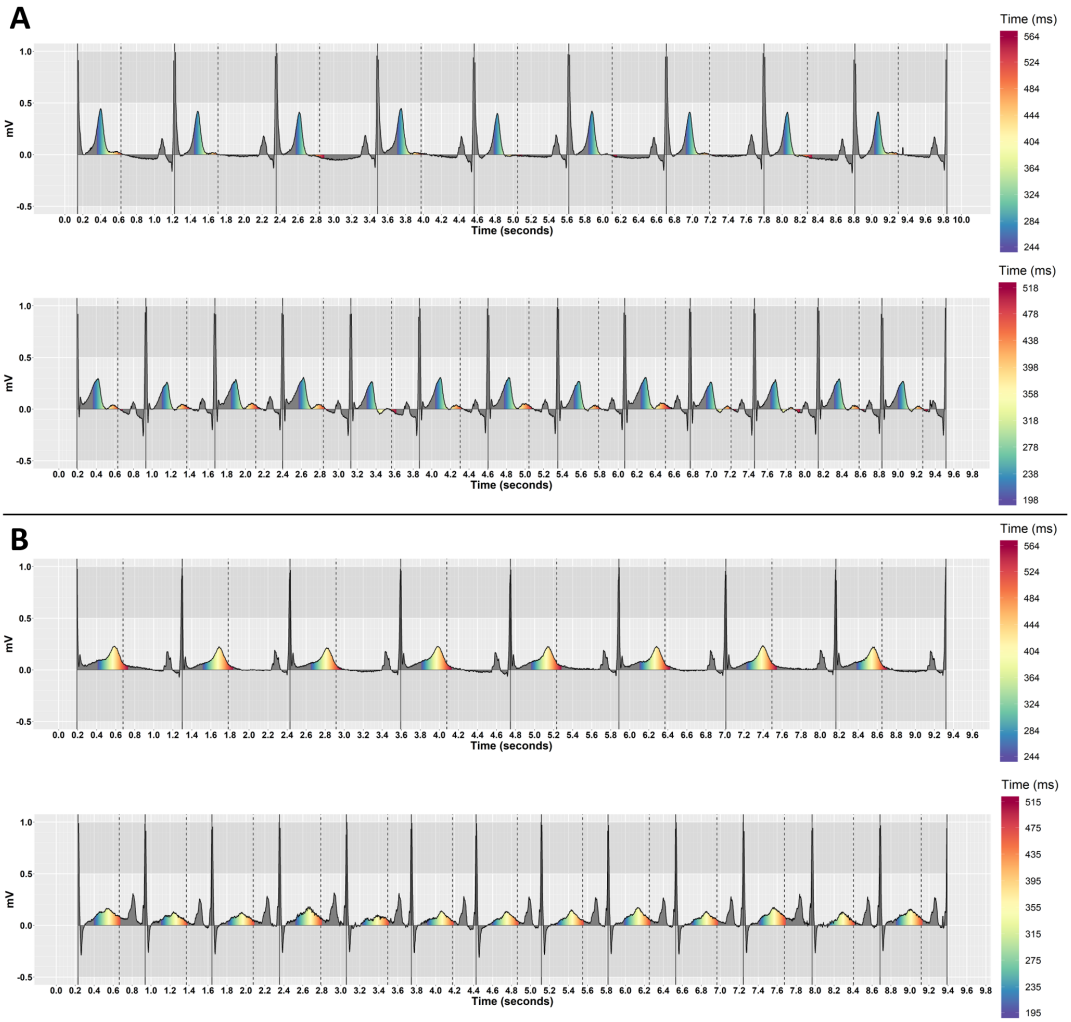


Figure 8.3: Pseudo-coloured ECGs that have different heart rates, but similar QT-interval TdP risk levels. The dashed lines represent the QT nomogram line value at risk of TdP. (A) QT-intervals are below the nomogram line by 120ms for both ECGs (no risk of TdP). The top stimulus has a low heart rate (HR = 55, QT = 361) and the bottom has a high heart rate (HR = 83, QT = 329). (B) QT-intervals are above the nomogram line in both ECGs (risk of TdP). The top stimulus has a low heart rate (HR = 52, QT = 579) and the bottom has a high heart rate (HR = 85, QT = 470).

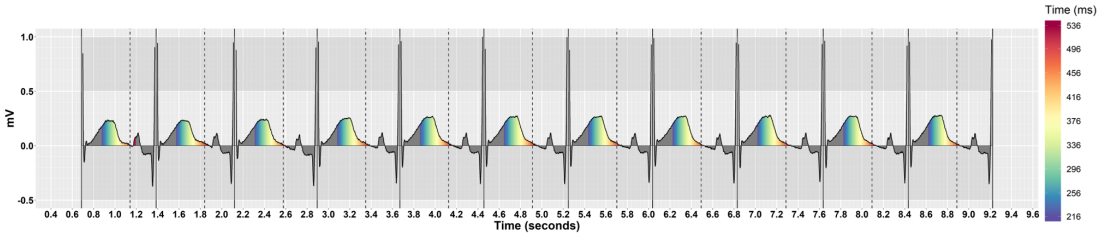


Figure 8.4: A pseudo-coloured ECG that has wide T-waves with prolonged J–T^{peak} and T^{peak}-T^{end} intervals but only 28.67% warm colours. The dashed lines represent the QT nomogram line value at risk of TdP.

ECGs, the algorithm was more accurate in detecting QT-prolongation at risk of TdP than current signal processing techniques, including threshold-based and slope-based QT-interval measurement methods. However, this pilot version of the algorithm has two significant limitations. Firstly, as it relies on the ratio of warm to cool pseudo-colour under the T-wave, it may inaccurately classify the QT-interval as normal where drugs prolong both early and late ventricular repolarisation, shown on the ECG as prolonged J–T^{peak} and T^{peak}-T^{end} intervals, or when a patient has an electrolyte abnormality such as hypercalcemia alongside QT-prolongation. In these cases, the width of the T-wave may increase, and thus contain proportionally more cool colours than warm. Figure 8.4 shows an example of this issue, where the pseudo-coloured ECG has wide T-waves with clinically prolonged J–T^{peak} and T^{peak}-T^{end} intervals, and the percentage of warm colours area is therefore only 28.67%—considerably below the 50% threshold required for a 'prolonged' result from the algorithm. Secondly, calculating the AUC alone, without inspecting the shape of the curve, makes it difficult to distinguish between T and U waves, particularly where there are significant T-wave morphology changes.

In this paper, we combine the information about QT-interval length provided by the pseudo-colour with additional knowledge about T-wave morphology to generate a set of visual features that form the basis of an automated algorithm that can determine risk of TdP across a wide range of cases. We empirically compare two approaches to building the algorithm: a manually-curated 'expert' algorithm that incorporates knowledge from the clinical literature; and a decision tree, which automates the generation of the rules from the data. In addition, we report the results of two focus groups—one consisting of patients, the other of clinicians—which explored the relevance of our approach to clinical practice.

8.3 Materials and methods

In this section, we provide a detailed description of how the algorithm was developed in 8.3.2 and explain the evaluation process in 8.3.3. The source code for the algorithm can be found in [ADVJ21].

8.3.1 ECG data acquisition

The ECG datasets ($n = 5050$) were acquired from a clinical trial that assessed the effect of four known QT-prolonging drugs versus placebo on healthy subjects [JVM$^+$14]. The 10-second lead-II recording was selected from each 12-lead ECG, as this is typically used to measure the QT-interval [CMY08]. The heart rates of the ECGs ranged from 40 to 96 beats per minute (bpm), and the QT-interval values ranged from 300 to 579 ms. The ECGs were from 22 subjects who received a single dose of a pure hERG potassium channel blocker ('Dofetilide'), and three drugs that block hERG and either calcium or late sodium currents ('Quinidine', 'Ranolazine', and 'Verapamil'), during a placebo-controlled cross-over trial.

As part of the clinical study methodology, QT-intervals and heart rates were calculated for all ECGs, and it is these QT/HR values that were used as ground truth for our subsequent evaluation of the algorithm. According to the QT/HR pair plots of all ECGs on the nomogram [CIKD07], 180 ECGs were on or above the nomogram line, showing risk of TdP, while the other ECGs ($n = 4870$) were below the nomogram line, as shown in Figure 8.5. The open ECG datasets are available online from the PhysioNet database [GAG$^+$00].

8.3.2 Algorithm development

Pseudo-colouring application

The R-peaks in the ECG signal were detected using an automated math function that finds the greatest peaks (maxima), according to regular pattern in the signal [Bor15]. The average RR-interval and heart rate were calculated. The TdP risk threshold was calculated for each heart rate using the nomogram line [CIKD07].

In clinical practice, the QT-interval is measured by counting the small squares (each representing 40 ms) on the standard ECG background grid from the beginning of the Q-wave to the end of the T-wave [CMY08]. An approximate time for the ventricular

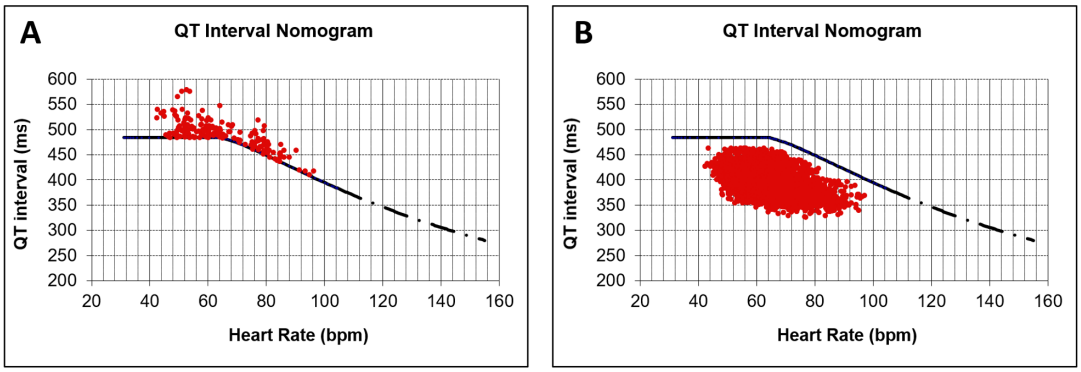


Figure 8.5: Assessment of the TdP risk for all ECGs ($n = 5050$) using the QT-nomogram. According to the acquired QT/HR pair values from the clinical trial study: (A) A total of 180 ECGs were on or above the nomogram line, showing risk of TdP. (B) The other ECGs ($n = 4870$) were below the nomogram line, showing no risk of TdP.

depolarisation and repolarisation cycle was calculated for each heartbeat from the R-peak minus 20 ms (which served as a proxy for the beginning of the Q-wave) to the maximum potential QT-prolongation at risk of TdP, which was estimated as the QT-nomogram line value at risk of TdP plus two small squares (80 ms). This formed an additional inter-heartbeat time dimension, to which the pseudo-colour could be mapped.

As the time period of interest is the duration of ventricular repolarisation, the pseudo-colouring sequence was applied to the area between the isoelectric line (where amplitude is zero) and the ECG signal, starting at the time of the QT-nomogram line value at risk of TdP plus two small squares (80 ms) to six small squares (240 ms) below the nomogram line. The time of the QT-nomogram line value at risk of TdP was mapped to dark orange, and values 40 ms and 80 ms above the nomogram line were mapped to red and dark red respectively, showing the higher risk of TdP. Time values below the nomogram line were mapped to progressively cooler colours, showing no risk of TdP. This resulted in nine indices on the pseudo-colouring scale, where each index was mapped to a colour code and represented a small square on the ECG. Figure 8.6 illustrates how the pseudo-colouring technique was applied according to the QT nomogram line and the standard ECG background grid.

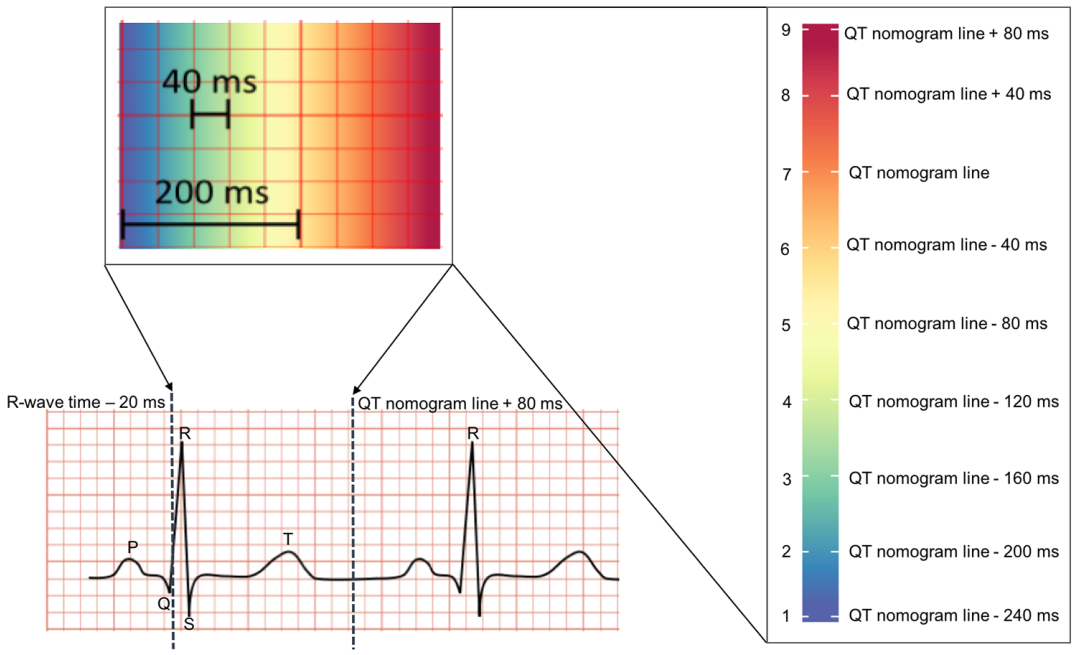


Figure 8.6: An illustration of how the pseudo-colouring technique was applied according to the QT nomogram line and the standard ECG background grid. A small square on the grid is equal to 40 ms.

Calculating the pseudo-coloured area under the curve (AUC) of the ECG signal

The pseudo-coloured area under the curve (AUC) of the ECG signal was calculated using the trapezoidal rule. In mathematics, and more specifically in integral calculus, the trapezoidal rule is a common method for approximating the area under the curve of a linear function, which works by dividing it into a number of intervals of equal width [Y+02]. That is, given a linear function $f(x)$ of a real variable x and an interval [a, b], the rule estimates the area under the graph of the function $f(x)$ as a trapezoid, calculating its area as follows:

$$\int_a^b f(x)dx \approx \frac{(f(a)+f(b))}{2}(b-a) \tag{8.1}$$

The time dimension of the raw ECG signal is represented by integer numbers ranging from 1 to 10000 milliseconds for a 10-second ECG recording. To maximise the precision of the calculation, we considered [a, b] to be the interval of two successive timestamps in the ECG signal, where the difference between them is equal to one millisecond. If we consider the ECG time interval of x_1 and x_2, the trapezoidal rule was applied by taking the average amplitude of the ECG signal on the Y-axis of this interval as $f(x_1)$ and $f(x_2)$, and multiplying it by the difference in time between x_1 and x_2, which is always equal to one. The trapezoidal rule was thus applied to each one millisecond subinterval, and the results were then summed for every 40 ms colour index on the pseudo-colouring scale to produce a total AUC value for each small square on the standard ECG background grid. The negative area below the ECG baseline was treated in the same way as the positive area, as the baseline is not usually fixed, and in some abnormal cases the T-wave can be negative in Lead-II, a condition known as 'T wave inversion'. As such, the absolute AUC was calculated.

As a spectrum-approximation pseudo-colouring sequence was used, the indices from 1 to 4 (purple, blue, green, lime) were considered to be cool spectral colours, while the indices from 5 to 9 (yellow, orange, dark orange, red, dark red) were considered warm spectral colours. By plotting the calculated AUC as a function of pseudo-colouring sequence, we obtain a global representation of the pseudo-coloured repolarisation morphology across the whole 10-second ECG signal. Figures 8.7 and 8.8 show examples of pseudo-coloured ECGs with normal (no risk of TdP) and prolonged (at

risk of TdP) QT-intervals, and their global representation plot, showing an approximation of the total pseudo-coloured AUC.

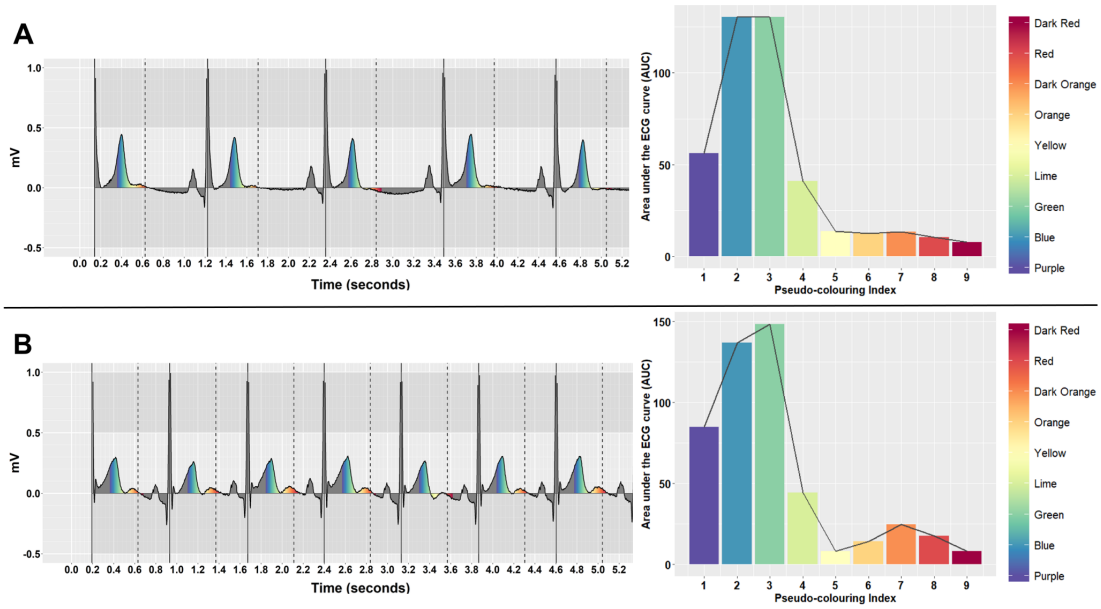
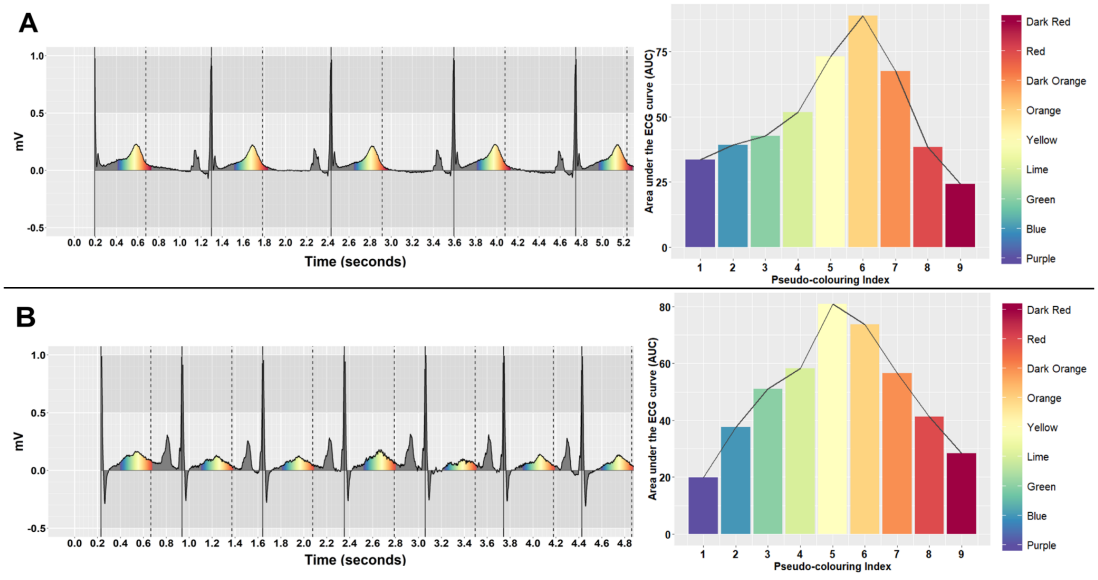


Figure 8.7: Pseudo-coloured ECGs with normal QT-intervals (no risk of TdP), and their global representation graph, showing the calculated $\sum AUC$ for each pseudo-colour across the whole ECG signal. (A) The ECG has a low heart rate (HR = 55, QT = 361), and displays a greater area of cool pseudo-colours. It also has a negative small red area below the EGC baseline, which was included in the AUC calculation. (B) The ECG has a high heart rate (HR = 83, QT = 329), with a cool pseudo-coloured T-wave and warm pseudo-coloured U-wave.

Locating intervals of concavity and inflection points

As T-wave morphology can change substantially and other ECG waves may present during ventricular repolarisation, including an augmented U-wave, a simple comparison of the relative amounts of cool to warm pseudo-colour in the AUC may not be sufficient for recognising a prolonged QT-interval. We therefore also consider the T-wave morphology by inspecting the shape of the ECG signal curve.

In differential calculus, the second derivative $f''(x)$ of a linear function $f(x)$ is the derivative of the derivative of $f(x)$, *i.e.* the derivative of the first derivative $f'(x)$. While the first derivative measures the rate of change of a quantity, *e.g.* the slope or gradient of a line that is tangential to the $f(x)$ curve at a particular point, the second



H

Figure 8.8: Pseudo-coloured ECGs with prolonged QT-intervals (at risk of TdP), and their global representation graph, showing the calculated $\sum AUC$ for each pseudo-colour across the whole ECG signal. (A) The ECG has a low heart rate (HR = 52, QT = 579), and displays a greater area of warm pseudo-colours. (B) The ECG has a high heart rate (HR = 85, QT = 470), with a notched T-wave that contains a greater area of warm pseudo-colours.

derivative measures how the rate of change of a quantity is itself changing. The tangent method commonly used to identify the end of the T-wave is equal to the first derivative of the ECG signal function at the T-peak point; the second derivative will measure the concavity of the T-wave at the same T-peak point, showing the rate at which the T-wave signal is increasing or decreasing over time. This provides a new approach to estimating the end of the T-wave, which builds on the widely utilised tangent method [PDJVdBW08].

The second derivative measures the concavity of the graph of $f(x)$, and identifies inflection points at which the shape of the function changes from concave down to concave up (or vice versa). If the second derivative of a function is positive, then the shape of this function is concave up (convex), and if it is negative, then its shape is concave down (concave). We thus determined the second derivative of the graph representing the global pseudo-coloured AUC, to estimate the location of the maximum concavity downward within the ECG signal and its subsequent inflection point, at which the shape of the signal changes to convex. This helps with identifying the location of the maximum rate of change of the maximum T-wave down slope on the pseudo-colouring scale, and the subsequent pseudo-colour index at which the T-wave signal starts to rise again, indicating its end.

When using Leibniz's notation for derivatives, the first and second derivatives of the graph representing the global pseudo-coloured AUC, considering the AUC as a dependent variable y with respect to an independent variable x representing the pseudo-colouring index, are written as follows:

$$f'(x) = \frac{dy}{dx} \; and \; f''(x) = \frac{d^2y}{dx^2} \tag{8.2}$$

The difference between any two successive indices dx on the pseudo-colouring scale is always equal to one. That is, the first derivative is measured by calculating the difference in the AUC between each two successive indices on the pseudo-colouring scale. This produced 8 derivatives. The second derivative was calculated by taking the derivative of these 8 derivatives, which produced further 7 derivatives. Then, the maximum concave down and its subsequent inflection point (if any) were identified with respect to the pseudo-colouring index. Figure 8.9 shows an illustration of this process. As the QT nomogram line represents the QT value at risk of TdP, only the first five second derivatives were used as predictors of at risk QT-prolongation, shown by the grey blocks in Figure 8.9. In addition, the inflection point index, defined as the

index between any two pseudo-colour indices at which the maximum concave down changes to convex, was also used as a predictor. As such, six predictors were calculated for each ECG.

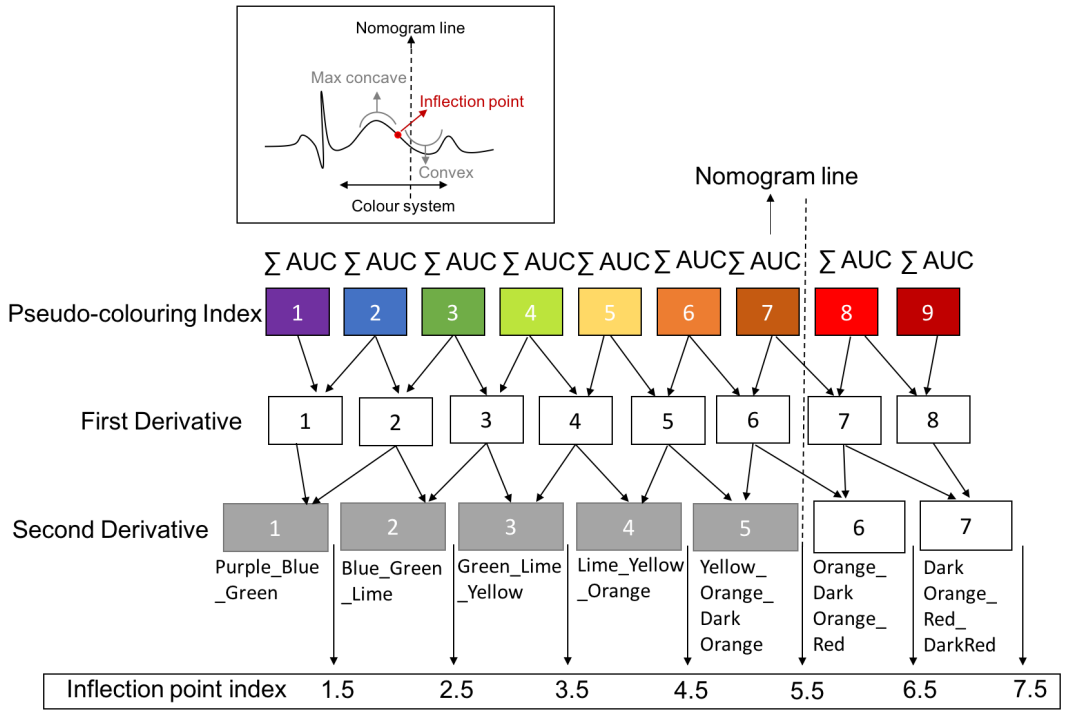


Figure 8.9: An illustration of how the second derivative of the pseudo-coloured AUC was measured and how the inflection point index was determined.

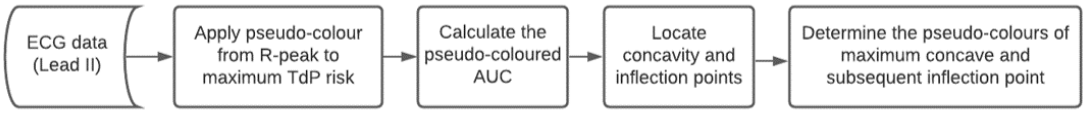


Figure 8.10: The expert algorithm's inference process.

Generating the expert rules

A rule-based algorithm was developed to classify the QT-interval as 'normal' (no TdP risk) or 'abnormal' (at risk of TdP), and rate the confidence of the classification on a 6-point scale of 'very likely normal' (1), 'probably normal' (2), 'possibly normal' (3), 'possibly abnormal' (4), 'probably abnormal' (5), and 'very likely abnormal' (6).

As the inflection point is a potential proxy for the end of the T-wave, we hypothesised that the probability of QT-prolongation at risk of TdP increases as the inflection point index increases. If the inflection point index is located within the cool pseudo-colours, *i.e.*, an index of 3.5 or less, this indicates a normal QT-interval (no TdP risk). Conversely, if the inflection point index is located within the warm pseudo-colours, *i.e.*, an index 5.5 or more, this indicates an abnormal QT-interval at risk of TdP.

An inflection point index of 4.5 is considered borderline, as the maximum concave down, representing the T-wave, is located between the cool and warm pseudo-colours of green, lime, yellow and orange. As the QT-interval may be very close to or on the nomogram line at this point, it is difficult to determine the risk of TdP precisely using the inflection point alone. In this situation, we thus also used the pseudo colouring index of the maximum concave down. If most of the maximum concave down colours are cool pseudo-colours, *i.e.* 50% or more of the concave is located within green, lime and yellow, then the QT-interval is not considered at risk of TdP; if 50% or more of the concave is located within lime, yellow, and orange, it is considered at risk. The expert algorithm's inference process is shown in Figure 8.10, and Figure 8.11 shows a flowchart of the expert rules. The confidence rating is also determined based on the inflection point index and the maximum concave location on the pseudo-colouring index. As the location of the maximum concavity on the pseudo-colouring scale increases, the likelihood of QT-prolongation at risk of TdP also increases, as illustrated in Figure 8.11.

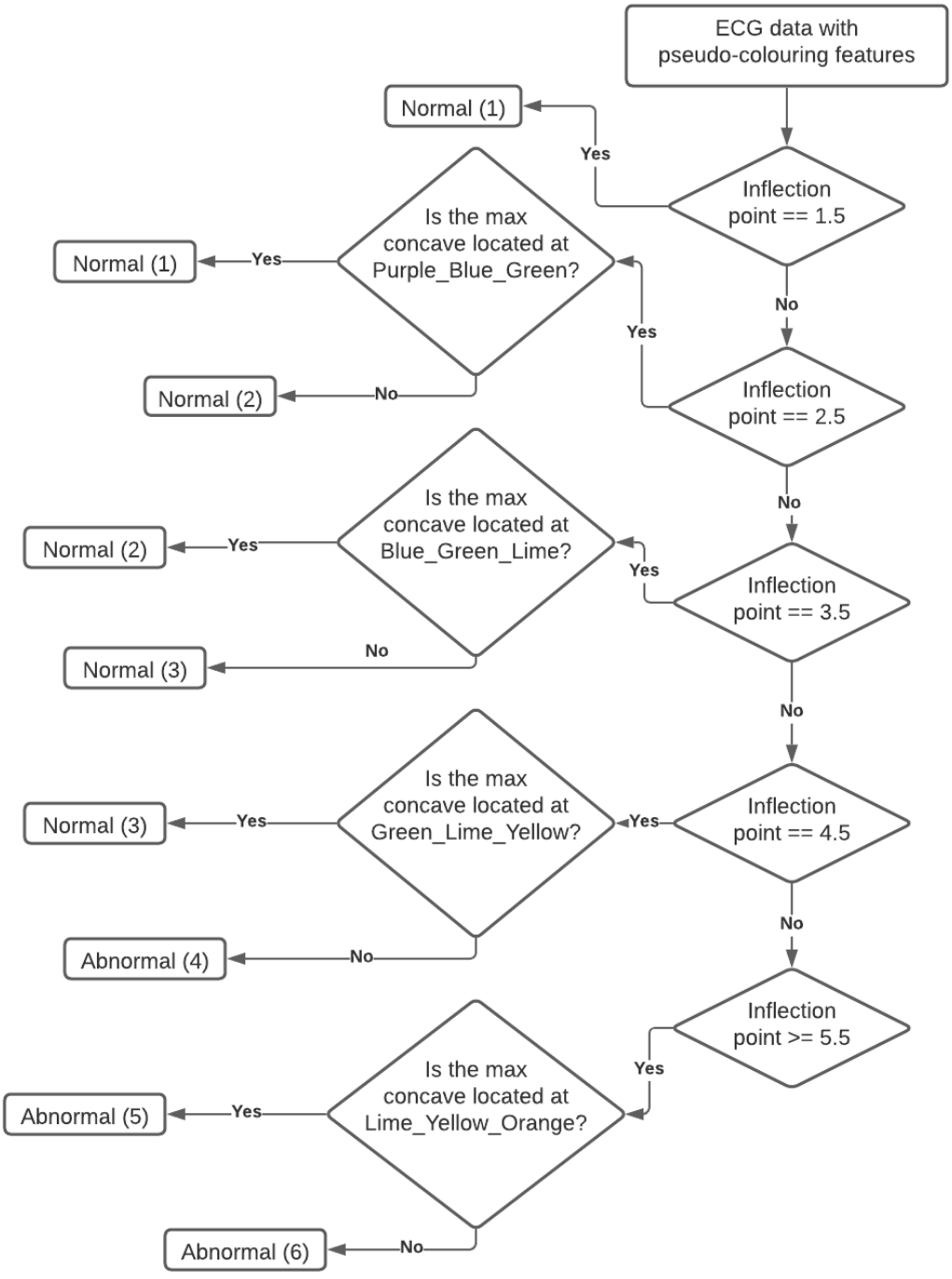


Figure 8.11: A flowchart of the expert IF-THEN rules. The number in round brackets represents the confidence rating of the algorithm.

Table 8.1: The confidence rating and the corresponding maximum concave pseudo-colours, inflection point index and the QT range relative to the pseudo-colouring index and the QT-nomogram.

Rating	Likelihood	Maximum concave pseudo-colours	Inflection point index	QT range (Pseudo-colouring index)	QT range (Nomogram Line = NL)
1	Very Likely Normal	Purple,Blue, Green	1.5	3 (Green) to 4 (Lime)	NL - 160ms to NL - 120ms
2	Probably Normal	Blue,Green, Lime	2.5	4 (Lime) to 5 (Yellow)	NL - 120ms to NL - 80ms
3	Possibly Normal	Green,Lime, Yellow	3.5	5 (Yellow) to 6 (Orange)	NL - 80ms to NL - 40ms
4	Possibly Abnormal	Lime,Yellow, Orange	4.5	6 (Orange) to 7 (Dark Orange)	NL - 40ms to NL
5	Probably Abnormal	Yellow,Orange, Dark Orange	5.5	7 (Dark Orange) to 8 (Red)	NL to NL + 40ms
6	Very Likely Abnormal	Yellow,Orange Dark Orange	6.5	8 (Red) to 9 (Dark Red)	NL + 40ms to NL + 80ms

The inflection point index can also help estimate the range in which the numerical value of the QT-interval is likely to be. This can support an ECG reader measuring the QT-interval, by providing an estimated 40 ms time window (representing a small square on the ECG grid) at which the T-wave probably ended. Figure 8.9 shows how the second derivative was calculated by examining the shape of the ECG curve within three successive pseudo-colours, and then moving one pseudo-colour at a time (*i.e.* inspecting the shape of the ECG curve again within the two backward pseudo-colours and one forward pseudo-colour). The inflection point index is located between the maximum concave down (corresponding to a negative second derivative that decreases within the three successive pseudo-colours) and the subsequent convex, which has a positive second derivative that is increasing within the previous two pseudo-colours and one subsequent pseudo-colour, representing an increasing curve, at which point the T-wave probably ended. If the maximum concave down in Figure 8.9 was located in the first grey block of the second derivative (*i.e.* within the purple, blue and green pseudo-colours), and the inflection point was 1.5, the T-wave was decreasing within the blue and green pseudo-colours, but increasing within the lime pseudo-colour. As each colour represents a 40 ms time scale, then the QT range can be estimated accordingly. Table 8.1 illustrates how the QT range relative to the pseudo-colouring index and the QT-nomogram was estimated.

'Explaining' the algorithm

A textual interpretation explaining the output of the algorithm was provided with the classification. It described what is visualised in the pseudo-coloured ECG signal and how the algorithm reached its decision. In some cases, the ECG may have an ST-elevation, a sign of myocardial infarction (heart attack), represented by the shape of the ECG signal as a concave (representing the ST-elevation), followed by a convex (representing the beginning of the T-wave, *i.e.* the J–T^{peak} interval), then a maximum concave down representing the T-wave. This information was included in the result text report if the ECG had this pattern. Below, we show examples of how the algorithm results were explained for four ECGs with different QT-intervals and heart rates. Figure 8.12 shows normal QT-intervals, and Figure 8.13 shows at risk QT-prolongation, where the ECG (A) also has an ST-elevation.

> “The QT-interval of this ECG is very likely normal, and the patient is not considered at risk for TdP. This decision has been made based on the assumption that the QT-interval is considered normal when the area under the T-wave contains cool pseudo-colours (purple to blue to green) and prolonged with risk of TdP when it contains warm pseudo-colours (yellow to orange to red). The maximum concave down in the pseudo-coloured area was considered to be the T-wave, and most of its colours are cool (purple to blue to green), with a greater amount of blue than green, indicating a normal QT-interval. The T-wave probably ends within the green colour region, which indicates that the QT/HR falls below the nomogram line by approximately 120 ms or more, showing no risk of TdP. Based on the pseudo-colouring scale, the estimated value of the QT-interval ranges from 310 to 350 ms, and the HR is 71.”
>
> *[Figure 8.12A]*

“ The QT-interval of this ECG is probably normal, and the patient is not considered at risk of TdP. This decision has been made based on the assumption that the QT-interval is considered normal when the area under the T-wave contains cool pseudo-colours (purple to blue to green) and prolonged with risk of TdP when it contains warm pseudo-colours (yellow to orange to red). The maximum concave down in the pseudo-coloured area was considered to be the T-wave, and most of its colours are cool colours (purple to blue to green), with a greater amount of green than blue, indicating a normal QT-interval. The T-wave is probably ends within the green region, which indicates that the QT/HR falls below the nomogram line by approximately 80 ms or more, showing no risk of TdP. Based on the pseudo-colouring scale, the estimated value of the QT-interval ranges from 343 to 383 ms, and the HR is 75. ”

[Figure 8.12B]

“ The QT-interval of this ECG is possibly abnormal, and the patient is considered at risk of TdP. This decision has been made based on the assumption that the QT-interval is considered normal when the area under the T-wave contains cool pseudo-colours (purple to blue to green) and prolonged with risk of TdP when it contains warm pseudo-colours (yellow to orange to red). The maximum concave down in the pseudo-coloured area was considered to be the T-wave, and most of its colours are warm colours (green to yellow to orange), with a greater amount of yellow than green, indicating an abnormal QT-interval. It looks like there is an ST-elevation as there is a purple-blue wave before the T-wave. The T-wave probably ends within the yellow-orange region, which indicates that the QT/HR falls on or very close to the nomogram line, showing risk of TdP. Based on the pseudo-colouring scale, the estimated value of the QT-interval ranges from 444 to 484 ms, and the HR is 53. ”

[Figure 8.13A]

> “The QT-interval of this ECG is very likely abnormal, and the patient is considered at risk of TdP. This decision has been made based on the assumption that the QT-interval is considered normal when the area under the T-wave contains cool pseudo-colours (purple to blue to green), and prolonged with risk of TdP when it contains warm pseudo-colours (yellow to orange to red). The maximum concave down in the pseudo-coloured area was considered to be the T-wave, and most of its colours are warm colours (orange to red), indicating an abnormal QT-interval. The T-wave probably ends within the orange-red region, which indicates that the QT/HR falls above the nomogram line, showing risk of TdP. Based on the pseudo-colouring scale, the estimated value of the QT-interval ranges from 496 to 536 ms, and the HR is 77.”
>
> *[Figure 8.13B]*

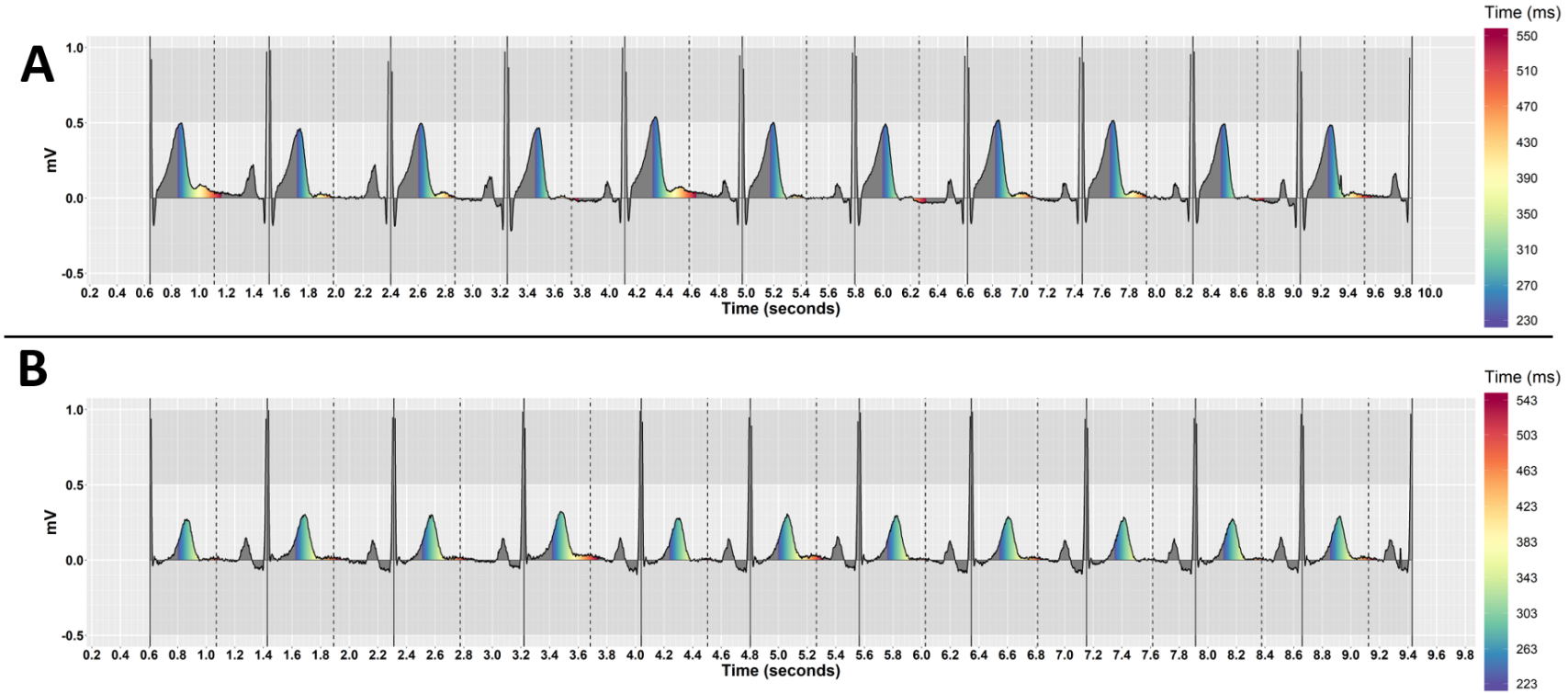


Figure 8.12: The pseudo-coloured ECG examples with normal QT-intervals showing no risk of TdP interpreted by the expert algorithm. Based on the actual QT/HR acquired from the clinical trial study: (A) The ECG has a normal QT-interval (QT = 345, HR = 71, Difference from the nomogram line = -125 , Drug = Ranolazine). (B) The ECG has a normal QT-interval (QT = 378, HR = 75, Difference from the nomogram line = -85 , Drug = Placebo).

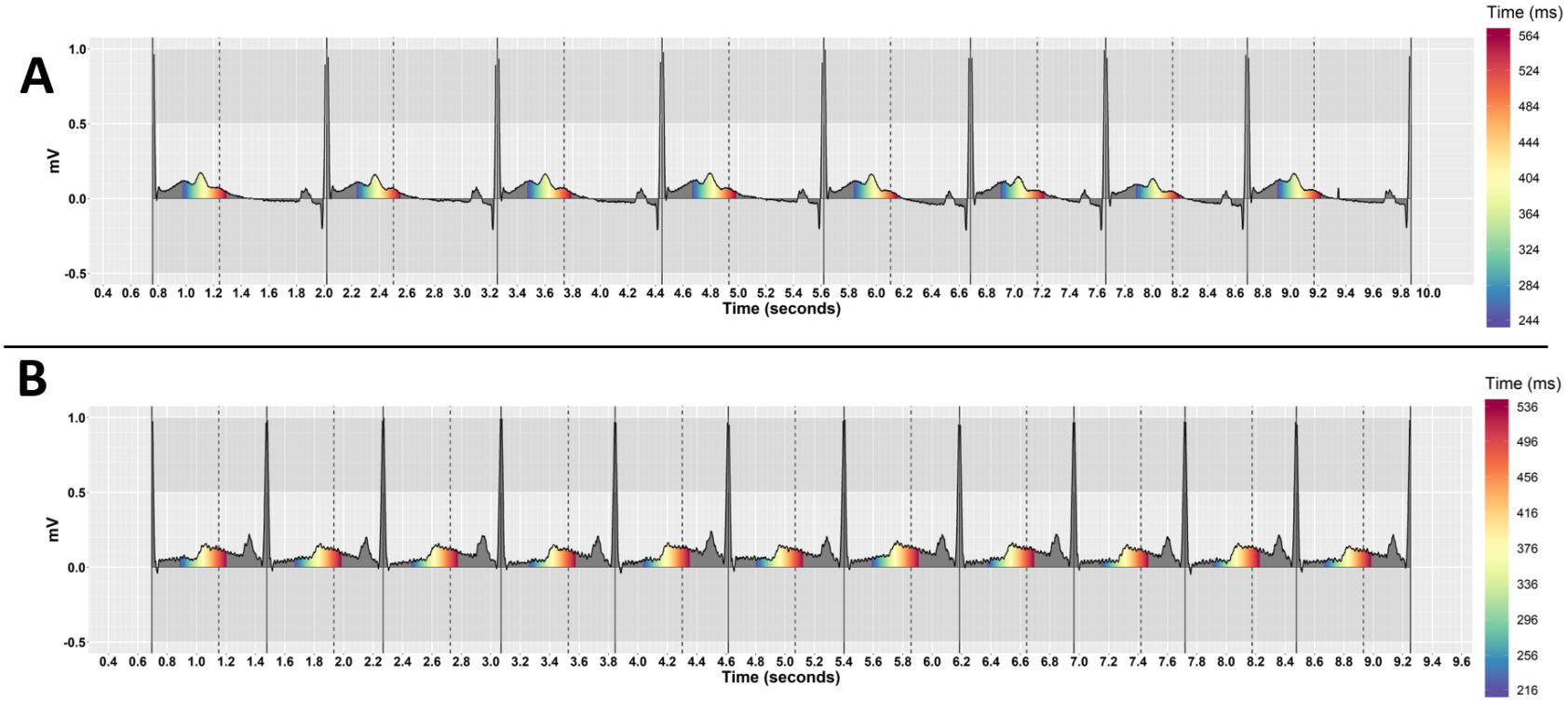


Figure 8.13: The pseudo-coloured ECG examples with prolonged QT-intervals at risk of TdP interpreted by the expert algorithm. Based on the actual QT/HR acquired from the clinical trial study: (A) The ECG has at risk QT-prolongation and ST-elevation (QT = 532, HR = 52, Difference from the nomogram line = + 48 , Drug = Dofetilide). (B) The ECG has at risk QT-prolongation (QT = 518, HR = 77, Difference from the nomogram line = + 62 , Drug = Quinidine).

8.3.3 Evaluation design

The expert algorithm was evaluated in four ways. Firstly, we evaluated the diagnostic accuracy of the algorithm for detecting QT-prolongation at risk of TdP on all ECG datasets ($n = 5050$). Secondly, as different types of QT-prolonging drug can affect the T-wave morphology in different ways, we systematically evaluated the effect of drug type on the algorithm's sensitivity to increases in the QT-interval using psychophysical methods. In vision science, psychophysical experiments investigate the relationship between the intensity of a physical stimulus and human perception, by systematically varying the properties of the stimulus along one or more physical dimensions [Ste17]. They are widely used in computer vision research to evaluate an algorithm's behaviour by measuring the exemplar by exemplar difficulty and modeling the algorithm's pattern of errors over different levels of object visibility and saliency, making the algorithm's classification inference more explainable (see examples in [RYKC$^+$18, GJS$^+$17, GWB13, ECS16, HSSB98]). In addition, we assessed how the drug type affected the algorithm's classification confidence score.

Thirdly, we compared the human expert-generated rules with a set of rules automatically generated through statistical machine learning. In this case the same six predictors used for the expert algorithm were used as inputs to train a C4.5 decision tree classifier. The comparison was performed under two conditions: (1) generating the decision tree model with imbalanced class datasets, *i.e.* using all 5050 ECGs, and (2) with balanced class datasets using down-sampling, which involves randomly removing datasets from the majority class. A 5-fold cross-validation with 20% holdout procedure was used to evaluate the rule-based decision tree performance under both conditions. That is, 80% of the data was used to train the model and this was tested using 5-fold cross-validation withholding 20% of the data, as shown in Figure 8.14. The source code of the decision tree classification model can be found in [ADVJ21]. Finally, we compared the classification accuracy of the expert algorithm with the results of a previous study conducted with humans [ADVJ20a].

8.3.4 Statistical analysis

The diagnostic accuracy of the algorithm was measured by calculating the area under the receiver operating characteristic (ROC) curve (AUC), sensitivity, specificity, precision, accuracy, F1-score and error rate. QT-intervals below the nomogram line were classified as negative (*i.e.* 'normal', no risk of TdP), and QT-intervals on or above the

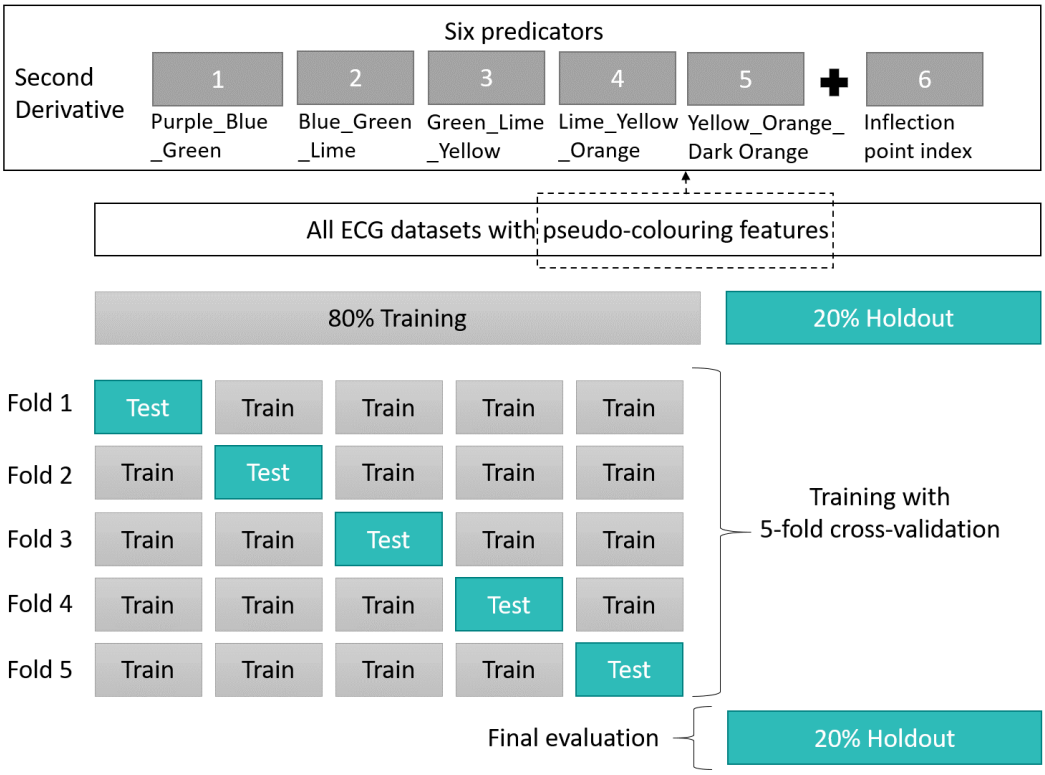


Figure 8.14: An illustration of the 5-fold cross-validation with 20% holdout procedure used to develop a rule-based decision tree classification model for pseudo-coloured ECGs.

nomogram line as positive (*i.e.* 'abnormal', at risk of TdP). As the ECG dataset has a class imbalance issue, *i.e.* the number of samples in the negative class ($n = 4870$) is much larger than the number of samples in the positive class ($n = 180$), we used the additional statistical measures of the balanced accuracy (*i.e.* the average of the sensitivity and specificity), area under the precision-recall curve (PR-AUC) and Matthews correlation coefficient (MCC), which have been shown to be particularly useful for evaluating imbalanced binary classification [CJ20, SR15].

The effect of drug type on the algorithm's sensitivity to increases in the QT-interval was modeled using the psychometric function and just noticeable difference (JND) threshold [LLM16]. The psychometric function is an inferential model used in psychophysical detection and discrimination tasks, here used to model the proportion of the ECGs classified as 'QT-prolongation at risk of TdP' by the expert algorithm, as a function of QT-interval increase across the four QT-prolonging drugs and placebo. The just noticeable difference (JND) threshold is defined in psychophysics as the minimum amount of a change in a stimulus required for it to be 'just noticeable' [P$^+$16]. In this study, we defined it as the minimum difference in the QT-interval from the nomogram line required for the TdP risk to be detectable. The JND threshold was determined by fitting the psychometric function using a logistic function with maximum likelihood estimation (MLE). The correlation between the QT-interval difference from the nomogram line and the algorithm's confidence score for each drug type and across all drugs was calculated using a Spearman's rank correlation.

The comparison with the rule-based decision tree model involved testing the expert algorithm on the same testing data used for each fold (then averaged across the five folds), and on the final 20% holdout testing datasets, under both conditions (imbalanced and balanced class datasets). Evaluation metrics included sensitivity, specificity, precision, balanced accuracy, F1-score, and Matthews correlation coefficient (MCC).

The comparison with human interpretation was conducted by testing the expert algorithm on ECGs ($n = 40$) used in a previous study conducted with humans to evaluate the pseudo-colouring technique [ADVJ20a]. The ECGs were acquired from the same clinical trial study used in the current paper, and were selected from multiple patients with different values of the QT-interval and heart rate; 20 ECGs were below the nomogram line (no TdP risk), and 20 ECGs were on or above the nomogram line, showing QT-prolongation at risk of TdP. Evaluation metrics included sensitivity, specificity and accuracy.

Table 8.2: The diagnostic accuracy results of the expert algorithm.

Evaluation Metric	Proportion
Accuracy	0.99
Balanced Accuracy	0.97
Recall/sensitivity	0.94
Specificity	0.99
Precision/positive predictive value (PPV)	0.83
F1-score	0.88
ROC (AUC)	0.98
Precision-recall (AUC)	0.88
Matthews correlation coefficient (MCC)	0.88
Error rate	0.01

8.4 Results

All pseudo-coloured ECGs and related metadata underpinning the findings reported in this article can be found in [ADVJ21].

8.4.1 Diagnostic accuracy

The results show that the expert algorithm is reliable and accurate in detecting QT-prolongation at risk of the TdP regardless of heart rate. The algorithm was tested on the 5050 ECGs (TdP risk $n = 180$, no risk $n = 4870$), with various values of QT-interval and heart rate across different QT-prolonging drugs, and a placebo. It achieved an accuracy of 0.99, balanced accuracy of 0.97, sensitivity of 0.94, specificity of 0.99, precision of 0.83, F1-score of 0.88, ROC (AUC) of 0.98, precision-recall (AUC) of 0.88, Matthews correlation coefficient (MCC) of 0.88, and error rate of 0.01. The full results are shown in Table 8.2. Figure 8.15 shows the algorithm's ROC and precision-recall curves.

The expert algorithm provides an estimated 40 ms range of the QT-interval value, and we found that 74% of the ECGs were within the algorithm's estimated range. In addition, the expert algorithm showed superior performance to the pilot version, which considered a QT-interval to be prolonged with risk of TdP if the warm pseudo-colours (yellow to orange to red) occupied more than 50% of the area under the ECG signal. Figure 8.16 shows a comparison of sensitivity, specificity and balanced accuracy, when classifying all ECGs ($n = 5050$), between the pilot version and the modified version reported in this paper. This indicates that locating concavity and inflection points with

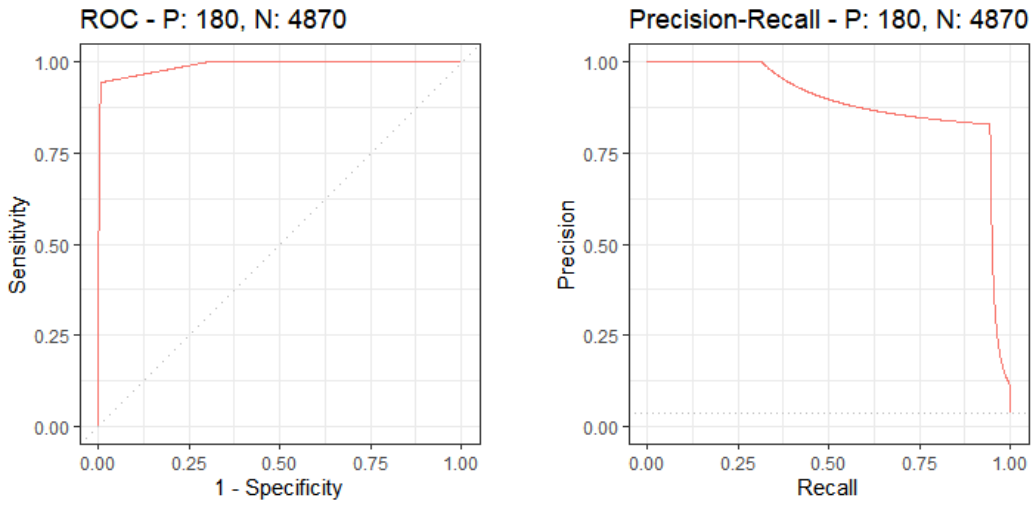


Figure 8.15: The algorithm's ROC and precision-recall curves when tested on all ECGs (Positive cases (P) $n = 180$, Negative cases (N) $n = 4870$). ROC (AUC) = 0.98 and precision-recall (AUC) = 0.88.

respect to the pseudo-colouring scale significantly increased sensitivity to detecting QT-prolongation.

8.4.2 Modeling the effect of drug type on QT-interval prolongation detection

The clinical trial study from which the ECGs were acquired assessed the effect of QT-prolonging drug type on T-wave morphology [VJM+15]. The results showed that patients on a pure hERG blocker (Dofetilide) developed flat, asymmetric, and notched T-waves, whilst patients on multi-channel blocking drugs (Quinidine and Ranolazine) had equal or greater T-wave morphology changes including distorted and bizarre T-waves. In particular, 'Dofetilide' and 'Quinidine' were shown to cause significant T-wave morphology changes, while substantially prolonging the QT-interval. No significant T-wave changes were observed for patients on 'Verapamil', or the placebo.

In this study, we examined the robustness of the expert algorithm to T-wave morphology changes by evaluating its sensitivity in detecting at risk QT-prolongation across drug types. The psychometric function was plotted as the proportion of the ECGs classified as 'QT-prolongation at risk of TdP' by the expert algorithm, as a function of the QT-interval difference from the nomogram line. The results show that the sensitivity of the expert algorithm in detecting at risk QT-prolongation increased as the

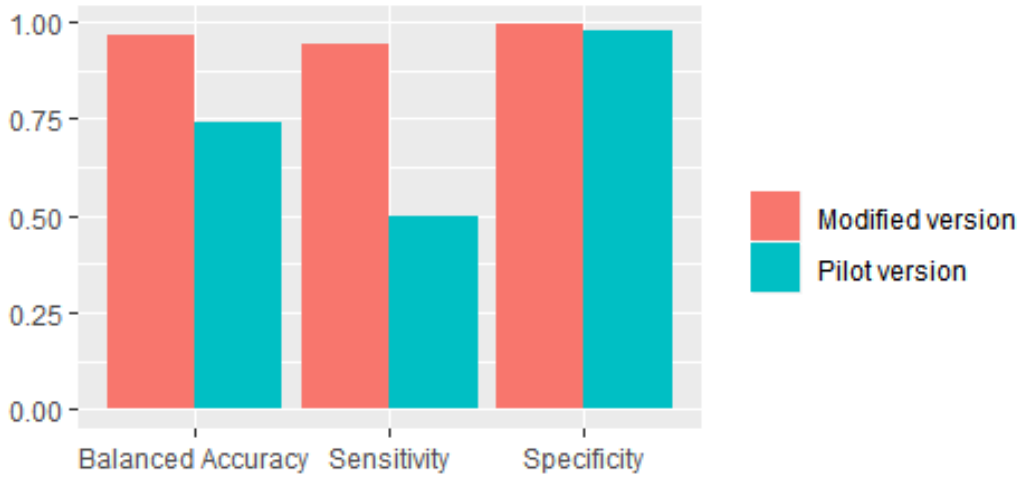


Figure 8.16: A comparison of the sensitivity, specificity and balanced accuracy, when classifying all ECGs ($n = 5050$), between the pilot version and the modified version of the expert algorithm.

QT-interval difference from the nomogram decreased, regardless of T-wave morphology changes (Figure 8.17). We estimated the 75% just noticeable difference (JND) threshold as the value of the QT-interval with respect to the nomogram line at which the proportion of the ECGs classified as 'QT-prolongation at risk of TdP' by the expert algorithm is equal to 0.75. It was estimated for the drugs that caused a QT-prolongation at risk of TdP, which were 'Dofetilide' and 'Quinidine'. Figure 8.18 shows the JND threshold results, which show that the expert algorithm was sensitive in detecting TdP risk for 75% of the ECGs at which the QT-interval value is below the nomogram line by approximately -9 and -3 milliseconds for 'Dofetilide' and 'Quinidine' respectively.

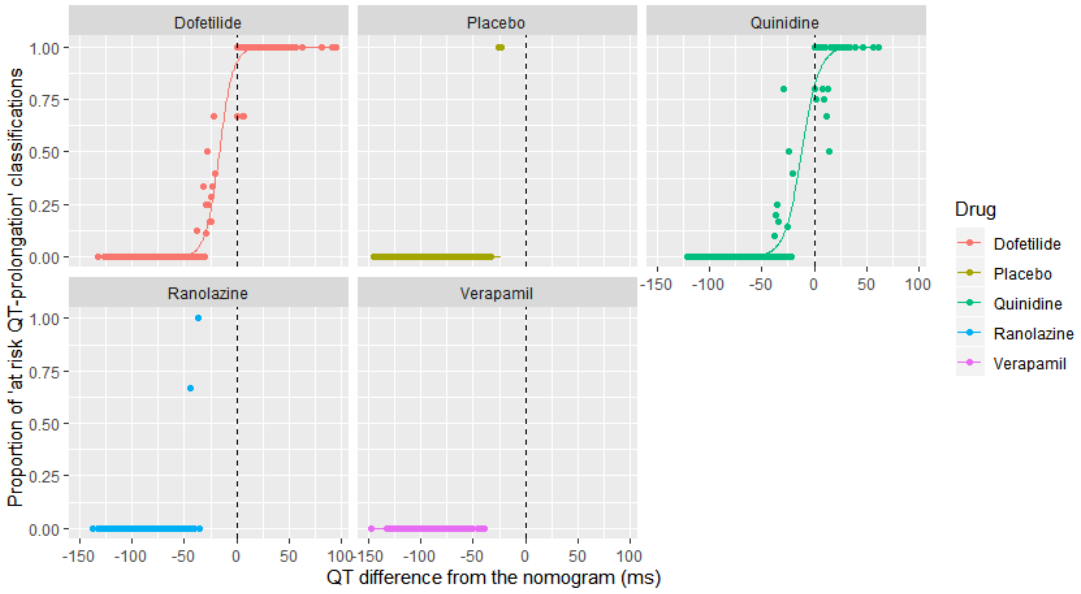


Figure 8.17: The psychometric function plot shows the proportion of the ECGs classified as 'QT-prolongation at risk of TdP' by the expert algorithm, as a function of the QT-interval difference from the nomogram line for each drug type. The QT value of the nomogram line is equal to 0 on the X-axis.

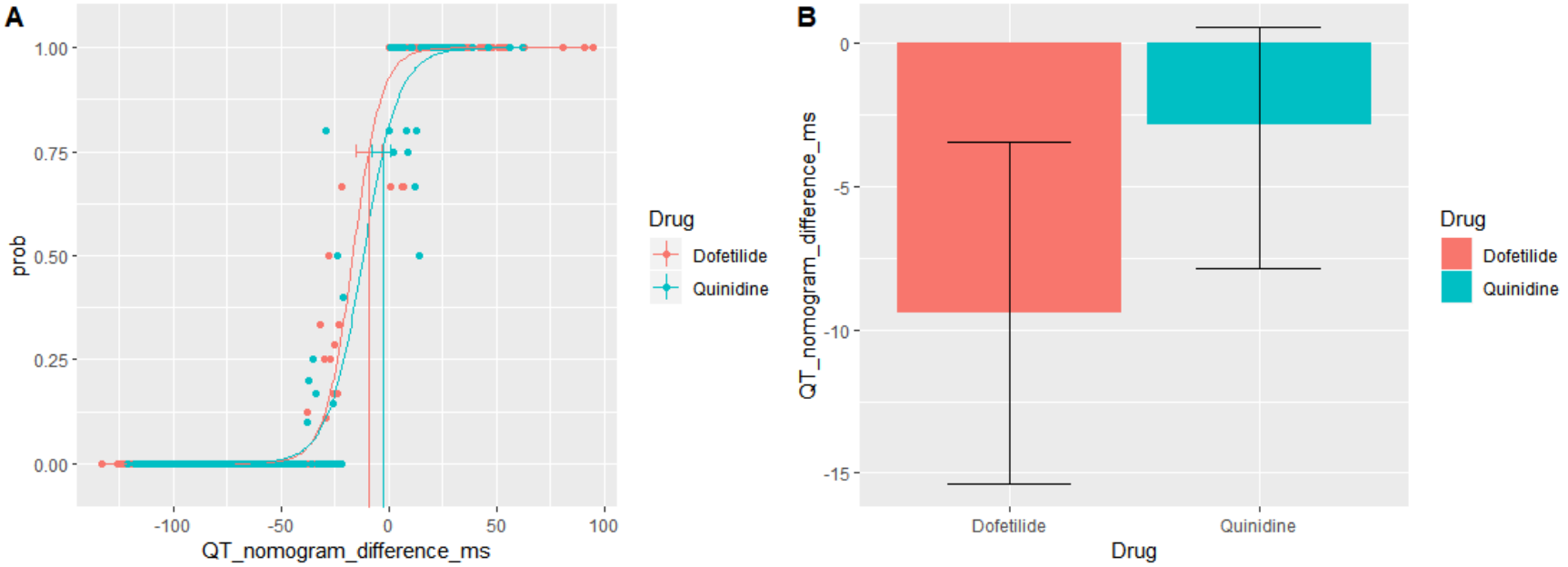


Figure 8.18: Psychophysical detection measures of the expert algorithm's sensitivity. (A) The fitted psychometric function plot shows the proportion of the ECGs classified as 'QT-prolongation at risk of TdP' by the expert algorithm, as a function of the QT-interval difference from the nomogram line for 'Dofetilide' and 'Quinidine'. The QT value of the nomogram line is equal to 0 on the X-axis. (B) The just noticeable difference (JND) thresholds plot. The error bars represent bootstrap confidence intervals.

Table 8.3: The results of the Spearman's rank correlation between the algorithm's confidence rating and the QT-interval difference from the nomogram line for each drug type and all.

Drug type	Spearman's rho (R)	P-value
Dofetilide	0.867	**<0.001**
Quinidine	0.824	**<0.001**
Ranolazine	0.768	**<0.001**
Verapamil	0.750	**<0.001**
Placebo	0.810	**<0.001**
All	0.828	**<0.001**

A Spearman's rank correlation was used to determine the relationship between the algorithm's confidence rating and the QT-interval difference from the nomogram line. The results show that there was a strong, positive correlation between the algorithm's confidence rating and the QT difference from the nomogram, which was statistically significant (R = 0.828, $p < 0.001$) across all drug types, and per drug (Table 8.3). Figure 8.19 shows how the algorithm's confidence rating score of TdP risk increased as the QT-interval difference from the nomogram line decreased (*i.e.* as the QT-value gets closer to the nomogram line). The results demonstrate the reliability of the algorithm as high risk scores (5 and 6) were given only for 'Dofetilide' and 'Quinidine' drugs, where the QT-values were above the nomogram line showing risk of TdP (Figure 8.19).

8.4.3 Comparison with rule-based decision tree classification model

Imbalanced classes are a common issue in machine learning, particularly with medical data, where rare conditions have a limited number of representative cases. Imbalance may introduce bias into the training data, but there is no one-size-fits-all solution, as in some cases using more data, even if not balanced, may improve the model performance [KAL20, LCKP19, AS18]. As such, we examined whether class balance affected the decision tree's rule generation and performance. We used the down-sampling technique as we have a very large number of inputs in the negative class. The re-sampled datasets have various values of QT-interval and heart rate, and were for multiple patients on multiple QT-prolonging drugs/placebo.

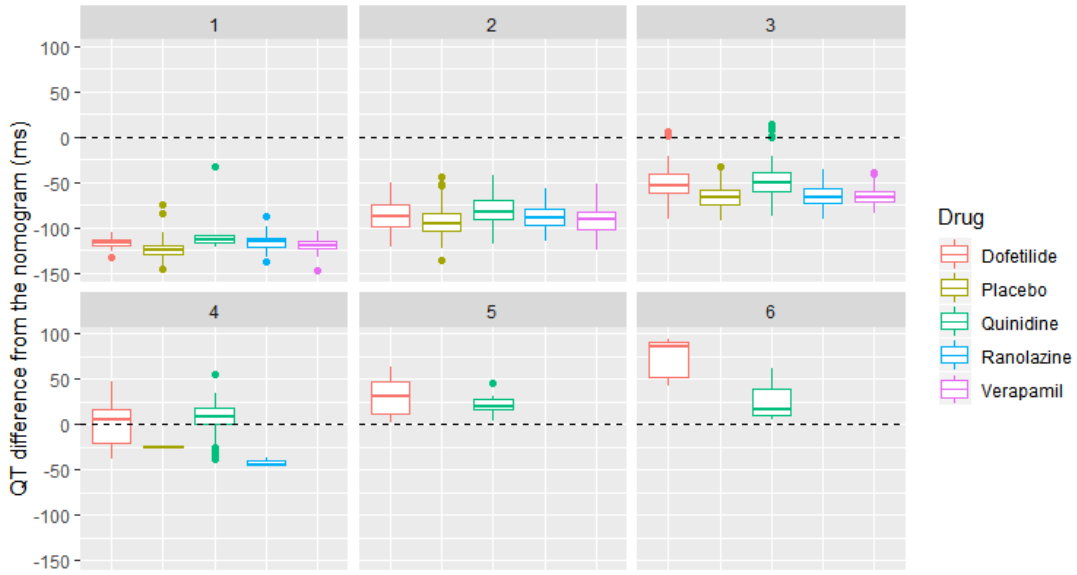


Figure 8.19: A boxplot shows the distribution of the ECG cases grouped by their confidence rating (1-6) as classified by the expert algorithm. The QT value of the nomogram line is equal to 0 on the Y-axis. The confidence ratings are 'very likely normal' (1), 'probably normal' (2), 'possibly normal' (3), 'possibly abnormal' (4), 'probably abnormal' (5), and 'very likely abnormal' (6).

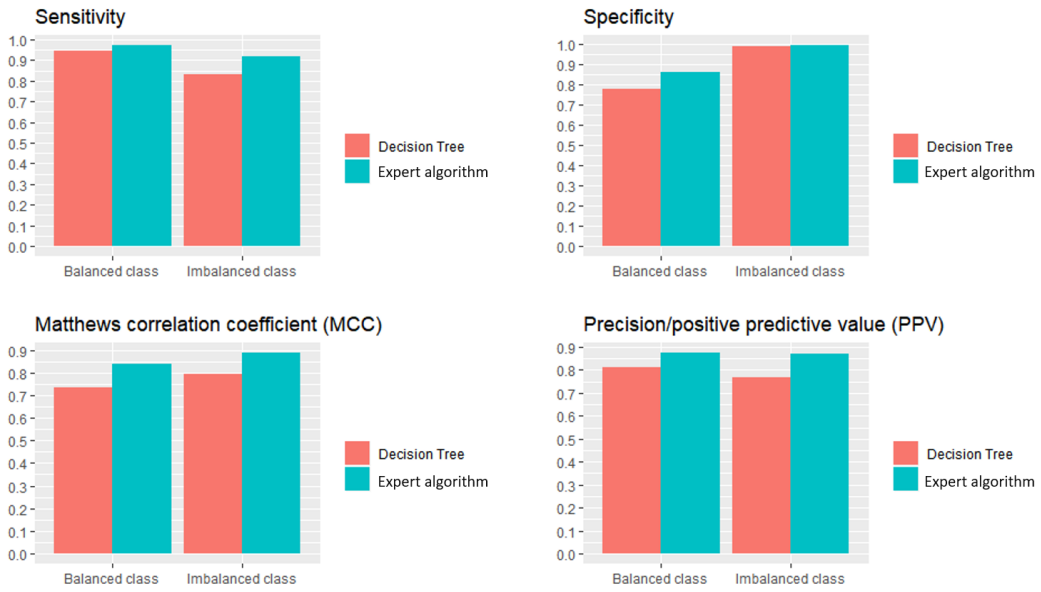


Figure 8.20: A comparison of the expert algorithm and the rule-based decision tree model, under imbalanced and balanced class conditions, when tested on 20% hold-out datasets. Evaluation metrics include sensitivity, specificity, Matthews correlation coefficient (MCC) and precision, also known as positive predictive value (PPV).

Under both the imbalanced and balanced class conditions, the results show that the expert algorithm was more accurate and generalisable to new ECG data than the decision tree model (Figure 8.20). The decision tree model was over-fitted to the training data under both conditions, showing higher average accuracy during cross-validation than the expert algorithm, but lower accuracy when tested on new data. Tables 8.4 and 8.5 show the evaluation results for the imbalanced and balanced class conditions respectively.

More rules were generated under the balanced class condition, which was expected given that the training dataset is smaller and thus the classification rules may be more difficult to infer. Figure 8.21 illustrates a flowchart of the decision tree IF-THEN classification rules for the imbalanced and balanced class conditions.

Table 8.4: A comparison of the rule-based decision tree model and the expert algorithm, under the imbalanced class condition, showing the evaluation metrics on each fold (K), averaged across the 5 folds, and on 20% holdout datasets. MCC = Matthews correlation coefficient.

		MCC	Precision	Recall/ Sensitivity	Specificity	Balanced Accuracy	F1-score
K1	Expert algorithm	0.882	0.837	0.939	0.992	0.965	0.885
	Decision tree model	0.919	0.937	0.909	0.997	0.953	0.923
K2	Expert algorithm	0.961	0.925	0.967	0.997	0.998	00.961
	Decision tree model	0.979	0.961	1	0.998	0.999	0.980
K3	Expert algorithm	0.828	0.730	0.950	0.991	0.970	0.826
	Decision tree model	0.903	0.863	0.950	0.996	0.973	0.904
K4	Expert algorithm	0.876	0.822	0.948	0.989	0.969	0.880
	Decision tree model	0.859	0.914	0.820	0.996	0.908	0.864
K5	Expert algorithm	0.839	0.774	0.923	0.991	0.957	0.842
	Decision tree model	0.943	0.928	0.962	0.997	0.980	0.945
Averaged across 5 folds	Expert algorithm	0.877	0.818	0.952	0.992	0.972	0.879
	Decision tree model	0.921	0.921	0.928	0.997	0.962	0.923
20% holdout datasets	Expert algorithm	0.888	0.868	0.917	0.995	0.956	0.892
	Decision tree model	0.793	0.769	0.833	0.991	0.912	0.800

Table 8.5: A comparison of the rule-based decision tree model and the expert algorithm, under the balanced class condition, showing the evaluation metrics on each fold (K), averaged across the 5 folds, and on 20% holdout datasets. MCC = Matthews correlation coefficient.

		MCC	Precision	Recall/ Sensitivity	Specificity	Balanced Accuracy	F1-score
K1	Expert algorithm	0.889	0.972	0.945	0.952	0.949	0.958
	Decision tree model	0.887	0.947	0.972	0.904	0.938	0.960
K2	Expert algorithm	0.734	0.800	0.965	0.750	0.857	0.875
	Decision tree model	0.806	0.828	1	0.785	0.892	0.906
K3	Expert algorithm	0.893	0.956	0.916	0.970	0.943	0.936
	Decision tree model	0.928	0.958	0.958	0.970	0.964	0.958
K4	Expert algorithm	0.894	0.960	0.923	0.967	0.945	0.941
	Decision tree model	0.894	0.960	0.923	0.967	0.945	0.941
K5	Expert algorithm	0.861	0.928	0.928	0.933	0.930	0.928
	Decision tree model	0.896	0.962	0.928	0.966	0.947	0.945
Averaged across 5 folds	Expert algorithm	0.854	0.923	0.935	0.914	0.925	0.927
	Decision tree model	0.882	0.931	0.956	0.919	0.937	0.942
20% holdout datasets	Expert algorithm	0.839	0.875	0.972	0.861	0.917	0.921
	Decision tree model	0.732	0.810	0.944	0.778	0.861	0.872

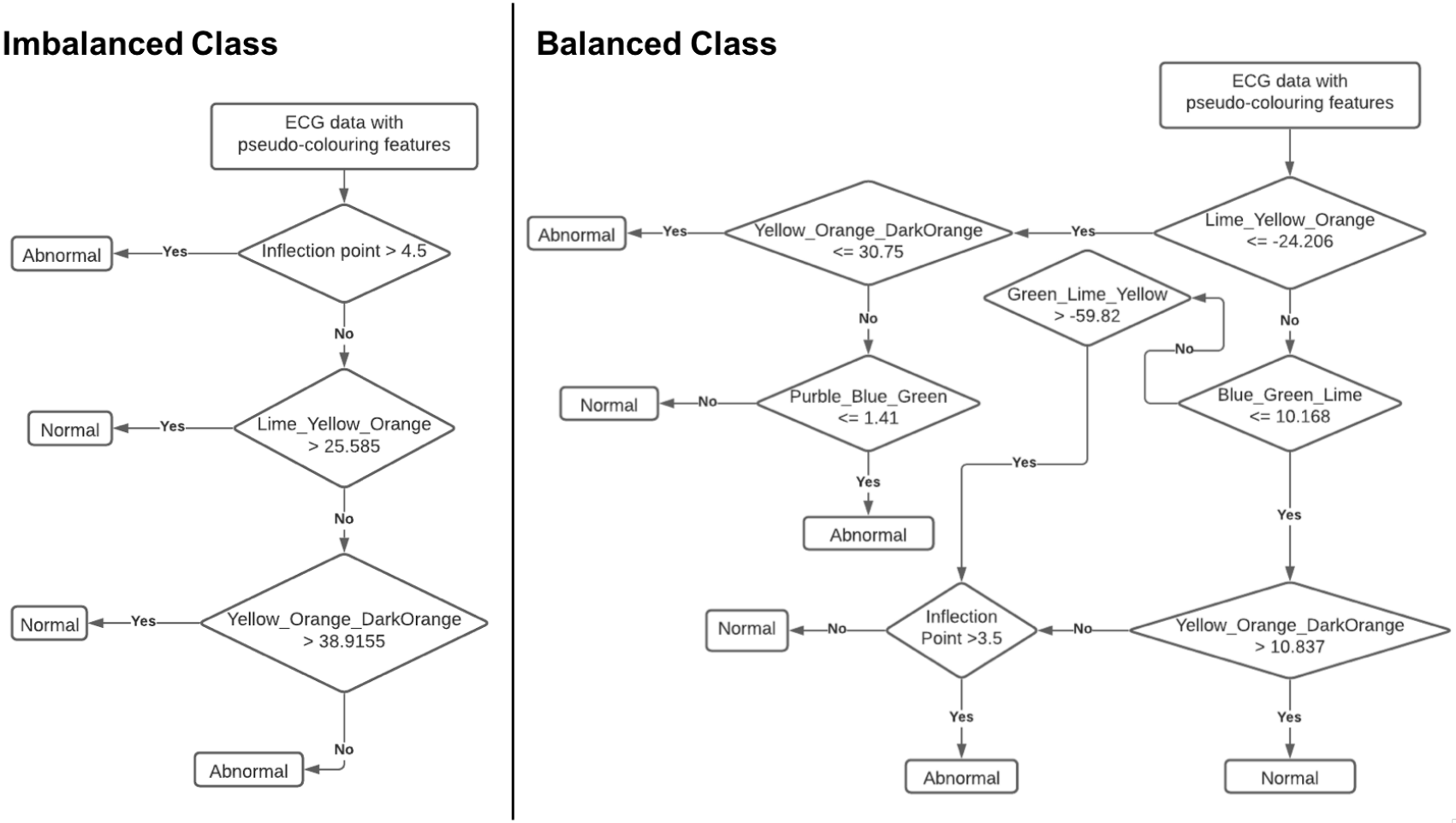


Figure 8.21: A flowchart of the decision tree IF-THEN rules when trained on imbalanced and balanced class data.

Table 8.6: The sensitivity, specificity, balanced accuracy, and area under the ROC curve of the expert algorithm and human participants (mean values) when classifying the 40 ECGs (TdP risk $n = 20$, no risk $n = 20$).

	Expert algorithm	Human	Difference
Sensitivity	1	0.83	0.17
Specificity	0.95	0.90	0.05
Balanced accuracy	0.98	0.87	0.11
ROC (AUC)	0.98	0.93	0.05

8.4.4 Comparison with human interpretation

A comparison with human interpretation shows that the expert algorithm was more accurate at classifying the 40 ECGs than the averaged human participants ($n = 43$). Figure 8.22 and Table 8.6 show the performance results of the expert algorithm and human interpretation in terms of sensitivity, specificity, balanced accuracy, and area under the ROC curve.

8.5 Focus group evaluation

We conducted two focus group discussions, one with patients and one with clinicians, to gather feedback about the potential usage of the pseudo-coloured ECG and automated algorithm within clinical practice. The focus groups were conducted as patient and public involvement/stakeholder engagement events, which are less formal than an interview study, and designed to gather feedback to direct an approach, rather than provide empirical evidence for it.

The patient focus group involved seven cancer patients (4 male and 3 female), who were recruited via the Patient and Public Involvement and Engagement (PPIE) coordinators of the Cancer Precision Medicine themes, within the NIHR Manchester Biomedical Research Centre (BRC), UK. Table 8.7 shows the patients' demographic information and Table 8.8 summarises the focus group discussion with patients using the short form of GRIPP2 (Guidance for Reporting Involvement of Patients and Public) [SBS+17].

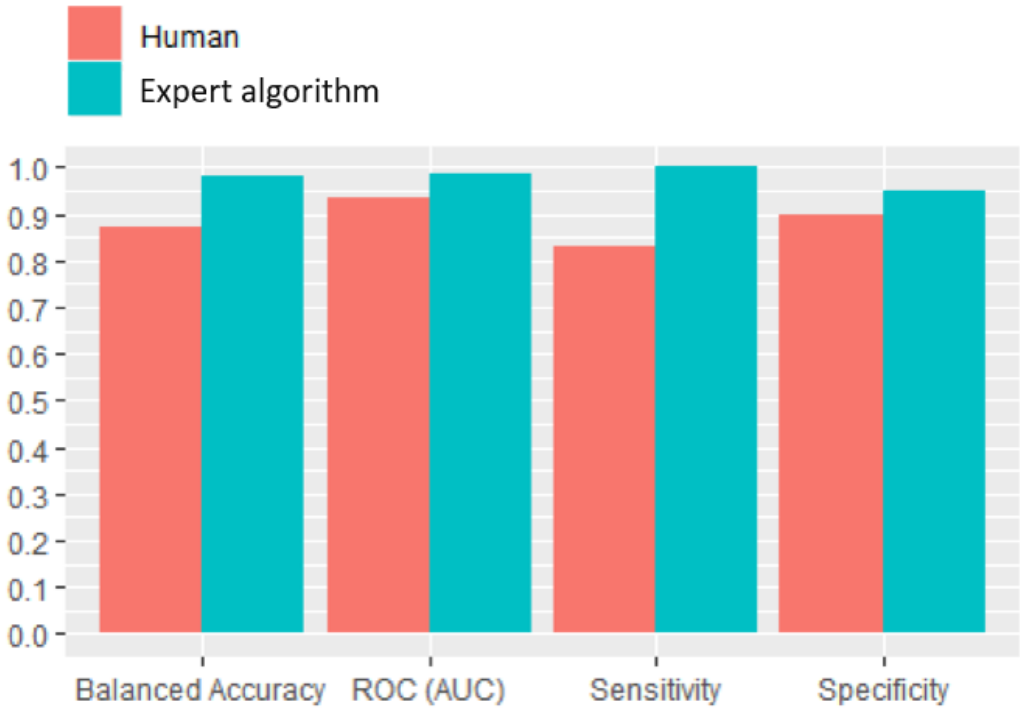


Figure 8.22: The sensitivity, specificity, balanced accuracy, and area under the ROC curve of the expert algorithm and human participants (mean values) when classifying the 40 ECGs (TdP risk $n = 20$, no risk $n = 20$).

Table 8.7: Patients' demographic information.

ID	Sex	Age	Education	Occupation
1	Female	50-59 years	Bachelors	Disabled, not working
2	Male	<30 years	Diploma	Working full time
3	Male	<30 years	Diploma	Working full time
4	Female	30-39 years	Masters	Unemployed
5	Male	50-59 years	Bachelors	Working part time
6	Female	50-59 years	Masters	Working full time
7	Male	50-59 years	Masters	Working full time

All patients had been regularly monitored for drug-induced ECG changes. Some patients ($n = 3$) had attempted to interpret their ECG signal results before, but they found them difficult to understand. The investigator gave a short introduction explaining drug-induced LQTS and how to detect it on both standard and pseudo-coloured ECGs.

All patients found the pseudo-colouring technique to be effective in distinguishing between normal and prolonged QT-intervals, and preferred the coloured ECG over the non-coloured one.

Most patients ($n = 5$) agreed that empowering people to self-monitor for drug-induced LQTS had the potential to save lives, provide a cost-effective healthcare solution, and support more informed shared decision-making between patients and clinicians. Three patients agreed a verbal/text-based explanation feature would support them in interpreting the pseudo-coloured ECG, and felt that this would be particularly helpful for elderly patients. Two patients believed that having a risk assessment feature that showed the severity of warm colours relative to the QT-prolongation level could help people with anxiety, so they know when not to worry, and when to seek medical advice.

While two patients raised concerns about trusting automated ECG interpretation, the other patients welcomed the idea of the expert rule-based ECG interpretation algorithm as a supportive tool that may help to overcome the potential challenges associated with human interpretation. Most patients ($n = 5$) wanted to understand why the algorithm made a certain decision, particularly if it suggested the QT-interval to be abnormal.

Table 8.8: The short form of GRIPP2 (Guidance for Reporting Involvement of Patients and Public) for the focus group discussion with patients.

Section and topic	Extra detail
Aim	To gather feedback and opinions about the potential usage of the pseudo-coloured ECG, and the automated algorithm that interprets it, within clinical practice.
Methods	An online focus group meeting with patients was held in May 2020. Seven patients who were under frequent ECG monitoring were approached by the PPIE coordinators for the Cancer Precision Medicine themes. A semi-structured focus group guide was followed, seeking views on 1) self-monitoring drug-induced LQTS using the pseudo-colouring technique, and 2) automated ECG interpretation. The meeting was led by one of the research team, facilitated by two members of the Manchester BRC's staff.
Study results	All patients perceived pseudo-colouring as an effective way of distinguishing between normal and prolonged QT-intervals, and preferred the coloured ECG over the non-coloured one. They said that empowering people to self-monitor ECGs is going to save lives, be cost-effective, and improve patient-clinician communication, supporting more informed shared decision-making. While two patients raised concerns about trusting automated ECG interpretation, other patients welcomed the idea of an automated algorithm that used the same process of reading the pseudo-colours as expert humans.
Discussion and conclusions	All patients had a positive attitude towards using the pseudo-colouring technique, and perceived the expert automated algorithm as a supportive tool that may help to overcome the potential challenges associated with human interpretation of the pseudo-colours.
Reflections and critical perspective	Involving patients demonstrated the potential of the expert algorithm as a risk assessment tool that could help to quantify the severity of QT-interval prolongation through the area of warm colours displayed, and highlighted the advantage of producing explainable results in a way that could support and guide the lay interpreter of the pseudo-coloured ECG.

Table 8.9: Clinicians' demographic information.

ID	Sex	Age	Education	Job title
1	Female	<30 years	Diploma	Clinical Research Nurse
2	Female	30-39 years	BNurs	Clinical Research Nurse
3	Male	30-39 years	MRes	Clinical Fellow

The focus group with clinicians involved two female nurses and one male doctor, who are working on early phase clinical trials of new experimental cancer drugs, within the Experimental Cancer Medicine Team (ECMT) at The Christie NHS Foundation Trust, Manchester, UK. They were recruited via our research collaborators within the digital ECMT. Table 8.9 shows the clinicians' demographic information.

As a part of their routine clinical practice, the nurses record patients' ECGs, but they do not interpret them, while the doctor interprets the ECGs and monitors QT-interval changes. The doctor stated that the standard approach to assessing the QT-interval in clinical trials is by using a rigid manual QT-interval measurement, and they do not rely on any kind of automated QT measurements at present. The clinicians identified the potential benefits of using the pseudo-colouring technique within clinical practice as speeding up the process of interpreting ECGs, and supporting QT-interval monitoring out of hours. All clinicians perceived the pseudo-colouring technique to be particularly useful for clinicians who have less training in ECG interpretation, as it can assist them in visually detecting drug-induced QT-prolongation.

The doctor, who frequently interprets ECGs, recommended having an automated expert algorithm that provided an alert about QT-prolongation, while showing the visualised pseudo-coloured ECG at the same time. He commented this would be particularly useful in reducing common errors associated with manual QT-interval measurement, and in resolving issues with inter-observer variability, particularly in clinical trial settings testing a new drug where accurate QT-monitoring is crucial. He said that well-known difficulties with measuring the QT-interval in clinical practice included slow and fast heart rates and T-wave morphology changes (*e.g.* the fused T-U complex). He felt that the expert algorithm would be very useful in clinical practice if it could detect QT-prolongation regardless of these issues. All clinicians believed that an expert, explainable automated algorithm would have a greater potential of being trusted and adopted in clinical practice than a fully automated algorithm where the basis for decisions was less clear.

8.6 Discussion

Many pharmacological drugs have been shown to prolong the QT-interval on the ECG and reported to cause drug-induced TdP and/or sudden cardiac death [Kha02, CMY08, ESTB20]. Despite its clinical importance, predicting the TdP risk for most of these drugs is difficult, even for clinicians who routinely read ECGs [VRS$^+$05, STN08]. Major challenges include measuring the QT-interval and determining the TdP risk at varying heart rates [YC03, CMY08]. While many modern ECG machines provide automated measurements of the QT-interval, these are usually correct only in noise-free normal sinus rhythm, in which the T-wave morphology is well defined [CMY08, SW17]. As QT-prolonging drugs often affect the morphology of the T-wave, this can easily invalidate automated QT measurement [TAS$^+$15, KBD$^+$18, EI13, GL13, TABW11, RSG09].

This study demonstrates that an automated ECG algorithm, developed with a *human-like* approach—using human perceptual heuristics to determine features, and expertise from the clinical literature to determine rules—is highly effective in detecting drug-induced QT-prolongation at risk of TdP regardless of heart rate (Table 8.2; Figure 8.15). The psychophysical detection measures show that the sensitivity of the expert algorithm in detecting TdP risk increased as the QT-interval approached the nomogram risk line, regardless of drug type (Figures 8.17 and 8.18). The JND threshold results indicate that although multi-channel blocking drugs are known to affect T-wave morphology to a greater extent than pure hERG blocking drugs [VJM$^+$15], the algorithm was in fact more sensitive with 'Quinidine' than 'Dofetilide' (Figure 8.18). Research has estimated that 2.0% to 8.8% of patients treated with 'Quinidine' will develop TdP, which is a high number [SW64, RWP86, BBH$^+$84, Dar01, CMY08]. Although the difference in the JND thresholds between the two drugs is very small ($\approx 6ms$), one possible reason behind it could be that 'Dofetilide' can cause greater T-U fusion/complexes, which are known to precede the development of TdP [JG14, DmFBR$^+$96, Zip87]. In fact, in this case measurement of the QU-interval is recommended [GMZ06, SOKS91, EsBH89]. This, in turn, will prolong the maximum concave and inflection point in the ECG signal, which potentially increases the sensitivity of the algorithm to TdP risk, whilst the QT-value is below, but very close, to the nomogram line. In addition, a Spearman's rank correlation shows a strong, positive relationship between the algorithm's confidence rating and the QT difference relative to the nomogram line regardless of drug type, demonstrating the reliability of the expert algorithm (Table 8.3; Figure 8.19).

The comparison with the rule-based decision tree model showed the value of using expert knowledge in the development of the algorithm. Both the expert and decision tree algorithms were highly effective, demonstrating the promise of using perceptual heuristics as features; whilst the decision tree was more effective under cross-validation, however, it was poorer at generalising than the expert algorithm. The decision tree inference mechanism depends on the 'value' of certain features to separate the two classes, which cannot be standardised across all ECGs. For example, under the imbalanced class condition in Figure 8.21, the second decision point separated the two classes based on the inflection point index and the value of the 'Lime_Yellow_Orange' feature. If the value of this feature is convex greater than 25.585, and the inflection point index is less than 4.5, then the QT-interval is classified as normal. However, the rate of change of a convex/concave depends on the amplitude of the ECG wave, which is known to differ substantially across individuals, and is affected by factors including age, race, sex and health status, body mass and electrode position [GMZ06, CMY08]. The decision of the expert algorithm is not based on inferred feature values, but rather on the relationships between features acquired from the medical knowledge used to detect risk of TdP. For instance, at the fourth decision point in Figure 8.11 the expert algorithm separates a normal from an abnormal QT-interval, despite the fact that they have the same inflection point index, based on the location of the maximum concave. As shown in Figure 8.9, if the inflection point index is equal to 4.5 and the maximum concave is located at 'Green_Lime_Yellow' and not 'Lime_Yellow_Orange', then this means the T-wave is farther from the nomogram line, and there is thus no TdP risk. Furthermore, we encountered the issue of cognitive and technical bias in selecting good, representative training datasets. Research has shown that this issue is relatively common when selecting training medical datasets [KAL20, LCKP19, AS18]. Despite numerous statistical re-sampling methods including up/downsampling and k-fold cross validation techniques, acquiring representative datasets that support the generalisability of machine learning algorithms, remains a challenging problem.

An additional benefit of the human-like approach is its inherent explainability. The expert algorithm is able to both precisely specify the reasons for a decision, and—due to the algorithm and human interpreter sharing the same, pseudo-coloured representation of the data—the reasons for the decision are straightforward for someone to perceive. The shared representation of the ECG data, and the shared model of how to interpret it, are also important when considering the application of this approach within

clinical practice. All new regulations for adopting artificial intelligence (AI)- and machine learning (ML)-based technologies in healthcare emphasise the importance of explainability [FA+19a, JM19]. Results from the focus groups with clinicians and patients also provided evidence for the value of the expert algorithm. The explanations it provides (Figures 8.12 and 8.13) may be helpful in training lay people to read pseudo-coloured ECGs to self-monitor QT-prolongation, as well as supporting clinicians in supervising and validating its automated results in clinical practice, using a human-in-the-loop approach. Clinicians commented that our approach would be particularly useful in reducing common errors associated with manual QT-interval measurement, and in resolving issues with inter-observer variability, particularly in clinical trial setting testing a new drug where accurate, frequent QT-monitoring is crucial. Whilst the decision tree algorithm is also technically explainable, its decisions cannot be mapped directly to the medical evidence, and are therefore harder to understand, and potentially spurious, if inferred from a biased dataset. Nevertheless, machine learning has been shown to have promise in many areas of medical decision making, and we highlight in particular its potential for efficiently inferring new knowledge, which, once validated, could be used in future expert algorithms.

8.6.1 Limitations and future work

In this study we only examined the detection of QT-prolongation at risk of TdP, and it is not clear whether the pseudo-colour heuristics would support ECG interpretation of other abnormalities that may increase the risk of TdP, including electrolyte imbalance (*e.g.* hypokalemia and hypocalcemia) and changes in ST-segment elevation.

This study showed the preliminary results of using human-like perceptual pre-processing of the ECG signal data to facilitate a single decision-tree machine learning algorithm, and future work should explore whether this can aid other machine learning techniques, including computer vision, where pseudo-colouring could be used to improve ECG information segmentation in a pre-processing step. Only a small number of people participated in the focus groups, and further usability evaluations are necessary to determine the utility of the technique in clinical practice and with more diverse clinical populations.

8.7 Conclusion

This study demonstrates that combining a data representation based on human perceptual heuristics with expert clinical knowledge results in accurate, reliable and explainable automated detection of drug-induced QT-prolongation at risk of TdP regardless of heart rate, and robust to T-wave morphology changes. The results indicate that a prolonged ventricular repolarisation area can be a significant risk predictor of TdP, and it is potentially easier and more reliable to automate detection of this than automating the measurement of the QT-interval distance *per se*.

Chapter 9

Summary, Synthesis and Future Work

9.1 Research overview

This thesis explores whether we can improve the early detection and assessment of drug-induced long QT syndrome (diLQTS) at risk of TdP by non-experts in ECG interpretation by incorporating knowledge of human visual perception into the design, development and evaluation of computer-based ECG interpretation approaches. As discussed in the literature review in Chapter 3, research in this area has focused mainly on congenital LQTS; few studies have investigated the use of computer-based ECG interpretation and visualisation methods to support diLQTS detection or TdP risk assessment. No previous work has focused on supporting non-experts in ECG interpretation or using knowledge of visual perception to enhance ECG interpretation.

The research used an exploratory process, investigating the broader context of the research problem with an interdisciplinary review approach across the fields of cardiac physiology, computer science and cognitive psychology (as presented in Chapter 2) to formulate the hypotheses, and using the findings of each study to inform the aims and methods of the next study. In particular, the research drew from the field of pre-attentive processing theory, which is fundamental to how we understand visual distinctiveness [Tre85]. This theory outlines a set of visual properties including colour (hue and intensity), form, and motion, known to be detected rapidly and accurately by the human eye [War21]. Empirical evidence appears to confirm the notion that complex visual image is broken via a primitive perceptual grouping process into these pre-attentive elements [War21]. Using these pre-attentive properties in visualisation design can improve both the effectiveness and efficiency of interpreting complex information by increasing the salience of target patterns [War21, HBE95, HBE96].

9.2 Summary of findings and contributions

Five empirical studies were conducted within this thesis, yielding five publications. The main chapters within this thesis (Chapter 4 to 8) are in the form of research papers. Here, we summarise the main findings of each study, draw out their unique contribution to the overall research problem, and describe how the papers are interrelated.

Chapter 4: The first study used methods from psychophysics and eye-tracking research to systematically examine the ability of non-experts to identify drug-induced QT-interval prolongation on an ECG. It also investigated whether the presentation of the ECG signal trace (as a single complex representing a single heartbeat or a 10-second rhythm strip showing more than one complex/heartbeat) affects interpretation accuracy. The results of this study show that the majority of laypeople were able to perceive a clinically significant difference in the QT-interval length (drug-induced QT $>$ 500ms with a potential risk of TdP) when compared to a 'normal' drug-free ECG baseline, where both ECGs have a heart rate of 60 bpm. It also demonstrated that the rhythm strip, which shows multiple complexes/heartbeats, is a better form of presentation than a single complex, as it is less susceptible to artefacts or changes in the ECG morphology. This study provided the empirical evidence that self-monitoring for diLQTS may be possible, laying the foundation for developing visualisation techniques to further support this process.

Chapter 5: This chapter describes a novel ECG visualisation technique, designed using a science-of-perception-based approach, which introduces pseudo-colour to the ECG to highlight QT-interval duration at a regular heart rate of 60 bpm. It investigated whether there is an interaction effect between the coordinate system (Cartesian *vs.* Polar) and the pseudo-colour on diLQTS detection. Psychophysical and eye-tracking methods were used to systematically evaluate the effectiveness of the proposed visualisation techniques on the detection of variable QT-interval increases. The results of this study show that introducing pseudo-colour significantly improves ECG interpretation accuracy, reduces the reaction time to notice an increase in the QT-interval, and increases laypeople's satisfaction in interpreting the ECG signal. The technique significantly improved non-experts' sensitivity to drug-induced QT-interval increases that are much smaller than a 1mm square on the standard ECG grid (which represents 40ms) when compared to a drug-free 'normal' QT-interval baseline, with the effect being strongest for Polar coordinates, even when T-wave morphology is abnormal.

Chapter 6: This chapter further developed the pseudo-colouring technique (which eliminates the need to measure the QT-interval to identify QT-prolongation) to work at any heart rate, and evaluated whether it could be interpreted without a drug-free 'normal' baseline. An enhanced version of the pseudo-colouring technique was developed to automatically adjust the QT-interval duration at risk of TdP according to heart rate based on a clinically reliable TdP risk assessment method known as 'QT-nomogram' [CIKD07]. The study used a multi-reader, multi-case (MRMC) receiver operating characteristic (ROC) design within a psychophysical paradigm. The results of this study showed that pseudo-colouring significantly improves laypeople's accuracy in detecting QT-prolongation at risk of TdP visually, as well as identifying 'normal' QT-intervals showing no risk of TdP, regardless of heart rate, T-wave morphology and coordinate system, and without comparing the ECG to a drug-free baseline. These results provide further evidence that self-monitoring ECGs for diLQTS is feasible, in particular with the aim of preventing the development of TdP arrhythmia attacks.

Chapter 7: Based on the evidence that pseudo-colouring significantly improves human ECG interpretation accuracy, this chapter explored whether modelling the perceptual heuristics used by humans when interpreting the pseudo-coloured ECG can also improve machine ECG interpretation accuracy. The potential of this new 'human-like' approach to ECG interpretation was investigated in an exploratory study, which developed and evaluated a simple, explainable QT-prolongation detection algorithm that yielded more accurate results than current signal processing techniques, and has the benefit of the human and machine sharing the same representation of the data.

Chapter 8: This chapter further evaluated the effectiveness of the 'human-like' ECG interpretation approach, developing and evaluating an enhanced version of the algorithm to detect diLQTS across a wide range of drug-induced ECG morphological cases, and testing it on a large number of ECGs ($n = 5050$) across multiple patients and QT-prolonging drugs. The study empirically compared two approaches to building the enhanced algorithm: a manually-curated 'expert' algorithm that incorporates knowledge from the clinical literature; and a decision tree, which automates the generation of the rules from the data. The effect of drug type on the algorithm's sensitivity to increases in the QT-interval was also modelled using psychophysical methods. The results of two focus groups—one consisting of patients, the other of clinicians—provided evidence about the acceptability of the approach within clinical practice. The results show that

combining the 'human-like' perceptual heuristics about QT-interval length provided by the pseudo-colour with additional clinical knowledge about T-wave morphology significantly improves the machine ECG interpretation accuracy. Whilst the decision tree is highly accurate when tested using cross-validation, it performs less well on unseen data than the 'expert' algorithm, due to the generation of spurious rules that lack clinical validity. As the 'expert' algorithm interprets the visualised pseudo-coloured ECG, it can be understood intuitively, which the focus groups demonstrated to be important for helping patients to self-monitor diLQTS, as well as ensuring clinicians can validate its results.

From both a clinical and computational point of view, the 'human-like' ECG interpretation algorithm explored in this thesis helps reveal a new clinically reliable approach for automatically detecting LQTS, showing that since the prolonged ventricular repolarisation area under the ECG signal curve (represented by large fused T-U waves) is a significant risk predictor of TdP, detection of this is potentially easier and more reliable to automate than measuring the QT-interval distance directly, by incorporating both human pattern recognition ability and clinical expertise. The heuristic of quantifying area, rather than identifying the interval *per se*, is effective because the T-wave generally has the largest area under the curve of the ECG signal.

Overall, the thesis contributes to both medicine and computer science research, as it demonstrates that it is possible to self-monitor for and automate the detection of a potentially lethal medical condition that has not been possible before. LQTS has challenged many clinicians for a long time, and supporting laypeople in self-monitoring their own ECG for LQTS may ultimately play a role in the secondary prevention of TdP arrhythmia attacks and sudden cardiac death. The thesis also advances the explainability of diagnostic algorithms, as our proposed algorithm is not just technically explainable (*i.e.* an expert human can understand how it works and verify its rules) but also 'intuitively' explainable to members of the public. This demonstrates the potential of a positive feedback loop between understanding how humans interpret data and using this knowledge to inform how machines should do it.

9.3 Synthesis of findings

Most of the studies within this thesis employed methods from psychophysics—a scientific approach to studying sensory capabilities which identifies perceptual responses to changes in physical stimuli [Ste17]—to systematically evaluate the effect of different signal representations on human and machine detection of QT-prolongation that may vary from small increases (few milliseconds) to large increases (more than 100ms). This enabled us to observe similarities and differences between human and machine performance across different signal representation methods. In the evaluation studies with humans, eye-tracking methods were used to contextualise our understanding of the psychophysical performance measures. Using these methods across the different studies allows us to synthesise their results.

An important insight from the first study with laypeople (Chapter 4) was that accurately detecting QT-prolongation requires reading more than one heartbeat (ECG complex), and this finding thus informed the design of all of our subsequent studies, including those focused on algorithm development. It has already been established that the morphology of the ECG signal is highly dynamic, variable across individuals and subject to signal noise (discussed in detail in Chapter 2); the results of our studies show that the ECG morphology can also be variable within an ECG of a single patient and even between successive heartbeats within a single ECG. Such variability affects the accuracy of the QT-interval assessment when interpreting a single heartbeat, and a global accumulated view of the ECG signal across multiple heartbeats provides a more accurate QT assessment for humans (Chapters 4, 5 and 6) and machines (Chapters 7 and 8).

The psychophysical and eye-tracking results of the second and third study (Chapters 5 and 6) show the effectiveness of pseudo-colour in exposing QT-interval duration, even when the interval is borderline and not yet prolonged. This improves on the results of the first study evaluating the standard ECG presentation method without pseudo-colour (Chapter 4), where people were not able to detect borderline QT-intervals, and they only detected a QT-prolongation at potentially dangerous level when comparing with a normal baseline. The borderline QT level is critical from a clinical perspective as just a 40ms increase from this point could put someone at risk of TdP. Clinical research has shown that even small ($\approx$ 10ms) QT-interval increases from the baseline should be regarded as a significant side effect of a QT-prolonging drug [RFF$^+$09, Bre10, FDA$^+$05]. This shows the power of the colour codes used

with the spectrum-approximation pseudo-colouring sequence, where warmer colours including orange and red help to attract attention to the early stages of abnormal QT-interval prolongation.

The studies also showed that pseudo-colour helped to focus visual attention on the T-wave area, which is crucial for detecting QT-prolongation at both a 60 bpm heart rate (Chapter 5) and faster and slower heart rates (Chapter 6). Focusing attention in this area is particularly useful, as long QT syndrome (LQTS) is associated with prolonging ventricular repolarisation [MK05, DSR08], which is represented by the T-wave on the ECG [GMZ06, Sch71]. As research examining the effects of QT-prolonging drugs is investigating precisely how the T-wave responds to them to provide further insights into drug ion channels interactions and TdP risk [VJM$^+$15, CMY08], pseudo-colour might ultimately help clinicians to determine which part of the T-wave—the initial half of the T-wave (J-T-peak) or the second half (T-peak T-end)—underpins the QT-prolongation.

We also found that there is an interaction effect between pseudo-colour, coordinate system and task type. In the second study (Chapter 5), which used a 'two alternative forced choice' (2AFC) psychophysical discrimination task asking the participant to compare two ECGs and select the ECG with a longer QT-interval, the coordinate system interacts with pseudo-colour, such that people were most accurate in detecting the condition where the ECG was presented using polar coordinates and pseudo-colour, and least accurate when presentation occurred with polar coordinates and no colour. Eye-tracking research has shown that people's initial eye movements are more commonly located in the center of the screen [Bin10]. According to the study's eye-tracking data, the warmer hues of the pseudo-colour helped to focus visual attention; as Polar coordinates concentrate more colour in the center of the screen than Cartesian coordinates, the increased salience may be easier to perceive in foveal vision. However, this interaction effect was not observed when we changed the task type to a 'one alternative forced choice' (1AFC) presenting a single ECG in the third study (Chapter 6), as there was no statistically significant difference in people's detection accuracy between the Cartesian and Polar coordinates when pseudo-colour was used. This may be due to the fact that people have to make a more careful judgement when reading a single pseudo-coloured ECG, where perceptual processes of colour are likely to differ from the 'pop-out' colour comparison between two pseudo-coloured ECGs. Note here we do not assume the heart rate has any additional effect, as the results show that pseudo-colour helps improve detection accuracy regardless of heart rate.

A further interesting insight is the value of using the dual-process perception model to enhance machine interpretation. Human perception, broadly speaking, involves two forms of processing: bottom-up processing, which is driven by incoming stimuli, and uses perceptual organisation to form a representation of an object; and top-down processing, which uses contextual information to aid the perception of patterns [EY97, CEY04, FAGB06]. The pseudo-colour technique works by harnessing bottom-up processing, drawing visual attention to the critical information contained within the ECG signal. The fourth study (Chapter 7) shows that using a 'human-like' model of this bottom-up processing can improve automated ECG interpretation accuracy. However, when testing this on a wide range of ECG cases, where the T-wave morphology is substantially affected, the fifth study (Chapter 8) demonstrated that the 'human-like' top-down processing that used contextual clinical knowledge, along with the 'human-like' bottom-up processing provided by the pseudo-colour, provided more accurate automated decisions.

9.4 Limitations and recommendations for future work

In this section, we discuss potential avenues for extending the work presented within this thesis, and articulate limitations of the current approach, as follows:

Applicability to wearable ECG devices: In all studies within this thesis, the ECG data were from a high-quality signal with little noise; they were acquired from a 12-lead ECG, not a wearable ECG device, where the signal is much more likely to be affected by noise. An interesting future research area is the applicability of this technique to wearable ECG devices, in terms of both the challenge of working with the signal data, and considering the practical aspects of ECG self-monitoring. Research questions include: how can this technique be developed and evaluated when the specifications of the wearable devices differ significantly from the 12-lead ECG devices, including number and type of leads (such as limbs vs chest leads), signal quality and accessibility to raw data?; from a clinical perspective, will applying this technique to wearables also improve LQTS detection and provide clinically reliable results?

One of the studies (Chapter 6) showed that Cartesian and Polar coordinates support the same level of accuracy for detecting QT-prolongation across several heart rates, but people expressed a preference for pseudo-colour displayed on Cartesian coordinates. Under certain circumstances, however, the ECG trace may be better presented with

Polar coordinates, for example on smaller screens like wearable smart watches, and future work should thus examine the effects of screen size and lighting setting on ECG interpretation accuracy.

Clinical evaluation of the technique: Whilst the research within this thesis demonstrates the potential for pseudo-colouring to assist in the self-monitoring of drug-induced QT-prolongation, all our studies with laypeople took place in a controlled laboratory setting with a limited number of ECG stimuli, and the transferability of the technique to practice remains an open question. An interesting area of research is comparing the performance of laypeople across different categories (*e.g.* patient *vs.* non-patient), as well as across different demographics (*e.g.* understanding the effect of age on colours perception within the ECG signal). Research has shown that a large number of clinicians lack the skills to interpret QT-prolongation accurately [VRS$^+$05, STN08]; the pseudo-coloured ECG, and the explainable algorithm that interprets it, could also be used to help clinicians, especially within emergency departments, to assess and monitor patients' QT-intervals before or during the provision of a QT-prolonging medication. Future work should evaluate our proposed ECG interpretation approach in clinical trials with more diverse clinical populations.

Future research should adopt a systematic approach for transferring the technique to clinical practice. A good example is adopting the Technology Readiness Level (TRL) approach, which includes one to nine levels showing the technology maturity roadmap, with nine being the most ready-to-launch technology [Str15]. The research within this thesis contributes to the transfer of technique to the third level, which is the 'Proof-of-Concept Demonstrated, Analytically and/or Experimentally'.

Generalizability to other ECG abnormalities: ECG interpretation primarily involves two tasks: (1) measuring the duration and amplitude of different waves and intervals (known to be challenging); and (2) recognition of abnormal signal patterns (acquired by expertise). The research in this thesis aimed to combine these two tasks into one perceptual task by substituting the measurement of the QT-interval itself with the perception of areas of colour, an 'intuitive' pattern recognition task requiring no prior expertise. This thesis only examined the perception of QT-interval prolongation. Whilst the proposed approach may generalise to the interpretation of other ECG abnormalities, such as changes in the ST-segment (*e.g.* representing the ST-depression with

a pseudo-colour sequence that progressively changes from blue to red as the increasing ST-elevation indicates a potential heart attack), confirming this requires further research.

The work within this thesis focused on dynamic changes in QT-interval and heart rate to assess the risk of TdP, and it is not clear whether the pseudo-colour heuristics would also support human-machine ECG interpretation of other abnormalities that may increase the risk of TdP, such as electrolyte imbalance (*e.g.* hypokalemia and hypocalcemia), which are also characterised by specific changes in the T-wave morphology. Pseudo-colour represents an ECG area (*i.e.* representing the duration and amplitude of the ECG signal), and it thus has the potential to also indicate the amplitude and morphology of the T-wave, a potential avenue for future research.

Evaluation of the science-of-perception-based engineering approach in developing and evaluating clinical diagnostic algorithms: Chapter 7 hypothesised that the 'human-like' algorithm development approach explored in this thesis, which exploits an understanding of human perception, has the potential to aid the design of other biologically-inspired algorithms such as deep Convolutional Neural Networks (CNN). For example, using pseudo-colouring to improve information segmentation in a pre-processing step may help to improve a CNN's accuracy in classifying medical images. In particular, the pre-attentive features that are salient to a human could be prioritised as input to the CNN, reducing the search space. Furthermore, we hypothesise that a human-like approach to the development of diagnostic algorithms—*i.e.* incorporating both human pattern recognition ability and clinical expertise—may have much wider utility.

The field of human visual perception has drawn the curiosity of computer scientists for many years, and has been a great source of inspiration for developing and improving machine perception and artificial intelligence [Top19a, OPN08]. A motivation for our approach was not only improving the accuracy of automated ECG interpretation, but also producing data representations that can be used to provide a transparent, understandable and explainable interpretation that keeps the human in the loop. Further exploration of the potential that human visual perception has for informing human-machine interpretation is a promising avenue for future research.

Bibliography

[AAKR+02] Mark E Anderson, Sana M Al-Khatib, Dan M Roden, Robert M Califf, Duke Clinical Research Institute, et al. Cardiac repolarization: current knowledge, critical gaps, and new approaches to drug development and patient management. *American heart journal*, 144(5):769–781, 2002.

[ADD+21] Alaa Alahmadi, Alan Davies, Katherine Dempsey, Markel Vigo, and Caroline Jay. Human-machine perception of complex signal data. In *Human-Like Machine Intelligence*. Oxford University Press, 2021.

[ADR+19] Alaa Alahmadi, Alan Davies, Jennifer Royle, Markel Vigo, and Caroline Jay. Evaluating the impact of pseudo-colour and coordinate system on the detection of medication-induced ECG changes. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2019.

[ADR+21] Alaa Alahmadi, Alan Davies, Jennifer Royle, Leanna Goodwin, Katharine Cresswell, Zahra Arain, Markel Vigo, and Caroline Jay. An explainable algorithm for detecting drug-induced QT-prolongation at risk of torsades de pointes (TdP) regardless of heart rate and T-wave morphology. *Computers in Biology and Medicine*, 2021.

[ADVJ18] Alaa Alahmadi, Alan Davies, Markel Vigo, and Caroline Jay. Can lay people identify a drug-induced QT-interval prolongation? A psychophysical and eye-tracking experiment examining the ability of non-experts to interpret an ECG. *Journal of the American Medical Informatics Association*, 2018.

[ADVJ20a] Alaa Alahmadi, Alan Davies, Markel Vigo, and Caroline Jay. Pseudo-colouring an ECG enables lay people to detect QT-interval prolongation regardless of heart rate. *PLoS One*, 15(8):e0237854, 2020.

[ADVJ20b] Alaa Alahmadi, Alan Davies, Markel Vigo, and Caroline Jay. Visualising ECGs with Pseudo-colouring to detect LQTS regardless of heart rate. Version 1.0.0. [Computer Software]. `https://doi.org/10.5281/zenodo.3940377`, 2020.

[ADVJ21] Alaa Alahmadi, Alan Davies, Markel Vigo, and Caroline Jay. An explainable algorithm for detecting drug-induced QT-prolongation at risk of torsades de pointes (TdP) regardless of heart rate and T-wave morphology. Version 1.0 [Computer Software]. `https://doi.org/10.5281/zenodo.4415791`, 2021.

[AJB16] Muhammad Adnan, Mike Just, and Lynne Baillie. Investigating time series visualisations to improve the user experience. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 5444–5455. ACM, 2016.

[AN09] Lawrence G Appelbaum and Anthony M Norcia. Attentive and pre-attentive aspects of figural processing. *Journal of Vision*, 9(11):18–18, 2009.

[Ana20] Image Analyst. Simplecolordetection(). `https://www.mathworks.com/matlabcentral/fileexchange/26420-simplecolordetection`, MATLAB Central File Exchange, Retrieved January 15, 2020.

[Ant07] Charles Antzelevitch. Ionic, molecular, and cellular bases of qt-interval prolongation and torsade de pointes. *Europace*, 9(suppl_4):iv4–iv15, 2007.

[Ant19] Charles Antzelevitch. Tpeak-tend interval as a marker of arrhythmic risk in early repolarization syndrome. *Journal of cardiovascular electrophysiology*, 30(10):2106–2107, 2019.

[AORD18] Ashish Atreja, Emamuzo Otobo, Karthik Ramireddy, and Allyssa Deorocki. Remote patient monitoring in ibd: current state and future directions. *Current gastroenterology reports*, 20(2):6, 2018.

[AS18] Adewole S Adamson and Avery Smith. Machine learning and health care disparities in dermatology. *JAMA dermatology*, 154(11):1247–1248, 2018.

[ASS+99] Geoffrey W Abbott, Federico Sesti, Igor Splawski, Marianne E Buck, Michael H Lehmann, Katherine W Timothy, Mark T Keating, and Steve AN Goldstein. Mirp1 forms ikr potassium channels with herg and is associated with cardiac arrhythmia. *Cell*, 97(2):175–187, 1999.

[AT06] P Augustyniak and R Tadeusiewicz. Assessment of electrocardiogram visual interpretation strategy based on scanpath analysis. *Physiological Measurement*, 27(7):597, 2006.

[ATOP14] Anand Ambhore, Swee-Guan Teo, Abdul Razakjr Bin Omar, and Kian-Keong Poh. Ecg series. importance of qt interval in clinical practice. *Singapore medical journal*, 55(12):607, 2014.

[Aug03] P Augustyniak. How a human perceives the electrocardiogram-the pursuit of information distribution through scanpath analysis. In *Computers in Cardiology, 2003*, pages 601–604. IEEE, 2003.

[AVJ20] Alaa Alahmadi, Markel Vigo, and Caroline Jay. The QT-interval ECG Visualisation and the Human-like Algorithm. Version 1.0.0. [Computer Software]. `https://doi.org/10.5281/zenodo.3866622`, 2020.

[B+18] Charles I Berul et al. Acquired long qt syndrome: Definitions, causes, and pathophysiology. *Downloaded from UptoDate at https://www-uptodatecom. proxy. cvtc. edu/contents/acquired-long-qt-syndrome-definitions-causes-andpathophysiology*, 2018.

[BBBD08] Michael Burch, Felix Bott, Fabian Beck, and Stephan Diehl. Cartesian vs. radial–a comparative evaluation of two visualization tools. In *International Symposium on Visual Computing*, pages 151–160. Springer, 2008.

[BBH+84] Jerry L Bauman, Robert A Bauernfeind, Julie V Hoff, Boris Strasberg, Steven Swiryn, and Kenneth M Rosen. Torsade de pointes due to quinidine: observations in 31 patients. *American heart journal*, 107(3):425–430, 1984.

[BELL17] Dominik Bramreiter, Markus Ebner, Philipp Leitner, and Christoph Lipautz. Accessible visualisation. 2017.

[BFN+13] Raymond R Bond, Dewar D Finlay, Chris D Nugent, George Moore, and Daniel Guldenring. Methods for presenting and visualising electrocardiographic data: from temporal signals to spatial imaging. *Journal of electrocardiology*, 46(3):182–196, 2013.

[BHW09] Marc D Binder, Nobutaka Hirokawa, and Uwe Windhorst. *Encyclopedia of neuroscience*, volume 3166. Springer Berlin, Germany, 2009.

[BI15] Ingrid Berling and Geoffrey K Isbister. The half rr rule: a poor rule of thumb and not a risk assessment tool for qt interval prolongation. *Academic Emergency Medicine*, 22(10):1139–1144, 2015.

[Bin10] Markus Bindemann. Scene and screen center bias early eye movements in scene viewing. *Vision research*, 50(23):2577–2587, 2010.

[BK03] Marlene Behrmann and Ruth Kimchi. What does visual agnosia tell us about perceptual organization and its relationship to object perception? *Journal of Experimental Psychology: Human Perception and Performance*, 29(1):19, 2003.

[BNNK18] Gerald D Buckberg, Navin C Nanda, Christopher Nguyen, and Mladen J Kocica. What is the heart? anatomy, function, pathophysiology, and misconceptions. *Journal of cardiovascular development and disease*, 5(2):33, 2018.

[Bor] Hans Werner Borchers. Pracma: practical numerical math functions. *R package version 2.1.8., year=2018*.

[Bor15] Hans Werner Borchers. Pracma: practical numerical math functions. *R package version*, 1(3), 2015.

[Bow83] JK Bowmaker. Trichromatic colour vision: why only three receptor channels? *Trends in Neurosciences*, 6:41–43, 1983.

[BP09] M Borjesson and A Pelliccia. Incidence and aetiology of sudden cardiac death in young athletes: an international perspective. *British journal of sports medicine*, 43(9):644–648, 2009.

[BPR+17] MS Bohnen, G Peng, SH Robey, C Terrenoire, V Iyer, KJ Sampson, and RS Kass. Molecular pathophysiology of congenital long qt syndrome. *Physiological reviews*, 97(1):89–134, 2017.

[BPW+38] Arlie R Barnes, Harold EB Pardee, Paul D White, Frank N Wilson, Charles C Wolferth, et al. Standardization of precordial leads: supplementary report. *American Heart Journal*, 15(2):235–239, 1938.

[BR99] Peter L Bonate and Tanya Russell. Assessment of qtc prolongation for non-cardiac-related drugs from a drug development perspective. *The Journal of Clinical Pharmacology*, 39(4):349–358, 1999.

[Bre10] Joanna M Brell. Prolonged qtc interval in cancer therapeutic drug development: defining arrhythmic risk in malignancy. *Progress in cardiovascular diseases*, 53(2):164–172, 2010.

[Bro04] AM Brown. Drugs, herg and sudden death. *Cell calcium*, 35(6):543–547, 2004.

[BSR+16] Troels N Bachmann, Morten W Skov, Peter V Rasmussen, Claus Graff, Adrian Pietersen, Bent Lind, Johannes J Struijk, Morten S Olesen, Stig Haunsø, Lars Køber, et al. Electrocardiographic tpeak–tend interval and risk of cardiovascular morbidity and mortality: Results from the copenhagen ecg study. *Heart Rhythm*, 13(4):915–924, 2016.

[BSW14] Shashank P Behere, Catherine D Shubkin, and Steven N Weindling. Recent advances in the understanding and management of long qt syndrome. *Current opinion in pediatrics*, 26(6):727–733, 2014.

[BVM97] Hans Brettel, Françoise Viénot, and John D Mollon. Computerized simulation of color appearance for dichromats. *Josa a*, 14(10):2647–2655, 1997.

[BY05] Rong Bai and Gan-Xin Yan. Accurate interpretation of the qt interval: a vital task that remains unaccomplished. *Heart Rhythm*, 2(6):575–577, 2005.

[BZF+14] RR Bond, Tingting Zhu, DD Finlay, B Drew, PD Kligfield, Daniel Guldenring, Cathal Breen, AG Gallagher, MJ Daly, and GD Clifford. Assessing computerized eye tracking technology for gaining insight

into expert interpretation of the 12-lead electrocardiogram: an objective quantitative approach. *Journal of electrocardiology*, 47(6):895–906, 2014.

[CAK07] Shin-Ho Chung, Olaf S Anderson, and Vikram V Krishnamurthy. *Biological membrane ion channels: dynamics, structure, and applications*. Springer Science & Business Media, 2007.

[CAM+06] Gari D Clifford, Francisco Azuaje, Patrick McSharry, et al. *Advanced methods and tools for ECG data analysis*. Artech house Boston, 2006.

[Can13] Virginio Cantoni. *Human and machine vision: analogies and divergencies*. Springer Science & Business Media, 2013.

[CCCK01] Huihua Kenny Chiang, Chao-Wei Chu, Gau-Yang Chen, and Cheng-Deng Kuo. A new 3-d display method for 12-lead ecg. *IEEE transactions on biomedical engineering*, 48(10):1195–1202, 2001.

[CCDS08] Lia Crotti, Giuseppe Celano, Federica Dagradi, and Peter J Schwartz. Congenital long qt syndrome. *Orphanet journal of rare diseases*, 3(1):1–16, 2008.

[CEY04] Charles E Connor, Howard E Egeth, and Steven Yantis. Visual attention: bottom-up versus top-down. *Current biology*, 14(19):R850–R852, 2004.

[CHA+15] Ehud Chorin, Ofer Havakuk, Arnon Adler, Arie Steinvil, Uri Rozovski, Christian van der Werf, Pieter G Postema, Guy Topaz, Arthur AM Wilde, Sami Viskin, et al. Diagnostic value of t-wave morphology changes during 'qt stretching' in patients with long qt syndrome. *Heart Rhythm*, 12(11):2263–2271, 2015.

[CIKD07] A Chan, GK Isbister, CMJ Kirkpatrick, and SB Dufful. Drug-induced qt prolongation and torsades de pointes: evaluation of a qt nomogram. *QJM: An International Journal of Medicine*, 100(10):609–615, 2007.

[CJ20] Davide Chicco and Giuseppe Jurman. The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation. *BMC genomics*, 21(1):6, 2020.

[CKA18] Christopher C Cheung, Andrew D Krahn, and Jason G Andrade. The emerging role of wearable technologies in detection of arrhythmia. *Canadian Journal of Cardiology*, 34(8):1083–1087, 2018.

[CKH+07] JP Couderc, S Kaab, M Hinterseer, S McNitt, X Xia, A Fossa, B Beckmann, S Polonsky, and W Zareba. Investigating the role of ventricular repolarization morphology in surface ecgs for identifying patients with a history of drug-induced arrhythmias. In *2007 Computers in Cardiology*, pages 337–340. IEEE, 2007.

[CM84] William S Cleveland and Robert McGill. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American statistical association*, 79(387):531–554, 1984.

[CM86] William S Cleveland and Robert McGill. An experiment in graphical perception. *International Journal of Man-Machine Studies*, 25(5):491–500, 1986.

[CMY08] A John Camm, Marek Malik, and Yee Guan Yap. *Acquired long QT syndrome*. John Wiley & Sons, 2008.

[CNC15] Kelvin CM Chua, Sandeep G Nair, and Sumeet S Chugh. Risk markers of sudden death on the 12-lead ecg: Tpeak-tend interval makes the cut. *Heart rhythm*, 12(8):1798–1799, 2015.

[CNG06] YC Chesnokov, D Nerukh, and RC Glen. Individually adaptable automatic qt detector. In *2006 Computers in Cardiology*, pages 337–340. IEEE, 2006.

[CNK+15] Min Soo Cho, Gi-Byoung Nam, Yong-Guin Kim, Ki-Won Hwang, Yoo Ri Kim, HyungOh Choi, Sung-Hwan Kim, Kyoung-Suk Rhee, Nam-Joon Kim, June Soo Kim, et al. Electrocardiographic predictors of bradycardia-induced torsades de pointes in patients with acquired atrioventricular block. *Heart Rhythm*, 12(3):498–505, 2015.

[CSDB+09] G Curigliano, G Spitaleri, F De Braud, D Cardinale, C Cipolla, M Civelli, N Colombo, A Colombo, M Locatelli, and A Goldhirsch. Qtc prolongation assessment in anticancer drug development: clinical and methodological issues. *ecancermedicalscience*, 3, 2009.

[CZM05] JP Couderc, W Zareba, and AJ Moss. Discrimination of herg carrier from non-carrier adult patients with borderline prolonged qtc interval. In *Computers in Cardiology, 2005*, pages 125–128. IEEE, 2005.

[CZSZ08] Jean-Philippe Couderc, Meijian Zhou, Nenad Sarapa, and Wojciech Zareba. Investigating the effect of sotalol on the repolarization intervals in healthy young individuals. *Journal of Electrocardiology*, 41(6):595–602, 2008.

[CZY+20] Cheng Chen, Jiandong Zhou, Haixu Yu, Qingpeng Zhang, Lianjun Gao, Xiaomeng Yin, Yingxue Dong, Yajuan Lin, Daobo Li, Yiheng Yang, et al. Identification of important risk factors for all-cause mortality of acquired long qt syndrome patients using random survival forests and non-negative matrix factorization. *Heart Rhythm*, 2020.

[DAF+10] Barbara J Drew, Michael J Ackerman, Marjorie Funk, W Brian Gibler, Paul Kligfield, Venu Menon, George J Philippides, Dan M Roden, and Wojciech Zareba. Prevention of torsade de pointes in hospital settings: a scientific statement from the american heart association and the american college of cardiology foundation. *Circulation*, 121(8):1047–1060, 2010.

[Dar01] B Darpö. Spectrum of drugs prolonging qt interval and the incidence of torsades de pointes. *European Heart Journal Supplements*, 3(suppl_K):K70–K80, 2001.

[DB98] Sara H Downs and Nick Black. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *Journal of Epidemiology & Community Health*, 52(6):377–384, 1998.

[DB14] Alok Deshpande and Yochai Birnbaum. St-segment elevation: distinguishing st elevation myocardial infarction from st elevation secondary to nonischemic etiologies. *World journal of cardiology*, 6(10):1067, 2014.

[DBB10] Stephan Diehl, Fabian Beck, and Michael Burch. Uncovering

strengths and weaknesses of radial visualizations—an empirical approach. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):935–942, 2010.

[DBDM20] J De Bie, I Diemberger, and JW Mason. Comparison of pr, qrs, and qt interval measurements by seven ecg interpretation programs. *Journal of Electrocardiology*, 63:75–82, 2020.

[DBM92] Donald D Dorfman, Kevin S Berbaum, and Charles E Metz. Receiver operating characteristic rating analysis: generalization to the population of readers and patients with the jackknife method. *Investigative radiology*, 27(9):723–731, 1992.

[DBV+16] Alan Davies, Gavin Brown, Markel Vigo, Simon Harper, Laura Horseman, Bruno Splendiani, Elspeth Hill, and Caroline Jay. Exploring the relationship between eye movements and electrocardiogram interpretation accuracy. *Scientific reports*, 6:38227, 2016.

[DCKH+19] Alan Davies, Marisa Cunha, Kamilla Kopec-Harding, Paul Metcalfe, James Weatherall, and Caroline Jay. Biomarker data visualisation for decision making in clinical trials. *International journal of medical informatics*, 132:104008, 2019.

[DDB+97] Claire Donger, Isabelle Denjoy, Myriam Berthet, Nathalie Neyroud, Corinne Cruaud, Mohammed Bennaceur, Guy Chivoret, Ketty Schwartz, Philippe Coumel, and Pascale Guicheney. Kvlqt1 c-terminal missense mutation causes a forme fruste long-qt syndrome. *Circulation*, 96(9):2778–2781, 1997.

[Des66] Francois Dessertenne. La tachycardie ventriculaire a deux foyers opposes variables. *Arch Mal Coeur*, 59:263–272, 1966.

[Dir83] GR Dirkin. Cognitive tunneling: Use of visual information under stress. *Perceptual and Motor Skills*, 56(1):191–198, 1983.

[DLD+03] Mehul Desai, Lang Li, Zeruesenay Desta, Marek Malik, and David Flockhart. Variability of heart rate correction methods for the qt interval. *British journal of clinical pharmacology*, 55(6):511–517, 2003.

[DLRH19] Abdellaziz Dahou, Dmitry Levin, Mark Reisman, and Rebecca T Hahn. Anatomy and physiology of the tricuspid valve. *JACC: Cardiovascular Imaging*, 12(3):458–468, 2019.

[DmFBR+96] Jean-Louis De´ molis, Christian Funck-Brentano, Jacques Ropers, Mathieu Ghadanfar, Donald J Nichols, and Patrice Jaillon. Influence of dofetilide on qt-interval duration and dispersion at various heart rates during exercise in humans. *Circulation*, 94(7):1592–1599, 1996.

[DMH+19] Alan Davies, Julia Mueller, Laura Horseman, Bruno Splendiani, Elspeth Hill, Markel Vigo, Simon Harper, and Caroline Jay. How do healthcare practitioners read electrocardiograms? a dual-process model of electrocardiogram interpretation. *British Journal of Cardiac Nursing*, 14(10):1–19, 2019.

[DNS06] Borje Darpo, Thierry Nebout, and Philip T Sager. Clinical evaluation of qt/qtc prolongation and proarrhythmic potential for nonantiarrhythmic drugs: the international conference on harmonization of technical requirements for registration of pharmaceuticals for human use e14 guideline. *The Journal of Clinical Pharmacology*, 46(5):498–507, 2006.

[DS14] Alan Davies and Alwyn Scott. *Starting to Read ECGs*. Springer, 2014.

[DSP90] Trevor Darrell, Stan Sclaroff, and Alex Pentland. Segmentation by minimal description. In *[1990] Proceedings Third International Conference on Computer Vision*, pages 112–116. IEEE, 1990.

[DSR08] John D Hunter DICM, Prashast Sharma, and Sarika Rathi. Long qt syndrome. *Continuing Education in Anaesthesia, Critical Care & Pain*, 8(2), 2008.

[Dub00] Dale Dubin. *Rapid interpretation of EKG's: an interactive course*. Cover Publishing Company, 2000.

[Duc07] Andrew T Duchowski. Eye tracking methodology. *Theory and practice*, 328, 2007.

[DVDL14] Nicolas Debue and Cécile Van De Leemput. What does germane load mean? an empirical contribution to the cognitive load theory. *Frontiers in psychology*, 5:1099, 2014.

[DVHJ16] Alan Davies, Markel Vigo, Simon Harper, and Caroline Jay. The visualisation of eye-tracking scanpaths: what can they tell us about how clinicians view electrocardiograms? In *2016 IEEE Second Workshop on Eye Tracking and Visualization (ETVIS)*, pages 79–83. IEEE, 2016.

[EBB04] Mark J Eisenberg, Anya Brox, and Alain N Bestawros. Calcium channel blockers: an update. *The American journal of medicine*, 116(1):35–43, 2004.

[ECS16] Sven Eberhardt, Jonah G Cader, and Thomas Serre. How deep is the feature analysis underlying rapid visual categorization? In *Advances in neural information processing systems*, pages 1100–1108, 2016.

[EI13] NA Mark Estes III. Computerized interpretation of ecgs: supplement not a substitute, 2013.

[ERR+19] Andre Esteva, Alexandre Robicquet, Bharath Ramsundar, Volodymyr Kuleshov, Mark DePristo, Katherine Chou, Claire Cui, Greg Corrado, Sebastian Thrun, and Jeff Dean. A guide to deep learning in healthcare. *Nature medicine*, 25(1):24–29, 2019.

[EsBH89] Nabil El-sherif, Soad-Saad Bekheit, and Raphael Henkin. Quinidine-induced long qtu interval and torsade de pointes: role of bradycardia-dependent early afterdepolarizations. *Journal of the American College of Cardiology*, 14(1):252–257, 1989.

[ESTB20] Nabil El-Sherif, Gioia Turitto, and Mohamed Boutjdir. Acquired long qt syndrome and electrophysiology of torsade de pointes. In *Cardiac Repolarization*, pages 201–216. Springer, 2020.

[EY97] Howard E Egeth and Steven Yantis. Visual attention: Control, representation, and time course. *Annual review of psychology*, 48(1):269–297, 1997.

[FA+19a] Food, Drug Administration, et al. Proposed regulatory framework for modifications to artificial intelligence/machine learning (ai/ml)-based software as a medical device (samd)-discussion paper, 2019.

[FA19b] Christopher Foth and Heidi Alvey. Ventricular tachycardia (vt, v tach). In *StatPearls [Internet]*. StatPearls Publishing, 2019.

[FAGB06] Mark J Fenske, Elissa Aminoff, Nurit Gronau, and Moshe Bar. Top-down facilitation of visual object recognition: object-based and context-based contributions. *Progress in brain research*, 155:3–21, 2006.

[FBG+19] Maarten Falter, Werner Budts, Kaatje Goetschalckx, Véronique Cornelissen, and Roselien Buys. Accuracy of apple watch measurements for heart rate and energy expenditure in patients with cardiovascular disease: Cross-sectional study. *JMIR mHealth and uHealth*, 7(3):e11889, 2019.

[FDA+05] Food, HHS Drug Administration, et al. International conference on harmonisation; guidance on e14 clinical evaluation of qt/qtc interval prolongation and proarrhythmic potential for non-antiarrhythmic drugs; availability. notice. *Federal register*, 70(202):61134, 2005.

[Fel97] Jacob Feldman. Regularity-based perceptual grouping. *Computational Intelligence*, 13(4):582–623, 1997.

[Fel16] Jacob Feldman. The simplicity principle in perception and cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 7(5):330–340, 2016.

[FFM+13] Johannes Fuchs, Fabian Fischer, Florian Mansmann, Enrico Bertini, and Petra Isenberg. Evaluation of alternative glyph designs for time series data in a small multiple setting. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 3237–3246, 2013.

[FG01] John M Findlay and Iain D Gilchrist. Visual attention: The active vision perspective. In *Vision and attention*, pages 83–103. Springer, 2001.

[FNY+13] Takahide Fujii, Masanao Nakano, Ken Yamashita, Toshihiro Konishi, Shintaro Izumi, Hiroshi Kawaguchi, and Masahiko Yoshimoto. Noise-tolerant instantaneous heart rate and r-peak detection using short-term autocorrelation for wearable healthcare systems. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 7330–7333. IEEE, 2013.

[fPMP+97] Committee for Proprietary Medicinal Products et al. Points to consider: The assessment of the potential for qt interval prolongation by non-cardivascular medicinal products. *CPMP/986/96*, 1997.

[Fra56] Ernest Frank. An accurate, clinically practical system for spatial vectorcardiography. *circulation*, 13(5):737–749, 1956.

[FSB18] Matthew Fenech, Nika Strukelj, and Olly Buston. Ethical, Social, and Political Challenges of Artificial Intelligence in Health. Technical report, Future Advocacy, 2018.

[FWM+05] Anthony A Fossa, Todd Wisialowski, Anthony Magnano, Eric Wolfgang, Roxanne Winslow, William Gorczyca, Kimberly Crimin, and David L Raunig. Dynamic beat-to-beat modeling of the qt-rr interval relationship: analysis of qt prolongation during alterations of autonomic state versus human ether a-go-go-related gene inhibition. *Journal of Pharmacology and Experimental Therapeutics*, 312(1):1–11, 2005.

[GAG+00] Ary L Goldberger, Luis AN Amaral, Leon Glass, Jeffrey M Hausdorff, Plamen Ch Ivanov, Roger G Mark, Joseph E Mietus, George B Moody, Chung-Kang Peng, and H Eugene Stanley. Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000.

[GCC17] Carl Gutwin, Andy Cockburn, and Ashley Coveney. Peripheral popout: The influence of visual angle and stimulus intensity on popout effects. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 208–219. ACM, 2017.

[GEB18] Anna Garcia-Elias and Begoña Benito. Ion channel disorders and sudden cardiac death. *International journal of molecular sciences*, 19(3):692, 2018.

[Gib79] James J Gibson. The theory of affordances. the ecological approach to visual perception, 1979.

[GJS+17] Robert Geirhos, David HJ Janssen, Heiko H Schütt, Jonas Rauber, Matthias Bethge, and Felix A Wichmann. Comparing deep neural networks against humans: object recognition when the signal gets weaker. *arXiv preprint arXiv:1706.06969*, 2017.

[GK13] Gösta H Granlund and Hans Knutsson. *Signal processing for computer vision*. Springer Science & Business Media, 2013.

[GL13] Anubhav Garg and Michael H Lehmann. Prolonged qt interval diagnosis suppression by a widely used computerized ecg analysis system. *Circulation: Arrhythmia and Electrophysiology*, 6(1):76–83, 2013.

[GM08] Ilan Goldenberg and Arthur J Moss. Long qt syndrome. *Journal of the American College of Cardiology*, 51(24):2291–2300, 2008.

[GMZ06] ILAN Goldenberg, Arthur J Moss, and Wojciech Zareba. Qt interval: how to measure it and what is "normal". *Journal of cardiovascular electrophysiology*, 17(3):333–336, 2006.

[GPAW14] Pieter G Postema and Arthur AM Wilde. The measurement of the qt interval. *Current cardiology reviews*, 10(3):287–294, 2014.

[GPBC10] Gaspare Galati, Gina Pelle, Alain Berthoz, and Giorgia Committeri. Multiple reference frames used by the human brain for spatial perception and memory. *Experimental brain research*, 206(2):109–120, 2010.

[GPR16] Andrew T Gomez, Jordan M Prutkin, and Ashwin L Rao. Evaluation and management of athletes with long qt syndrome: An evolved paradigm. *Sports health*, 8(6):527–535, 2016.

[GPS04] FG Gascon and D Peralta-Salas. On the construction of global coordinate systems in euclidean spaces. *Nonlinear Analysis: Theory, Methods & Applications*, 57(5-6):723–742, 2004.

[GSS15] Ian J. Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and Harnessing Adversarial Examples. In *ICLR*, pages 1–11, 2015.

[GWB13] Holly E Gerhard, Felix A Wichmann, and Matthias Bethge. How sensitive is the human visual system to the local statistics of natural images? *PLoS computational biology*, 9(1), 2013.

[GWY+11] Jun Guo, Tingzhong Wang, Tonghua Yang, Jianmin Xu, Wentao Li, Michael D Fridman, John T Fisher, and Shetuan Zhang. Interaction between the cardiac rapidly (ikr) and slowly (iks) activating delayed rectifier potassium channels revealed by low k+-induced herg endocytic degradation. *Journal of Biological Chemistry*, 286(40):34664–34674, 2011.

[HBE95] Christopher G Healey, Kellogg S Booth, and James T Enns. Visualizing real-time multivariate data using preattentive processing. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 5(3):190–221, 1995.

[HBE96] Christopher G Healey, Kellogg S Booth, and James T Enns. High-speed visual estimation using preattentive processing. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 3(2):107–135, 1996.

[HBM08] Stephen L Hillis, Kevin S Berbaum, and Charles E Metz. Recent developments in the dorfman-berbaum-metz procedure for multireader roc study analysis. *Academic radiology*, 15(5):647–661, 2008.

[HBPK17] Andreas Holzinger, Chris Biemann, Constantinos S Pattichis, and Douglas B Kell. What do we need to build explainable AI systems for the medical domain? *arXiv preprint arXiv:1712.09923*, 2017.

[HBV+20] Ben JM Hermans, Frank C Bennis, Arja S Vink, Tijmen Koopsen, Aurore Lyon, Arthur AM Wilde, Dieter Nuyens, Tomas Robyns, Laurent Pison, Pieter G Postema, et al. Improving long qt syndrome diagnosis by a polynomial-based t-wave morphology characterization. *Heart Rhythm*, 17(5):752–758, 2020.

[HC94] PD Higham and RW Campbell. Qt dispersion. *British heart journal*, 71(6):508, 1994.

[HCLF12] Howard C Hughes, Gideon Paul Caplovitz, Rebecca A Loucks, and Robert Fendrich. Attentive and pre-attentive processes in change detection and identification. *PloS one*, 7(8):e42851, 2012.

[HE85] Gary Hatfield and William Epstein. The status of the minimum principle in the theoretical analysis of visual perception. *Psychological Bulletin*, 97(2):155, 1985.

[Hea96] Christopher G Healey. Choosing effective colours for data visualization. In *Visualization'96. Proceedings.*, pages 263–270. IEEE, 1996.

[Hen92] John M Henderson. Visual attention and eye movement control during reading and picture viewing. In *Eye movements and visual cognition*, pages 260–283. Springer, 1992.

[HES92] MOHAMMAD A HABBAB and NABIL EL-SHERIF. Tu alternans, long qtu, and torsade de pointes: clinical and experimental observations. *Pacing and Clinical Electrophysiology*, 15(6):916–931, 1992.

[HG06] Richard H Hongo and Nora Goldschlager. Status of computerized electrocardiography. *Cardiology clinics*, 24(3):491–504, 2006.

[HGM+97] Patrick Hoffman, Georges Grinstein, Kenneth Marx, Ivo Grosse, and Eugene Stanley. Dna visual and analytic data mining. In *Visualization'97., Proceedings*, pages 437–441. IEEE, 1997.

[Hil07] Stephen L Hillis. A comparison of denominator degrees of freedom methods for multiple observer roc analysis. *Statistics in medicine*, 26(3):596–619, 2007.

[HLCL12] Min Huang, Haoxue Liu, Guihua Cui, and M Ronnier Luo. Testing uniform colour spaces and colour-difference formulae using printed samples. *Color Research & Application*, 37(5):326–335, 2012.

[HNA+11] Kenneth Holmqvist, Marcus Nyström, Richard Andersson, Richard Dewhurst, Halszka Jarodzka, and Joost Van de Weijer. *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford, 2011.

[Hoc57] Julian E Hochberg. Effects of the gestalt revolution: The cornell symposium on perception. *Psychological Review*, 64(2):73, 1957.

[HOSB05] Stephen L Hillis, Nancy A Obuchowski, Kevin M Schartz, and Kevin S Berbaum. A comparison of the dorfman–berbaum–metz and obuchowski–rockette methods for receiver operating characteristic (roc) data. *Statistics in medicine*, 24(10):1579–1607, 2005.

[HRK19] Habib Hajimolahoseini, Damian Redfearn, and Andrew Krahn. A deep learning approach for diagnosing long qt syndrome without measuring qt interval. In *Canadian Conference on Artificial Intelligence*, pages 440–445. Springer, 2019.

[HS18] Stephen L Hillis and Kevin M Schartz. Multireader sample size program for diagnostic studies: demonstration and methodology. *Journal of Medical Imaging*, 5(4):045503, 2018.

[HSB+18] Ben JM Hermans, Job Stoks, Frank C Bennis, Arja S Vink, Ainara Garde, Arthur AM Wilde, Laurent Pison, Pieter G Postema, and Tammo Delhaas. Support vector machine-based assessment of the t-wave morphology improves long qt syndrome diagnosis. *EP Europace*, 20(suppl_3):iii113–iii119, 2018.

[HSSB98] Mike Heath, Sudeep Sarkar, Thomas Sanocki, and Kevin Bowyer. Comparison of edge detectors: a methodology and initial study. *Computer vision and image understanding*, 69(1):38–54, 1998.

[HST+16] Katerina Hnatkova, Peter Smetana, Ondrej Toman, Georg Schmidt, and Marek Malik. Sex and race differences in qrs duration. *Ep Europace*, 18(12):1842–1849, 2016.

[HTWW07] Jia-Bin Huang, Yu-Cheng Tseng, Se-In Wu, and Sheng-Jyh Wang. Information preserving color transformation for protanopia and deuteranopia. *IEEE Signal Processing Letters*, 14(10):711–714, 2007.

[Hun05] Anthony Charles Hunt. Accuracy of popular automatic QT interval algorithms assessed by a'gold standard'and comparison with a novel method: computer simulation study. *BMC cardiovascular disorders*, 5(1):29, 2005.

[Hut00] Edwin Hutchins. Distributed cognition. *International Encyclopedia of the Social and Behavioral Sciences. Elsevier Science*, 138, 2000.

[HVB+17] Ben JM Hermans, Arja S Vink, Frank C Bennis, Luc H Filippini, Veronique MF Meijborg, Arthur AM Wilde, Laurent Pison, Pieter G Postema, and Tammo Delhaas. The development and validation of an easy to use automatic QT-interval algorithm. *PloS one*, 12(9):e0184352, 2017.

[HW97] William R Hendee and Peter NT Wells. *The perception of visual information*. Springer Science & Business Media, 1997.

[IG96] Ross Ihaka and Robert Gentleman. R: a language for data analysis and graphics. *Journal of computational and graphical statistics*, 5(3):299–314, 1996.

[IMZ12] Ziad Issa, John M Miller, and Douglas P Zipes. *Clinical Arrhythmology and Electrophysiology: A Companion to Braunwald's Heart Disease E-Book: Expert Consult: Online and Print*. Elsevier Health Sciences, 2012.

[IP13] Geoffrey K Isbister and Colin B Page. Drug induced qt prolongation: the measurement and assessment of the qt interval in clinical practice. *British journal of clinical pharmacology*, 76(1):48–57, 2013.

[IRHH98] RH Ireland, RTCE Robinson, SR Heller, and ND Harris. Qt measurement for tu fused ecg morphology as exhibited during hypoglycaemia. In *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Vol. 20 Biomedical Engineering Towards the Year 2000 and Beyond (Cat. No. 98CH36286)*, pages 240–242. IEEE, 1998.

[ITD+20] Praveen Indraratna, Daniel Tardo, Madeline Delves, Richard Szirt, and Ben Ng. Measurement and management of qt interval prolongation for general physicians. *Journal of general internal medicine*, 35(3):865–873, 2020.

[JA09] Jonathan N Johnson and Michael J Ackerman. Qtc: how long is too long? *British journal of sports medicine*, 43(9):657–662, 2009.

[JA13] Jonathan N Johnson and Michael J Ackerman. Return to play? athletes with congenital long qt syndrome. *British journal of sports medicine*, 47(1):28–33, 2013.

[JBCW91] James F Juola, Don G Bouwhuis, Eric E Cooper, and C Bruce Warner. Control of attention around the fovea. *Journal of Experimental Psychology: Human Perception and Performance*, 17(1):125, 1991.

[JG14] Abhishek Jaiswal and Seth Goldbarg. Dofetilide induced torsade de pointes: mechanism, risk factors and management strategies. *Indian heart journal*, 66(6):640–648, 2014.

[JM19] I Joshi and J Morley. Artificial intelligence: How to get it right. putting policy into practice for safe data-driven innovation in health and care. 2019.

[JVM+14] Lars Johannesen, Jose Vicente, JW Mason, Carlos Sanabria, Kristin Waite-Labott, Mira Hong, Ping Guo, John Lin, Jens Stampe Sørensen, Loriano Galeotti, et al. Differentiating drug-induced multichannel block on the electrocardiogram: randomized study of dofetilide, quinidine, ranolazine, and verapamil. *Clinical Pharmacology & Therapeutics*, 96(5):549–558, 2014.

[Kah11] Daniel Kahneman. *Thinking, fast and slow*. Farrar, Straus and Giroux, New York, 2011.

[KAL20] Amit Kaushal, Russ Altman, and Curt Langlotz. Geographic distribution of us cohorts used to train deep learning algorithms. *JAMA*, 324(12):1212–1213, 2020.

[Kat10] Arnold M Katz. *Physiology of the Heart*. Lippincott Williams & Wilkins, 2010.

[KBD+18] Paul Kligfield, Fabio Badilini, Isabelle Denjoy, Saeed Babaeizadeh, Elaine Clark, Johan De Bie, Brian Devine, Fabrice Extramiana, Gianluca Generali, Richard Gregg, et al. Comparison of automated interval measurements by widely used algorithms in digital electrocardiographs. *American heart journal*, 200:1–10, 2018.

[KBM20] Anthony H Kashou, Hajira Basit, and Ahmad Malik. St segment. *StatPearls [Internet]*, 2020.

[KBR+14] Paul Kligfield, Fabio Badilini, Ian Rowlandson, Joel Xue, Elaine Clark, Brian Devine, Peter Macfarlane, Johan de Bie, David Mortara, Saeed Babaeizadeh, et al. Comparison of automated measurements of electrocardiographic intervals and durations by computer-based algorithms of digital electrocardiographs. *American heart journal*, 167(2):150–159, 2014.

[KBS+13] Tomasz Kameczura, Leszek Bryniarski, Sławomir Surowiec, Maryla Kocowska, Kalina Kawecka-Jaszcz, and Danuta Czarnecka. Myocardial infarction caused by pharmacological substances–case description and literature review. *Postępy w Kardiologii Interwencyjnej= Advances in Interventional Cardiology*, 9(3):250, 2013.

[KE14] Peter Y Kim and Michael S Ewer. Chemotherapy and qt prolongation: overview with clinical perspective. *Current treatment options in cardiovascular medicine*, 16(5):303, 2014.

[KFBW09] Paulus Kirchhof, Michael R Franz, Abdennasser Bardai, and Arthur M Wilde. Giant t–u waves precede torsades de pointes in long qt syndrome: A systematic electrocardiographic analysis in patients with acquired and congenital qt prolongation. *Journal of the American College of Cardiology*, 54(2):143–149, 2009.

[KGK+96] Josef Kautzner, Yi Gang, Ravi Kishore, Xavier Copie, Tomas Janota, Hirokazu Nagayoshi, A John Camm, and Marek Malik. Interobserver reproducibility of qt interval measurement and qt dispersion in patients after acute myocardial infarction. *Annals of Noninvasive Electrocardiology*, 1(4):363–374, 1996.

[KGS+12] Eleftherios M Kallergis, Christos A Goudis, Emmanuel N Simantirakis, George E Kochiadakis, and Panos E Vardas. Mechanisms, risk factors, and management of acquired long qt syndrome: a comprehensive review. *The Scientific World Journal*, 2012, 2012.

[Kha02] Ijaz A Khan. Clinical and therapeutic aspects of congenital and acquired long qt syndrome. *The American journal of medicine*, 112(1):58–66, 2002.

[Kha08] M Gabriel Khan. *Rapid ECG interpretation*. Springer Science & Business Media, 2008.

[KHBP05] Ruth Kimchi, Batsheva Hadad, Marlene Behrmann, and Stephen E Palmer. Microgenesis and ontogenesis of perceptual organization: Evidence from global and local processing of hierarchical patterns. *Psychological science*, 16(4):282–290, 2005.

[Kir16] Andy Kirk. *Data visualisation: A handbook for data driven design*. Sage, 2016.

[KK09] Sameh N Khouzam and Rami N Khouzam. Long qt syndrome misdiagnosed and mistreated as a seizure disorder for eight years. *The Canadian journal of cardiology*, 25(3):166, 2009.

[KMS00] Ron Kimmel, Ravi Malladi, and Nir Sochen. Images as embedded maps and minimal surfaces: movies, color, texture, and volumetric medical images. *International Journal of Computer Vision*, 39(2):111–129, 2000.

[KMSPB] Stephen L. Hillis Kevin M. Schartz, Lorenzo L. Pesce, and Kevin S. Berbaum. Or-dbm mrmc 2.51, medical image perception laboratory department of radiology, the university of iowa. `https://perception.lab.uiowa.edu/OR-DBM-MRMC-program-manual`.

[KNK+10] Sayaka Kurokawa, Shinichi Niwano, Michiro Kiryu, Masami Murakami, Shoko Ishikawa, Yoshihiro Yumoto, Masahiko Moriguchi, Hiroe Niwano, Tomoko Kosukegawa, and Tohru Izumi. Importance of morphological changes in tu waves during bepridil therapy as a predictor of ventricular arrhythmic event. *Circulation Journal*, 74(5):876–884, 2010.

[KOF08] Giovane R Kuhn, Manuel M Oliveira, and Leandro AF Fernandes. An efficient naturalness-preserving image-recoloring method for dichromats. *IEEE transactions on visualization and computer graphics*, 14(6):1747–1754, 2008.

[KOO+11] Yuji Kasamaki, Yukio Ozawa, Masakatsu Ohta, Akira Sezai, Takashi Yamaki, Mutsuo Kaneko, Ichiro Watanabe, Atsushi Hirayama, and

Tomohiro Nakayama. Automated versus manual measurement of the qt interval and corrected qt interval. *Annals of Noninvasive Electrocardiology*, 16(2):156–164, 2011.

[KRA04] Ruth Kimchi and Irene Razpurker-Apfeld. Perceptual grouping and attention: Not all groupings are equal. *Psychonomic Bulletin & Review*, 11(4):687–696, 2004.

[KS09] Bruce A Koplan and William G Stevenson. Ventricular tachycardia and sudden cardiac death. In *Mayo clinic proceedings*, volume 84, pages 289–297. Elsevier, 2009.

[Kus20] Fred Kusumoto. *ECG interpretation: from pathophysiology to clinical application*. Springer Nature, 2020.

[KvH08] Jan A Kors and Gerard van Herpen. Mirror image electrocardiograms and additional electrocardiographic leads: new wine in old wineskins? *Journal of electrocardiology*, 41(3):245–250, 2008.

[KW12] Teri M Kozik and Shu-Fen Wung. Acquired long qt syndrome: frequency, onset, and risk factors in intensive care patients. *Critical care nurse*, 32(5):32–41, 2012.

[LAT+09] Alessandro Liberati, Douglas G Altman, Jennifer Tetzlaff, Cynthia Mulrow, Peter C Gøtzsche, John PA Ioannidis, Mike Clarke, Philip J Devereaux, Jos Kleijnen, and David Moher. The prisma statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Journal of clinical epidemiology*, 62(10):e1–e34, 2009.

[LCKP19] Yun Liu, Po-Hsuan Cameron Chen, Jonathan Krause, and Lily Peng. How to read articles that use machine learning: users' guides to the medical literature. *Jama*, 322(18):1806–1816, 2019.

[Lib91] Lynn S Liben. Adults' performance on horizontality tasks: Conflicting frames of reference. *Developmental Psychology*, 27(2):285, 1991.

[LL20] Chengyu Liu and Jianqing Li. *Feature Engineering and Computational Intelligence in ECG Monitoring*. Springer, 2020.

[LLL+14] Won Kyu Lee, Hong Ji Lee, Jeong Su Lee, Hee Nam Yoon, Soo Young Sim, Yong Gyu Lim, and Kwang Suk Park. Validation of algorithm with noise tolerance methods to detect r-wave. In *The 15th International Conference on Biomedical Engineering*, pages 539–542. Springer, 2014.

[LLM16] Daniel Linares and Joan López-Moliner. quickpsy: An r package to fit psychometric functions for multiple groups. *The R Journal, 2016, vol. 8, num. 1, p. 122-131*, 2016.

[LM05] Daniel Lowd and Christopher Meek. Adversarial learning. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, pages 641–647, 2005.

[LMG+16] Nicole Lowres, Georgina Mulcahy, Robyn Gallagher, Saul Ben Freedman, David Marshman, Ann Kirkness, Jessica Orchard, and Lis Neubeck. Self-monitoring for atrial fibrillation recurrence in the discharge period post-cardiac surgery using an iphone electrocardiogram. *European Journal of Cardio-Thoracic Surgery*, 50(1):44–51, 2016.

[LMJM04] Shen Luo, Kurt Michler, Paul Johnston, and Peter W Macfarlane. A comparison of commonly used qt correction formulae: the effect of heart rate on the qtc of normal ecgs. *Journal of electrocardiology*, 37:81–90, 2004.

[Loc17] Emanuela T Locati. New directions for ambulatory monitoring following 2017 hrs-ishne expert consensus. *Journal of electrocardiology*, 50(6):828–832, 2017.

[Low12] David Lowe. *Perceptual organization and visual recognition*, volume 5. Springer Science & Business Media, 2012.

[LPL04] Pallavi Lanjewar, Vaishali Pathak, and Yash Lokhandwala. Issues in qt interval measurement. *Indian pacing and electrophysiology journal*, 4(4):156, 2004.

[LS52] Eugene Lepeschkin and Borys Surawicz. The measurement of the qt interval of the electrocardiogram. *Circulation*, 6(3):378–388, 1952.

[LSWA78] Robert L Lux, Creig R Smith, Roland F Wyatt, and JA Abildskov. Limited lead selection for estimation of body surface potential maps in electrocardiography. *IEEE Transactions on Biomedical Engineering*, (3):270–276, 1978.

[LSZ+09] Gaëlle Leroux, Jeanne Spiess, Laure Zago, Sandrine Rossi, Amélie Lubin, Marie-Renée Turbelin, Bernard Mazoyer, Nathalie Tzourio-Mazoyer, Olivier Houdé, and Marc Joliot. Adult brains don't fully overcome biases that lead to incorrect performance during cognitive development: an fmri study in young adults completing a piaget-like task. *Developmental Science*, 12(2):326–338, 2009.

[LVL+18] Kai Li, Kathy Vo, Byron K Lee, Newton Addo, and Zlatan Coralic. Effect of a single dose of iv ondansetron on qtc interval in emergency department patients. *The Bulletin of the American Society of Hospital Pharmacists*, 75(5):276–282, 2018.

[LZ01] Harvey Lodish and S Lawrence Zipursky. Molecular cell biology. *Biochem Mol Biol Educ*, 29:126–133, 2001.

[Mad04] John E Madias. The 13th multiuse ecg lead: Shouldn't we use it more often, and on the same hard copy or computer screen, as the other 12 leads? *Journal of electrocardiology*, 37(4):285–287, 2004.

[Mak18] Dominique Makowski. The psycho package: an efficient and publishing-oriented workflow for psychological science. *Journal of Open Source Software*, 3(22):470, 2018.

[Mal01] Marek Malik. Problems of heart rate correction in assessment of drug-induced qt interval prolongation. *Journal of cardiovascular electrophysiology*, 12(4):411–420, 2001.

[Mal11] Daniel Malacara. Color vision and colorimetry: theory and applications. Spie Bellingham, WA, 2011.

[MBA14] Beth A Medford, J Martijn Bos, and Michael J Ackerman. Epilepsy misdiagnosed as long qt syndrome: it can go both ways. *Congenital heart disease*, 9(4):E135–E139, 2014.

[MCbA01] Misha D Miller, J Porter Co-burn, and Michael J Ackerman. Diagnostic accuracy of screening electrocardiograms in long qt syndrome i. *Pediatrics*, 108(1):8–12, 2001.

[MCM95a] NB McLaughlin, RWF Campbell, and A Murray. Influence of t wave amplitude on automatic qt measurement. In *Computers in Cardiology 1995*, pages 777–780. IEEE, 1995.

[MCM95b] Neil B McLaughlin, RW Campbell, and Alan Murray. Comparison of automatic qt measurement techniques in the normal 12 lead electrocardiogram. *Heart*, 74(1):84–89, 1995.

[MCM96] NB McLaughlin, RW Campbell, and A Murray. Accuracy of four automatic qt measurement techniques in cardiac patients and healthy subjects. *Heart*, 76(5):422–426, 1996.

[MGH+19] Marek Malik, Christine Garnett, Katerina Hnatkova, Jose Vicente, Lars Johannesen, and Norman Stockbridge. Implications of individual qt/rr profiles—part 1: inaccuracies and problems of population-specific qt/heart rate corrections. *Drug safety*, 42(3):401–414, 2019.

[MK05] Arthur J Moss and Robert S Kass. Long qt syndrome: from channels to cardiac arrhythmias. *The Journal of clinical investigation*, 115(8):2018–2024, 2005.

[MMC+09] Judith M MacCormick, Hugh McAlister, Jackie Crawford, John K French, Ian Crozier, Andrew N Shelling, Carey-Anne Eddy, Mark I Rees, and Jonathan R Skinner. Misdiagnosis of long qt syndrome as epilepsy at first presentation. *Annals of emergency medicine*, 54(1):26–32, 2009.

[MMDY94] PW Macfarlane, SC McLaughlin, B Devine, and TF Yang. Effects of age, sex, and race on ecg interval measurements. *Journal of Electrocardiology*, 27:14–19, 1994.

[MOF09] Gustavo M Machado, Manuel M Oliveira, and Leandro AF Fernandes. A physiologically-based model for simulation of color vision deficiency. *IEEE transactions on visualization and computer graphics*, 15(6):1291–1298, 2009.

[Mor01] J Morganroth. Focus on issues in measuring and interpreting changes in the qtc interval duration. *European Heart Journal Supplements*, 3(suppl_K):K105–K111, 2001.

[MPS+19] Micaela Morettini, Chiara Peroni, Agnese Sbrollini, Ilaria Marcantoni, and Laura Burattini. Classification of drug-induced herg potassium-channel block from electrocardiographic t-wave features using artificial neural networks. *Annals of Noninvasive Electrocardiology*, 24(6):e12679, 2019.

[MS99] Joel Morganroth and Steven S Silber. How to obtain and analyze electrocardiograms in clinical trials: focus on issues in measuring and interpreting changes in the qtc interval duration. *Annals of Noninvasive Electrocardiology*, 4(4):425–433, 1999.

[MSC+85] Arthur J Moss, Peter J Schwartz, Richard S Crampton, Emanuela Locati, and Eric Carleen. The long qt syndrome: a prospective international study. *Circulation*, 71(1):17–21, 1985.

[MSC+91] Arthur J Moss, Peter J Schwartz, Richard S Crampton, Dan Tzivoni, Emanuela H Locati, Jean MacCluer, W Jackson Hall, Lowell Weitkamp, G Michael Vincent, and A Garson Jr. The long qt syndrome. prospective longitudinal study of 328 families. *circulation*, 84(3):1136–1144, 1991.

[MSSS17] Mario Merone, Paolo Soda, Mario Sansone, and Carlo Sansone. Ecg databases for biometric systems: A systematic review. *Expert Systems with Applications*, 67:189–202, 2017.

[Mun15] Albert Henry Munsell. *Atlas of the Munsell color system*. Wadsworth, Howland & Company, Incorporated, Printers, 1915.

[Nam10] Narayanan Namboodiri. Bradycardia-induced torsade de pointes–an arrhythmia less understood. *Indian pacing and electrophysiology journal*, 10(10):435, 2010.

[NAS12] Senthil Nachimuthu, Manish D Assar, and Jeffrey M Schussler. Drug-induced qt interval prolongation: mechanisms and clinical management. *Therapeutic advances in drug safety*, 3(5):241–253, 2012.

[NKM+04] Stefan P Nelwan, Jan A Kors, Simon H Meij, Jan H van Bemmel, and Maarten L Simoons. Reconstruction of the 12-lead electrocardiogram from reduced lead sets. *Journal of electrocardiology*, 37(1):11–18, 2004.

[Not93] Hans-Christoph Nothdurft. The role of features in preattentive vision: Comparison of orientation, motion and color cues. *Vision research*, 33(14):1937–1958, 1993.

[NSK+12] Masafumi Nakayama, Atsushi Saito, Hitoshi Kitazawa, Minoru Takahashi, Masahito Sato, Koichi Fuse, Masaaki Okabe, Kou Hoshino, Nobuhiro Tanaka, Akira Yamashina, et al. Papaverine-induced polymorphic ventricular tachycardia in relation to qtu and giant tu waves in four cases. *Internal Medicine*, 51(4):351–356, 2012.

[NWM+11] Magnus O Nimmermark, John J Wang, Charles Maynard, Mauricio Cohen, Ian Gilcrist, John Heitner, Michael Hudson, Sebastian Palmeri, Galen S Wagner, and Olle Pahlm. The impact of numeric and graphic displays of st-segment deviation levels on cardiologists' decisions of reperfusion therapy for patients with acute coronary occlusion. *Journal of electrocardiology*, 44(5):502–508, 2011.

[OAD+12] Alice J O'Toole, Xaiobo An, Joseph Dunlop, Vaidehi Natu, and P Jonathon Phillips. Comparing face recognition algorithms to humans on challenging tasks. *ACM Transactions on Applied Perception (TAP)*, 9(4):16, 2012.

[OCT+09] Kevin E O'Brien, Maria L Cannarozzi, Dario M Torre, Alex J Mechaber, and Steven J Durning. Training and assessment of ecg interpretation skills: results from the 2005 cdim survey. *Teaching and learning in medicine*, 21(2):111–115, 2009.

[OJRJ95] Nancy A Obuchowski Jr and Howard E Rockette Jr. Hypothesis testing of diagnostic accuracy for multiple readers and multiple tests an anova approach with dependent observations. *Communications in Statistics-simulation and Computation*, 24(2):285–308, 1995.

[OMK10] Chikaya Omichi, Yoshio Momose, and Shigemi Kitahara. Congenital

long qt syndrome presenting with a history of epilepsy: misdiagnosis or relationship between channelopathies of the heart and brain? *Epilepsia*, 51(2):289–292, 2010.

[Opi04] Lionel H Opie. *Heart physiology: from cell to circulation*. Lippincott Williams & Wilkins, 2004.

[OPN08] Alice J O'Toole, P Jonathon Phillips, and Abhijit Narvekar. Humans versus algorithms: Comparisons from the face recognition vendor test 2006. In *2008 8th IEEE International Conference on Automatic Face & Gesture Recognition*, pages 1–6. IEEE, 2008.

[OS00] Aude Oliva and Philippe G Schyns. Diagnostic colors mediate scene recognition. *Cognitive psychology*, 41(2):176–210, 2000.

[P+16] Nicolaas Prins et al. *Psychophysics: a practical introduction*. Academic Press, 2016.

[Par85] CSE Working Party. Recommendations for measurement standards in quantitative electrocardiography. *European Heart Journal*, 6(10):815–825, 1985.

[PAS+16] Alex Page, Mehmet K Aktas, Tolga Soyata, Wojciech Zareba, and Jean-Philippe Couderc. "qt clock" to improve detection of qt prolongation in long qt syndrome patients. *Heart Rhythm*, 13(1):190–198, 2016.

[PDJVdBW08] Pieter G Postema, Jonas SSG De Jong, Ivo AC Van der Bilt, and Arthur AM Wilde. Accurate electrocardiographic assessment of the qt interval: teach the tangent. *Heart Rhythm*, 5(7):1015–1018, 2008.

[PGLS95] Ronald M Pickett, Georges Grinstein, Haim Levkowitz, and Stuart Smith. Harnessing preattentive perceptual processes in visualization. In *Perceptual issues in visualization*, pages 33–45. Springer, 1995.

[PHC05] François Portet, Alfredo I Hernández, and Guy Carrault. Evaluation of real-time qrs detection algorithms in variable contexts. *Medical and Biological Engineering and Computing*, 43(3):379–385, 2005.

[PK01] Rod Passman and Alan Kadish. Polymorphic ventricular tachycardia, long qt syndrome, and torsades de pointes. *Medical Clinics of North America*, 85(2):321–341, 2001.

[PK11] Konstantinos Papadopoulos and Eleni Koustriava. Piaget's water-level task: The impact of vision on performance. *Research in developmental disabilities*, 32(6):2889–2893, 2011.

[PKN+09] Gopi Krishna Panicker, Dilip R Karnad, Mili Natekar, Snehal Kothari, Dhiraj Narula, and Yash Lokhandwala. Intra-and interreader variability in qt interval measurement by tangent and threshold methods in a central electrocardiogram laboratory. *Journal of electrocardiology*, 42(4):348–352, 2009.

[PMX+17] Alex Page, Scott McNitt, Xiaojuan Xia, Wojciech Zareba, and Jean-Philippe Couderc. Population-based beat-to-beat qt analysis from holter recordings in the long qt syndrome. *Journal of electrocardiology*, 50(6):787–791, 2017.

[Põd07] Endel Põder. Effect of colour pop-out on the recognition of letters in crowding conditions. *Psychological Research*, 71(6):641–645, 2007.

[Pod17] Philip J Podrid. Ambulatory ecg monitoring. *UpToDate, Waltham, MA. Accessed*, 6:05–18, 2017.

[Pri10] Dallas Price. How to read an electrocardiogram (ecg). part 1: Basic principles of the ecg. the normal ecg. *South Sudan Medical Journal*, 3(2):26–31, 2010.

[PRUE+11] Ragesh Panikkath, Kyndaron Reinier, Audrey Uy-Evanado, Carmen Teodorescu, Jonathan Hattenhauer, Ronald Mariani, Karen Gunson, Jonathan Jui, and Sumeet S Chugh. Prolonged tpeak-to-tend interval on the resting ecg is associated with increased risk of sudden cardiac death. *Circulation: Arrhythmia and Electrophysiology*, 4(4):441–447, 2011.

[PSCA15] Alex Page, Tolga Soyata, Jean-Philippe Couderc, and Mehmet Aktas. An open source ecg clock generator for visualization of long-term cardiac monitoring data. *IEEE Access*, 3:2704–2714, 2015.

[PSN+98] Silvia G Priori, Peter J Schwartz, Carlo Napolitano, Laura Bianchi, Adrienne Dennis, Maurizio De Fusco, Arthur M Brown, and Giorgio Casari. A recessive variant of the romano-ward long-qt syndrome? *Circulation*, 97(24):2420–2425, 1998.

[PSSH+17] Andreu Porta-Sánchez, David R Spillane, Louise Harris, Joel Xue, Pat Dorsey, Melanie Care, Vijay Chauhan, Michael H Gollob, and Danna A Spears. T-wave morphology analysis in congenital long qt syndrome discriminates patients from healthy individuals. *Clinical Electrophysiology*, 3(4):374–381, 2017.

[PTH07] Jeremy Pinnell, Simon Turner, and Simon Howell. Cardiac muscle physiology. *Continuing Education in Anaesthesia, Critical Care and Pain*, 7(3):85–88, 2007.

[Quw09] Mahmoud K Quweider. Adaptive pseudocoloring of medical images using dynamic optimal partitioning and space-filling curves. In *2009 2nd International Conference on Biomedical Engineering and Informatics*, pages 1–6. IEEE, 2009.

[Rau07] Pentti M Rautaharju. The birth of computerized electrocardiography: Hubert v. pipberger (1920-1993). *Cardiology journal*, 14(4):420–421, 2007.

[RF18] Diogo Ramalho and João Freitas. Drug-induced life-threatening arrhythmias and sudden cardiac death: A clinical perspective of long qt, short qt and brugada syndromes. *Revista Portuguesa de Cardiologia*, 37(5):435–446, 2018.

[RFF+09] Edwin P Rock, John Finkle, Howard J Fingert, Brian P Booth, Christine E Garnett, Stephen Grant, Robert L Justice, Richard J Kovacs, Peter R Kowey, Ignacio Rodriguez, et al. Assessing proarrhythmic potential of drugs when optimal studies are infeasible. *American heart journal*, 157(5):827–836, 2009.

[RFP19] Vincenzo Randazzo, Jacopo Ferretti, and Eros Pasero. Ecg watch: A real time wireless wearable ecg. In *2019 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pages 1–6. IEEE, 2019.

[RFTT02] Guillaume A Rousselet, Michèle Fabre-Thorpe, and Simon J Thorpe. Parallel processing in high-level categorization of natural images. *Nature neuroscience*, 5(7):629–630, 2002.

[RHA+16] Tobias Raupach, Sigrid Harendza, Sven Anders, Nikolai Schuelper, and Jamie Brown. How can we improve teaching of ecg interpretation skills? findings from a prospective randomised trial. *Journal of electrocardiology*, 49(1):7–12, 2016.

[RKE+89] Shereif Rezkalla, Robert A Kloner, J Ensley, M Al-Sarraf, S Revels, A Olivenstein, S Bhasin, S Kerpel-Fronious, and ZG Turi. Continuous ambulatory ecg monitoring during fluorouracil therapy: a prospective study. *Journal of Clinical Oncology*, 7(4):509–514, 1989.

[RPM+05] Maurizio Recanatini, Elisabetta Poluzzi, Matteo Masetti, Andrea Cavalli, and Fabrizio De Ponti. Qt prolongation through herg k+ channel blockade: current knowledge and strategies for the early prediction during drug development. *Medicinal research reviews*, 25(2):133–166, 2005.

[RSG09] Pentti M Rautaharju, Borys Surawicz, and Leonard S Gettes. Aha/accf/hrs recommendations for the standardization and interpretation of the electrocardiogram: part iv: the st segment, t and u waves, and the qt interval a scientific statement from the american heart association electrocardiography and arrhythmias committee, council on clinical cardiology; the american college of cardiology foundation; and the heart rhythm society endorsed by the international society for computerized electrocardiology. *Journal of the American College of Cardiology*, 53(11):982–991, 2009.

[RV05] Dan M Roden and Prakash C Viswanathan. Genetics of acquired long qt syndrome. *The Journal of clinical investigation*, 115(8):2025–2032, 2005.

[RVL+07] Partha Pratim Roy, Eduard Vazquez, Josep Lladós, Ramon Baldrich, and Umapada Pal. A system to segment text and symbols from color maps. In *International Workshop on Graphics Recognition*, pages 245–256. Springer, 2007.

[RWP86] Dan M Roden, Raymond L Woosley, and R Kirby Primm. Incidence and clinical features of the quinidine-associated long qt syndrome: implications for patient care. *American heart journal*, 111(6):1088–1093, 1986.

[RYK17] Einat Rashal, Yaffa Yeshurun, and Ruth Kimchi. Attentional requirements in perceptual grouping depend on the processes involved in the organization. *Attention, Perception, & Psychophysics*, 79(7):2073–2087, 2017.

[RYKC+18] Brandon RichardWebster, So Yon Kwon, Christopher Clarizio, Samuel E Anthony, and Walter J Scheirer. Visual psychophysics for making face recognition algorithms more explainable. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 252–270, 2018.

[SANC14] Walter J Scheirer, Samuel E Anthony, Ken Nakayama, and David D Cox. Perceptual annotation: Measuring human vision to improve computer vision. *IEEE transactions on pattern analysis and machine intelligence*, 36(8):1679–1686, 2014.

[SAW03] Stephen M Salerno, Patrick C Alguire, and Herbert S Waxman. Competency in interpretation of 12-lead electrocardiograms: a summary and appraisal of published evidence. *Annals of Internal Medicine*, 138(9):751–760, 2003.

[SBOR06] Pawan Sinha, Benjamin Balas, Yuri Ostrovsky, and Richard Russell. Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proceedings of the IEEE*, 94(11):1948–1962, 2006.

[SBS+17] Sophie Staniszewska, Jo Brett, Iveta Simera, Kate Seers, Carole Mockford, Susan Goodlad, DG Altman, David Moher, Rosemary Barber, Simon Denegri, et al. Gripp2 reporting checklists: tools to improve reporting of patient and public involvement in research. *Research involvement and engagement*, 3(1):13, 2017.

[SC08] Maully Shah and Christopher Carter. Long qt syndrome: A therapeutic challenge. *Annals of pediatric cardiology*, 1(1):18, 2008.

[SCGM00] John Stasko, Richard Catrambone, Mark Guzdial, and Kevin McDonald. An evaluation of space-filling information visualizations for depicting hierarchical structures. *International journal of human-computer studies*, 53(5):663–694, 2000.

[Sch71] Leo Schamroth. *An introduction to electrocardiography*. Wiley-Blackwell, 1971.

[Sch20] Peter J Schwartz. 1970–2020: 50 years of research on the long qt syndrome—from almost zero knowledge to precision medicine. *European Heart Journal*, 2020.

[SDPF+14] Arash Sadrieh, Luke Domanski, Joe Pitt-Francis, Stefan A Mann, Emily C Hodkinson, Chai-Ann Ng, Matthew D Perry, John A Taylor, David Gavaghan, Rajesh N Subbiah, et al. Multiscale cardiac modelling reveals the origins of notched t waves in long qt syndrome type 2. *Nature communications*, 5:5069, 2014.

[Sem18] Noura A Semary. Aproposed hsv-based pseudo-coloring scheme for enhancing medical images. In *Proc. Comput. Sci. Inf. Technol.*, pages 251–262, 2018.

[SF18] Michael Stettler and Gregory Francis. Using a model of human visual perception to improve deep learning. *Neural Networks*, 104:40–49, 2018.

[SG31] Thomas Smith and John Guild. The cie colorimetric standards and their use. *Transactions of the optical society*, 33(3):73, 1931.

[SGH+17] Thomas Solosko, Stacy Gehman, Earl Herleikson, Thomas Lyster, Shannon Fong, Kim Hansen, Jon Bishay, Chuni Kao, Brett Cross, Krishnakant Nammi, et al. Continuous outpatient ecg monitoring system, April 11 2017. US Patent 9,615,793.

[Sha81] TRD Shaw. Recurrent ventricular fibrillation associated with normal qt intervals. *QJM: An International Journal of Medicine*, 50(4):451–462, 1981.

[Sha02] Rashmi R Shah. The significance of qt interval in drug development. *British journal of clinical pharmacology*, 54(2):188–202, 2002.

[SHL19] Seung Yong Shin, Jun Young Hong, and Dong Hoon Lee. Delayed diagnosis of long qt syndrome in a patient with seizures. *Hong Kong Journal of Emergency Medicine*, 26(3):190–193, 2019.

[SKP+11] Vaibhav Salvi, Dilip R Karnad, Gopi Krishna Panicker, Mili Natekar, Pooja Hingorani, Vaibhav Kerkar, Arumugam Ramasamy, Michiel de Vries, Troy Zumbrunnen, Snehal Kothari, et al. Comparison of 5 methods of qt interval measurements on electrocardiograms from a thorough qt/qtc study: effect on assay sensitivity and categorical outliers. *Journal of electrocardiology*, 44(2):96–104, 2011.

[SKQ+15] Alan Sugrue, Vaclav Kremen, Bo Qiang, Seth H Sheldon, Christopher V DeSimone, Yehu Sapir, Bryan L Striemer, Peter Brady, Samuel J Asirvatham, Michael J Ackerman, et al. Electrocardiographic predictors of torsadogenic risk during dofetilide or sotalol initiation: utility of a novel t wave analysis program. *Cardiovascular drugs and therapy*, 29(5):433–441, 2015.

[SMM14] L Sathyapriya, L Murali, and T Manigandan. Analysis and detection r-peak detection using modified pan-tompkins algorithm. In *2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies*, pages 483–487. IEEE, 2014.

[SMY+05] Takafumi Saito, Hiroko Nakamura Miyamura, Mitsuyoshi Yamamoto, Hiroki Saito, Yuka Hoshiya, and Takumi Kaseda. Two-tone pseudo coloring: Compact visualization for one-dimensional data. 2005.

[SNK+17] Alan Sugrue, Peter A Noseworthy, Vaclav Kremen, J Martijn Bos, Bo Qiang, Ram K Rohatgi, Yehu Sapir, Zachi I Attia, Peter Brady, Pedro J Caraballo, et al. Automated t-wave analysis can differentiate acquired qt prolongation from congenital long qt syndrome. *Annals of Noninvasive Electrocardiology*, 22(6):e12455, 2017.

[SOKS91] WATARU SHIMIZU, TOHRU OHE, TAKASHI KURITA, and KATSURO SHIMOMURA. Differential response of qtu interval to exercise, isoproterenol, and atrial pacing in patients with congenital long qt syndrome. *Pacing and Clinical Electrophysiology*, 14(11):1966–1970, 1991.

[SP08] Matthias Schonlau and Ellen Peters. Graph comprehension: An experiment in displaying data as bar charts, pie charts and tables with and without the gratuitous 3rd dimension. *Social Science Research Network Working Paper Series*, pages 1–16, 2008.

[SPA19] Syed Raza Shah, Ki Park, and Richard Alweis. Long qt syndrome: a comprehensive review of the literature and current evidence. *Current Problems in Cardiology*, 44(3):92–106, 2019.

[Spe00] Nicholas Sperelakis. *Heart physiology and pathophysiology*. Elsevier, 2000.

[Spe12] Nicholas Sperelakis. *Cell physiology source book: essentials of membrane biophysics*. Elsevier, 2012.

[SPS+19] Christian Steinberg, François Philippon, Marina Sanchez, Pascal Fortier-Poisson, Gilles O'Hara, Franck Molin, Jean-François Sarrazin, Isabelle Nault, Louis Blier, Karine Roy, et al. A novel wearable device for continuous ambulatory ecg recording: proof of concept and assessment of signal quality. *Biosensors*, 9(1):17, 2019.

[SR15] Takaya Saito and Marc Rehmsmeier. The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. *PloS one*, 10(3):e0118432, 2015.

[SSS66] Leo Schamroth, Colin Schamroth, and Leo Schamroth. An introduction to electrocardiography. 1966.

[Sta19] Michael W Stadler. Thinking, experiencing and rethinking mereological interdependence. *Gestalt Theory*, 41(1):31–46, 2019.

[Ste57] Stanley S Stevens. On the psychophysical law. *Psychological review*, 64(3):153, 1957.

[Ste17] Stanley Smith Stevens. *Psychophysics: Introduction to its perceptual, neural and social prospects*. Routledge, 2017.

[STN08] Luke Solomons, Adrian Treloar, and Ryan Noronha. Competence of psychiatric clinicians in interpreting electrocardiograms and qt intervals: can they do this? does it matter? *Psychiatric Bulletin*, 32(8):291–294, 2008.

[Sto16] Maureen Stone. *A field guide to digital color*. CRC Press, 2016.

[Str15] Jeremy Straub. In search of technology readiness level (trl) 10. *Aerospace Science and Technology*, 46:312–320, 2015.

[SW64] Arthur Selzer and H WESLEY WRAY. Quinidine syncope: paroxysmal ventricular fibrillation occurring during treatment of chronic atrial arrhythmias. *Circulation*, 30(1):17–26, 1964.

[SW12] Lothar Spillmann and John S Werner. *Visual perception: The neurophysiological foundations*. Elsevier, 2012.

[SW17] Jürg Schläpfer and Hein J Wellens. Computer-interpreted electrocardiograms: benefits and limitations. *Journal of the American College of Cardiology*, 70(9):1183–1192, 2017.

[SWA+19] Jonathan R Skinner, Annika Winbo, Dominic Abrams, Jitendra Vohra, and Arthur A Wilde. Channelopathies that lead to sudden cardiac death: clinical and genetic aspects. *Heart, Lung and Circulation*, 28(1):22–30, 2019.

[TABW11] Benoît Tyl, Sara Azzam, Nathalie Blanco, and William Wheeler. Improvement and limitation of the reliability of automated qt measurement by recent algorithms. *Journal of electrocardiology*, 44(3):320–325, 2011.

[Tac63] Bruno Taccardi. Distribution of heart potentials on the thoracic surface of normal human subjects. *Circulation Research*, 12(4):341–352, 1963.

[TAS+15] Soheila Talebi, Azhir Alaleh, Zuber Sam, Soman Sandeep, Ferdinand Visco, Holly Totouom-Tangho, Hossein Kalantari, and Getaw Worku Hassen. Underestimated and unreported prolonged qtc by automated ecg analysis in patients on methadone: can we rely on computer reading? *Acta cardiologica*, 70(2):211–216, 2015.

[TBM+04] Lara Testai, AM Bianucci, I Massarelli, MARIA CRISTINA Breschi, E Martinotti, and Vincenzo Calderone. Torsadogenic cardiotoxicity of antipsychotic drugs: a structural feature, potentially involved in the

interaction with cardiac herg potassium channels. *Current medicinal chemistry*, 11(20):2691–2706, 2004.

[TCX+17] Yong-hong Tian, Xi-lin Chen, Hong-kai Xiong, Hong-liang Li, Li-rong Dai, Jing Chen, Jun-liang Xing, Xi-hong Wu, Wei-min Hu, Yu Hu, et al. Towards human-like and transhuman perception in AI 2.0: a review. *Frontiers of Information Technology & Electronic Engineering*, 18(1):58–67, 2017.

[TFS08] Christian Tominski, Georg Fuchs, and Heidrun Schumann. Task-driven color coding. In *2008 12th International Conference Information Visualisation*, pages 373–380. IEEE, 2008.

[The94] Jan Theeuwes. Stimulus-driven capture and attentional set: selective search for color and visual abrupt onsets. *Journal of Experimental Psychology: Human perception and performance*, 20(4):799, 1994.

[The95] Jan Theeuwes. Temporal and spatial characteristics of preattentive and attentive processing. *Visual Cognition*, 2(2-3):221–233, 1995.

[The13] Jan Theeuwes. Feature-based attention: It is all bottom-up priming. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1628):20130055, 2013.

[Thi18] Gaetano Thiene. Sudden cardiac death in the young: a genetic destiny? *Clinical Medicine*, 18(Suppl 2):s17, 2018.

[THTA07] Nathaniel W Taggart, Carla M Haglund, David J Tester, and Michael J Ackerman. Diagnostic miscues in congenital long-qt syndrome. *Circulation*, 115(20):2613–2620, 2007.

[Tis16] James E Tisdale. Drug-induced qt interval prolongation and torsades de pointes: role of the pharmacist in risk assessment, prevention and management. *Canadian Pharmacists Journal/Revue des Pharmaciens du Canada*, 149(3):139–152, 2016.

[TKK06] D Thomas, CA Karle, and J Kiehn. The cardiac herg/ikr potassium channel as pharmacological target: structure, function, regulation, and clinical applications. *Current pharmaceutical design*, 12(18):2271–2283, 2006.

[TLT+20] Kuo-Kun Tseng, Jiaqian Li, Yih-Jing Tang, Ching Wen Yang, and Fang-Ying Lin. Healthcare knowledge of relationship between time series electrocardiogram and cigarette smoking using clinical records. *BMC Medical Informatics and Decision Making*, 20(3):1–11, 2020.

[TMH+13] BKA Thomson, B Momciu, SHS Huang, CT Chan, B Urquhart, A Skanes, A Krahn, G Klein, and RM Lindsay. Ecg machine qtc intervals are inaccurate in hemodialysis patients. *Nephron Clinical Practice*, 124(1-2):113–118, 2013.

[TOB10] Jan Theeuwes, Christian NL Olivers, and Artem Belopolsky. Stimulus-driven capture and contingent capture. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(6):872–881, 2010.

[Top19a] Eric Topol. *Deep medicine: how artificial intelligence can make healthcare human again*. Hachette UK, 2019.

[Top19b] Eric Topol. The Topol review: prepating the healthcare workforce to deliver the digital future. Technical Report February, Health Education England, London, 2019.

[TRB16] G Thiene, S Rizzo, and C Basso. Pathology of sudden death, cardiac arrhythmias and conduction system. In *Cardiovascular pathology*, pages 361–433. Elsevier, 2016.

[Tre83] Anne Treisman. The role of attention in object perception. In *Physical and biological processing of images*, pages 316–325. Springer, 1983.

[Tre85] Anne Treisman. Preattentive processing in vision. *Computer vision, graphics, and image processing*, 31(2):156–177, 1985.

[TRR+07] Ian Topilski, Ori Rogowski, Rafael Rosso, Dan Justo, Yitschak Copperman, Michael Glikson, Bernard Belhassen, Marek Hochenberg, and Sami Viskin. The morphology of the qt interval predicts torsade de pointes during acquired bradyarrhythmias. *Journal of the American College of Cardiology*, 49(3):320–328, 2007.

[TZGM10] Thomas Töllner, Michael Zehetleitner, Klaus Gramann, and Hermann J Müller. Top-down weighting of visual dimensions: Behavioral

and electrophysiological evidence. *Vision research*, 50(14):1372–1381, 2010.

[VBC18] Effy Vayena, Alessandro Blasimme, and I. Glenn Cohen. Machine learning in medicine: Addressing ethical challenges. *PLoS Medicine*, 15(11):4–7, 2018.

[vdH15] Peter van der Helm. Simplicity in perceptual organization. 2015.

[vdH17] Peter A van der Helm. Human visual perceptual organization beats thinking on speed. *Attention, Perception, & Psychophysics*, 79(4):1227–1238, 2017.

[VDPL+18] Adriana N Vest, Giulia Da Poian, Qiao Li, Chengyu Liu, Shamim Nemati, Amit J Shah, and Gari D Clifford. An open source benchmarked toolbox for cardiovascular waveform and interval analysis. *Physiological measurement*, 39(10):105004, 2018.

[Vel10] Rosemarie Velik. Towards human-like machine perception 2. 0. *International Review on Computers and Software*, 5(4):476–488, 2010.

[vEPvH+07] Henk J Ritsema van Eck, Pieter G Postema, Gerard van Herpen, Arthur AM Wilde, and John A Kors. Dynamic changes of the tu complex in the electrocardiogram. *Journal of Electrocardiology*, 40(1):S22–S25, 2007.

[Ves91] Iris Vessey. Cognitive fit: A theory-based analysis of the graphs versus tables literature. *Decision sciences*, 22(2):219–240, 1991.

[VG91] Iris Vessey and Dennis Galletta. Cognitive fit: An empirical study of information acquisition. *Information systems research*, 2(1):63–84, 1991.

[VJM+15] Jose Vicente, Lars Johannesen, Jay W Mason, William J Crumb, Esther Pueyo, Norman Stockbridge, and David G Strauss. Comprehensive t wave morphology assessment in a randomized clinical study of dofetilide, quinidine, ranolazine, and verapamil. *Journal of the American Heart Association*, 4(4):e001615, 2015.

[VRS+05] Sami Viskin, Uri Rosovski, Andrew J Sands, Edmond Chen, Peter M Kistler, Jonathan M Kalman, Laura Rodriguez Chavez, Pedro Iturralde Torres, Osmar A Centurión, Akira Fujiki, et al. Inaccurate electrocardiographic interpretation of long qt: the majority of physicians cannot recognize a long qt when they see one. *Heart Rhythm*, 2(6):569–574, 2005.

[VTVAV15] Mihailo Vukmirović, Irena Tomašević-Vukmirović, Lazar Angelkov, and Filip Vukmirović. Emotional stress as a cause of syncope and torsade de pointes in patients with long qt syndrome. *Vojnosanitetski pregled*, 72(2):192–195, 2015.

[WA84] Michele Andrisin Wittig and Mary J Allen. Measurement of adult performance on piaget's water horizontality task. *Intelligence*, 8(4):305–313, 1984.

[War88] Colin Ware. Color sequences for univariate maps: Theory, experiments and principles. *IEEE Computer Graphics and Applications*, 8(5):41–49, 1988.

[War12] Colin Ware. *Information visualization: perception for design*. Elsevier, 2012.

[War21] Colin Ware. *Information visualization: perception for design*. Elsevier, 2021.

[Wat75] YOSHIO Watanabe. Purkinje repolarization as a possible cause of the u wave in the electrocardiogram. *Circulation*, 51(6):1030–1037, 1975.

[WBA+14] Greg Wood, Jeremy Batt, Andrew Appelboam, Adrian Harris, and Mark R Wilson. Exploring the impact of expertise, clinical history, and visual search on electrocardiogram interpretation. *Medical Decision Making*, 34(1):75–83, 2014.

[Wer23] Max Wertheimer. Laws of organization in perceptual forms. *A source book of Gestalt Psychology*, 1923.

[WFG+12] Johan Wagemans, Jacob Feldman, Sergei Gepshtein, Ruth Kimchi, James R Pomerantz, Peter A Van der Helm, and Cees Van Leeuwen. A

century of gestalt psychology in visual perception: Ii. conceptual and theoretical foundations. *Psychological bulletin*, 138(6):1218, 2012.

[WGG+10] W Stephen Waring, Ann Graham, Julie Gray, Allen D Wilson, Catherine Howell, and D Nicholas Bateman. Evaluation of a qt nomogram for risk assessment after antidepressant overdose. *British journal of clinical pharmacology*, 70(6):881–885, 2010.

[Wil80] Jos L Willems. A plea for common standards in computer aided ECG analysis. *Computers and Biomedical Research*, 13(2):120–131, 1980.

[Wil85] Leonard J Williams. Tunnel vision induced by a foveal load manipulation. *Human factors*, 27(2):221–227, 1985.

[WK17] Eric Welch and S Kobourov. Measuring symmetry in drawings of graphs. In *Computer Graphics Forum*, volume 36, pages 341–351. Wiley Online Library, 2017.

[WLX+16] Xiangkui Wan, Yan Li, Chong Xia, Minghu Wu, Jin Liang, and NA Wang. A t-wave alternans assessment method based on least squares curve fitting technique. *Measurement*, 86:93–100, 2016.

[WO14] Joanne M Wood and Cynthia Owsley. Useful field of view test. *Gerontology*, 60(4):315–318, 2014.

[Wri67] WD Wright. Color science, concepts and methods. quantitative data and formulas. *Physics Bulletin*, 18(10):353, 1967.

[WRY19] Taiyang Wu, Jean-Michel Redouté, and Mehmet Yuce. A wearable, low-power, real-time ecg monitor for smart t-shirt and iot healthcare applications. In *Advances in Body Area Networks I*, pages 165–173. Springer, 2019.

[WS20] Raymond L Woosley and Peter J Schwartz. Drug-induced long qt syndrome and torsades de pointes. In *Cardiac Repolarization*, pages 185–200. Springer, 2020.

[WU19] Jeremy M Wolfe and Igor S Utochkin. What is a preattentive feature? *Current opinion in psychology*, 29:19–26, 2019.

[Wys07] Diane K Wysowski. Surveillance of prescription drug-related mortality using death certificate data. *Drug Safety*, 30(6):533–540, 2007.

[XGC+18] Ke Xu, Shunan Guo, Nan Cao, David Gotz, Aiwen Xu, Huamin Qu, Zhenjie Yao, and Yixin Chen. Ecglens: Interactive visual exploration of large scale ecg data for arrhythmia detection. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 663. ACM, 2018.

[XHC+19] Cong Xue, Wei Hua, Chi Cai, Li-Gang Ding, Hong-Xia Niu, Xiao-Han Fan, Zhi-Min Liu, Min Gu, Yun-Zi Zhao, and Shu Zhang. Predictive value of tpeak-tend interval for ventricular arrhythmia and mortality in heart failure patients with an implantable cardioverter-defibrillator: A cohort study. *Medicine*, 98(49), 2019.

[XR98] Qiuzhen Xue and Shankara Reddy. Algorithms for computerized qt analysis. *Journal of electrocardiology*, 30:181–186, 1998.

[Y+02] Shi-Tao Yeh et al. Using trapezoidal rule for the area under a curve calculation. *Proceedings of the 27th Annual SAS® User Group International (SUGI'02)*, 2002.

[YA82] John I Yellott and Albert Ahumada. Visual perception: Theory and practice. *Journal of Mathematical Psychology*, 25(1):82–86, 1982.

[YA98] Gan-Xin Yan and Charles Antzelevitch. Cellular basis for the normal t wave and the electrocardiographic manifestations of the long-qt syndrome. *Circulation*, 98(18):1928–1936, 1998.

[YC03] Yee Guan Yap and A John Camm. Drug induced qt prolongation and torsades de pointes. *Heart*, 89(11):1363–1372, 2003.

[Zar06] Wojciech Zareba. Genotype-specific ecg patterns in long qt syndrome. *Journal of electrocardiology*, 39(4):S101–S106, 2006.

[ZD14] Daniele Zavagno and Olga Daneyko. Perceptual grouping, and color. *Encyclopedia of Color Science and Technology*, pages 1–5, 2014.

[ZF19] Zhenglong Zhou and Chaz Firestone. Humans can decipher adversarial images. *Nature Communications*, 10(1), 2019.

[ZH15] Liang Zhou and Charles D Hansen. A survey of colormaps in visualization. *IEEE transactions on visualization and computer graphics*, 22(8):2051–2069, 2015.

[Zip87] Douglas P Zipes. Proarrhythmic effects of antiarrhythmic drugs. *The American journal of cardiology*, 59(11):E26–E31, 1987.

[ZSST11] Zahra Zahedi, Saied Sadri, Mohammad Soltani, and Masoud Kavosh Tehrani. Breast diseases detection and pseudo-coloring presentation for gray infrared breast images. In *2011 Asia Communications and Photonics Conference and Exhibition (ACP)*, pages 1–8. IEEE, 2011.

Appendix A

Supplement for Chapter 4

A.1 Participant information sheet and consent form

MANCHESTER 1824
The University of Manchester

School of Computer Science

Can lay people perceive a prolonged QT interval in the 'raw' ECG data

Participant Information Sheet

You are being invited to take part in a research study. Before you decide, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Please ask if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for taking the time to read this.

Who will conduct the research?

Alaa Alahmadi (PhD student).

What is the purpose of the research?

This research aims to examine whether lay people can perceive a prolonged QT interval in the 'raw' data (ECG) signal, and whether this can be enhanced by employing data visualisation methods.

Why have I been chosen?

We are inviting students and university staff, who have no experience and training in ECG interpretation.

MANCHESTER 1824
The University of Manchester

What would I be asked to do if I took part?

You will sit in front of an eye tracker (which looks like a desktop monitor), which will help us to know exactly where your eyes are focused. Two selected ECG signals (only two complexes) will appear on the screen. You should look at the two ECGs images and decide whether there is any difference between them. You may take as long as you need to, but press either the 'Yes' or 'No' button, as soon as you have made your decision. You should press the 'Yes' button if you think there is a difference, and the 'No' button if you do not think there is a difference.

What happens to the data collected?

All data will be encrypted and stored securely.

Any information that could be used to identify individuals will be anonymised.

How is confidentiality maintained?

All data will be encrypted and stored securely. Participants will be allocated a ID/code, which will be used to identify their responses and eye tracking data. Names will not be associated with the data at any point.

What happens if I do not want to take part or if I change my mind?

It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part you are still free to withdraw from the research at any point without giving prior reason.

Will I be paid for participating in the research?

No

What is the duration of the research?

The study will take from 1 to 2 hours to complete.

Where will the research be conducted?

In the University of Manchester premises.

Will the outcomes of the research be published?

The outcomes will be published in journals and/or conference proceedings.

Contact for further information.

For further information, please contact either myself or my supervisor (details below)

Alaa Alahmadi: alaa.alahmadi@postgrad.manchester.ac.uk

Dr Caroline Jay: Caroline.Jay@manchester.ac.uk

What if something goes wrong?

If a participant wants to make a formal complaint about the conduct of the research they should
contact the Head of the Research Office, Christie Building, University of Manchester, Oxford
Road, Manchester, M13 9PL.

This Project Has Been Approved by the University of Manchester's Research Ethics Committee [Ref: 2017-2714-3948]

MANCHESTER 1824
The University of Manchester

School of Computer Science

Can lay people perceive a prolonged QT interval in the 'raw' ECG data

CONSENT FORM

If you are happy to participate please complete and sign the consent form below:

Please Initial Box

1. I confirm that I have read the attached information sheet on the above project and have had the opportunity to consider the information and ask questions and had these answered satisfactorily. ☐
2. I understand that my participation in the study is voluntary and that I am free to withdraw at any time without giving a reason. ☐
3. I understand that the session will be audio recorded and an eye-tracker will be used. ☐
4. I agree to the use of anonymous quotes. ☐

Name of participant	**Date**	**Signature**

Name of person taking consent	**Date**	**Signature**

Appendix B

Supplement for Chapter 5

B.1 Participant information sheet and consent form

Can visualisation properties influence perception of QT prolongation?

Participant Information Sheet (PIS)

You are being invited to take part in a research study to understand how lay people, who have no prior knowledge in ECG interpretation, can self-monitor abnormal changes in ECG when compared to a normal ECG (Baseline). We focused on one type of significant change which is the QT-interval prolongation. The ECG data will be presented to you using different visualization methods, which differ in their display properties or attributes. We aim to investigate the effect of such visualization properties on lay people's interpretation of QT-interval prolongation. Before you decide whether to take part, it is important for you to understand why the research is being conducted and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Please ask if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for taking the time to read this.

Who will conduct the research?

Alaa Alahmadi (PhD student) is the researcher from the Interaction Analysis and Modelling Lab (IAM), School of Computer Science, University of Manchester.

What is the purpose of the research?

People take medication that may cause abnormal change in ECG including prolongation of the QT interval, which can lead to sudden death if not discovered early. It is not possible for them to have frequent enough ECGs interpreted by their clinicians to identify the problem; if they can self-monitor outside of clinical environment, and spot when problem is arising, timely clinical intervention may save their life. This study aims to examine whether lay people's perception of QT

interval prolongation in ECG can be influenced by the way we visualise the ECG data. In this study, we are going to present different types of data visualisation methods to lay people and compare their performance in interpretation across those visualisations.

Why have I been chosen?

We are inviting lay participants, who have no prior experience and training in ECG interpretation, including students and university staff to understand how they interpret this type of data and what is the best way to present to them. We need around 60 participants.

What would I be asked to do if I took part?

At the beginning, you will have a quick training session to understand what is a QT-interval and where its location is on an ECG image. Then, you will sit in front of an eye tracker (which looks like a desktop monitor), which will help us to know exactly where your eyes are focused. Three selected ECG images will appear on the screen, one of them shows a normal ECG and will be used as a comparator. You should look at the other two ECGs images and compare it with the normal ECG image and decide whether which of the two ECGs has a prolonged QT-interval. You will be presented with different visualization methods, and you will be asked to rate whether they were effective and helped you in perceiving the prolonged QT-interval. After the experiment, you will have a quick semi-structured interview that will be run by the researcher to ask you about what do you think about interpreting this type of data and the types of visualisation used.

What will happen to my personal information?

In order to undertake the research project we will need to collect the following personal information/data about you:

- Eye-tracking data will be recorded during the experiment using a Tobii eye- tracker. This is a non-invasive device attached to the computer screen.
- A university audio recorder will be used during the experiment and in the post-experimental interview to record your comments.
- Observation data including your comments and suggestions during the experiment will be written down on a paper sheet by the researcher.

Only the research team will have access to this information.

We are collecting and storing this personal information in accordance with the General Data Protection Regulation (GDPR) and Data Protection Act 2018 which legislate to protect your personal information. The legal basis upon which we are using your personal information is "public interest task" and "for research purposes" if sensitive information is collected. For more information about the way we process your personal information and comply with data protection law please see our Privacy Notice for Research Participants.

The University of Manchester, as Data Controller for this project, takes responsibility for the protection of the personal information that this study is collecting about you. In order to comply with the legal obligations to protect your personal data the University has safeguards in place such as policies and procedures. All researchers are appropriately trained and your data will be looked after in the following way:

The study team at the school of computer science, interaction and modelling lab, the University of Manchester will have access to your personal identifiable information, that is data which could identify you, but they will anonymise it after one day from finishing the experiment. All data will be encrypted and stored securely. You will be allocated a ID/code, which will be used to identify your responses/answers and eye tracking data. Your name will not be associated with the data at any point. However, your consent form, contact details, etc will be

retained for five years and kept in paper formats in a locked drawer, at the interaction and modelling lab.

You have a number of rights under data protection law regarding your personal information. For example you can request a copy of the information we hold about you, including audio recordings or eye-tracking data. This is known as a Subject Access Request. If you would like to know more about your different rights, please consult our privacy notice for research and if you wish to contact us about your data protection rights, please email dataprotection@manchester.ac.uk or write to The Information Governance Office, Christie Building, University of Manchester, Oxford Road, M13 9PL. at the University and we will guide you through the process of exercising your rights.

You also have a right to complain to the Information Commissioner's Office, Tel 0303 123 1113

Will my participation in the study be confidential?

Your participation in the study will be kept confidential to the study team and those with access to your personal information as listed above. The following shows in details how the data will be kept confidential:

- The audio recordings will be used to create digital transcripts, by the researcher (Alaa Alahmadi). The transcripts will be stored securely in the interaction and modelling lab' server, and they will be allocated a ID/code, which will be used to identify the participant.
- After five years the recordings will be destroyed and digitally removed from the server.
- Only the research team (the main researcher -Alaa Alahmadi- and the main supervisor-Caroline Jay- will have access to the data).
- All the data will be encrypted and stored securely in the interaction and modelling lab' server. Participants will be allocated a ID/code, which will be used to identify their responses/answers and eye tracking data. Names

will not be associated with the data at any point. The consent form, contact details, etc will be retained for five years and kept in paper formats in a locked drawer, at the interaction and modelling lab.

What happens if I do not want to take part or if I change my mind?

It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part you are still free to withdraw at any time without giving a reason and without detriment to yourself. However, it will not be possible to remove your data from the project once it has been anonymised and forms part of the dataset as we will not be able to identify your specific data. This does not affect your data protection rights.

The audio recording is desirable to your participation in the study. You should be comfortable with the recording process at all times and you are free to stop recording at any time.

Will my data be used for future research?

Your anonymous data may be used for health informatics research in the future. Your information will only be used by organisations and researchers to conduct research in accordance with the UK Policy Framework for Health and Social Care Research.

This information will not identify you and will not be combined with other information in a way that could identify you. The information will only be used for the purpose of health and care research, and cannot be used to contact you regarding any other matter or to affect your care. It will not be used to make decisions about future services available to you.

Will I be paid for participating in the research?

Yes, a £10 shopping voucher will be provided to you after completing the study.

What is the duration of the research?

The study will take a maximum of one hour to complete, divided as follows:

- 10 minutes before the experiment for introduction and short training session.
- 40 minutes for the experiment.
- 10 minutes after the experiment for the semi- structured interview.

Where will the research be conducted?

In the University of Manchester premises, at the Interaction Analysis and Modelling Lab, School of Computer Science, University of Manchester.

Will the outcomes of the research be published?

The outcomes will be published in journals and/or conference proceedings.

Who has reviewed the research project?

The project has been reviewed by "Proportionate University Research Ethics Committee (UREC) Review" (School of Computer Science).

What if I want to make a complaint?

Minor complaints

If you have a minor complaint then you need to contact the researcher(s) in the first instance.

- The researcher: Alaa Alahmadi
  Email: alaa.alahmadi@postgrad.manchester.ac.uk
  Telephone number: 0161-275-7821
- The supervisor: Dr Caroline Jay
  Email: Caroline.Jay@manchester.ac.uk
  Telephone number: 0161-275-0677

Formal Complaints

If you wish to make a formal complaint or if you are not satisfied with the response you have gained from the researchers in the first instance then please contact

The Research Governance and Integrity Manager, Research Office, Christie Building, University of Manchester, Oxford Road, Manchester, M13 9PL, by emailing: research.complaints@manchester.ac.uk or by telephoning 0161 275 2674.

What Do I Do Now?

If you about have any queries the study or if you are interested in taking part then please contact the researcher(s)**.**

- The researcher: Alaa Alahmadi
  Email: alaa.alahmadi@postgrad.manchester.ac.uk
  Telephone number: 0161-275-7821

- The supervisor: Dr Caroline Jay
  Email: Caroline.Jay@manchester.ac.uk
  Telephone number: 0161-275-0677

This Project Has Been Approved by the University of Manchester's Research Ethics Committee [Ref: 2017-2714-3948]

MANCHESTER 1824
The University of Manchester

Can visualisation properties influence perception of QT prolongation?

Consent Form

If you are happy to participate please complete and sign the consent form below

	Activities	Initials
1	I confirm that I have read the attached information sheet (**Version 1.0, Date 10/06/2018**) for the above study and have had the opportunity to consider the information and ask questions and had these answered satisfactorily.	
2	I understand that my participation in the study is voluntary and that I am free to withdraw at any time without giving a reason and without detriment to myself. I understand that it will not be possible to remove my data from the project once it has been anonymised and forms part of the data set. I agree to take part on this basis.	
3	I agree to the experiment itself and the interviews after the experiment being audio recorded.	
4	I agree that my eye-moments will be recorded using eye-tracker device.	
5	I agree that any data collected may be published in anonymous form in academic books, reports or journals.	
6	I agree to take part in this study	

Data Protection

The personal information we collect and use to conduct this research will be processed in accordance with data protection law as explained in the Participant Information Sheet and the Privacy Notice for Research Participants.

_____	_____	_____
Name of participant	Signature	Date
_____	_____	_____
Name of the person taking consent	Signature	Date

Appendix C

Supplement for Chapter 6

C.1 Participant information sheet and consent form

MANCHESTER 1824
The University of Manchester

Can data visualisations support ECG Interpretation?

Participant Information Sheet (PIS)

An ECG (electrocardiogram) is a graph that shows the electrical activity of the heart, picked up via sensors placed on the skin. It is commonly used in medicine to diagnose heart conditions, and other illnesses.

You are being invited to take part in a research study to understand how lay people, who have no prior experience of ECG interpretation, can be supported in understanding changes in an ECG. This will ultimately help patients to monitor their own ECG at home, allowing the early detection of abnormalities.

You do not need any medical knowledge to take part, and will receive training on how to read the ECG at the start of the study. The ECG data will be presented to you using the current standard format, and using alternative formats. The aim of the study is to understand how the format affects ECG interpretation. Please note that you are not going to interpret your own ECGs, as all ECGs will be belonging to other people.

Before you decide whether to take part, it is important for you to understand why the research is being conducted and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Please ask if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for taking the time to read this.

Who will conduct the research?

Alaa Alahmadi, a PhD Student from the Interaction Analysis and Modelling (IAM) lab, School of Computer Science, University of Manchester.

MANCHESTER 1824
The University of Manchester

What is the purpose of the research?

A number of common medications can cause abnormal changes in the ECG, which can lead to serious illness or sudden death if not discovered early. It is often not possible for patients to have frequent enough ECGs to identify the problem; if they can self-monitor at home, and spot when a problem arises, timely clinical intervention may save their life. This study aims to examine whether lay people's perception and interpretation of some patterns of serious ECG changes can be influenced by the way we visualise the ECG data.

Why have I been chosen?

We are inviting lay participants, who have no prior experience of ECG interpretation, to take part.

What would I be asked to do if I took part?

At the start of the study, you will have a quick training session to help you identify patterns in the ECG signal, and distinguish between those that are normal, and those that are abnormal. Then, you will sit in front of an eye tracker (which looks like a desktop monitor), which will help us to know exactly where your eyes are focused. A series of ECGs will appear on the screen. You should look at these ECGs, and rate how likely you think it is that they contain an abnormality on a rating scale.

What will happen to my personal information?

In order to undertake the research project we will need to collect the following information/data:

- Eye-tracking data will be recorded during the experiment using a Tobii eye- tracker. This is a non-invasive device attached to the computer screen. It will not video your eyes, only record their positions on the screen.
- Your responses on the rating scale.

Only the research team will have access to this information.

We are collecting and storing this information in accordance with the General Data Protection Regulation (GDPR) and Data Protection Act 2018 which legislate to protect your personal information. The legal basis upon which we are using your personal information is "public interest task" and "for research purposes" if sensitive information is collected. For more information about the way we process your personal information and comply with data protection law please see our Privacy Notice for Research Participants.

The University of Manchester, as Data Controller for this project, takes responsibility for the protection of the personal information that this study is collecting about you. In order to comply with the legal obligations to protect your personal data the University has safeguards in place such as policies and procedures. All researchers are appropriately trained and your data will be looked after in the following way:

All data will be encrypted and stored securely. You will be allocated a ID/code, which will be used to identify your responses/answers and eye tracking data. Your name will not be associated with the data at any point. However, your consent form will be retained for five years and kept in paper formats in a locked drawer, in the School of Computer Science.

You have a number of rights under data protection law regarding your personal information. For example you can request a copy of the information we hold about you, including audio recordings or eye-tracking data. This is known as a Subject Access Request. If you would like to know more about your different rights, please consult our privacy notice for research and if you wish to contact us about your data protection rights, please email dataprotection@manchester.ac.uk or write to The Information Governance Office, Christie Building, University of Manchester, Oxford Road, M13 9PL. at the University and we will guide you through the process of exercising your rights.

You also have a right to complain to the Information Commissioner's Office, Tel 0303 123 1113

MANCHESTER 1824
The University of Manchester

Will my participation in the study be confidential?

Your participation in the study will be kept confidential to the study team and those with access to your personal information as listed above. The following shows in details how the data will be kept confidential:

- All the data will be encrypted and stored securely on a server at the University of Manchester. Participants will be allocated an ID/code, which will be used to identify their rating answers and eye tracking data. Names will not be associated with the data at any point. The consent form will be retained for five years and kept in paper formats in a locked drawer, in the School of Computer Science.

What happens if I do not want to take part or if I change my mind?

It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part you are still free to withdraw at any time without giving a reason and without detriment to yourself. However, it will not be possible to remove your data from the project once it has been anonymised and forms part of the dataset as we will not be able to identify your specific data. This does not affect your data protection rights.

Will my data be used for future research?

We will make the anonymised data open access. This is in accordance with the Open Research Framework principles, and means that other researchers can check our results, or use them in further research. This maximises the value of the data. It will not be possible to identify any individual from this dataset.

Will I be paid for participating in the research?

Yes, a £10 amazon voucher will be given to each participant.

MANCHESTER 1824
The University of Manchester

What is the duration of the research?

The study will take a maximum of one hour to complete, divided as follows:

- 20 minutes before the experiment for introduction and short training session.
- 40 minutes for the experiment.

Where will the research be conducted?

In the Kilburn Building, School of Computer Science, University of Manchester.

Will the outcomes of the research be published?

The outcomes will be published in journals and/or conference proceedings.

Who has reviewed the research project?

The project has been reviewed by "Proportionate University Research Ethics Committee (UREC)
Review" (School of Computer Science).

What if I want to make a complaint?

Minor complaints

If you have a minor complaint then you need to contact the researcher(s) in the first instance.

- The researcher: Alaa Alahmadi
  Email: alaa.alahmadi@postgrad.manchester.ac.uk
  Telephone number: 0161-275-7821

- The supervisor: Dr Caroline Jay
  Email: Caroline.Jay@manchester.ac.uk
  Telephone number: 0161-275-0677

Formal Complaints

If you wish to make a formal complaint or if you are not satisfied with the response you have gained from the researchers in the first instance then please contact

The Research Governance and Integrity Manager, Research Office, Christie Building, University of Manchester, Oxford Road, Manchester, M13 9PL, by emailing: research.complaints@manchester.ac.uk or by telephoning 0161 275 2674.

What Do I Do Now?

If you about have any queries the study or if you are interested in taking part then please contact the researcher(s)**.**

- The researcher: Alaa Alahmadi
  Email: alaa.alahmadi@postgrad.manchester.ac.uk
  Telephone number: 0161-275-7821

- The supervisor: Dr Caroline Jay
  Email: Caroline.Jay@manchester.ac.uk
  Telephone number: 0161-275-0677

This Project Has Been Approved by the University of Manchester's Research Ethics Committee [Ref: 2019-6122-9941]

Can data visualisations support ECG Interpretation?

Consent Form

If you are happy to participate please complete and sign the consent form below:

	Activities	Initials
1	I confirm that I have read the attached information sheet (**Version 1.0, Date 20/03/2019**) for the above study and have had the opportunity to consider the information and ask questions and had these answered satisfactorily.	
2	I understand that that I am free to withdraw at any time without giving a reason and without detriment to myself. I understand that it will not be possible to remove my data from the project once it has been anonymised and forms part of the data set. I agree to take part on this basis.	
4	I agree that my eye-movements will be recorded using eye-tracker device.	
5	I agree that any data collected may be published in anonymous form in academic books, reports or journals.	
6	I agree that anonymised data will be accessible for other researchers after the end of the study (i.e. the anonymised data will be made open access).	
7	I agree that the principal investigator (Alaa Alahmadi) and the main supervisor (Caroline Jay) will have access to the data throughout the life-cycle of the research project, while the main supervisor (Caroline Jay) will be also the overall custodian (responsible beyond the end of the project).	
8	I agree to take part in this study.	

Data Protection
The personal information we collect and use to conduct this research will be processed in accordance with data protection law as explained in the Participant Information Sheet and the Privacy Notice for Research Participants.

_____	_____	_____
Name of participant	Signature	Date
_____	_____	_____
Name of the person taking consent	Signature	Date