Optimising Route Comfort Indices for Neonatal Transfers by Road

Tom J. Partridge, MEng.

Thesis submitted to The University of Nottingham
for the degree of Doctor of Philosophy

June 2021

Abstract

The risk of severe brain injuries in sick premature infants increases when transferred between hospitals. Causality is uncertain, but stress levels are elevated during ambulance journeys; potentially due to excessive levels of noise and vibration. It has been proposed that reducing these levels would reduce the risk, with one prospective method being comfort-optimised navigation.

An Android app was developed that logs noise level, Inertial Measurement Unit (IMU) and location data during journeys, sampling at the fastest rates possible depending on the hardware and firmware. The smartphone used during development was found to sample noise levels accurate to 0.3 dB up to 80 dB(A) and accelerations accurate to 10% up to 40 Hz, although considerable jitter was present in the IMU sampling. Recorded data were shown to be repeatable for multiple passes over the same stretch of road (acceleration interquartile range (IQR): 0.14 $m \cdot s^{-2}$; noise IQR: 2.8 dB). Data were influenced by both supplementary audio and the smartphone model so an initial idea of gathering data through public engagement was determined unsuitable.

Controlled collection of data was planned, utilising the neonatal ambulances operated by CenTre Neonatal Transport (CenTre). A new smartphone model was identified that was capable of sampling accelerations at a sufficient rate to comply with the "Evaluation of human exposure to whole-body vibration" standard, ISO 2631. This model also had greater processing power than the previous model used during initial testing, resulting in reduced jitter, and was found to provide more accurate accelerations (within 5% up to 55 Hz). Logging of periods before and after each journey was added along with meta-data describing each journey.

Journeys performed by CenTre were recorded over the course of 12 months. Recorded variables were supplemented by calculation of ISO-weighted vibration parameters. The final dataset comprises 1,487 journeys over 81,901 km and 1,318 hours. Strong similarities between meta-data and officially reported transport data suggested there was no bias in the journeys that the staff recorded.

Roads driven between Nottingham City Hospital (NCH) and Leicester Royal Infirmary (LRI) were chosen as a case study. Data from 588 journeys contributed towards the analysis.

A range of metrics, derived from previous studies and adult standards, were used to assess the roads of the NCH to LRI network. Both speed and road classification were found to influence vibration and noise level, however the influence could not be separated due to the inherent link between both parameters. All routes involved either use of motorway or a concrete A-road, with the latter producing worse vibration. Although individual road sections varied, differences were reduced between the routes.

Assessments were also performed of the metrics at each of the 42 hospitals (36 departing; 38 arriving) present in the data. Results were similar between hospitals, but differed between loading and unloading phases. High magnitude shocks were more abundant during the loading phases, whereas low impact vibrations were more frequent during unloading. Both phases registered greater shocks than those found during journeys.

In summary, this work provides a low-cost method of obtaining large amounts of data describing the ambulance environment without requiring any technical knowledge to operate. The theory that the physical environment could be altered through routing has also been confirmed. The data collected during this work could be utilised in the future to aid determination of neonatal responses and subsequently establish optimal routes.

Publications Arising from this Thesis

The following papers and conference presentations have been based on part of this work:

Papers

Partridge TJ, Gherman L, Morris DE, Light RA, Leslie A, Sharkey D, McNally DS and Crowe JA. Smartphone monitoring of in-ambulance vibration and noise. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine* 2021.

Conferences

Partridge TJ, Morris DE, Light RA, Leslie A, Sharkey D, Crowe JA and McNally DS. *Towards reducing vibration and noise encountered by neonates during ambulance transfers by using intelligent routing*. Presented at the UK National Neonatal Transport Group Meeting, Southampton, UK, November 2019.

Partridge TJ, Morris DE, Light RA, Leslie A, Sharkey D, Crowe JA and McNally DS. Finding Comfortable Routes for Ambulance Transfers of Newborn Infants. In: *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Montreal, Canada, July 2020, pp.5905–5908.

Acknowledgements

I would like to thank Dr. Ed Morris and Prof. Donal McNally for their guidance and support throughout this whole process. I would also like to thank Prof. John Crowe for all the support he provided, along with the many challenges that made me a better researcher. An additional thank you to Dr. Roger Light for his supervision during the first couple of years and the greatly appreciated recording of his morning commutes during the development stage.

The input of Dr. Andy Leslie was instrumental in the arrangement of the ambulance data collection, as well as the understanding of clinical practices, and for that I am very grateful. I would also like to express my gratitude to Dr. Don Sharkey for providing medical grounding to this work. I must also express a huge thank you to the staff at CenTre Neonatal Transport for incorporating the use of my app into their routines; without them this thesis would be extremely short!

Next, I would like to thank Dr. Lorelei Gherman, my honorary supervisor, for always making herself available to answer a seemingly endless stream of questions. Additionally, thank you to all of my other friends for keeping me (relatively) sane throughout this process, including those at badminton and volleyball that provided some much-needed stress relief.

This work would not be possible without the funding of both the Engineering and Physical Sciences Research Council and Jaguar Land Rover, as part of an Industrial CASE studentship, to whom I am eternally grateful.

Lastly, I would like to thank my family for their continual support, especially through the write up process. And yes, it is finished now!

Contents

List of Abbreviations

r.m.s. Root-Mean-Square

AAP American Academy of Pediatrics

ABSS Anderson Behavioural State Scoring System

ANCOVA Analysis of Covariance

ANOVA Analysis of Variance

API Application Programming Interface

CAN Controller Area Network

CB COMFORT Behaviour scale

CenTre CenTre Neonatal Transport

CI Confidence Interval

CPU Central Processing Unit

CSV Comma-Separated Value

dB(A) A-weighted dB

ELBW Extremely Low Birth Weight

EMNODN East Midlands Neonatal Operational Delivery Network

EU European Union

GA Gestational Age

HR Heart Rate

HRV Heart Rate Variability

HSE Health and Safety Executive

HTTP Hypertext Transfer Protocol

HTTPS Hypertext Transfer Protocol Secure

ID Identifier

IMU Inertial Measurement Unit

IQR Interquartile Range

IRI International Roughness Index

ISO International Organisation for Standardisation

IVH Intraventricular Haemorrhage

NICU Neonatal Intensive Care Unit

NTG UK Neonatal Transport Group

OS Operating System

OSM OpenStreetMap

PHP PHP: Hypertext Preprocessor

PIPP Premature Infant Pain Profile

PIPP-R Premature Infant Pain Profile-Revised

PSD Power Spectral Density

RAM Random Access Memory

RR Respiration Rate

SCANNER Surface Condition Assessment of the National Network of Roads

SI units International System of Units

SPL Sound Pressure Level

TRIPS Transport Risk Index of Physiologic Stability

URL Uniform Resource Locator

USB Universal Serial Bus

UTC Coordinated Universal Time

VDV Vibration Dose Value

VISA Virtual Instrument Software Architecture

VLBW Very Low Birth Weight

WBV Whole-Body Vibration

XML Extensible Markup Language

List of Hospital Codes

AH Addenbrooke's Hospital

BCH Birmingham Children's Hospital

BHH Birmingham Heartlands Hospital

BPH Boston Pilgrim Hospital

BQH Burton Queen's Hospital

BRI Bradford Royal Infirmary

BWH Birmingham Women's Hospital

CRH Chesterfield Royal Hospital

FPH Frimley Park Hospital

GEH George Eliot Hospital

GH Glenfield Hospital

GOS Great Ormond Street Hospital

HH Hinchingbrooke Hospital

HRI Hull Royal Infirmary

JR John Radcliffe Hospital

KGH Kettering General Hospital

KH Kingston Hospital

KMH King's Mill Hospital

LAD Luton & Dunstable University Hospital

LAH Liverpool Alder Hey Children's Hospital

LCH Lincoln County Hospital

LGH Leicester General Hospital

LRI Leicester Royal Infirmary

MKH Milton Keynes University Hospital

NCH Nottingham City Hospital

NGH Northampton General Hospital

NNH Norfolk & Norwich University Hospital

PCH Peterborough City Hospital

QMC Queen's Medical Centre

RDH Royal Derby Hospital

RGH Rotherham General Hospital

RHH Russells Hall Hospital

RSH Royal Stoke University Hospital

SCH Sheffield Children's Hospital

SMH Stoke Mandeville Hospital

UCL University College London

UHC University Hospital Coventry

WH Warrington Hospital

WMH Walsall Manor Hospital

WNC New Cross Hospital

WRH Worcestershire Royal Hospital

WWH Warwick Hospital

CHAPTER 1

Introduction

1.1 Project Aims

The overall aim of this work is to explore whether it is possible to reduce the discomfort that neonatal infants undergoing ambulance transfer experience by navigating along an appropriate network of low vibration and noise inducing roads.

Neonatal infants that are transferred have an increased risk of severe brain injury compared to those born at hospitals with the required care facilities. The general consensus is that the levels of vibration and noise that occur during transfers are contributing factors to the observed increase, however the true influence is not understood.

One proposed method of reducing the levels of noise and vibration inside the transport incubator is to decrease the levels that enter the ambulance. This work intends to determine whether this can be performed through routing along smoother and quieter roads.

Data on the levels experienced as a result of driving along different roads are required to be able to carry out comparisons. Data logging equipment is typically expensive, complex and consists of multiple components. The use of the self-contained sensors built into smartphones will therefore be investigated as a means of collecting large-scale data, with a new app being developed for use in this work and available for future research. This app will subsequently be used for data collection with the aim of gathering more data than previous studies while also providing greater levels of detail.

Maximising the amount of data involved in the assessments of roads is vital to provide increased confidence in the levels of noise and vibration experienced by reducing the impact of anomalous events. This requires the formation of a network that consists of unique sections of road where the nodes are the junctions at which journeys diverged. The use of a bespoke algorithm to perform segmentation automatically upon reception

of new recordings will therefore be investigated. This would enable analysis to be maximised during the data collection period where new routes may be used, as well as facilitating continual analysis in future collections.

Comparisons of the levels of noise and vibration that occur during both road and hospital sections of neonatal transfers will be conducted to visualise the scope for improvement. Although the impact on the infants cannot be determined, predictions can be formulated, influences can be identified and, ultimately, the data will be able to serve as a thorough reference of the potential environmental conditions experienced for future research in the area.

As a final point, this work will aim to provide a set of recommendations that future work on the physical environment within ambulances should follow to ensure data are accurate, fully understood and comparable between studies.

1.2 Project Structure

Before collecting any data, the concept of neonatal transport is introduced in Chapter 2, covering the extent of noise and vibration during transfers, the potential effects on the infants and an overview of the methods that have been employed to reduce them. An overview of the existing methods of capturing the relevant data along roads is also presented along with the capabilities of smartphones as logging devices.

The development of an app to record noise, vibration and location data during journeys is detailed in Chapter 3. Chapter 4 subsequently presents validation of the data sampled by the app and investigates the effects of factors such as road surface, background noise and smartphone model.

Chapter 5 outlines the steps taken towards a controlled collection of data within neonatal ambulances. This includes the selection and subsequent validation of an optimal smartphone as well as the physical and software modifications implemented. The extent of the data collection is shown in Chapter 6 as well as the stages to the preparation of this data for analysis.

Chapter 7 provides a detailed description of the algorithm created to divide roads into unique sections. The effectiveness and accuracy of this algorithm is assessed for a selected group of roads, with the final segmented network also detailed.

Levels of noise and vibration are assessed in Chapter 8 for the roads of the network identified within Chapter 7. This involves the definition of a range of metrics to be used before investigating potential influencing factors and the spread of values be-

tween routes.

In Chapter 9, the metric results at different hospitals are compared during both the loading and unloading sequences. This involves the comparison both between hospitals and between the two stages of transfer at a single site.

The final chapter provides a summary of the work conducted here, along with a discussion of how the data compared to previous work, what the potential implications may be, and how the work should be furthered.

CHAPTER 2

Background

2.1 Introduction

As introduced in the previous chapter, this work is concerned with the levels of discomfort that are experienced during neonatal transport. To begin addressing the problem, the population that are involved and their susceptibilities need to be understood.

Next, the concept of neonatal transport and the need for it in modern practice will be detailed. The aim of this work is to identify whether the comfort of transport can be improved and therefore the concerns that have arisen require outlining along with the potential physiological connection.

The neonatal response to physical stimuli is not fully understood, however there is a general consensus that these affect the physiology of the infants undergoing ambulance transport. Previous work investigating the physiological inputs of vibration and noise, regarding both the extent at which they occur during transfers and the current knowledge on how neonatal infants are affected, will therefore be explored.

Methods of reducing levels of both vibration and noise have begun to be investigated due to the consensus that they are harmful to the infants. The work on the topic, both experimental and theoretical, will be reviewed to assess the feasibility and effectiveness of the different options. This section will summarise the methods with the aim of justifying the interest in routing of this thesis.

Data are ultimately required on the levels of vibration and noise on different roads to establish the scope that routing may provide in decreasing discomfort. The final section will therefore outline the various methods that have been used to gather similar data with an aim of deciding on the optimal choice for this work.

2.2 Neonatal infants

First, it is necessary to understand the population that are most at risk during ambulance transportation. These are neonatal infants, also known as neonates, which are defined as infants between birth and 28 days of age [1], although hospital staff designate an infant as neonatal until they are released home. Between 600,000 and 700,000 infants are born each year in the UK [2], with around 90% of these allowed to go home shortly after birth. The remaining 60–70,000 neonates, however, require some degree of specialised care [3].

Neonatal infants born before 37 weeks Gestational Age (GA) are termed 'preterm' or 'premature' and tend to require the most care. Preterm births account for roughly 8% of all births [4], with most of these occurring after 32–36 weeks GA although a quarter of these occur before 32 weeks GA. The lack of time in the womb results in the underdevelopment of preterm neonates, with the risk of adverse outcomes decreasing with gestational length.

All preterm infants are at greater risk of both mortality (death) and morbidity (illness) than those born after the full 40 weeks gestation. The rates of both death and major morbidity, in the form of Intraventricular Haemorrhage (IVH), necrotising enterocolitis and seizures, increase below 32 weeks GA and are at the highest levels at 25 weeks GA [5]. Multiple studies have shown that being born premature can lead to hardship in later life — in the form of neurodevelopmental disabilities [6, 7], low intelligence [8] or neuromotor impairments [9] — with IVH being a key contributor.

While improvements in care have led to a reduction in rates, mortality and morbidity are still prevalent. Yeo et al and Chen et al examined the trends in outcomes of infants born after fewer than 32 weeks GA. Both studies reported similar decreases in mortality of around 4% (11% to 7% in Australia and New Zealand [10]; 12.6% to 8.2% in Switzerland [11]) and decreases in severe IVH (4.0% to 2.8% [10]; 7.8% to 5.8% [11]). Although the mortality and morbidity rates are dropping, they are not eradicated and therefore more work is required.

Studies have shown that access to a Neonatal Intensive Care Unit (NICU), and the specialised care which is provided, results in reduced mortality rates [12–15] although no improvement in morbidity [13]. Ideally, all premature infants would be born at a hospital with an established NICU — as recommended by Watson et al. [12] — this is often not the case due to the centralisation of neonatal care.

In 2003, the UK Department of Health set out some recommendations on improving the care of neonatal infants by concentrating specialist care in fewer hospitals [16]. This de-

cision was made due to a lack of staff, either experts or otherwise, available to operate wards effectively at all hospitals. Instead, networks were formed of groups of hospitals, with these hospitals designated a level from 1 to 3 depending on the facilities present. A large number of hospitals fall under level 1 (primary), which possess facilities to provide both "routine" and "special" care, or level 2 (secondary), which also have the ability to provide short-term intensive care [16]. Long term intensive care (NICUs) are only present at level 3 (tertiary) hospitals [16], of which there are only a few in each network. While dividing up the hospitals this way led to improvements in the delivery of specialist neonatal care, it also led to an increase in the number of neonatal transfers between hospitals within 28 days of birth [17].

2.3 Introduction to neonatal transport

Neonatal infants require moving from a non-tertiary hospital to a tertiary hospital in the event that intensive treatment is required, due to the availability of the relevant facilities. The need for organised inter-hospital transfers of these infants was recognised immediately after the recommendation for centralisation [18]. The series of considerations which were outlined ultimately led to the formation of the UK Neonatal Transport Group (NTG) [19] in 2006.

The NTG consists of 16 teams around the UK, which operate with the common goal of improving the outcomes of critically ill infants who are not born at a tertiary hospital and therefore require transportation [20]. Collectively, these teams perform between 15,000 & 16,000 neonatal transfers each year [21], where half of the transfers involve moving infants to specialist facilities and the other half are to create space in the NICU by either returning a stable infant to the hospital of birth or moving the infant to an available cot elsewhere.

Although the NTG have focussed their efforts on improving the efficiency and organisation of transfers, along with the medical practices, a worrying trend has been identified in the outcomes of transferred infants.

2.3.1 The problem with neonatal transport

The key problem with the inter-hospital transportation of neonatal infants is the impact it has on the medical outcomes. This is predominantly with extremely preterm neonates that were born after fewer than 28 weeks GA [22, 23]. Although the transfer of infants was found to have no effect on the mortality rate, Helenius et al identified

levels of severe brain injury occurring more than twice as often in transferred infants compared to those born at hospitals with adequate facilities [22]. Shipley et al corroborated these findings when focusing specifically on the rates of IVH, finding a 184% increase in infants with fewer than 28 weeks gestation which reduced to a 69% increase when including all infants up to 32 weeks GA [23].

An increase in IVH rates is not solely a UK concern, with several other countries conducting investigations. In the USA, Mohamed and Aly found that the rates of IVH increased when examining the outcomes of Very Low Birth Weight (VLBW) neonates [24]. Infants are classed as VLBW when they weigh less than 1500 g at the time of birth. All low birth weights are a result of preterm birth. Mohamed et al found that the transfer of VLBW infants led to a 44% increase in severe IVH and 75% in overall IVH levels [24]. The study was inconclusive as to whether infants that weighed less than 1000 g at birth — termed Extremely Low Birth Weight (ELBW) — were at greater risk during transfers compared to VLBW, as they recorded a greater increase in overall IVH levels (91% vs 47%) but a lower increase in severe IVH (36% vs 60%). What is confirmed is the negative impact of the inter-hospital transfers.

2.3.2 Stress during neonatal transport

Although it is possible that the medical needs for which the infants required intensive care, and therefore the need for transportation, could be a cause of the increased IVH rates [25], the studies by Helenius et al. [22], Shipley et al. [23] and Mohamed and Aly [24] made attempts to control for these factors. In an effort to show that the transportation itself is the problem, several studies analysed the levels of stress and discomfort experienced by the neonatal infants during the journeys.

Multiple different indicators of stress were implemented during the studies on the effect of ambulance transportation on neonatal infants. One of which was the Premature Infant Pain Profile (PIPP) — a score calculated from a combination of GA, physical observations and changes in both Heart Rate (HR) and O_2 saturation, with a higher score indicating greater discomfort [26] — used by Harrison and McKechnie to assess the impact during 140 separate transfers [27]. They found that the PIPP scores were at their highest during the transfer compared to when being handled in and out of the transport incubator as well as while in the unit before and after, with PIPP score increasing with lower GA.

Zwissig et al presented similar findings using the Premature Infant Pain Profile-Revised (PIPP-R) — a modified version of the PIPP addressing clinical feedback and ensuring

scores are valid for GAs of <28 weeks [28] — showing marked increases during the transfer of 20 neonates in Switzerland [29]. The COMFORT Behaviour scale (CB) [30] was also used in the study by Zwissig et al, however movement within the hospital outscored the transfer itself. This may be because the CB is both observational-only and designed for general paediatrics, and therefore may not be as accurate an indicator as the neonatal-specific PIPP and PIPP-R.

An additional assessment of the outcome of ambulance journeys which can be conducted is the Transport Risk Index of Physiologic Stability (TRIPS), which was designed specifically for neonatal transfers [31]. This scoring system was constructed with the aim of predicting the likelihood of mortality after transfer, either within 7 days or the full length of the infant's NICU stay, and scores the variables of temperature, respiratory status, blood pressure and stimulus response. This has been shown to be highly effective, with high TRIPS scores also significantly indicating a high risk of adverse outcomes [32]. Snedec et al. [33] implemented the TRIPS score in their study on the Heart Rate Variability (HRV) outcomes of transport. Although the TRIPS scores were near-identical, it was found that a higher HRV on arrival was linked to a reduced HR and a shorter duration within the NICU, suggesting that HRV is an adequate indicator of discomfort during transport.

Finally, Grosek et al assessed stress levels by recording both the HR and leukocyte counts of the transferred neonates [34]. Although the effect of transfer on individual infants was not distinguishable, Grosek et al identified a clear increase in both HR and leukocyte counts — indicating heightened stress — when transported by ambulance during daytime (7am to 7pm) compared to night-time which may provide some insight into potential causality. Typically, there is less traffic on the road at night which could result in smoother driving as traffic lights often change sequencing, for example. Similarly, the general noise level may be reduced due to fewer vehicles, and their engines, are in the vicinity. Transporting infants by helicopter during the daytime was also investigated and found to be similar to the road ambulance at night [34], however it is unknown which factors may have contributed towards the stress, or lack of it, as the two environments are mismatched.

The general consensus has shown that transferring neonatal infants by ambulance results in an increase in the stress levels, which may in turn be an explanation for the increased rates of IVH [35]. With the link between ambulance transfers and stress clear, the next step is to understand what factors may be affecting the infants.

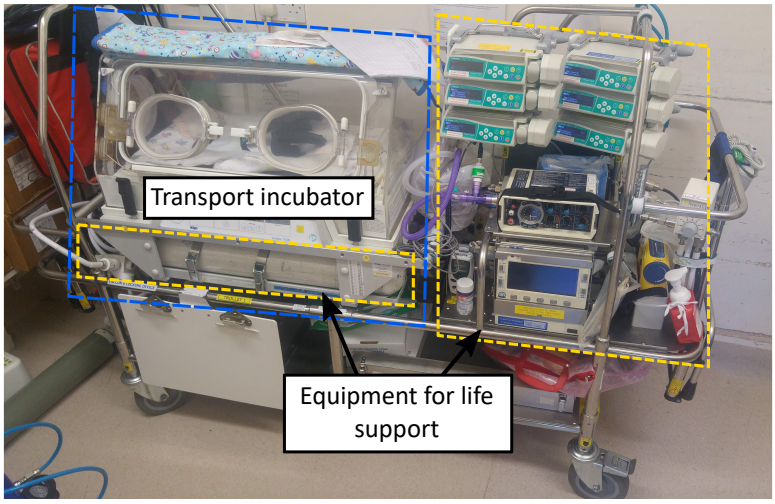

Figure 2.1: Components which form a trolley for neonatal transport in the UK.

2.4 Potential causes of discomfort during transport

Before the environmental stressors can be explored, the physical components of neonatal transport must first be outlined. This thesis will focus on the use of road ambulances, as they are used for 98.8% of all transfers of neonatal infants in the UK [21], with the remainder conducted by either fixed-wing or rotary aircraft.

The transport equipment used for neonatal transfers consists of a transport incubator, along with other medical equipment, attached to a trolley (Figure 2.1). All equipment involved in the transport are subject to the British standard 13976-1:2018 Rescue systems - Transportation of incubators [36] which specifies that the safety of other passengers must be ensured if subjected to a shock of 10 g. Passengers are deemed safe if the components, including the patient, are displaced by a maximum of 150 mm. Manufacturers ensure compliance with this standard by rigidly attaching all equipment to the trolley, and then the trolley to the floor of the ambulance. It is believed that this rigid attachment could be a contributor to the discomfort of neonatal transfers.

Driving along roads is known to subject vehicles, and their occupants, to potentially distressing levels of vibration and noise. Although manufacturers implement measures to reduce the levels which penetrate the passenger cabin by using sophisticated suspension and soundproofing materials, the most effective measures tend to be limited to expensive, luxurious vehicles [37]. Instead, more general consumer vehicles resort to compensation methods such as increasing the volume of the audio system to account for the engine and road noise produced at higher speeds [38, 39]. Ambulances are essentially commercial vehicles which have been adapted to carry, and care for, patients without consideration for what effect the journey may have. Any vibration

or noise which leaks into the ambulance cabin may therefore be a factor in the stress experienced by the neonates.

The following sections will, separately, explore the levels of vibration and noise which occur during neonatal transport, along with the possible effects they may have. These will then be summarised and compared to ascertain probable cause of discomfort.

2.4.1 Vibration during transfers

Multiple studies have investigated the levels of vibration and acceleration during the transport of neonates, both in comparison to physiological results and in terms of solely quantifying the potential environment. The following sections will explore the links between vibration and discomfort, first by comparing the recorded infant response where available and then through different approximation methods.

2.4.1.1 Comparison of vibration and physiological outcomes

The response of neonatal infants when subjected to vibration is not fully understood due, in part, to the ethical considerations in conducting systematic testing. Instead, there are two available routes for assessing the impact of vibration during transport: comparing against physiological data or approximating using the known adult response. Although having the potential to identify the scale of the harm to infants, physiological comparisons were only performed during three studies [40–42], with the majority recording acceleration values alone.

Both Bailey et al. [42] and Karlsson et al. [40] recorded TRIPS scores during their studies of vibration inside neonatal ambulances. Ultimately, the use of TRIPS did not yield any clear results. This was either due to no significant changes witnessed when scoring in 15 minute intervals [42] or because TRIPS was investigated alongside vibration as a contributing factor rather than an outcome [40].

More significant results were achieved by comparing the recorded vibration levels with the HR [40]. In their study, Karlsson et al found that higher magnitudes of vibration appeared to reduce the HR, and therefore the stress, of the neonatal infants, although this was a study of only 16 patients. No significant evidence, however, was found during the study to suggest any relationship between vibration and HRV. Hall also recorded HR, along with other physiological data, during her study on the effect of speed and road conditions [41], however the analysis was limited. Comparisons were made with generalised road classifications, however the presented results appeared to overlap and no statistical analyses were performed. Hall did note that reducing the

amount of braking and accelerating — often a characteristic of roads in urban areas — during a transfer would be optimal as it led to patient stabilisation, although the final population of her study was even lower than Karlsson et al and greater patient numbers would improve confidence in the findings.

Studies which involved neonatal patients during the data collection but did not record, or at least did not report, the physiological response do not help the limited knowledge in this field. The extra sets of data could have reduced some of the uncertainty in the previous studies or even provided some new context as to what links may exist with vibration. One study explicitly stated that patient data was not collected in order to bypass the need for ethical approval [43] which can affect projects with limited capacities for time. Although Blaxter et al. [44][1] said that the aim of their study was to provide baseline exposure levels, the addition of physiological data would have further added to the usefulness of the work. Ethical approval was given to the study, which involved placing custom accelerometers in contact with the skin of patients, and therefore the lack of additional data recording was presumably due to some other factor.

The most common reason for the lack of physiological information in the studies on vibration during neonatal transport is because no neonatal patients were being transferred when measurements were taken. Several studies simply used manikins as a placeholder for a real infant during feasibility test runs and comparisons between equipment configurations. Although a wide range of test dummies have been designed for automotive assessment, they are not designed for recumbent positions and the smallest is a 6 week old healthy newborn which, at almost 3.5 kg in mass, does not represent the population at risk. Studies have therefore either used medical resuscitation devices [44–46] such as the Life/form Micro-Preemie Simulator (Nasco, Fort Atkinson, WI, USA) — designed to simulate an infant after 25 weeks GA — or constructed their own dummy to represent the expected weight distribution [47]. A further study used two manikins with masses of 300 & 2,000 g, however it is not clear whether these were medical or custom-made [48].

2.4.1.2 Health impact of vibration during transport

In the absence of clear known links between vibration and the response of neonatal infants, the response could be approximated using the adult response, which has been the focus of numerous studies over the years and is outlined in the International Organ-

[1] The journal article by Blaxter et al [44] is referenced on multiple occasions throughout this work. It is important to note that several of the authors were also involved in parts of the work here, however care was taken to remain impartial in its interpretation and assessment.

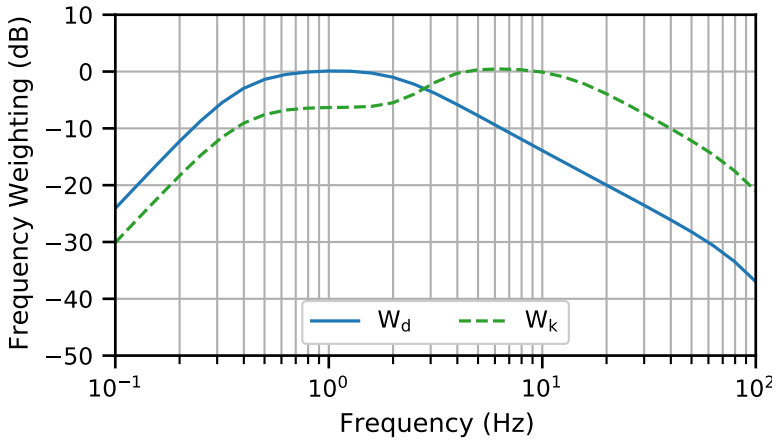

Figure 2.2: Frequency weighting curves for the W_d and W_k weightings used in ISO 2631 [49].

isation for Standardisation (ISO) standard 2631 [49]. This standard, entitled the "evaluation of human exposure to whole-body vibration", sets out multiple components for Whole-Body Vibration (WBV) assessment. The first of these are the frequency weightings which adjust the magnitude of recorded accelerations according to the known effects depending on both the position of the body involved as well as the contact point for the vibration. Different weightings apply depending on whether the accelerations are in the horizontal (W_d) or vertical (W_k) planes, with the main frequencies of interest occurring between 0.6 & 10 Hz (Figure 2.2), derived by assessing adult perception during various studies. These weighted accelerations are then summarised by calculating the r.m.s. (Root-Mean-Square) value, or an additional measure of Vibration Dose Value (VDV) that accounts for peaks in vibration, which can then be evaluated in terms of either health (or comfort, which will be the focus of the following section).

Evaluating the health impact of vibration is performed against recommended thresholds for an 8-hour workday. By combining the r.m.s. values with the durations in which they occurred to produce, an exposure value is calculated, termed as $A(8)$ or written in units of $\mathrm{m{\cdot}s^{-2}}$ $A(8)$. The risk to health can then be ascertained by comparing the calculated exposure values to the action ($0.5\ \mathrm{m{\cdot}s^{-2}}$) and limit ($1.15\ \mathrm{m{\cdot}s^{-2}}$) values stipulated by a European Union (EU) directive on health and safety [50]. This directive also sets out the equivalent thresholds to be used with VDV results, although no studies were found which included this measure when investigating neonatal transport. The $A(8)$ exposure thresholds are enforced by the UK's Health and Safety Executive (HSE) [51], typically to protect the health of workers while using power tools and heavy machinery.

Only Blaxter et al. [44] implemented the $A(8)$ exposure in their study on ambulance vibration levels. Accelerations were recorded at locations on the forehead and torso of

both neonatal infants and a resuscitation dummy. Of the 32 sets of forehead vibration, only two (6%) registered exposure levels above the recommended action limit. This increased to 4 (11%) when examining the torso exposure. The discrepancy between the two sites could be due to the locations, however it is likely that the additional 2 transfers which breached the action limit at the torso coincided with 2 of the 3 transfers where the forehead data became corrupted. This is not clear, or even discussed, from the presented results, although it was stated that both the transfers which contained failed data and the transfers which exceeded the action limit involved manikins instead of actual infants. The use of manikins may also explain the elevated exposure levels, possibly due to different physical characteristics compared to a live infant or due to drivers, consciously or unconsciously, taking less care with manoeuvres.

A further approximation of the health impact of ambulance transport was performed by Shah et al. [52]. In this study, the physiological response of neonatal infants was estimated by subjecting Sprague-Dawley rats to a form of acceleration recorded earlier by an affiliated group [45]. Shah et al determined that the input to the rats clearly resulted in degraded respiratory functionality, and therefore this is what human infants would experience during transport. Multiple flaws in the methodology employed by these two studies cast uncertainty on these findings.

The main problem with the work by Shah et al. [52], which began in their preceding work [45], is the form of acceleration to which the rats were subjected. Shah et al defined their own measure of acceleration and termed it "impulse" [45]. In physics, and therefore engineering, impulse is a set quantity defined as the integral of a force and is given in units of N·s [53], however the measure defined by Shah et al was in terms of $m \cdot s^{-2}$ per minute. Working back from the resultant values, combined with how the sampling was reported, this supposed "impulse" appears to actually be the sum of the average (presumably r.m.s., but not stated) acceleration values in each second:

$$\text{impulse (according to Shah et al. [45])} = \sum_{1}^{60} \sqrt{\frac{1}{n} \sum_{i}^{n} a_i^2} \tag{2.4.1}$$

a_i : instantaneous acceleration

n : accelerometer sampling frequency

Due to the unstandardised, and incomprehensible, measure of acceleration it is not possible to understand how the recorded acceleration is comprised. It is also highly uncertain how this aggregated measure was then converted into an input to which the rats were subjected, ultimately meaning the study cannot be replicated or used in any comparisons.

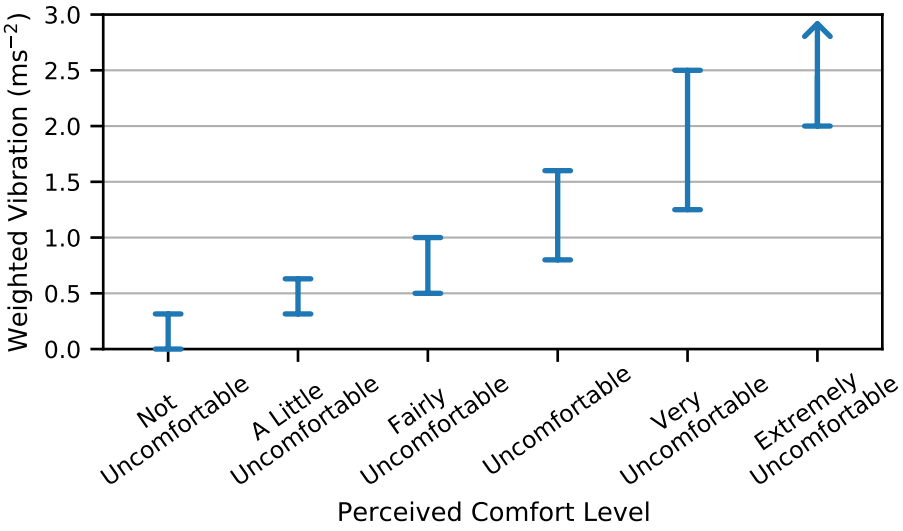

Figure 2.3: Perceived comfort levels and the vibration ranges at which they occur, as defined in the ISO 2631 standard [49].

2.4.1.3 Comfort impact of vibration during transport

Comfort is far more ambiguous than health due to relying on individual's perception of vibration, which could vary, rather than deriving it from physiological responses. Nonetheless, guidance is given in ISO 2631 [49] as to how it can be assessed. Six different levels, termed as "reactions" in the standard, are defined and range from "not uncomfortable" through to "extremely uncomfortable". Each level corresponds to a range of vibrational magnitudes as illustrated in Figure 2.3. It is important to note that — apart from the "not uncomfortable" level — all magnitude bounds overlap. This is due to the aforementioned ambiguity in assessing comfort. A further item of note is that ISO 2631 states that the true perceived comfort varied with respect to duration and other factors [49] and therefore these levels are only guidelines.

Studies tend to locate accelerometers either on the outside frame of the incubator [43, 44] or on the inside, underneath the mattress [40, 42, 48] during ambulance transport. These locations have the benefit of allowing more permanent attachment, unlike sensors attached to the infant itself which would require removal at the end of the journey. A further benefit is the lack of requirement for ethical approval due to involving zero interaction with the patient. Both Bouchut et al. [43], on the outside of the incubator, and Karlsson et al. [40], on the inside, registered similar levels of vibration with the average magnitudes within 0.05–0.10 $m{\cdot}s^{-2}$ of each other. This was either classed as "not uncomfortable" [40] or occurred towards the bottom of the "a little uncomfortable" range [43]. This similarity in levels was expected due to being connected to the same rigid structure, even if at different positions.

Although Bailey et al. [42] also used an accelerometer under the mattress, there are significant errors in their study that warranted exclusion from comparison. It was reported that vibration levels were recorded in accordance with ISO 2631 (at least in terms of axis orientation), however the given results suggested that the weightings were either applied incorrectly or not at all. Average WBV values were reportedly in the range of 66 $m \cdot s^{-2}$, around 200 times greater than the other studies and 33 times greater than the lower limit for "extremely uncomfortable" vibration. Further evidence for the incorrect handling of the accelerations is found in the vertical components where the average values are around 10 $m \cdot s^{-2}$. This suggests that the gravitational component of acceleration was not filtered out, and therefore the weightings could not have been applied.

The final two studies, performed by Gajendragadkar et al. [48] and Blaxter et al. [44], did not report the recorded levels directly, but instead performed comparisons with the neonatal head and torso.

Vibration levels at the head of the neonate were not directly reported by any study. Blaxter et al did record the accelerations at the head, but took the decision to report the levels expected from a combination of the recorded accelerations on the frame of the transport incubator and a determined gain relationship [44]. This relationship confirmed the findings of Gajendragadkar et al. [48] — that vibration magnitudes increased between the incubator and the head of the neonatal patient — and was used to facilitate comparison between different equipment configurations that Blaxter et al were investigating. An average magnitude of 0.5 $m \cdot s^{-2}$ was calculated for neonatal transfers, with the driving style when manikins were transported increasing this by 0.2–0.3 $m \cdot s^{-2}$. The results from the transfers fall firmly in the "a little uncomfortable" perceived level, although the harsher driving moved the vibration into "fairly uncomfortable" territory.

Blaxter et al also presented levels of vibration expected at the torso of a neonate during transport [44]. These levels were towards the upper limit of the "a little uncomfortable" vibration range for the driving where patients were on board, increasing to "fairly uncomfortable" when with manikins. Levels at the torso were at slightly greater magnitudes than those at the head, but fell into the same bounds for comfort perception. Values were also reported from inside the torso of a custom manikin during a study conducted on a runway [47]. Although it was reported that the accelerations were weighted according to ISO 2631, the magnitudes were far greater than expected and ranged from the upper limit of "uncomfortable" to the upper limit of "very uncomfortable" (which is also deemed "extremely uncomfortable").

The vast difference in results compared to the other studies [40, 43, 44] suggests that the values presented by Prehn et al. [47] do not represent the levels in actual transport, although it is unknown what caused the discrepancies. The journeys recorded by Prehn et al during their study were all conducted on a runway, however these are typically designed to be smooth [54] and would presumably result in reduced levels compared to normal roads. A more plausible explanation for the increased vibration is the use of American truck-style ambulances as opposed to the modified vans used in the UK and around Europe. Unfortunately, there are no further studies with which to compare against as both other studies in these ambulances must be discounted. This is because it is highly likely that vehicle suspension improved in the 24 years since the study by Shenai et al, who registered unweighted vibrations between 2 & 6 $m \cdot s^{-2}$ [55], while the results of Bailey et al. [42] could not be trusted as explained earlier.

2.4.1.4 Presence of shocks

Shocks are termed as sudden changes in movement [56] and in the worst cases can cause brain injury in adults [57]. Similarly, shocks from vehicular accidents and domestic violence have been shown to cause complications during pregnancy [58]. This is often due to problems with the placenta, however both increases in distress and reductions in oxygen levels of the fetus have been noted [58]. A wide range of short, sharp accelerations can be classed as shock, and therefore there is no standardised measurement.

The lack of standardisation in shock measurement led to studies reporting different quantities which could not be compared. Karlsson et al reported peak WBV, weighted in accordance with ISO 2631, over an undefined period of 2.85 $m \cdot s^{-2}$ [40] which would have been comparable with the results from Bailey et al. [42] if it were not for the errors in the latter work. Bouchut et al also recorded peak ISO 2631 1-second WBV, however this was reported as the number of events (44) with a magnitude of 2 $m \cdot s^{-2}$ [43]. Finally, Blaxter et al showed the distribution of unfiltered accelerations of less than 100 ms with magnitudes of at least 2 g [44]. The majority of mid-transfer shocks were less than 3 g and were sustained for 20 ms. Interestingly, two studies [40, 44] reported greater shocks during the loading and unloading of the transport incubator, suggesting this is an area which could be improved.

2.4.1.5 Summary of vibrational impact during transport

Links between neonatal response and the levels of vibration which they are exposed to during transport remain uncertain. Although two studies presented findings that were deemed significant, they were only conducted with populations of 16 [40] and 7 [41] respectively. More studies are required which record both physiological and vibratory data to remove any doubt in the reported results.

In lieu of a known neonatal response, studies resorted to comparing vibration levels to an adult standard [49]. This standard provided both health and comfort guidance. No neonatal transfers exposed the patient to levels of vibration at or above the action value [51]. Perceived comfort fared slightly worse, with vibration of the incubator typically around the "not uncomfortable"/"a little uncomfortable" levels and the projected infant vibration reaching "fairly uncomfortable" magnitudes at times.

Similar to the lack of studies linking vibration with neonatal response, more research is required to achieve a greater consensus on the levels within ambulances. The multiple factors which could affect vibration, such as the driving style and the road quality, can vary greatly between team members and routes, respectively. This suggests that more than the current maximum of 35 total journeys [44], or 10 hours when looking at true transfers [40, 43], are needed to showcase the range of values that could be expected.

As well as the need for more knowledge of both the levels of vibration during transport, and the potential risk, there is a clear need for standardisation in the field. While the studies which have been performed may have experienced similar vibration, the lack of consistency in the reporting significantly reduced the ability to compare their results. All studies should be required to report the equipment used for sensing accelerations in full, along with the rate at which data was sampled, in order to know that the same information was collected. Following on from this, any weightings used should be explicitly outlined. Ideally, to broaden comparability, vibration levels should be reported in both the raw sampled form as well as with any post-processing to encompass previous studies and potentially create links between them.

A key part of the problem with the field of vibration with regards to neonates is that it is interdisciplinary by nature. The majority of studies have been performed by medical professionals, for which their insight in understanding physiological effects is clearly needed. Often overlooked, however, is the engineering expertise required for the vibratory analysis. This has subsequently led to questionable units being used [45] and visibly incorrect values being reported [42], both of which hinder the expansion of knowledge in this limited area. It is therefore recommended that all future studies are

conducted collaboratively between both medics and engineers to ensure all of the required information is reported. Unfortunately, the process of peer review cannot be a substitute for collaboration as these studies are often submitted to journals which specialise in only one of these fields and, as shown, still resulted in the publication of the aforementioned works containing errors.

2.4.2 Noise during transfers

The levels of noise during neonatal transfers have been investigated by several studies that, similar to those which investigated vibration, either compared the results with physiological data or reported values to prompt further work. Here, the links between the recorded environment and distress will be explored in terms of both average levels and peak events.

2.4.2.1 Background on noise with respect to neonatal infants

Before exploring the levels, and subsequent effects, of noise during transport, the general response of neonates should be explored. There are multiple studies on the effect of noise on neonatal infants, along with guidelines for recommended levels. Research on noise is likely to be more abundant than that investigating vibration due both to the presence of noise on the NICU (where vibration is unlikely due to limited movement), the ability to recreate sounds without specialist equipment, and the non-contact nature of sound-waves (removing the need for equipment to interact with the infant, and therefore negating some of the ethical hurdles).

Potentially hazardous noise levels are typically reported with units of A-weighted dB (dB(A)), where frequencies have been filtered to approximate the human response to noise. This weighting was created to resemble the perceived loudness of different frequencies [59] and is defined in the British standard on the specifications of sound level meters [60]. A-weighted noise levels are used to assess the average exposure in the workplace alongside C-weighted levels, which include lower frequency components, that are used to assess peak values [61].

The American Academy of Pediatrics (AAP) issued a recommendation that noise on NICUs is kept to an upper limit of 45 dB(A), as any level above this is deemed "of concern" [62]. This level is not given any medical backing and is simply repeating the general advice given in 1974 by the American Environmental Protection Agency [63] that was to promote effective communication and undisturbed sleep. Despite this 45 dB(A) guideline, the average sound levels within a functioning incubator were greater by

between 2 and 13 dB [64–66], depending on the study. Although the ambient levels varied, both Knutson [64] and Altuncu et al. [65] found that additional events could increase the amount of noise significantly.

The fact that excessive and sudden noises occur on the NICU meant that there has been considerable interest in the effects on infants, while studies are easier to conduct as they are not constrained by time or location (as with neonatal transport which are subject to requirement). Research into the effects of noise found that both increased ambient levels [66] and sudden spikes [67, 68] caused alterations in cerebral blood flow, although these studies reported contrasting relationships which may have been due to their small population sizes. Similarly, both full-term [69] and very preterm neonates [70] have displayed behavioural reactions to sudden stimuli, adding further evidence to the need for noise control.

Noise appears to have an effect on neonatal infants, however more work is needed to produce definitive relationships due to the small sample sizes, possible differences in response at various GAs, and insufficient randomisation of studies [71]. Nonetheless, the studies which have been performed can serve as a useful reference for which to compare the noise during transport and the potential harm it may cause.

2.4.2.2 Noise within the ambulance environment

A recommended noise threshold of 60 dB(A) is given for the inside of transport incubators by the system requirements part to British standard 13976 [72]. This is similar to the NICU guideline given by the AAP [62] in that the rationale for the choice in value is not provided, however it has served as another reference point against which studies can compare.

Two studies recorded physiological measurements during neonatal transport alongside the noise levels. Bailey et al noted TRIPS scores at 15 minute intervals, however witnessed no significant changes [42]. Karlsson et al. [40] produced more significant findings which suggested a link between greater noise levels and an increased HR, similar to the NICU studies on peak events [67, 68], although no clear link was found between noise and HRV.

Both the NICU and transport thresholds were exceeded in all studies which recorded noise inside transport incubators, with the reported average levels occurring between 64.5 & 74.1 dB(A) [40, 42, 43, 47, 73]. Each study determined that the noise was caused by the ambulance movement, as levels were lower both before and after journeys. Buckland et al. [73] also provided a comparison of urban and rural roads, with the

latter resulting in an average increase of 9 dB which the authors suggested could have been due to either the road surface or difference in vehicle speed.

Peak events were recorded by several studies, however it is not as standardised for neonates as the average levels and therefore it was reported in different measures. The most common measure of peak noise used was maximum A-weighted values. Results varied from as low as 71 dB(A) [42] up to almost 88 dB(A) [47], with an increase again noted along rural roads compared to urban [73]. Bouchut et al. [43] also recorded peak levels, however these were unable to be compared as only the number of events above an arbitrary threshold of 85 dB(A) was reported. Finally, Buckland et al. [73] reported peak C-weighted values of 109.4 & 115.4 dB(C) for urban and rural roads, respectively, although it is unclear how this may relate to infant response.

2.4.2.3 Summary of the impact of noise during transport

Noise levels during neonatal transfers are nearly always above the recommended levels. There is a clear need to reduce the noise to which the infants are exposed, however further work is required to ascertain an ideal level due to the lack of grounding in the guidelines.

It is also apparent that noise can cause distress to neonatal infants, however causal links are uncertain. Further work investigating correlations between noise levels and physiological outcomes are required due to the small population sizes and limited number of stimuli assessed. Similarly, a greater amount of data should be collected on the noise levels experienced inside ambulances to provide insight both into exactly what the infant is exposed to, as well as dissect trends in relation to road types, vehicle speed, and other variables.

One positive aspect to the study of noise in relation to neonatal infants is the use of standardised measurements. This facilitated straightforward comparisons between studies performed both in NICU settings and in transport, unlike the vibration discussed earlier. Most studies also reported the details of the device used, including which frequencies were sampled, which further increased confidence in the comparability of data.

There is a need for collaboration between disciplines that should again be recommended for future studies involving noise. Although the standardised weightings for human hearing were applied in all studies, it is unclear whether the researchers fully understood how these values are calculated and therefore how they can then be processed. It appeared as if values were simply received from the sound meters (or whichever device was used) and then treated as a linear amount, as shown by the use of the term "mean"

when the papers discuss averages. In actuality, each noise level which gets reported to the researchers is the logarithm of the r.m.s. of all recorded Sound Pressure Levels (SPLs) within the preset time period [61], set as values ranging from 1 second [43] to 1 minute [73] in the studies here. Instead of presenting the averages as the mean of the provided values, each noise level should instead be converted back to SPLs, combined by taking the r.m.s., and then converted into decibels again. The true averages would therefore be slightly larger than if the mean was used (although slightly smaller than calculating the r.m.s. of the decibel values).

2.4.3 Overall state of the in-ambulance environment

Consensus from all studies has shown that ambulance transfers subject infants to levels of vibration and noise which are not potentially damaging. The full extent of the discomfort experienced by neonates due to these physical stressors is not fully understood, however the noise levels are above the issued guidelines while the vibration would be termed at least "a little uncomfortable" for adults.

Until there is conclusive evidence, either showing that vibration and noise are harmful to the infants or not, it must be assumed that the recorded levels perform some role in increasing the risk of IVH. Therefore, reducing these levels may then result in fewer instances of adverse outcomes in the transferred neonates.

2.5 Reducing the negative impact of ambulance transfers

There is a clear need to reduce the effect that ambulance transfers have on neonatal infants, where the ideal solution would be to negate the need for these transfers in the first place. By transferring infants in-utero, the rates of severe neonatal morbidity have been found to be reduced [74]. This suggests that women identified as being pregnant with a potentially at-risk child should be transferred to a specialist hospital before giving birth, as opposed to the infant being transferred after the event. It is often not possible to perform these antenatal transfers, however, due to time constraints in preparation and other factors. Therefore, improvements to the transfers themselves are required.

The following sections will give an overview of the methods — both experimental and theorised — which could be implemented to reduce the harshness of the in-ambulance environment. As earlier, methods will be separated into whether they concern vibration or noise.

2.5.1 Reducing vibration during transport

Reducing the amount of vibration inside vehicles is a well-established field of research within the automotive sector. Reductions are performed mainly through the use of suspension systems between the vehicle axles and the main body, although components such as the tyre size also have an effect [75, 76]. These modifications can vary in both complexity and cost, but are ultimately the responsibility of the manufacturers as they can affect how the vehicle handles during driving. Reducing vibration by selecting a vehicle with improved suspension would be both expensive (due to the cost of purchase and then installation of all required equipment) and slow (because of the vast number of ambulances, both in the UK and worldwide, which would need replacing). Instead, methods of reducing vibration levels to the infant are required which would work both in existing ambulances, as well as any future ambulances, and could be implemented relatively quickly.

Methods largely fall into two groupings: those focused on the inside of the incubator and those focused on the outside. These methods will be discussed below to determine their feasibilities and effectivenesses.

2.5.1.1 Suppressing vibration within the incubator

A popular theory for reducing neonatal vibration is through the use of an optimal mattress for the infant to lay upon. Tests have been performed of several different mattresses of various construction, filled with standard foam, memory foam, gel or air [44, 45, 47, 48]. Manikins were used for all comparison studies as researchers were unsure whether vibration would improve or get worse, and therefore did not want to cause extra distress to patients, but also because several mattresses used were not designed for use in transport.

Methods varied between studies, however the results were largely in agreement. The accelerations recorded inside ambulances at the incubator level were found to have two main energy concentrations related to the suspension characteristics (1–2 Hz) and wheel hop (10–15 Hz) with most accelerations occurring below 20 Hz [44, 48]. All studies also found that vibrations were amplified from the incubator to the manikin, regardless of whether a mattress was in use [47, 48] or of the specific configuration being tested [44, 47, 48].

The Squishon gel mattress (Philips Healthcare, Amsterdam, the Netherlands) was observed to reduce vibration levels during transport the most of any single-mattress configuration [44, 47, 48], although combining the Squishon in series with either a

foam [48] or an air [47] mattress reduced magnitudes further. In a similar study, Shah et al. [45] explored the effect of different mattress (and pillow) configurations during movement solely within the hospital. Combining foam and air mattresses was reported to provide the least vibration, however the applicability of this work is questionable as the vibrations inside an ambulance are undoubtedly different to those in hospital corridors, and the presented quantities can not be compared to the other studies due to the units used (see Section 2.4.1.2).

It is important to note that neonatal ambulances transport infants of various masses, from the ELBW infants at less than 1 kg to full-term and grown infants weighing over 4 kg. The effectiveness and characteristics of a mattress will vary depending on the mass on top, as shown by Gajendragadkar et al. [48]. Substituting the 2 kg manikin, used to determine the optimal mattress, with a 0.3 kg manikin shifted the natural frequency of the setup from between 8 & 10 Hz to around 15 Hz. This higher natural frequency closely matched the peaks caused by the wheel hop and therefore resulted in an even greater increase in magnitude vibration compared to the 2 kg manikin. Ultimately, it would not be possible to obtain a mattress which would be optimal for all prospective patients due to the variation in mass which would need to be taken into account.

Although the choice of mattress can influence, and subsequently reduce, the levels of vibration which a neonate may experience, it is not a feasible solution to the problem. This is partly because a mattress which improves levels for one infant may worsen them for another, but mainly because the mattress choice did not reduce vibration compared to the incubator itself. Instead of attempting to reduce the gain through use of a mattress, efforts would be better served reducing the inputs to the incubator in the first place. Eliminating vibration at the incubator level would negate the need to search for incrementally improved gains by causing the input to disappear and thereby eliminate the vibrations experienced by the patients.

2.5.1.2 Isolating the incubator from external vibrations

Reducing vibration transmission through the trolley would result in smaller exposures for neonatal infants. As mentioned earlier, all components of the transport trolley are rigidly attached to comply with British standard 13796 [36], and therefore vibration is passed directly from the chassis to the incubator. The standard does not explicitly state that components must be rigidly attached, but rather that they must not displace more than 150 mm in the event of a 10 g collision. This suggests that interfacing methods could be introduced to dampen vibrations, as long as they also comply with the

maximum allowable displacement.

Targetting the whole isolator instead of solely the infant, as is the case with the choice of mattress, reduces the variability of the system and enables much more focused vibration reduction. Whereas the mattresses had to adapt to mass differences of over 400% from the smallest infants to the largest, the mass of the incubator should render the difference between infants negligible. The incubator system, which comprises of the standard transport incubator used by teams around the UK and worldwide — the TI500 (Dräger, Lübeck, Germany) — plus two gas cylinders which load into the base, has a total mass of over 62 kg. This would then mean that the size differences between infants would only alter the system mass by a maximum of around 8%, which is a far more manageable amount.

Two projects appear to have investigated the feasibility of reducing vibration to the infant through isolating the incubator. Bailey-Van Kuren and Shukla [77] performed a study which utilised an array of air springs between the trolley and the incubator, determining that a system with one air spring in each corner would provide the best passive isolation. It is unclear what values of reduction were obtained, however the researchers had sufficient confidence in their implementation to patent the system [78]. Similarly, the manufacturers of the TI500 also filed a patent incorporating vibration damping in a full transport incubator system [79], however there is no sign of it on the consumer market 6 years later.

Although Bailey-Van Kuren and Shukla [77] also signalled an interest in solving the vibration problem by applying magneto-rheological damping, there is no evidence of any further research being performed. Although magneto-rheological dampers, along with other active control systems, can be used to actively reduce vibration transmission and are used in multiple consumer vehicle models, they are not suited for use as part of the transport trolley due to the limited supply of power. British standard 13976 [36] stipulates that the transport incubator system consumes too much electricity and could result in problems regarding lack of power to the whole ambulance or overheated converters.

Research has also been conducted with the aim of reducing vibration to the whole trolley, rather than between the trolley and incubator. Different materials designed for vibration isolation were positioned under the wheels of the incubator system by Prehn et al. [47] in an attempt to decouple vibration from the ambulance chassis. Minimal numerical evidence was provided, however the authors reported that none of the chosen materials produced significant reductions.

Ultimately, work on reducing vibration within the transport trolley system is severely

lacking in both amount and in detail. Further investigation is recommended before any conclusions can be accurately determined.

2.5.1.3 Reducing accelerations through road choice

One method of potentially reducing vibration would be to select a route to drive along which would result in a smoother ride. This would be able to be implemented with near-immediate effect and with no physical modifications to any equipment, unlike any method involving dampening. The idea of routing was briefly explored by Blaxter et al. [44], however the researchers were unable to compare motorways with lower class roads. It was suggested that minor roads result in a harsher ride compared to 'A' roads and that there may be an optimal vehicle speed to minimise vibrations, however further work is required to draw meaningful conclusions.

Similarly, Hall [41] found that infants responded positively to being driven without interruption. This could suggest that, as well as an optimal speed or road type, route choices should make an effort to minimise the number of traffic lights, tight junctions and slow roundabouts.

2.5.2 Reducing noise during transport

Noises inside ambulances are likely generated by a combination of the interaction between the vehicle and the surroundings, the equipment on the transport trolley, and conversation between staff members. Each of these inputs have their own challenges, but produce levels which add up to those recorded during transport.

Although consumer vehicle manufacturers aim to reduce the noise levels, whether from the engine, tyres or air passing the vehicle, through the implementation of materials and minimal bodywork gaps, it is unlikely commercial vehicles put much care into the loading area (which gets converted for medical practice). It is also uncertain whether the technicians who convert the vans into ambulances place any thought on noise suppression alongside the electrical, mechanical and storage systems required. Regardless of whether considerations for noise were made in the construction process of the ambulance, it is clear that further effort could be made.

Similarly, Knutson [64] found that noise levels inside the NICU incubator were in the vicinity of the set transport threshold of 60 dB(A) [72]. Considering the addition of incubator components vibrating, the transport incubator is likely at similar levels before the addition of the ambulance noise. Reduction of this equipment noise is largely the responsibility of the manufacturer, who should aim to produce quieter machinery and

employ more rigid construction techniques.

Speech between staff members will undoubtedly increase the levels within the ambulance, although this is largely a factor of the existing environment and cannot be helped. Communication in medical environments is essential to facilitate effective care and by putting restrictions in place may cause more harm than good. Besides, it is commonly known that more effort is required to make someone understand speech when in an already loud environment [80], and therefore the other inputs to the ambulance should be suppressed in order to minimise what effect may arise from staff communication.

As with the above vibration, studies can be grouped based on whether they are internal or external with regards to the incubator. The different methods will be set out and assessed over the next sections.

2.5.2.1 Noise suppression within the incubator

Noise has a presence both on NICUs and during transport, and therefore there are studies investigating potential methods of reduction in both settings. One of the most popular methods of preventing noise from reaching the infant is through the use of earmuffs, also recommended by British standard 13976 [72]. Earmuffs are effectively an insulated cup which forms a seal over the ear to prevent sound from penetrating, with one of the most commonly used with neonates being the MiniMuffs (Natus, San Carlos, CA, USA). The MiniMuffs reportedly reduce sound levels by at least 7 dB and reduce the SPL by at least 50%, although no frequency analysis is provided [81].

Studies have been conducted to assess the effect the use of MiniMuffs has on the neonatal response, to determine if the claims by the manufacturer are valid. Duran et al. [82] and Khalesi et al. [83] each investigated the potential behavioural changes using the Anderson Behavioural State Scoring System (ABSS), which gives values of 1 to 12. Both studies reported a significant difference between infants wearing the earmuffs and those without, with the earmuffs improving the ABSS from "active sleep" (3–4) to "quiet sleep" (1–2).

Disagreement has arisen, however, in the physiological effects of earmuffs. Although studies have recorded improvements in HR (both in a NICU [83] and during transport [40]), Respiration Rate (RR) [83] and oxygen saturation [83] in relation to earmuffs, Duran et al. [82] observed no significant changes.

While earmuffs appear to have a positive effect on infants, the true relationship is not known. One study recorded the noise levels both inside and outside the incubator and determined the input for all observations was similar [82]. Little detail is available,

however, regarding the noise levels during both other studies on earmuff effectiveness [40, 83]. Similarly, while Duran et al. [82] did record the noise inside the incubator (58 dB(A)), it is unknown how much of this noise reached the ears of the infants. Assuming the earmuffs reduced levels by the minimum stated in their specifications, this would still subject the infants on the NICU to more noise than recommended [62] and presumably greater still during transfers where the ambient levels are higher.

A related study was performed which investigated the effect of silicone earplugs on the development of low birth weight neonates. Abou Turk et al. [84] inserted Insta-Putty Silicone Earplugs (Insta-Mold, West Boulder, CO, USA) directly into the ear canals of a randomised selection of infants, leaving them in place until either 35 weeks had passed or the infant was discharged. No significant difference was observed in the growth of infants with earplugs fitted versus those without, possibly due to the low population sizes (24 total participants; 12 ELBW infants assessed after 18 to 22 months). The noise levels which reached the ear canal of each infant is also uncertain, although the average NICU levels were around 56 dB(A) and the earplugs were determined to reduce noise by at least 7 dB (similar to the MiniMuffs).

Altuncu et al. [65] took a vastly different approach to the previously discussed studies and instead looked at the use of sound absorbent panelling to reduce noise attenuation inside the NICU incubator. The results suggested that insulating the interior walls of the incubator could reduce some noise events, however none of the events were reduced to the recommended 45 dB(A). It is also important to note that the use of the sound absorbent panelling restricts the view of the incubator interior and could therefore disrupt the standards of care. Sound absorbent panelling does possess promising characteristics compared to both earmuffs and earplugs, such as the possibility of thicker construction that would block an increased range of wavelengths, but a less permanent installation would be preferred.

2.5.2.2 Preventing noise from entering the incubator

Preventing noise from reaching the incubator is more of a challenge than vibration due to the omnidirectional nature of sound, whereas the vibration can only be transmitted through the trolley. One potential solution may be in the form of quilt-like covers which many neonatal transfer teams place over the transport incubators, with the reasons for use varying. Studies within the NICU environment have investigated the use of covers to block stimulating light levels [85], although no significant difference was shown in infant response [86], while Macnab et al. [87], on the other hand, recommended the use of covers with thermal properties to reduce the potential for cold stress during

transports. It is uncertain, however, how well the covers designed for these purposes suppress noise.

Both Prehn et al. [47] and Kellam et al. [88] have performed experiments using covers which the researchers themselves created with the purpose of preventing noise from reaching the incubator[2]. Each team used different constructions (densified polyester versus acoustic foam) and different scenarios (transport versus NICU). Testing during transport was found to present similar levels of noise both with and without the acoustic cover [47], however tests were brief and performed on a runway so likely do not represent true road conditions and variability. The NICU study appears a lot more effective at higher frequencies, likely due to the similarity between the wavelengths of the sound and the chosen material thickness, but the overall differences seem to be minimal and the procedures are not completely clear [88].

Some potential is evident in the use of acoustic materials for reducing noise levels within an incubator. More research is required in the area to determine the frequencies which are present during transport as well as the effectiveness of acoustic covers when subjected to expected ambulance conditions.

2.5.2.3 Reducing noise levels through road choice

The influence of road was noted by Buckland et al. [73] in their study on the levels of vibration in neonatal ambulances. Eight of the transfers recorded in their study happened to be along rural roads, as opposed to the remainder which were short urban trips, and this provided the opportunity to investigate potential differences. Metrics were limited to the average and maximum A-weighted levels along with the peak C-weighted levels, with rural roads performing worst in each. This clearly suggested that road type had an effect, however it is unclear whether this was due to the roughness, vehicle speed or another factor.

Driving along a different set of roads would alter the noise which enters the ambulance and thus the incubator. This could negate, or at least reduce, the need for noise reduction methods relating to the infant or incubator. More research is therefore required to determine links between road choice and the resultant noise levels so suggestions could be provided for optimal routing.

[2]Note that, while Kellam et al. [88] indicate a copyright of the term "Acoustical Incubator Covers", there is no evidence for a commercially available product. Other studies do exist which have also used the specific terminology, however they were performed by affiliated researchers.

2.5.3 Summary of physical stressor reduction techniques

There are three main categories in which methods of reducing the levels of vibration and noise can be placed: those inside the incubator, those between the incubator and the surroundings, and those which alter the input to the ambulance.

The main problem with implementations within the incubator is the fighting of a losing battle. Vibrations have been shown to increase from the trolley and, while the amount of increase can be reduced, tackling the problem earlier in the transmission from the chassis would be more effective. Similarly, there is not much room inside the incubator in which noise suppression materials can be placed. Earmuffs have been shown to have potential, however the evidence is not conclusive and the amount of noise which does get transmitted through to the ear canal is unknown and may still be considered stressful.

Isolating the incubator from the vibrations which travel up through the trolley has been both theorised and patented but not implemented commercially. Quantitative studies of the results of these systems are lacking, while the implementation on all neonatal transport trolleys would require significant time and investment. Blocking noise from entering the incubator, on the other hand, would cost money depending on the materials, but would be a very quick implementation. Unfortunately, the amount of noise which could be reduced is not fully known and nor is the difference with the existing covers each transport team may already be using during standard practice.

Ideally, ambulances would be constructed with suspension systems which provided optimal ride conditions and the bodywork would dull the amount of noise which enters the patient compartment. Until ambulances are designed in mind of comfort, there is the possibility of driving an alternative route between hospitals to alter the levels of vibration and noise which are transmitted inside. The theory is unproven, however there has been evidence suggesting the roads taken could be an influence.

Of the possible mitigation strategies, directing ambulances along optimal routes appears to be the most beneficial. Unlike the other options, this method of intervention has an added boost of a relatively rapid implementation in the future due to the lack of physical alterations required. All ambulance drivers have the option of selecting a different road to use and therefore only information on the ideal route to use in each situation would be required.

Producing an optimal route for neonatal transfers would require identification of a set of roads which minimise a cost function. The specific cost function would need to be determined, however it is likely, from the results of previous studies, that it would

involve a combination of factors such as the total journey duration, the levels of vibration (frequency, magnitude, shocks, etcetera) and noise (overall level, sudden bursts, etcetera). In order to assess these metrics, data would therefore be required that connect the road input to the resultant in-ambulance environment. Whilst vehicle parameters such as suspension and soundproofing would affect the expected environment, these would not be easily modified. Instead, factors related directly to the road should be focused upon, ranging from the surface condition (roughness, potholes and speed bumps, etcetera) to the usage (speed, amount of traffic, etcetera).

The aim of this thesis is to assess the feasibility of improving neonatal outcomes after ambulance transfers by means of driving along a less stressful route. Achieving this will therefore require knowledge to be collected on the comfort of different roads.

2.6 Detection of the input from roads

Although a vast amount of data on the classification and allowable speeds on roads are both known and readily accessible through mapping datasets such as OpenStreetMap (OSM) [89], this does not include information on the road surface, let alone the resultant noise and vibration levels one might experience. Therefore, in order to assess which route may produce the least discomfort, this data needs to be collected. This section will explore the methods used by both industry and research to acquire information about road conditions, along with their potential suitability for the problem at hand.

2.6.1 Maintenance infrastructure

The traditional reason for gathering data on roads is to identify any flaws which may need fixing. As such, this is typically the responsibility of either the local or national government, depending on the road classification [90].

Local road networks are assessed most frequently using visual inspections [91]. These can be performed as frequently as monthly for major roads, but reduce to annually for smaller, less traversed, roads. The majority of the surveys conducted are focused on the general safety of the roads, whether due to obstacles or road deterioration, while more thorough "service" inspections are also performed. Although these service inspections aim to identify problems which may affect comfort (as well as other interests), these are also performed visually while being driven along. Visual inspections, however effective they may be, would not provide an accurate account of the levels of either vibration or noise which may be encountered.

A more promising method of road assessment are the automated collections of data using equipment attached to vehicles. These are either used for monitoring the structural condition or the grip of the road [90], however it is the former which will be the focus here due to unevenness being most likely to influence the in-ambulance environment. An example of the data collected on the road condition is with the SCANNER (Surface Condition Assessment of the National Network of Roads) surveys [92] performed in the UK. The SCANNER specification sets out a range of parameters which are assessed, including the longitudinal and transverse profiles of the road, texture depth and details on any surface cracking. Although the longitudinal profile (for example) is reported in both 3 m and 10 m summaries which represent the state of the road, work would be required on translating these values into the levels of noise and vibration that would be experienced.

Automated surveys are only conducted on an annual basis and are mainly carried out on major roads such as motorways and key distributor links [90], highlighting a further problem with using existing data collection methods to determine route comfort. These restrictions are likely due to the significantly increased cost compared to visual inspections [93], making them unsuitable for all but the most frequently traversed roads. This has prompted researchers to investigate cheaper methods of recording the road surface, both to reduce spending (in order for more of the budget to go towards repair works) and to facilitate more frequent data collections to account for any rapid deterioration or performed repairs.

2.6.2 Sensors embedded inside consumer vehicles

Recording information regarding road condition would be cheaper to perform if it was conducted during regular journeys, rather than on purpose-made trips as with SCANNER surveys. Recording roads passively — in terms of letting the journey dictate which roads are assessed, rather than actively driving along specific sections for the intent of assessment — would therefore have savings benefits both in fuel and staffing costs, as the journey would have been conducted regardless. Setting up general consumer vehicles to gather data would provide data more frequently, however there would inevitably be more collections on major roads compared to minor.

In 2008, Eriksson et al. [94] utilised seven taxis to collect data in the local area during their regular routes. The road was sensed by a single accelerometer inside each taxi, fixed to the underside of the dashboard after testing the effect of positioning within the cabin, along with a satellite receiver attached to the roof. Whilst a large amount of data was collected during the study, and was subsequently used to identify road arte-

facts such as potholes, the expertise and investment required in setting up the vehicles would hinder more widespread usage.

Data quality was increased by studies which implemented the use of accelerometers, built into the vehicle and accessed via the CAN (Controller Area Network) bus, positioned on the suspension and in the general chassis [95, 96]. González et al. [95] were able to use the data to provide an estimate of roughness using the calculated Power Spectral Density (PSD) of accelerations according to the standard for road surface profiles (ISO 8608 [97]), providing general verbose classifications. Li et al. [96] took the use of the multiple accelerometers further, recording variables including body yaw (angling front-to-back) and roll (side-to-side) to facilitate a more complex feature identification. Although gathered data showed promise for the collection of large amounts of information on the roads, along with the accuracy which comes with accelerometers recording the exact suspension inputs, the scale of possible collection using the CAN bus is doubtful and would largely rely on the vehicle manufacturers themselves.

Neither of the methods mentioned here facilitate easy scalability, regardless of the data recorded. This is either due to the need for potentially (to the untrained participant) complex equipment and set up required [94] or the variation in available data accessible from the vehicle CAN bus [95, 96]. While it is possible to purchase a cheap device to read from sensors based in cars, the amount of sensors which are publicly accessible are both severely limited and their positioning and quality would depend on both the make and model of the vehicles. Additionally, it is possible in some circumstances to modify vehicle behaviour through the CAN bus which could be a cause of concern in volunteers and result in lack of participation.

It is clear that meaningful information relating to roads could be collected using consumer vehicles, however a more layperson-friendly method is required to record the most data. One possible route would be to use smartphones, which have become ubiquitous in modern society.

2.6.3 Detection using smartphones

Smartphones are communication devices that are highly prevalent in everyday life, with surveys reporting that 84% of UK adults owned such a device in 2019 [98], rising to 95–98% in those under 55 years of age. Although initially for communication via audio, more and more functionality was progressively added that turned smartphones into compact personal computers. Increased methods of communicating data were added (such as wi-fi, Bluetooth and mobile access to the internet) along with both a

profusion of built-in sensors and APIs (Application Programming Interfaces) to aid their utilisation.

Sensors and communication methods were added to smartphones to improve user interactions and increase functionality. For example, accelerometers were initially included to help the smartphone detect changes in orientation and modify the display accordingly [99]. This progressed to being used by games associating device tilt with the turning of a steering wheel, etcetera, for increased intuition and immersion. The presence of these accelerometers, along with access to them with APIs [100], suggests that smartphones may have the ability to act as relatively low cost (depending on the chosen device) vibration data loggers for research.

Along with accelerometers for vibration detection, smartphones contain microphones due to their base function as communication devices. These are also accessible through APIs and therefore smartphones have the capability of logging noise levels.

The prevalence of smartphones, combined with the useful built-in features, has led to interest in their use in the mass collection of data by utilising public volunteers (colloquially termed "crowdsourcing") [101, 102]. Combining the recording of vibration and noise levels with the scalability of crowdsourcing could enable more rapid detection of any alterations to road quality, however previous work should first be explored to assess the validity of the data available.

2.6.3.1 Smartphone acceleration sensing

The accelerometers built into smartphones were not implemented with the goal of logging data, but instead are to enhance the general user experience. As such, while the accelerometer modules themselves may be accurate, the sampling methods provided by APIs can be sporadic [103]. Despite this, smartphones have been shown to record waveforms within a few percent of specialist reference accelerometers [104, 105].

Many researchers have observed the potential of using smartphones to record the vibration levels within vehicles using custom-made apps. The various studies performed have all been focused upon the monitoring of roads due to the potential reduction in costs of existing systems. Although these studies do not directly report the exposure levels within vehicles (which may then be experienced by transported infants) they provide an important insight into the ability of smartphones to detect inputs from the road.

Road detection studies using smartphones can be separated into two categories, with the first containing those which aimed to identify road artefacts. Multiple different

sample frequencies (ranging from 5 Hz [106] to 100 Hz [107]) and processing techniques (high-pass filtering [106, 108], low-pass filtering [107, 109], or simply comparing raw accelerations with user-inputted event locations [110]) were implemented by the various researchers. All studies concluded that smartphones were both capable of registering artefacts and were doing so on repeated passes over the same surfaces, although the events tended to be defined using trends within the data rather than the values themselves.

The second group of studies concerned the roughness of the road surface. Roads were classified either by approximating the commonly used IRI (International Roughness Index) [54, 111, 112] or developing a new method of differentiating surfaces [113, 114]. These studies further confirmed the ability to record the road surface using smartphones, doing so with reasonable accuracy compared to purpose-built equipment (similar to the SCANNER surveys [92] mentioned earlier), however they did not provide any meaningful information on the resultant levels that would be experienced.

Levels of vibrations recorded by the accelerometers built into smartphones have been shown to have the potential for accuracy and can have a role in the low cost monitoring of roads. Whilst studies have reported the various conditions of the roads, no studies have shown an interest in comparing either acceleration frequencies or magnitudes to the potential health impact. Although the roughness values from some of these apps can be correlated to comfort using vague verbose descriptions [54] or by suggesting an expected ISO 2631 comfort level when including the vehicle speed as a factor [115], these all discard the majority of data. Therefore, for the purposes of identifying roads of potential harm to infants, there is a clear need for another app which assesses the in-vehicle acceleration levels directly.

2.6.3.2 Smartphone noise detection

Many apps have been developed for smartphones which advertise the ability of reporting noise levels. Due to the vast costs associated with purchasing and maintaining a professional sound meter that meets the various standards, there has unsurprisingly been an interest as to whether these apps can provide equivalent values using the devices readily available.

Several studies investigated the precision and accuracy of the various apps. Kardous and Shaw [116] conducted a thorough evaluation of the apps available at the time of the study, eventually assessing the results of 10 apps for the iOS Operating System (OS) (Apple, Cupertino, CA, USA) and 4 apps for the Android OS (Google, Mountain

View, CA, USA). These apps were tested on a range of smartphones, with both different manufacturers and models, and compared to a precision sound meter at various SPLs of pink noise. Of the multiple apps, four were consistently within ±2 dB of the reference which showcased the potential for smartphones. The app commended as the most accurate during this study was later used by Murphy and King [117] in real world measurements, however, and found a greater difference of 4.4 dB between the smartphone and sound meter values. This difference may have been due to the concentration of frequencies experienced outside compared to in the lab, or possibly because of hardware differences although the latter study did not specify the exact model used.

Recorded noise levels were found to vary between devices using the same app, presumably linked to a combination of different internal sensors and processing methods. The largest differences were identified between Android devices [116] which is likely due to the multiple manufacturers building the smartphones and the variety of components that are inside. iOS smartphones were not perfect, however, and did vary between models albeit not to the same degree [116]. Murphy and King [118] suggested that iOS apps reported more accurate noise levels than Android, although Ventura et al. [119] showed that it was highly dependent on the device with some Android smartphones producing near-perfect results.

A couple of methods were investigated to further improve results after choosing an optimal smartphone and app. Serpanos et al. [120] suggested the use of a calibration signal improved the subsequent results above 50 dB(A), however the sounds below this level remained inaccurate and only a single smartphone was tested. A more significant improvement was made by Celestina et al. [121], whereby an external microphone was inserted into the audio port of the smartphone. Although results matched the professional sound meter more closely, the use of an external microphone would not be feasible during data collection as a key purpose of using a smartphone was reducing the complexity of recordings.

While there are studies which have used smartphones to record the noise exposure of certain roads [117, 122], these readings have been taken at the street level rather than inside vehicles. Utilising the microphones inside smartphones may, however, provide an efficient and effective method of crowdsourcing noise data inside vehicles. The devices themselves have been shown to be accurate [116, 119] — subject to both app selection and smartphone model — and would be present in the most journeys due to being owned by the vast majority of drivers, suggesting that this is a bountiful resource ready to be tapped.

2.6.4 Summary

Currently, road monitoring equipment is only focused on determining characteristics of the surface for maintenance goals. While this equipment is highly accurate, it is slow, expensive, and does not give any direct information of the levels of vibration and noise which may be experienced inside the passenger compartment.

One option for recording the data on the potential physical stressors is via the sensors built into modern smartphones. These ubiquitous devices have been shown to have the capability of recording both vibration and noise levels to a reasonable degree of accuracy and are present in the majority of regular journeys.

It is important to note that, while recorded data from smartphones can be accurate, it can vary both with the device used and the way an app has been programmed. Therefore, all outputs from all devices which are used for collection require calibration against professional-grade equipment to ensure the values are interpreted correctly.

While there are apps which record either vibration levels or noise exposure, there are few that record both and none that combine with the location data which would be required to identify the causal road. Equally, the variation in accuracy between available apps has been evident and suggests that more control is required by the researchers.

2.7 Conclusions

The background to neonatal transportation has been presented, outlining the vulnerability of the population. This subsequently led to an overview of the concern that undergoing ambulance transfer presents infants with a greater risk of severe brain injury. Although the causality is unclear, this increase in risk is thought to be influenced by the stress experienced by the infants that is potentially caused by excess noise and/or vibration.

The levels of vibration and noise during ambulance journeys were explored and compared between the limited number of studies. This showed that both quantities exceeded recommended thresholds, however these thresholds either had no medical backing or were intended for use with adults. There was a minimal amount of research that investigated the neonatal physiological response to these physical stimuli, however the results of these were inconclusive or presented from a restricted sample of data.

Until the effects on neonatal infants are fully understood, work should be undertaken to minimise levels as a precaution. Several studies have investigated various methods of both noise and vibration suppression, in both experimental and theoretical capaci-

ties. The majority of methods did not produce sizeable reductions. This is largely due to the limited amount of physical space in which methods can be implemented along with the need to ensure patient care is unimpeded. One option that was suggested, but not explored, was the selection and use of optimal roads during journeys. This possessed the potential to reduce the stimuli before entering the ambulance, thereby reducing both the levels experienced by the infants and the amount of suppression required by methods based on the trolley. It is for these reasons, along with the theoretically minimal delay in which results could be introduced to transport, that this thesis will investigate routing for comfort.

Comparing the levels of vibration and noise that occur along different routes naturally require data to be available for all roads that may be used. This data is not readily available and would therefore require collecting. Prospective methods for the acquisition of this data were reviewed and compared. Ultimately, it was found that there was no equipment that was designed for the purpose of gathering both noise and vibration data, with the bulk of methods conveying an interest in road monitoring for maintenance and using vibration alone. Clear scope was identified in the use of smartphones as a form of mass data collection, with the built-in sensors exhibiting strong promise. There are no smartphone apps that sample all of the required information or that have been thoroughly validated, and thus a bespoke app needs to be developed.

CHAPTER 3

App Development

3.1 Introduction

High-risk premature infants are more than twice as likely to experience severe brain injury when transferred by ambulances for specialist care compared infants not undergoing transfer [22, 23]. Multiple studies have suggested that this increase may be due to the excessive levels of both vibration [43, 44, 47] and noise [73] to which the neonatal infants are exposed. It therefore follows that the risk of ambulance transfer may be reduced if the vibration and noise were more appropriate.

Neonatal transports are currently undertaken using an incubator mounted on a trolley, also supporting life support equipment, which is clamped rigidly to the ambulance floor. The British standard for the transportation of incubators does recommend that vibration is kept to a minimum [72], however it also specifies that the trolley must be attached rigidly to the vehicle chassis [123]. Rigid attachment is to ensure no items — whether persons, medical equipment or otherwise — become projectiles should the ambulance be involved in a vehicle collision, rather than for any clinical reason. This therefore results in shocks from road artefacts being directly coupled to the trolley.

Similarly, the standard also states that equipment should emit less than 60 A-weighted dB (dB(A)) (frequency-weighted to approximate the adult response) of noise and that ear defenders must be used during transfers [72] although the currently available products are not capable of reducing levels appropriately. Defenders designed for transfers are only able to reduce sound levels by up to 7 dB [81] and studies on their impact and effectiveness are largely inconclusive [40, 82, 83]. Ultimately, despite being stipulated in the guidelines, noise levels during transport remain significantly higher than recommended [40, 62].

Redesigning the ambulance and transport trolley would be the optimal solution to re-

duce harmful levels of exposure, however this would require both time and significant investment to implement. Instead, improvements may be possible by redirecting ambulances along more comfortable roads. This would require data regarding the noise and vibration that are linked to the road surface (such as roughness, potholes and rapid changes of surface) as well as vehicle parameters (such as speed, suspension and soundproofing).

Analysing the state of roads traditionally is performed by Highways England on an annual basis [90]. This is due to the high cost of the equipment-laden vehicles, with one example being the UK SCANNER survey [92] which measures the texture, rut depth and roughness of a road along the left-hand wheel path. Local authorities do carry out surveys as frequently as monthly [91], however these are limited to visual checks focusing on safety. As roads can rapidly deteriorate, and subsequently be repaired, a more regular means of monitoring roads is required.

The ubiquity of smartphones may provide a simpler, more cost effective method of obtaining road information. Numerous groups have shown that the built-in accelerometers within smartphones are capable of identifying road roughness, either by attempting to approximate the IRI (International Roughness Index) [54] from vertical accelerations [111, 112] or by creating their own classification system [114]. One study also investigated the relationship between roughness, vehicle speed and the resultant vibration [113]. Road artefacts, such as potholes and speed bumps, have also been identified using raw [110], low pass [107, 109], and high pass [106, 108] accelerations, all at different sample rates. All of the above studies focused on the sensing of roads from a maintenance standpoint.

The aim here is to develop a smartphone application which combines the use of accelerometers with analysis of noise to measure the effect of road surface, and speed, on the environment within a vehicle rather than the road surface itself. This application can then be used to gather large amounts of road data, which can then be used to route ambulances accordingly.

3.2 Design Specification

3.2.1 Requirements

The most prevalent vibrational frequencies in ambulances have been shown to occur between 5 and 20 Hz [44]. According to the Nyquist-Shannon sampling theorem, the accelerometer needs to sample at a minimum of 40 Hz to enable full analysis of these

frequencies [124].

As well as registering the frequencies, it is important that acceleration magnitudes are accurately measured to ensure comfort is assessed correctly.

Audio data needs to accurately register sound levels. To prevent possible invasion of privacy, raw audio files should not be stored [125].

Accurate device locations are required to associate noise and vibration with a specific road. The satellite receiver therefore needs to reliably give the location to an accuracy that enables roads to be identified.

To ensure participants use — and continue to use — the app, the user interface needs to be straightforward. A method of automatically retrieving recorded data is also required to minimise participant effort.

3.2.2 Benefits of the Android Operating System

At the time of writing, two smartphone operating systems dominate the market: Android and iOS. Although similar, each operating system has its benefits.

As the developer of iOS, Apple, restricts the operating system to devices of its own manufacture, the smartphones work seamlessly. Similarly, only a handful of different iOS smartphone models are released each year so devices in circulation have identical hardware. This ensures that an app developed on one smartphone model will work identically on others.

Conversely, due to its open-source nature, there are countless different manufacturers of smartphones running Android operating systems. This has its pros and cons. Because there are so many different manufacturers of Android smartphones, the sensors on all of them can vary. Even smartphones made by the same manufacturer can vary greatly, due to targeting different price ranges in the market. Although this complicates app development as multiple different use-cases need to be taken into account, competition between manufacturers results in Android smartphones being available at a fraction of the cost of iOS devices [126].

Android also provide a fully-integrated development environment for free, along with thorough documentation. Releasing an app to market on Android devices is also subject to less stringent requirements than on iOS. Due to the ease of app development and release, along with the wider availability of smartphones, the app was developed for Android operating systems in the Java language.

Figure 3.1: Map of the 3.5 km route used during development for testing the app (© OpenStreetMap contributors).

3.3 Development Process

The following sections outline the functionality of the app, as well as detailing any modifications made and their reasons. Each section will focus on a separate component (such as user interface or individual sensors) for simplification, although in reality many modifications were concurrent.

Development began by targetting the 24th Android API (Application Programming Interface), which was released in August 2016, as well as supporting all API levels down to 17. This ensured that the app would work on the majority of devices released after November 2012 and therefore would not restrict the potential user base. New versions of the Android operating system were released during development and the code was modified to reflect any changes in functionality, ensuring brand new devices were not excluded from participation.

In-vehicle behaviour of the app was continually assessed during development by using a Moto E2 smartphone (Motorola, Chicago, IL, USA) to record a section of road. This enabled the functionality to be quantified and amended where necessary. A daily commute was logged between 17th January 2017 and 15th January 2018. A portion of road which was often travelled along (Figure 3.1) was chosen to assess the repeatability of data recorded by the app. This route section was chosen because it contained known features (most notably a large pothole) and was within close proximity to the university which enabled visual checks to be performed in order to correlate data with said road features.

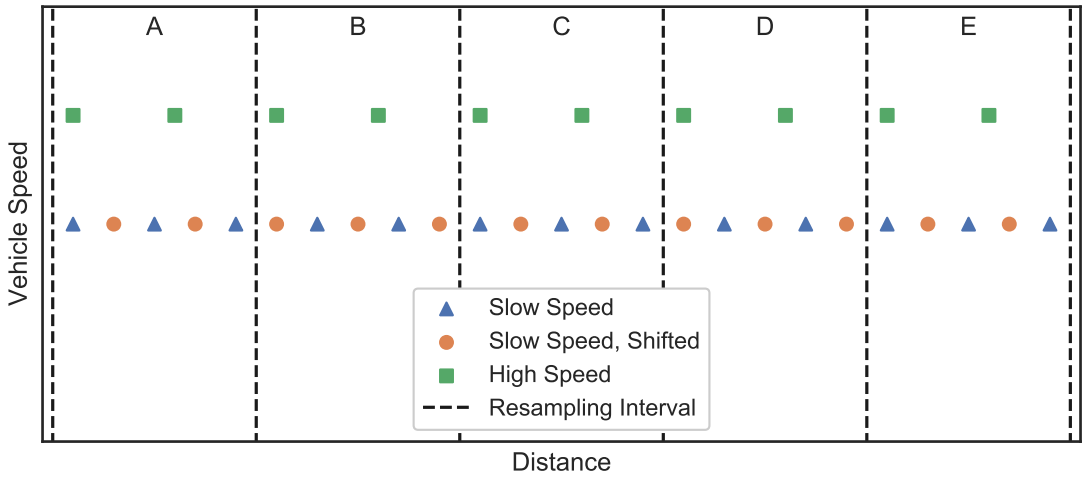

Figure 3.2: Illustration of how different journeys along the same route may not register samples at the same positions along a road due to the difference in sample time and vehicle speed.

3.3.1 Resampling by Distance

Throughout the development process there was a need to compare recordings made during different journeys to check whether data showed similar features. Unfortunately, data was not directly comparable between recordings due to the time-based sampling.

Two recordings would only be comparable when both journeys contained identical vehicle speeds throughout and the recording was started at the same exact location. This can be explained using the simplified graphic in Figure 3.2. Assuming constant and identical vehicle speeds, a shifted start to a recording (within a specific section of road) would result in zero samples matching position between two journeys as shown by the triangle and circle markers. Similarly, when two journeys start at the same position but have varying speeds, only the first samples of each recording can be guaranteed to match up.

Processing recorded data to approximate distance-based sampling would provide a means of comparing multiple journeys albeit at a cost of resolution. The conversion to distance-based sampling would require aggregating the time-sampled data over set intervals to ensure data was available at the same locations. Choosing a distance interval which was too small would result in too few samples and could result in some journeys not having any data at each position, whereas an interval which was too large would reduce the resolution — the ability to identify trends and features — more than was necessary.

An interval of 10 m was chosen as it provided a reasonably high level of resolution

while ensuring each interval contained enough data samples. At the initial Inertial Measurement Unit (IMU) sampling frequency of 50 Hz (which is outlined within Section 3.6) just under 16 acceleration signals would be logged per interval at the maximum UK speed limit of 70 mph, with the number of samples increasing as speed drops.

Location samples were only available at a rate of 1 Hz (outlined in Section 3.5) and would therefore provide coordinates at an interval of 31.3 $\mathrm{m{\cdot}s^{-1}}$ at 70 mph. Therefore, location coordinates required interpolating before any resampling could be conducted to ensure each IMU sample could be attributed to the correct distance interval. Interpolation of coordinates was conducted linearly under the assumption of a constant speed between location samples.

Next, calculation of the distance between each set of coordinates was required. This was computed using the haversine formula [127] (Equation 3.3.1) to calculate the surface angle between locations before multiplying by the average radius of the Earth. Distances were calculated in a cumulative fashion for each journey before being aligned for comparison with other recordings.

$$
\begin{aligned}
hav &= \sin^2\left(\frac{\delta\,\mathrm{lat}}{2}\right) + \cos\left(\mathrm{lat}_1\right)\cdot\cos\left(\mathrm{lat}_2\right)\cdot\sin^2\left(\frac{\delta\,\mathrm{lon}}{2}\right) \\
\text{surface angle} &= 2\cdot\mathrm{atan2}\left(\sqrt{hav},\sqrt{1-hav}\right) \\
\text{distance} &= \text{surface angle}\times r_{\mathrm{Earth}} \\
\mathrm{atan2} &: \text{two-argument arctangent} \\
hav &: \text{haversine between two coordinates} \\
\mathrm{lat,\,lon} &: \text{latitude and longitude coordinates} \\
r_{\mathrm{Earth}} &: \text{average radius of Earth} = 6,371\mathrm{km}
\end{aligned}
\tag{3.3.1}
$$

Resampling alone would not result in comparable information as the starting position of each journey, combined with any shifting between lanes in the road, would result in varying starting intervals within a specific route section. The offsets between journey intervals were therefore minimised by adjusting the calculated distances for each journey by the distance from the average starting coordinates within the area of interest.

The final component of the resampling by distance procedure was to aggregate all samples within each interval, on a journey-by-journey basis, to result in a single value. This aggregation consisted of either taking the mean (used for location variables) or the r.m.s. (Root-Mean-Square) (used for accelerations and noise) of all samples.

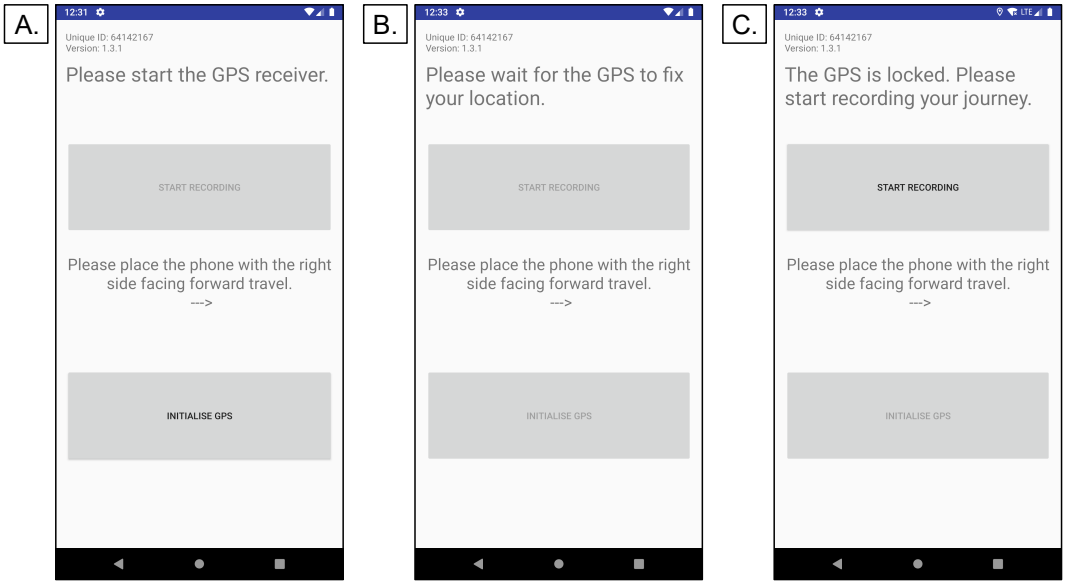

Figure 3.3: Screenshots of the initial app displaying the home screen (A), satellite initialisation (B) and start recording screen (C).

3.4 Operation

3.4.1 Initial User Interface

Initially, the app was controlled by a combination of two buttons: one for initialising the satellite receiver, and the other to start the recording of sensors. User input was required to start recording sensors as the app was unable to determine when a journey was being started or if the phone was in position. This button, logically labelled "Start Recording", was disabled until the satellite receiver had a location fix, as sensor data would be useless without being able to link it to a road (Figure 3.3A). Therefore, the user needed to first press the "Initialise GPS" button to start the receiver (Figure 3.3B). The app timed out after a minute if the satellite receiver did not achieve a fix, to save on battery when there was limited access to the sky. Once the receiver had a location fix, the "Start Recording" button became enabled (Figure 3.3C) and the user could record their journey.

Text boxes were utilised to convey information about the process of using the app. One text box, at the top of the screen, was used for user prompts to begin the initialisation ("Please start the GPS receiver"), wait for a fix before starting the journey ("Please wait for the GPS to fix your location"), and finally to begin the recording ("The GPS is locked. Please start recording your journey"). A second text box was placed at the bottom of the screen containing an arrow along with a message to suggest which orientation to place the device relative to the vehicle. This was to try and ensure that the same axes were

used for the direction of travel and vertical vibration, to simplify processing of data. The text boxes and buttons were equally sized and filled both the height and width of the screen to ensure everything was clearly visible and the buttons were easily selected.

3.4.2 Use of Services

A new 'activity' [128] was initially created to handle the sensor interaction and logging when the user started the recording as described in the previous section. An activity in Android is a collection of code which is accompanied by a linked window for user interaction. The code for this new activity was responsible for accessing each sensor and logging the data to a file. Visually, the activity provided a single button, for the user to stop the recording, along with some text boxes which updated with sensor values as proof of operation. This activity, along with the recording, was stopped when the user pressed the button and the starting screen was returned to view with a final message of appreciation: "Thank you for recording your journey".

The recording of data was moved from within a separate activity to within a 'service' because the device should be left alone during a journey and there was no need for the user to see any values sensed. A service in Android is defined as "an application component that can perform long-running operations in the background" [129] which made it ideal for recording data as no user interaction was required. The base service class runs code on the main thread and was found to affect the button behaviour, therefore a specific class of service which operates on a dedicated thread [130] was used. Upon starting the recording, the user was notified that the service has started and the app was working by the message changing ("Please turn your phone off for the duration of the journey") and the "Start Recording" button text changing to "Stop Recording", reflecting the change in its use.

The app was designed to prevent the recording being stopped by the Android system when the user was not interacting with the device. This was achieved by implementing a partial 'wake lock' [131] which ensured that the device CPU (Central Processing Unit) was not stopped when the screen was turned off. Without this feature, sensor data would only be registered and logged when the device screen was active, which would not be beneficial to the battery life.

3.4.3 Single Button Operation

App operation was simplified to a single, large button (Figure 3.4A) for starting and stopping recordings as the use of two buttons was deemed confusing. Instead of press-

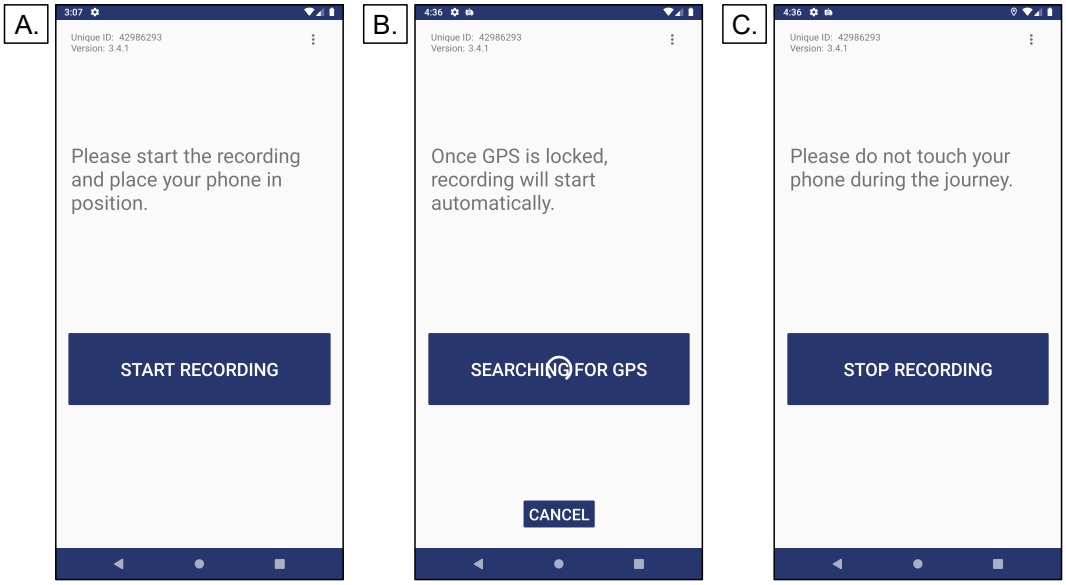

Figure 3.4: Screenshots of the final app displaying the home screen (A), satellite initialisation (B) and recording screen (C).

ing one button to initialise the satellite receiver, before then starting the recording when a location fix was achieved, the single button was used to first initialise the receiver and then automatically start the recording. Recording was delayed for at least 10 seconds (longer if the receiver struggled to get a fix) from the press of the button to allow for positioning and the screen to be switched off. This reduced the amount of user interaction and therefore vibrations which would need to be removed from analysis. The user was informed that the app was searching for a satellite fix by an animated icon (Figure 3.4B) to signify it had not frozen. Finally, the button changed to show the user that recording had begun (Figure 3.4C).

It can be seen in Figure 3.4B that a "cancel" button is visible. This was provided for the occasions when the user mistakenly started the app before being outside as otherwise they would need to either terminate the app, wait for the location receiver to time out due to lack of satellite signal (which would take 60 seconds) or let the recording start before immediately stopping it.

Settings were added to enable the user to choose preferred time delays for both positioning the device at the start of the recording and for the location receiver timeout. The positioning delay of 10 seconds was chosen due to needing extra time when validating the app output (which will be outlined in the next chapter), however different users may find it takes them longer. Similarly, users may wish to reduce the receiver delay to save on battery usage.

3.4.4 Foreground Services

Occasional interruptions of recordings were traced to the Android operating system forcing the service to stop in order to reduce memory usage. This was prevented by telling the system that the service should run in the foreground. Running a service in the foreground is meant for user interaction and therefore the system does not stop it as it would give a bad user experience. Instead, the service can only be stopped by the user, either by finishing the recording or manually killing the app.

3.4.5 Test Mode

A modified build variant of the app was created for use in calibrating sensors. The use of build variants — effectively different versions — enables certain elements of the app to be controlled depending on the build while maintaining all modifications to the code. For the test variant of the app, an option to disable the location receiver was added. The location receiver was essential for the recording of road data, however it would hinder the analysis of sensors in indoor environments where satellites may not be visible to the device. By providing the option to turn off the location receiver, time could be spent on testing the device rather than trying to get a satellite signal.

3.5 Location Receiver

3.5.1 Setting up

Geographic location was accessed through a built-in class at a rate of 1 Hz. As mentioned in Section 3.4, recording of data did not start until the satellite receiver had a fix on the device location. This was because an initial position fix was not instantaneous, and therefore resulted in some sensor values being recorded without geographic information. To prevent this, the receiver was registered in the code of the main activity upon first button press. A system interrupt was called when a new set of location data was available, enabling the processing as required. Recording was enabled once three consecutive location samples were reliably received at 1 Hz intervals, suggesting that the device has locked on to satellites.

A slight delay was identified between the recording starting and location data being logged. This was because the logging of the data was performed in a separate service which required a new satellite receiver to be registered. Unregistering the satellite receiver in the main activity before starting the recording of location data resulted

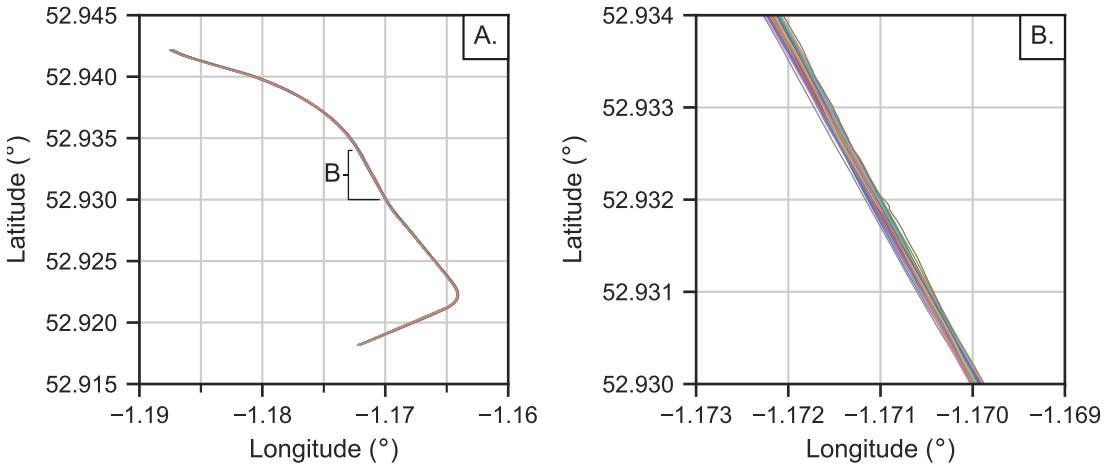

Figure 3.5: Unprocessed coordinates recorded using a Motorola Moto E2 along the full test route (A) along with a magnified section (B) (n = 132 journeys).

in this delay as a position fix would need to be found again, despite less than a few milliseconds passing between unregistering the old and registering the new receivers. Maintaining access to the receiver within the main activity while the recording was initialised eliminated this delay. The receiver within the main activity was unregistered several seconds after recording had begun, leaving enough time for the new receiver to be set up while removing unnecessary processes.

Sampling of location data was ultimately moved to a dedicated thread to ensure that the system interrupts from the satellite receiver did not interfere with any other sensor callbacks. Setting up a separate service for receiving location data both provided the new thread and also enhanced the readability of the code.

Initialisation of the satellite receiver was moved to a separate service to reduce the amount of code within the main activity, which became more convoluted over time as extra functionality was added. A countdown timer was also implemented to end the service 1 second after starting the recording. This reduced the memory usage of the app while still providing a seamless start to the recording.

3.5.2 Recorded variables

Location data was initially only logged in terms of latitude and longitude, enabling roads to be linked with data from the other sensors. It was thought that coordinates would provide enough information, with any other variables (such as vehicle speed) being derivable. Coordinates were found to be visually similar along the route (Figure 3.5).

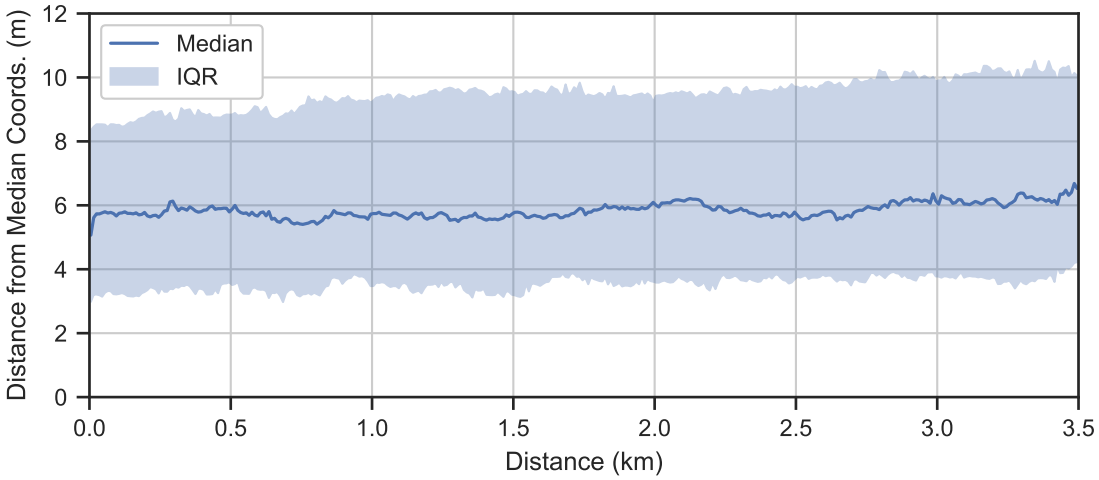

Figure 3.6: Variation in resampled journey coordinates recorded by a Motorola Moto E2 along the test route, compared to the average for each interval (n = 132 journeys).

Calculating the average coordinates, in 10 m intervals, for each journey confirmed that the recorded location data was highly repeatable. Journeys varied by an average of 5.8 m compared to the average coordinates for each interval (Figure 3.6). Although this is large relative to the interval size, the differences can be in any direction and are affected by both differences in, and perpendicular to, the direction of travel.

Knowing the vehicle speeds during recordings would enable any effect on vibration and noise to be analysed. This was available directly from the location receiver which reduced the need for later calculation. Vehicle speeds along the route were very similar (Figure 3.7), with an interquartile range of ±6.5%. The highly consistent speeds along the test route should provide appropriate data to assess the repeatability of the sensed vibration and noise.

Logging was expanded to include all available location variables encompassing altitude, horizontal accuracy, bearing, and Coordinated Universal Time (UTC). These were included due to the potential future uses in processing and insignificant storage impact. For example, horizontal accuracy could give a good indication of the smartphone proximity to hospital buildings, while bearing could be used to ensure journeys are travelling in the same direction before comparison. The units and data types of all logged location variables can be found in Table 3.1.

3.5.3 Stationary Detection

Battery drain, along with excessive storage use, could occur on the occasions when a user forgets to stop a recording at the end of a journey. This could be prevented by

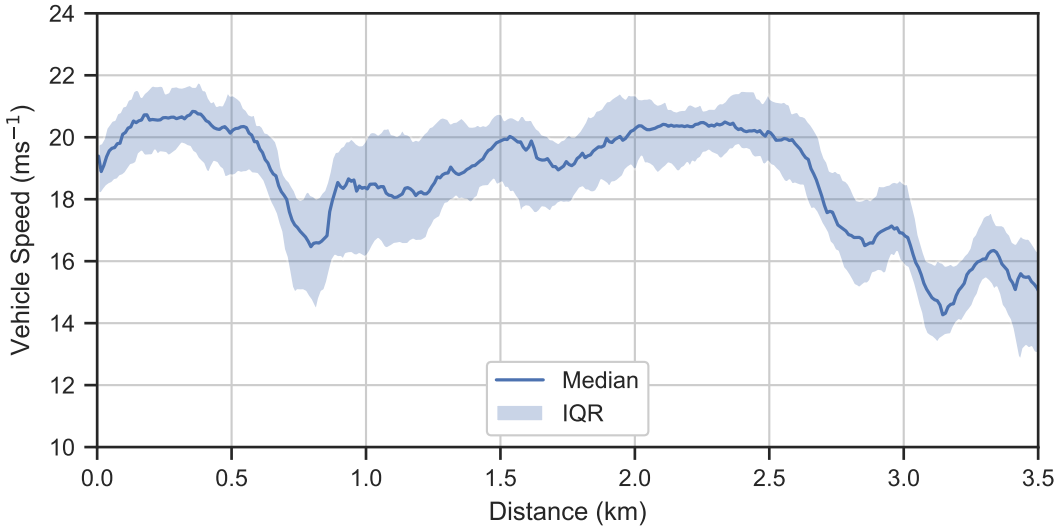

Figure 3.7: Median and I.Q.R. vehicle speeds along the test route using a Motorola Moto E2 (n = 40 journeys).

Table 3.1: Final location variable units and data types.

Variable	Units	Data Type
Latitude	°	64-bit floating-point
Longitude	°	64-bit floating-point
Speed	m·s^{-1}	32-bit floating-point
Altitude	m	64-bit floating-point
Accuracy	m	32-bit floating-point
Bearing	°	32-bit floating-point
UTC	milliseconds	64-bit integer

stopping the recording when a lack of movement was registered. Journeys were identified as having finished when 10 consecutive minutes had been spent at speeds below 2 $m \cdot s^{-1}$ (4.5 mph). A counter was started when the registered speeds drop below this threshold, and increment with every new sample. If speeds increased to over 10 $m \cdot s^{-1}$ (22.4 mph), or at least 10 seconds had been spent at speeds between 2 and 10 $m \cdot s^{-1}$, the counter was reset as the vehicle was moving again. Otherwise, when the counter reached 600 samples (10 minutes of 1 Hz sampling) the recording was automatically finished.

Encountering a traffic jam during recording could result in a false identification of a finished journey due to decreased movement. Therefore, a check for connected wi-fi was included in the auto-stop feature to provide a secondary verification that the journey has been completed. A receiver was registered to detect changes in the wi-fi state and call a specific method accordingly. If the wi-fi was connected to a known network, such as at home or a workplace, the device was determined to have arrived at a destination and a global boolean was set to 'true'. This boolean was then checked for each new location sample, once the counter had reached the above limit of "stationary" speeds, and the recording was stopped as before.

A safety net was required as some recordings were not identified as stationary because the device did not connect to wi-fi. Recordings of multiple hours without movement occurred when the test device was left in a vehicle, as it was out of range of any network. Therefore, an ultimate stationary limit of 20 minutes was set, regardless of network connectivity. Although this would mean that the recordings of journeys stuck in traffic jams for longer than this duration would be interrupted, the data from these journeys would likely skew any analysis of roads.

Android API 26 disabled the ability to register a receiver which was connected to the wi-fi status, resulting in the requirement of a new method to perform the wi-fi check. The first sample below the 2 $m \cdot s^{-1}$ threshold now scheduled a 'Job' with a single condition of requiring a wi-fi connection. Android automatically ran this job when the condition was met and the wi-fi boolean was set to 'true' as before. The boolean was then set to 'false' if the device disconnected from wi-fi or the speeds went above 10 $m \cdot s^{-1}$.

The use of a stationary sample counter was found to be ineffective when there was a loss of satellite fix. This was usually attributed to entering enclosed areas such as buildings and would result in recordings continuing unnecessarily. Checking of the duration spent while stationary was therefore changed to a countdown timer, as it did not require location samples unlike the original counter. This timer was started after the first speed sample below the set threshold. As before, after the timer passed 10 minutes

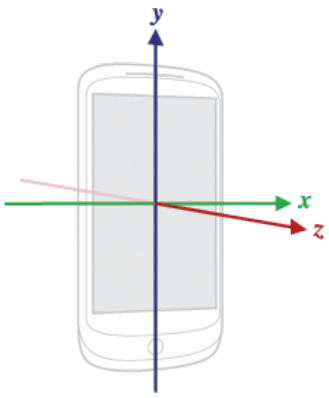

Figure 3.8: Android device coordinates, reproduced from `https://source.android.com/devices/sensors/sensor-types`

the recording would be stopped once the device was connected to wi-fi. When no wi-fi was detected, recording would be stopped once the timer reached 20 minutes. Movement, as defined above, resulted in the timer being cancelled.

Although they accounted for a lack of samples while running, the timers were only started after a speed was registered below 2 $m \cdot s^{-1}$. This meant that there was nothing in place for a lack of satellite fix before the speed threshold was met. Therefore, the stationary timers were moved to a dedicated service so they could be called either by the location receiver, as before, or by the logging operation when no new location data was sent for over 60 seconds.

3.6 Inertial Measurement Unit

The app was designed to access three components, when available, from the built-in IMU with these being the accelerometer, gyroscope and magnetometer. The core sensor was the accelerometer which would record the levels of vibration during recordings. The inclusion of gyroscope and magnetometer data, on smartphones which possess the sensors, were useful for reorienting the accelerometer values. Reorientation was necessary in order to convert from device coordinates (Figure 3.8) into horizontal and vertical values. If a magnetometer was present, accelerometer values could be reoriented to world coordinates (North, East and vertical down), where the North and East values could be adjusted to forwards and sideways relative to the vehicle using location data. For smartphones which only have an accelerometer present, reorienting could be done using vertical coefficients for each axis found by applying a low pass filter. All data outputted from the IMU were in the form of 32-bit floating point values, in units of $m \cdot s^{-2}$ (accelerometer), $rad \cdot s^{-1}$ (gyroscope) and µT (magnetometer).

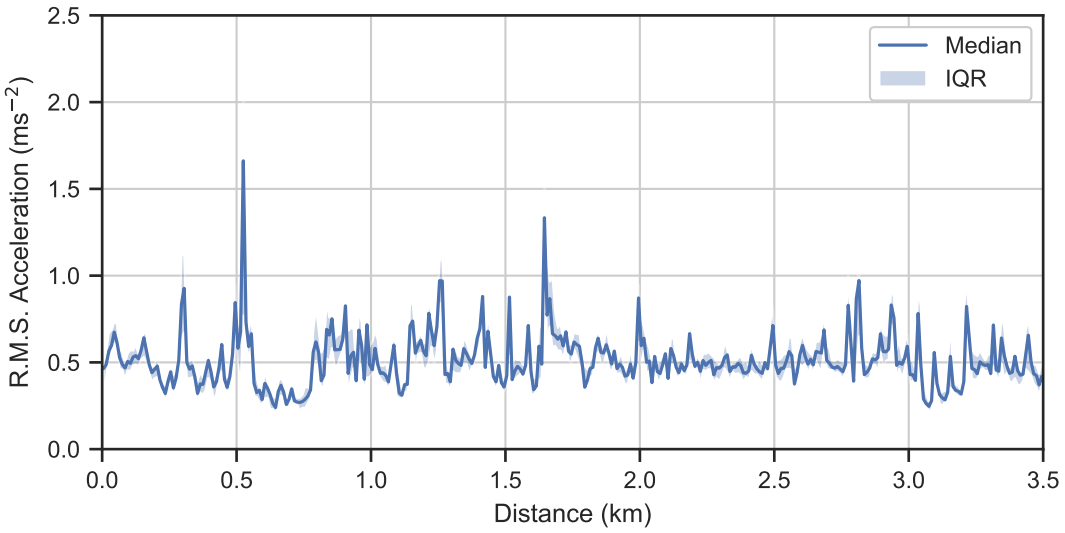

Figure 3.9: Median and inter-quartile range of 10 m resampled r.m.s. high-pass (0.5 Hz cut-off frequency) vertical acceleration recorded along the test route, sampling at 50 Hz, plotted against distance (n = 25 journeys).

In Android, the sampling frequency is set by specifying a desired delay between samples. This can be specified manually, or by selecting one of the four default data delays, specified in terms of microseconds: 'normal' (200,000 µs), 'UI' (60,000 µs), 'game' (20,000 µs) or 'fastest' (0 µs) [103]. This results in consistent sampling frequencies across devices for the first three presets, averaging at 5 Hz ('normal'), 15 Hz ('UI') and 50 Hz ('game'). The sampling frequency that corresponds to the 'fastest' preset is dependent on the manufacturer configuration, with the specified delay of 0 µs indicating a lack of restriction rather than a true duration between samples. It is important to note that whichever delay is chosen, it is not fixed and the actual time between samples can fluctuate [103].

The 'game' preset was used initially as sampling at 50 Hz would cover analysis of the 5 – 20 Hz range previously found to be amplified inside transport incubators [44]. 25 journeys were recorded along the test route while recording accelerations at 50 Hz. All recordings produced similar waveforms (Figure 3.9), clearly showing repeated events such as a large pothole around 0.5 km along with 4 expansion joints between 3.0 & 3.3 km. Acceleration magnitudes were reasonably consistent, with half of all data within ±12% of the medians at each 10 m interval.

Sampling was increased to the fastest possible rate to ensure all available data was captured and that frequencies were not being mistakenly ignored. For most current smartphones this varied between 100 and 200 Hz, with the higher frequencies typically being offered by the more high-end devices. The Motorola Moto E2, as a budget

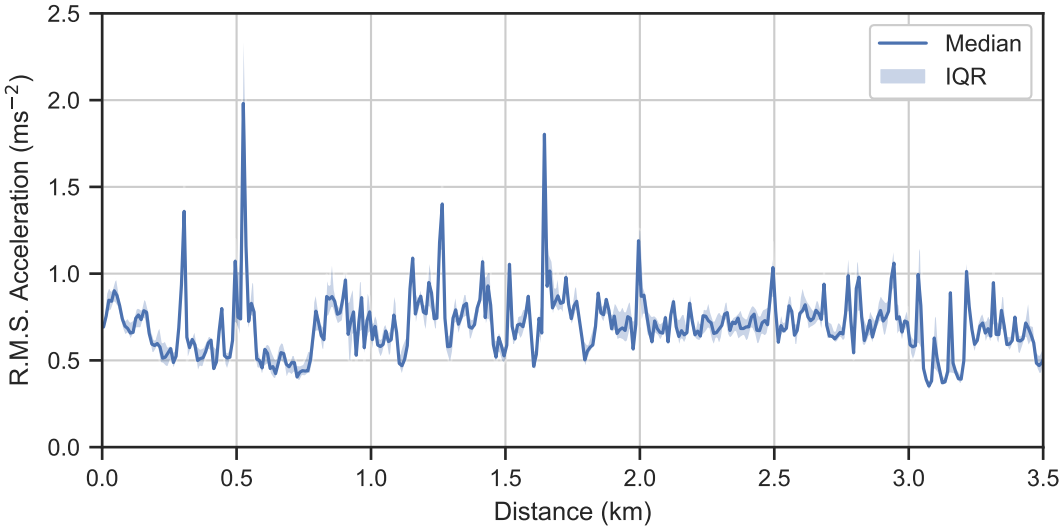

Figure 3.10: Median and inter-quartile range of 10 m resampled r.m.s. high-pass (0.5 Hz cut-off frequency) vertical acceleration recorded along the test route, sampling at 100 Hz, plotted against distance (n = 19 journeys).

smartphone, had an average sample rate of 102.5 Hz.

As with the data recorded at 50 Hz, the 19 journeys sampling at the fastest rate were highly repeatable with an interquartile range between −10 and + 11% (Figure 3.10). The resampled waveforms followed similar trends to the 50 Hz data, clearly showing the same features such as the large pothole and the expansion joints, but with an average magnitude increase of 39%. This increase is unlikely to have been caused by variation in road input as no physical changes were visually observed along the route and the magnitude differences exceeded the identified ±10% variation between journeys.

A comparison plot of the average Power Spectral Density (PSD) for the 50 & 100 Hz recordings was created to see if the increase in magnitudes may have been due to the greater range of frequencies included in the analysis. Figure 3.11 shows the result of applying Welch's method [132] to each journey before taking the mean of the power spectra at each frequency. It can be seen that the power at each frequency follows similar trends for both sample rates up to around 13 Hz, further suggesting that the road inputs were repetitive and any discrepancies were mainly due to the sampling. Examining the 100 Hz sampling closer shows greater magnitudes being registered at both frequencies of 13 & 20 Hz, as well as a new resonant peak at 40 Hz, all of which appear to be the cause of the 39% increase in the time domain. Reduced magnitudes above 13 Hz in the data sampled at 50 Hz suggests the presence of an anti-aliasing filter, with no evidence of an aliased 40 Hz peak.

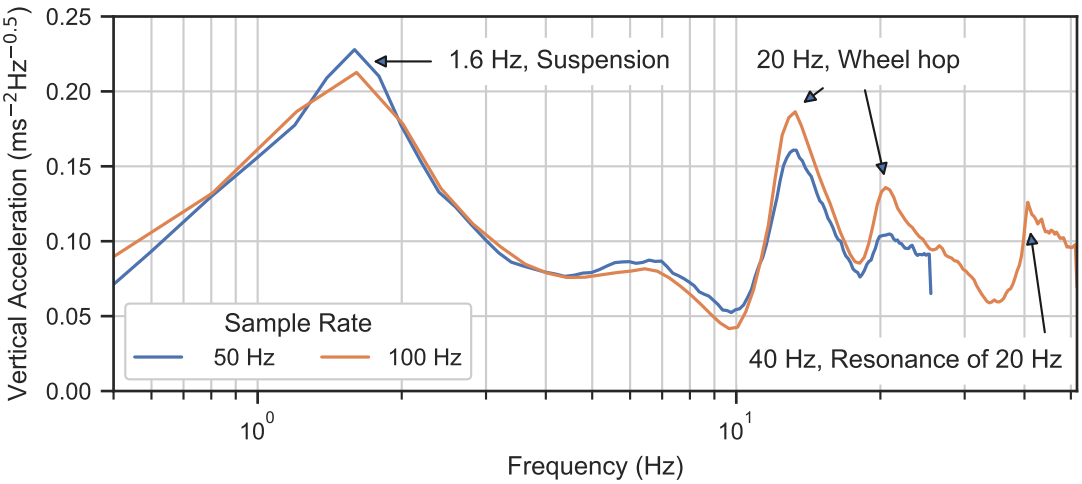

Figure 3.11: Comparison of vertical acceleration spectrum along the test route when sampling at 50 and 100 Hz.

3.7 Microphone

Once reading from the IMU was confirmed as working, a method of logging noise levels was required. Audio data in its raw format was not recorded to ensure privacy and to minimise the smartphone's storage use. Instead, the maximum sound level from the microphone was recorded inside a new service by calling the built-in method 'getMaxAmplitude' of the 'MediaRecorder' class. Documentation simply states that this method returns integers representing the "maximum absolute amplitude" [133], with a maximum value of 32,767 suggested within an official Android Open Source Project repository [134]. There is no mention of units, however it may be possible to convert the integer outputs to decibels using the standard formula:

$$\text{noise level} = 20\ \log_{10}(\text{result of getMaxAmplitude}) \tag{3.7.1}$$

The MediaRecorder needed to be setup as if a file was going to be written, despite only using a single method. The audio source, output format and audio encoder were set to the default values. Different codecs were tested using the method outlined in the next chapter and found to make no difference to the output, so the values were left to the defaults. As the raw audio output was not required, the output file was set to "/dev/null" which the system recognised as an allowable placeholder without actually storing potentially invasive data.

Repeated calls to getMaxAmplitude were required as the method is a single response call rather than a registered listener as with location and IMU data. A timer was setup to call the method every 10 ms to match the 100 Hz sampling frequency of the IMU on the test device.

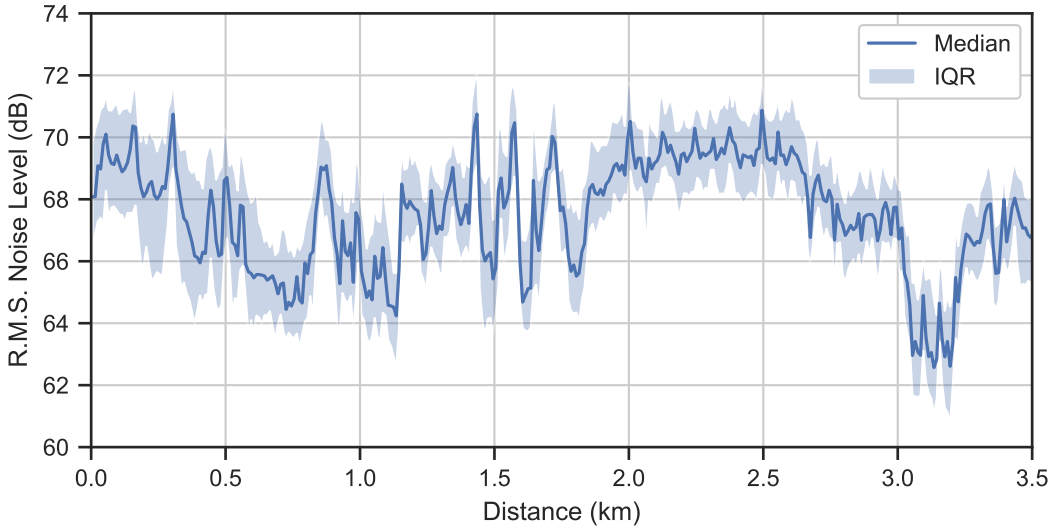

Figure 3.12: Median and inter-quartile range of 10 m resampled r.m.s. noise levels recorded along the test route, plotted against distance (n = 58 journeys).

Noise sampling was reduced to 50 Hz due to multiple false values. The documentation for getMaxAmplitude states that '0' is returned "when called for the first time" [133], however it was found to occur every other sample when calling at 100 Hz. These zero readings were classified as false because they were occurring in an office environment with a constant ambient noise level, which should have resulted in similar values. 50 Hz was chosen as the final sample rate as it was the fastest frequency found to not result in any false readings. The source code of the method was not found, so the reason for the sampling bottleneck could not be determined.

A slight delay in registering the first noise levels was traced to the initialisation time of the MediaRecorder class. Initialisation was therefore moved to coincide with searching for a satellite fix, ensuring noise levels were immediately available when recording begins.

Similar to acceleration, noise levels along the route were found to be highly repeatable for the 57 journeys using the Motorola Moto E2 (Figure 3.12). Features can be seen which line up with the vibration, such as the distinct noise bursts caused by expansion joints between 3.0 & 3.3 km. 10 m interval values varied by ±2.1% between recordings.

Table 3.2: Example of initial components of app-recorded filenames.

Start Time		Device ID	
2018-10-24_08-53-000	-ID	82057280	.csv

3.8 Logging

3.8.1 Filename choice

As the surface of a road, and therefore comfort, changes over time it is important to know when each recording takes place. Filenames include internal time of the smartphone at which recording was started, to millisecond accuracy (Table 3.2). While the probability of two recordings starting at the same exact time were very low, the more devices recording data the greater the risk. Each phone was assigned a random 8-digit Identifier (ID) when first opening the app which was included in the filename to further minimise the odds and enable device-by-device analysis.

3.8.2 Method of writing to file

Recorded data was stored in Comma-Separated Value (CSV) files. The use of a human-readable format enabled faster debugging of recordings as any problems within the file could be viewed within a simple text editor (or more structured within a spreadsheet such as Microsoft Excel) without the need for any intermediary parsing. There are also established methods of importing data from CSV files for most common programming languages, which aided analysis. Although writing in binary would result in reduced file sizes, the relatively small storage use of the CSVs, along with the extra processing required to access the file contents on demand, meant that it was not worthwhile implementing at this stage.

Initially, a new row was written to file after every new accelerometer sample from the IMU. Each row consisted of a comma-separated string of all variables. Because the acceleration, audio and location data were all accessed at different sampling rates and from different sensors, the timings did not match. New data from both the microphone and satellite receiver were therefore assigned to the same row as the next subsequent IMU sample, as this had the highest sampling frequency. Data from non-IMU sensors were accessed by means of global variables which could transfer data between services without the need for a method call. Timestamps were provided by the internal system time, accessed when a new accelerometer sample was available, and was logged as the

difference in milliseconds since the recording began.

Logging of location and noise data was only performed when a new sample was available and was blank otherwise. This reduced the filesize by reducing unnecessary repetition. Sample counters were implemented for both sensors to identify the presence of a new sample while accounting for the possibility of identical consecutive data.

Writing to file was moved to a dedicated service to reduce the amount of work done inside the accelerometer callback method. Variables were still combined into a comma-separated string after each new accelerometer sample, as this was required in order to assign timestamps to the data, however this string was simply added to a global queue. The queue contained each row as a separate element and operated in a first-in-first-out manner which ensured the sample order was maintained. Inputting the comma-separated string to a queue instead of writing to a file kept it in the random access memory rather than the on-board flash storage. A thread-safe queue [135] was implemented to ensure that data could be written within the IMU service without affecting the ability of reading within the logging service. The logging service then contained a timer which wrote the content of this queue as new lines in the file every 5 seconds.

3.8.3 Storage use

Files were stored in a "Recording" directory while being written to, before being moved to a "Finished" directory on completion. This was to enable files to be accessed during development without mistakenly disrupting any ongoing recording.

Compression was implemented to reduce the amount of storage used while maintaining the readability of the CSV format. Initially, this was performed using the Zip file format upon moving to the "Finished" directory. This was modified to using the 'gzip' functionality [136] which can compress the files while writing and therefore reduced the storage impact of long recordings.

3.9 Data retrieval

3.9.1 Uploading to a remote server

After recording accelerations and noise levels of roads, data needed to be transferred from the phone to a PC to facilitate analysis. Retrieving data was conducted manually via USB (Universal Serial Bus) cable during the early development stages, however

this would have required extra input from the users who would then need to email the recorded file. Instead, recorded files were uploaded automatically to a remote server, which could then be accessed from a PC without the need for user input.

A personal directory was created on a server owned and operated by the university, providing a secure location for recorded files accessible only to myself and the server administrators.

A script written in the PHP (PHP: Hypertext Preprocessor) programming language was required on the server to receive data files. First, the script checked that both a file and its name had been sent with the correct keys. These steps, along with checking the filename suffix, were an attempt to ensure no malicious uploads got sent to the server. The script then decompressed the recorded data and stored it on the server within an "uploads" directory. An HTTP (Hypertext Transfer Protocol) response was then returned to indicate that the file had been successfully uploaded.

Uploading of data from the app was handled within a dedicated thread, utilising the OkHttp HTTP client [137], which cycled through the files within the "Finished" directory. An HTTP request was created for each file consisting of the filename and the file itself. This request was posted to the URL (Uniform Resource Locator) of the server PHP script, using HTTPS (Hypertext Transfer Protocol Secure) to protect the data being transmitted, which worked as above. Once a response was received, the uploaded file got moved to an "Uploaded" directory and the next file was prepared to be sent. A simple notification was sent to the user after all files had been processed to indicate the number of successful uploads.

Files were only uploaded via wi-fi to ensure the app did not use any of a user's mobile data. Initially, uploads were performed when the phone connected to a known wi-fi, which alerted a receiver. This receiver then started the upload service if the app was not recording new data and there were files waiting to be uploaded. The receiver also stopped the upload service on the occasions that wi-fi was disconnected. During development, the receiving of wi-fi connection information was removed from the latest Android versions and the scheduling method needed to be modified.

The JobScheduler class [138] built into Android enables code (a 'job') to be run when a predefined set of parameters are met. In the case of uploading the files, this simply meant setting the required network type to "unmetered" and the job to "persisted" which meant that it remained in the scheduler even if the device was rebooted. A new service, tied in to the specific job, was started when the wi-fi condition was met to decide whether to upload files. As before, the upload service was started if the app was not recording and there were files available. The job service implemented a built-in

```
<recordings>
    <phone>
        <id>82057280</id>
        <time>2018-10-24_08-53-000</time>
    </phone>
</recordings>
```

Figure 3.13: Example entry in the XML file containing fully uploaded recordings.

method which took parameters defining the upload job as well as a boolean which defines whether the job needs to be scheduled or not. This enabled simple rescheduling the job when the app was recording. Additionally, a second built-in method was called when the device disconnected from wi-fi as the job conditions were no longer met, which stopped the upload service and rescheduled the job.

3.9.2 Limiting storage use

Successfully uploaded files were periodically deleted from each device to avoid using too much storage space. This required extra code on both the server, to identify which files could be deleted, and the app, to query the server and delete the files accordingly.

An XML (Extensible Markup Language) file was set up on the server to log the start times for each successfully uploaded recording, grouped by device ID. Using an XML file provided a simple method of filtering the filenames. Each unique device ID was added as a sub-element of a 'phone' element inside the root element 'recordings' (Figure 3.13). The log file was updated with the latest uploads as they were being moved to the "sync" directory.

Checking of the log file was required for the app to know which recorded files on the device could be safely deleted. A PHP script was implemented which could be polled by the app and return a response to say whether a recording could be deleted or not. This script required two inputs from the app: the device ID and the start time of the file in question. First, the script cycled through the 'phone' elements until an 'id' sub-element matched the inputted device ID. The 'time' sub-elements of this matching 'phone' element were then compared to the inputted start time. A response was then returned to inform the device whether the recording was present in the log or not.

A job to handle the checking and deleting of app recordings was scheduled after all files had been uploaded from the "Finished" directory. This was set up with the same parameters as the upload job, however started a different service when the conditions

were met. This new service created a new thread if the app was not currently uploading files and there were files within the "Uploaded" directory. Each of the files in the directory were cycled through in alphabetical order by the new thread and the unique start times were separated from the rest of the filenames. The device ID and each unique start time, in turn, were submitted to the script on the server with the returned response informing the next step. Where the ID and start time were in the server log, a method was started the file within the "Uploaded" directory which corresponded to this start time. Receiving a negative response resulted in the corresponding journey file being reuploaded, as this suggested not all data were uploaded successfully.

3.9.3 Retrieval from server

Recorded files needed to be downloaded from the server in order to process the contents. This was initially conducted manually using the Secure File Transfer Protocol client WinSCP, however this would not be suitable for mass data collection due to the time consumption. Instead, an automatic method was required to download and sort the files.

A shell script was created so that two commands could be called consecutively. This script was called daily at midnight to ensure minimal interruption of work. The first command was a call to the rsync application which downloaded all files from the "sync" directory on the server and placed them in a temporary folder on the PC. After all files were successfully copied across they were then deleted from the server.

The second command called a PHP script to organise the downloaded files based on the device ID. Files were moved to the corresponding directory, defaulting to the device ID as a name, and the ID was removed from the filename to leave only the start time. More logical, user-friendly, directory names could be assigned using a simple switch-case function where a known device ID was assigned a specific string value.

3.10 Conclusion

This chapter outlined the steps in development of an Android app to record the vibration and noise levels of roads.

Location data was logged at a rate of 1 Hz with a highly repeatable position between journeys. Vehicle speeds recorded during testing varied by less than 7%, suggesting input from the road should be comparable.

Accelerations were recorded from the built-in accelerometer at the fastest sample rate

possible. Magnitudes for every 10 m interval were found to be repetitive, and showed clear road features, indicating they reflect the road surface and not simply the vehicle characteristics.

Noise levels were logged as the maximum amplitude at a rate of 50 Hz. Multiple journeys along the same route resulted in levels within 5% of each other. As with acceleration, some road features were visible within the noise waveforms.

Operation of the app was contained to a single button to minimise the amount of interaction required from the user which should help facilitate continuous participation in data collection. This was furthered by the automatic retrieval of recorded journeys via wi-fi.

Although the recorded data has been shown to be both repeatable and influenced by the road surface, the accuracy of the detected vibrations and noise levels, along with the effect of additional factors, needs to be assessed before true data collection can be conducted.

CHAPTER 4

App Validation

4.1 Introduction

Recording the in-vehicle environment using the developed Android application required the sensor outputs to be fully understood before any analysis of comfort can take place. It was shown during the development stages that the app-recorded vibration and noise reflect the input from the road, rather than simply the vehicle, as values both varied for a single journey along the test route and were similar for multiple journeys. These recorded values next needed assessing to determine whether they reflect the true environment.

Calibration was performed for both the IMU accelerometer and the microphone in turn. These calibrations subjected the smartphone to a controlled input and compared the recorded results with those of a professional, purpose-built device. The result of the calibrations provided the frequency response of the smartphone accelerometer and the true interpretation of the built-in 'getMaxAmplitude' method used to return noise levels. Outputs from the app could then be adjusted to reflect the true values.

Although the repeatability of recordings along a given route was shown, it was imperative that the app could detect any changes in the road surface over time. Changes may result in a new optimal route being identified. The ability of the app to detect these changes was assessed by comparing data from before and after a known surface modification.

Crowdsourcing could present a problem with the recording of noise data related to the road. Listening habits of a user, such as whether they have the radio on or not, could affect the total noise levels registered by the app. Volunteers cannot be expected to constantly drive in silence, therefore the impact of additional audio inside the vehicle was assessed.

Endless variety in both manufacturer and model of Android device are available. It is inevitable that crowdsourcing volunteers own different devices. Therefore, the differences, if any, which could arise from the use of various smartphones was also assessed.

4.2 Validation Constants

The majority of this chapter examines the data recorded by the Moto E2 smartphone (Motorola, Chicago, IL, USA), as used during development in Chapter 3.

Calibration of app outputs were performed using the "testing" variant, as described in Section 3.4.5. This differed from the main app only in the removal of the location receiver. This enabled data to be recorded within buildings without requiring a satellite fix and also without the recordings being automatically stopped due to a lack of movement (Section 3.5.3).

Analysis involving data recorded on a road was recorded along the 3.5 km route used during development and described in Section 3.2.

4.3 Calibration of Smartphone Accelerometer

Accelerations registered by the smartphone were found to be repeatable for multiple journeys along a route (Figure 3.10), although the values may not have been accurate. Testing was therefore required to ensure the values outputted by the IMU were understood and the subsequent analysis would be meaningful.

4.3.1 Frequency Response

Frequency response is rarely listed in the specifications of smartphones, unlike the processor speed or graphics capability. This is because the primary function of the built-in IMU is to enhance the user experience, typically by identifying a change in orientation (portrait to landscape, for example) and updating the device display accordingly, or for enabling gesture inputs such as the "shake to give feedback" feature in several apps. These features require only relative thresholds to be serviceable and therefore do not implicitly require a sensor with high accuracy. Although the IMU in a smartphone is also used to provide input for mobile gaming, this tends to be limited to gyroscopic motions rather than the accurate accelerations that are required to record the in-vehicle environment. Therefore, the sensed values of a smartphone needed to be compared to those of a known reference accelerometer.

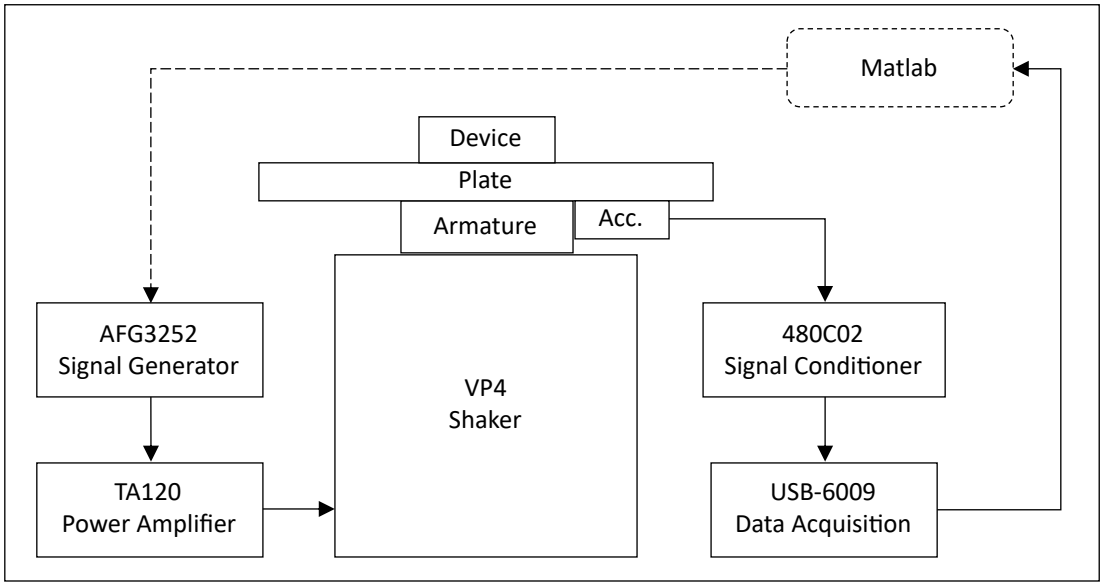

Figure 4.1: Diagrammatic representation of the frequency response experimental setup (not to scale).

Calibration of a smartphone IMU can be performed in accordance with the international standard which specifies the instrumentation required for measuring the human response to vibration [139]. This standard outlines the validation of research instruments which were not designed for measuring vibration (such as smartphones) as well as specifying the steps for manufacturers of the professional instruments. Research devices can be tested for compliance with the specifications in the standard using an electrodynamic shaker. This is achieved by comparing the output of the smartphone with that of a calibrated reference accelerometer at set frequencies and r.m.s. magnitudes.

4.3.1.1 Experimental Setup

A working setup was required to enable accurate calibration to be conducted. The components of the setup were connected as shown in Figure 4.1, which is outlined below.

The shaker equipment consisted of a VP4 electromagnetic vibrator powered by a TA120 solid state power amplifier (both of Derritron, Hastings, UK). The VP4 was rated to support a static load of 1.8 kg and had a specified peak sine force of 133 N, both of which were more than sufficient for the shaking of the 145 g smartphone. The 70 mm diameter armature also offered a large area for supporting the smartphone, along with a 3x3 grid of tapped M6 holes for attachment. Having a substantially massive construction (52.2 kg) helped to isolate the testing from any environmental vibrations in lieu of

a dedicated anti-vibration table.

Although large, the armature was too small to attach smartphones directly and therefore an additional platform was required. A plate was formed from 10 mm thick acrylic with 4 countersunk holes to locate with the corners of the grid of tapped holes on the armature. The plate, at 300 by 300 mm, was designed to enable all manner of devices — whether smartphones or tablets — to have their sensor outputs assessed. In hindsight, a device-specific plate would have enabled more robust attachment and may have increased stiffness of the system.

The smartphone was positioned geometrically in the centre in an attempt to distribute the load evenly during testing, although both the weight and IMU may not be at the centroid. Generic electrical tape was used to provide a secure locating method without damaging the smartphone.

A USB-controllable signal generator (AFG3252, Tektronix, Beaverton, OR, USA) enabled programmable tests to be performed. A Matlab script was written to change the frequencies of the sinusoidal signals using the Virtual Instrument Software Architecture (VISA) API, implemented using the Instrument Control Toolbox [140]. Each frequency was held for 10 seconds via the 'pause' function in Matlab. A 0.7 millisecond discrepancy was identified between the expected and recorded times at each frequency — possibly caused by a delay between Matlab transmitting to the signal generator and the frequency actually changing. The 10 second 'pause' was reduced to 9.9993 seconds to account for this transmission delay and ensure the duration at each frequency is correct. The start times, since the script sent a VISA signal to turn the signal generator output on, were recorded for each frequency tested in case the durations did vary.

Reference equipment comprised of a calibrated 352C65 accelerometer (PCB Piezotronics, Depew, NY, USA) which connected, via a 480C02 signal conditioner (PCB Piezotronics), to a USB-6009 data acquisition device (NI, Austin, TX, USA). The accelerometer was positioned on the underside of the acrylic plate, within 5 mm of the armature, by locating the American standard 5-40 mounting stud of the control sensor inside a ⅛" clearance hole and adhering it in place. Accelerations sensed should then have been representative of the whole shaker plate.

Logging of the accelerometer was implemented at a sample rate of 2,000 Hz within Matlab, using the Data Acquisition Toolbox [141], along with the signal control outlined above. At the time, Matlab was used for all analysis of data of the shaker-based experiments. Therefore, recording accelerations within Matlab would enable easier analysis as the timestamps which correspond to the frequencies of interest could now be logged instead of identified through processing. The accelerations were read by Matlab in the

background to ensure the main thread, which controlled the signal generator, was not interrupted.

Test magnitudes of 1.00 $m \cdot s^{-2}$ and 9.81 $m \cdot s^{-2}$ (1 g) were chosen for the calibrations to reflect the typical and extreme magnitudes one might expect along roads [44]. These also complied with some of the parameters set out in International Organisation for Standardisation (ISO) 8041 [139] for whole-body vibration.

Sinusoidal frequencies between 1 and 100 Hz were run in 1 Hz intervals to assess the full frequency range of the IMU inside the Moto E2, as well as the effect of any anti-aliasing filter which may be present.

4.3.1.2 Feedback Control

Constant voltage would result in different magnitudes from the shaker depending on the inputted sinusoidal frequency. Therefore, the voltage would need to be modified to obtain the required constant r.m.s. acceleration across all frequencies. This required a feedback loop because clear links between the inputted voltage, sinusoidal frequency and the resultant magnitude were not found.

Feedback was added in the form of proportional control, under the assumption that the input voltage correlated linearly with the resultant r.m.s. magnitude at a given sinusoidal frequency. The shaker was allowed to enter steady state after a frequency and initial voltage had been set by pausing the Matlab script for a certain number of cycles, depending on the frequency. Average r.m.s. magnitude was then calculated from the 0.5 Hz high-pass filtered accelerations of the reference accelerometer over the final 75% of the paused duration. When the r.m.s. value was within a threshold of 0.005 $m \cdot s^{-2}$ from the required magnitude, the time was recorded and the shaker was left to continue for a set delay before recording the time at the end. These times could then be used to grab the required data when it came to analysis.

Proportional control was implemented when the threshold was exceeded. Gain was calculated by dividing the required magnitude by the computed r.m.s. value. This gain was then multiplied by the initial voltage to obtain a new input voltage. Voltages were rounded to 3 decimal points, the maximum precision of the signal generator, and compared with the initial voltage. On the occasions where the new and old voltages were equal, 0.001 V was either subtracted (when gain > 1) or added (otherwise) to the new voltage before sending it to the signal generator. The process of waiting for steady state, computing r.m.s. magnitude and adjusting the input voltage continued until the shaker value was within the threshold.

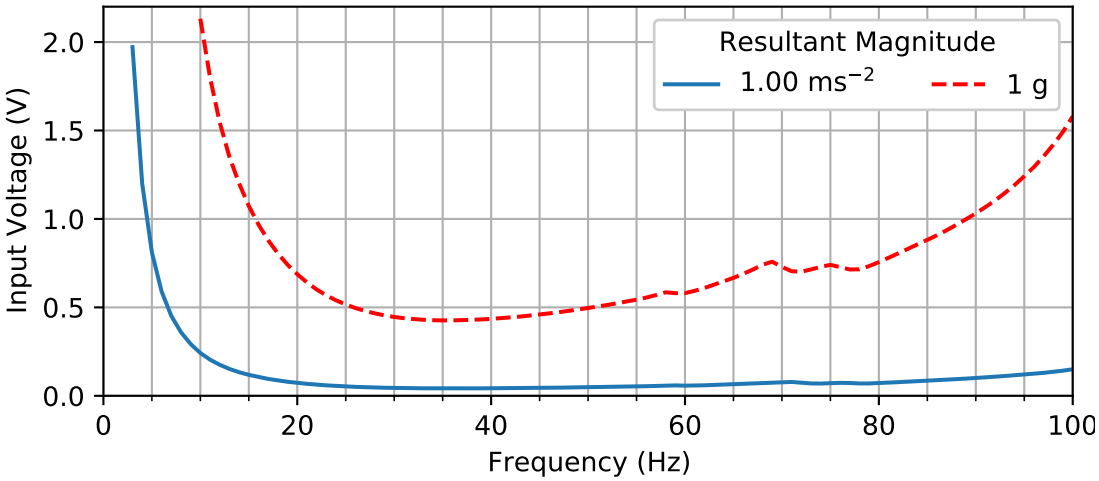

Figure 4.2: Input voltage required to achieve the desired r.m.s. magnitude at each tested frequency using the VP4 shaker.

Final voltages for each frequency were stored for each run to serve as the initial voltage for the next run. This reduced the time until the required r.m.s. magnitude was achieved. Feedback was operational on all runs as the room in which the system was based fluctuated in temperature. The need for feedback was further shown by very inconsistent voltage requirements to obtain the same r.m.s. magnitude at the same frequency between days, even when moved to an air-conditioned laboratory.

It was found during preliminary tests that the feedback control did not have enough sensitivity and the maximum voltage required was less than 2.5 V. The AFG3252 signal generator was limited to outputting voltages between 0.05 and 5.00 V (peak-to-peak) and the degree of precision was not able to be adjusted. Therefore, the gain on the power amplifier was changed from 100% to 50% to effectively double the sensitivity.

4.3.1.3 Results

Voltage gain was required to be set at 50% on the power amplifier both to produce small enough values for the full range of frequencies at 1 m·s^{-2} and for increased precision in obtaining the desired magnitude accelerations. Frequencies from 27 – 51 Hz required voltages less than the 0.05 V minimum output from the signal generator (Figure 4.2). Reducing the gain to below 100% enabled these frequencies to be assessed along with those both smaller and larger. Ideally, the gain would be programmable along with the signal generator, however it was simply a physical dial without graduations and required a voltmeter to be accurately set.

Calibration at low frequencies was limited by the available range of voltages which could be supplied by the signal generator. Although the 50% gain was required for

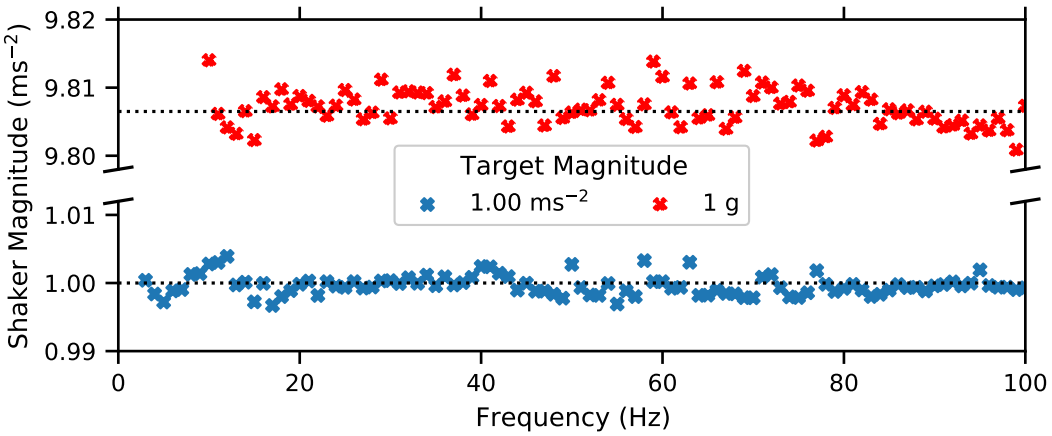

Figure 4.3: Average acceleration magnitude obtained from the VP4 shaker using the feedback control.

intermediate frequencies, it restricted the minimum frequencies to 3 and 10 Hz for the 1.00 and 9.81 m·s^{-2} tests respectively (Figure 4.2). Required input voltage increased exponentially as frequencies decreased below 32 Hz. Increasing the voltage gain on the power amplifier to 100%, the maximum obtainable value, would only have enabled a single additional frequency to be analysed during calibration at each of the chosen magnitudes. Although maximising gain would have theoretically enabled additional testing frequencies, these would have been impossible to obtain due to the maximum peak-to-peak shaker travel of 6.35 mm.

Feedback control of the shaker was highly successful at achieving the desired acceleration magnitudes at each frequency, despite only implementing proportional control and not including integral or derivative components. Shaker control was more effective in terms of acceleration at the lower magnitude runs (average deviation of 0.0012 m·s^{-2}) compared to that of the higher magnitude (average deviation of 0.0022 m·s^{-2}), as shown in Figure 4.3. Maximum deviations from the target magnitudes were 0.0039 m·s^{-2} (0.39%) and 0.0075 m·s^{-2} (0.08%) for the calibrations at 1.00 and 9.81 m·s^{-2} respectively, which more than satisfied the allowable variation [139].

Subject to sinusoidal excitation, the Motorola smartphone gave a different response at a magnitude of 1.00 m·s^{-2} compared to 9.81 m·s^{-2} (Figure 4.4). The response at 1.00 m·s^{-2} matched to within 10% of the reference accelerometer up to 40 Hz, as opposed to only 30 Hz at a magnitude of 9.81 m·s^{-2}. Above these frequencies, the recorded magnitudes at both amplitudes diverged rapidly from the reference.

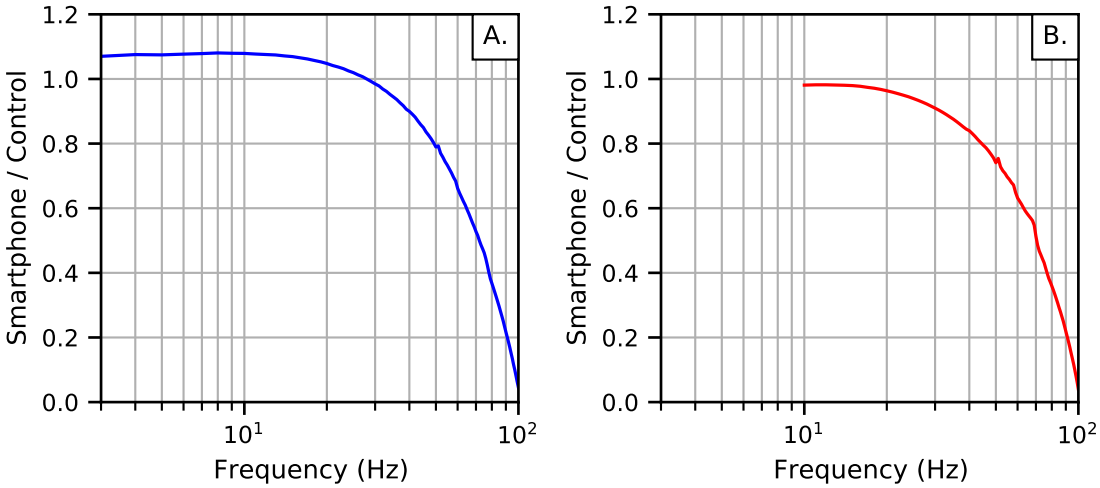

Figure 4.4: Sinusoidal frequency response of Motorola Moto E2 compared to a PCB Piezotronics 352C65 at constant magnitudes of 1.00 $m \cdot s^{-2}$ (A) and 9.81 $m \cdot s^{-2}$ (B).

4.3.2 Sampling Jitter

As mentioned in Section 3.6, the time intervals between subsequent IMU samples were not constant. Ideally, any sampling jitter was normally distributed and could be assumed negligible. If the jitter was significant, post-processing methods would be required to correct the timestamps for all recordings.

To quantify the sampling jitter, a 90-minute recording was made using the Motorola Moto E2. Resolution was increased for the test by modifying the app to log timestamps in nanoseconds instead of milliseconds. At the end of the recording, the differences between each timestamp were calculated and binned in 0.1 millisecond intervals (Figure 4.5). Although the median interval between samples was within 1% of the expected 10 ms, the distribution level of jitter was high with half of the intervals recorded having a spread of 5.4 ms. Large unexplained spikes were visible at intervals below 7.5 ms, however they did not appear to affect the acceleration frequencies analysed in the previous section.

4.4 Calibration of Smartphone Noise

Documentation regarding the model or specifications of the microphones built into smartphones is rarely published by manufacturers. There also was no built-in method available for the Moto E2[1] with which to access any microphone data, unlike with the

[1] A 'MicrophoneInfo' class was introduced with Android 9 (API 28), in August 2018, which provides methods for obtaining the microphone frequency response, sensitivity and max/min sound pressure lev-

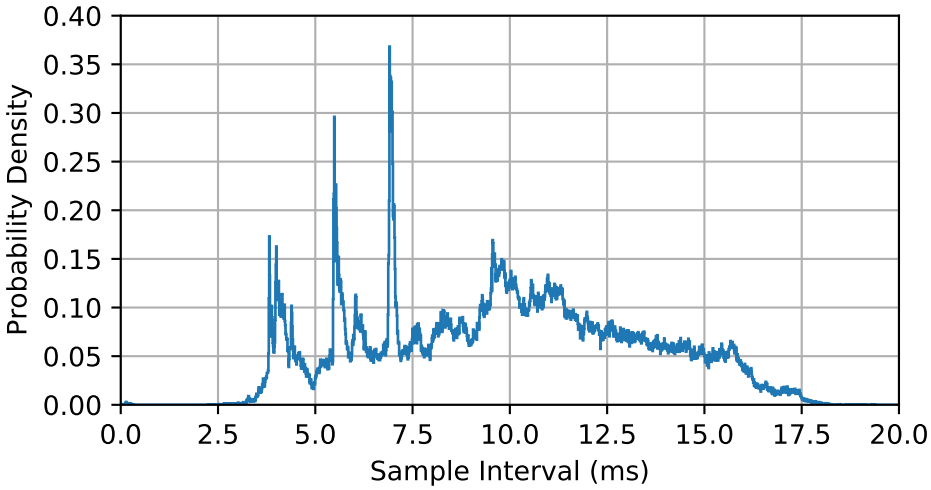

Figure 4.5: Distribution of IMU sample intervals, recorded by the Motorola Moto E2 over a 90-minute period (bin size = 1×10^{-5} s).

IMU. Instead, the output of the smartphone microphone needed to be compared to that of a calibrated sound meter to see how they compared.

Similarly, the output of the built-in method used within the app to obtain noise levels is not fully defined. As outlined in Section 3.7, the method returns an integer representing the maximum absolute amplitude, however in unspecified units. Confirmation of the theory that the output can be converted to decibels using the standard formula (Equation 3.7.1) could be obtained through comparison with the noise levels from a professional meter.

4.4.1 Method

Calibration testing was performed in the loudspeaker lab belonging to the Hearing Sciences research group. The lab consisted of a heavily soundproofed, but not anechoic, room containing 24 VX 6 speakers (Tannoy, Coatbridge, UK), in a circular array of approximately 4 m diameter, which were controllable from the outside. Each speaker was positioned at the same height, equidistant from one another and directed at the centre of the circle. This was capable of generating a diffuse sound field in the centre with an equal sound pressure level in every direction. As well as minimising the directionality of the noise input, a diffuse field removed the possibility of any standing waves. Therefore, by positioning the smartphone in close proximity with a precision sound meter (2260 Investigator with Type 4189 free-field microphone, Brüel & Kjær, Nærum, Denmark) in the centre of this array (Figure 4.6) both devices should have been subjected to identical noise and thus directly comparable.

els, however the Moto E2 only has support up to Android 6 (API 23).

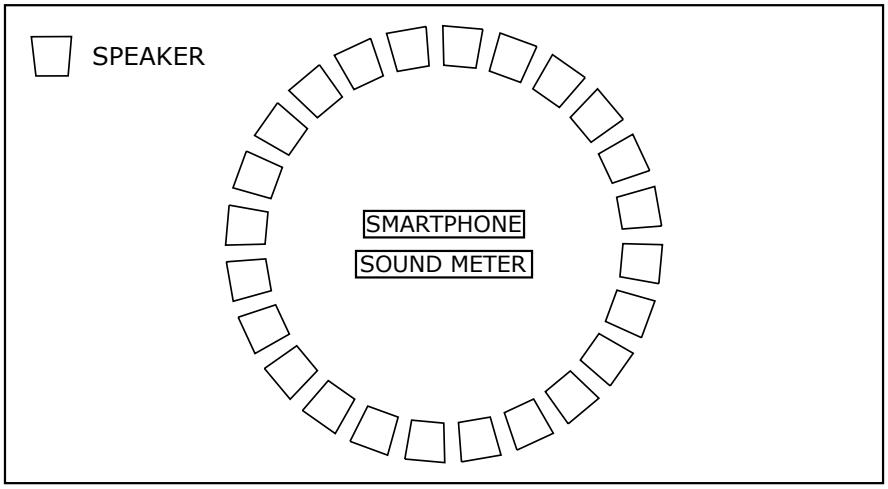

Figure 4.6: Diagrammatic representation of the loudspeaker positioning for noise calibration (not to scale).

Frequency analysis could not be performed on the noise levels recorded by the app due to the getMaxAmplitude method returning a single summary value. Instead, pink noise [142] was used because it is a combination of a range of frequencies with equal power per octave, simulating sounds one may expect to encounter while driving. Pink noise is recommended for testing purposes, especially in circumstances where an ideal free-field is unavailable [143]. Having a range of frequencies within the signal further reduced the risk of standing waves forming as the outputted wavelengths were constantly changing and therefore any reflected waves would not match those with which it may have come in contact.

A Matlab function, created and provided by Hearing Sciences, was used to output a pink noise approximation to the 24 loudspeakers. This function enabled the user to input the desired frequency range, noise level and duration of the outputted pink noise, with the suggested default values of 100 – 10,000 Hz, 60 dB and 5 s respectively. A simple 'for' loop was encoded for the calibration of the smartphone to produce 5 seconds of pink noise at a range of levels from 50 to 75 dB, chosen based on the recorded levels during development (Figure 3.12), in 5 dB steps with 3 second pauses between to aid differentiation.

As noise levels set by Matlab may not have matched the output of the loudspeakers due to the amplifier gain setting and any discrepancies which may have been present in the speakers themselves, the decibel offset between the Matlab input and the output needed to be identified and applied to the input to generate the desired levels. Calibration of the input was performed by conducting some dummy tests while standing inside the soundproofed room, but outside of the circular array of speakers, and making note of the A-weighted noise levels (dB(A)) registered by the sound meter ap-

Table 4.1: Comparison of desired pink noise output, as set within Matlab, and the resultant levels recorded by the 2260 sound meter.

Desired Level (dB)	Recorded Range (dB(A))
50	50.0 – 50.5
55	55.1 – 55.3
60	59.9 – 60.2
65	64.8 – 65.1
70	70.0 – 70.1
75	74.9 – 75.3

plying a slow time weighting. An offset increase of 3 dB was identified during the first test run. Feeding this back into the Matlab script by simply subtracting 3 dB from the desired level resulted in a near perfect match between input and output noise levels (Table 4.1). The levels registered by the sound meter were observed to vary slightly, most likely due to the different frequency components of the signal at any one time, hence the provided range in the table. No part of the setup was adjusted after the input was calibrated, therefore the output was assumed to be perfect and the noise levels recorded by the app were directly compared to those inputted via Matlab.

The Moto E2 smartphone was positioned flat atop a standard tripod, ensuring the built-in microphone was on the same plane as the 2260 sound meter and the centre of the loudspeaker drivers. After running through the six bursts of increasingly loud pink noise, the app-recorded values were assessed using the Python programming language. Periods without pink noise were identified by taking the median of all recorded noise levels (Figure 4.7), as a greater amount of time was spent not running the tests (due to setting up, leaving the room, pauses between levels, and returning to the room). This meant that the time spent above this level was likely due to the loudspeaker input. The six periods of pink noise were extracted by identifying the pauses in between as at least 3 s below the median. Average noise level for each period was then calculated as the r.m.s. of all values within that time frame.

4.4.2 Results

Values registered by the Moto E2 correlated perfectly (R^2 = 1.00) with the sound meter at each noise level tested (Figure 4.8). This proved that directly converting the result of 'getMaxAmplitude' to decibels does correspond to true noise levels as theorised.

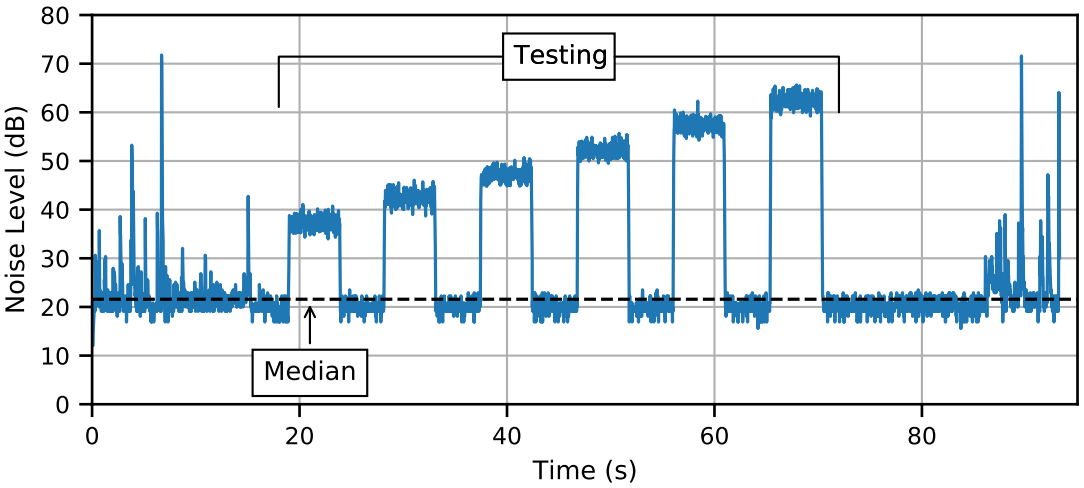

Figure 4.7: Noise levels recorded by the app running on the Moto E2, showing the median value and the period of pink noise bursts.

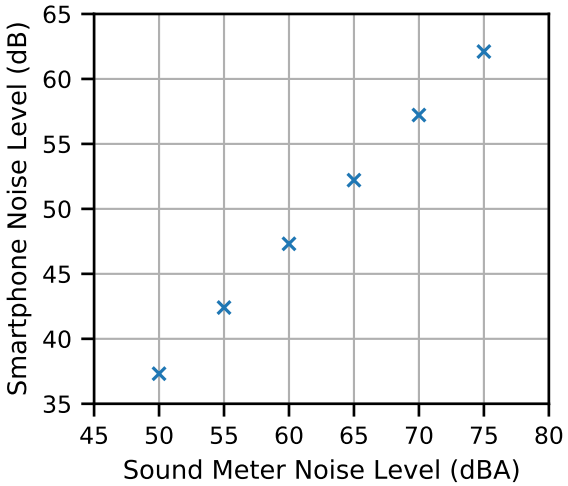

Figure 4.8: Plot of average r.m.s. noise level recorded by the Motorola Moto E2 vs A-weighted noise level from a sound meter, subject to pink noise containing frequencies from 100 – 10,000 Hz.

Although the smartphone noise levels correlated well, they were lower than those of the sound meter by an average of 12.74±0.25 dB. Therefore, any assessment of roads using data from this device would need to be adjusted by this offset and should then reflect the true noise levels.

The Moto E2 was theoretically capable of reporting accurate noise levels up to 103.0 dB(A). This was calculated by converting the maximum output value of the 'getMaxAmplitude' function, 32767, to decibels and then applying the above offset. All inputs above 103 dB(A) would be reported equally.

4.5 Detecting Road Surface Changes

One of the drawbacks of traditional road monitoring is the infrequent nature in which they are performed. Continuous monitoring using the app should have dual benefits of increased confidence in recorded values and enabling surface changes, and their effects on the in-vehicle environment, to be detected.

Road surfaces can change over time due to both the amount of traffic and the weather. Traffic, especially Heavy Goods Vehicles and buses, can produce rutting in roads due to the pressure exerted by the tyres. Weather, specifically precipitation, can instead result in potholes where water has infiltrated the surface. This water can either slowly erode particles away or, in the winter, freeze and expand before thawing to leave cracks and holes. These changes tend to occur gradually over an extended period of time, therefore it was imperative that the data from the app could identify these changes as a particular road could degrade and become worse for the outcome of a transferred infant.

Whereas degradation is typically a slow progress, repairs to roads can result in great changes in surface quality over the course of only a matter of days. The chosen test route was subject to partial resurfacing during the development of the app. This was fortunate as several journeys had already been recorded which could then be compared against any new data to check for any changes. Knowing a specific date for when the road surface changed enabled easier assessment of the recording ability of the app compared to trying to identify degradation over time.

The driver recording journeys along the test route reported that repair works had been carried out between the 24th and 31st May 2017. This consisted of resurfacing specific stretches of carriageway between 0.5 & 1.5 km along the route. One of the stretches corresponded to the large pothole (Figure 4.9A) for which the test route was partially chosen, as it was a clear reference point which should have been (and was) evident

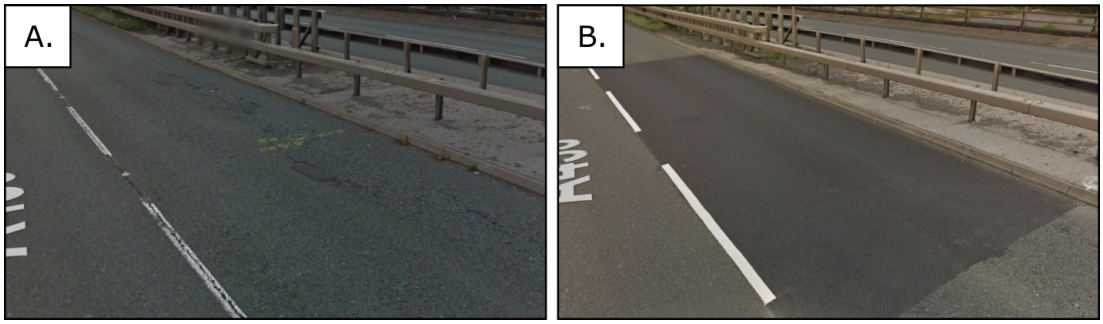

Figure 4.9: Photographs of part of the test route before (A: August 2016) and after (B: August 2017) resurfacing works (© Google Street View).

in the data. It was thought that this same pothole was one of the main reasons for the repair on this high traffic route as only select portions of road were resurfaced. Improvements to the road surface could clearly be seen by eye (Figure 4.9B), however it needed to be seen if they were also visible in the app-recorded data.

4.5.1 Resurfacing Comparison

Ninety journeys were recorded along the test route after the resurfacing works were carried out, which could be compared to the 19 journeys recorded before (excluding the earlier journeys sampling the IMU at the 50 Hz "game" sampling rate). This provided a reasonable dataset for comparing the vibrations before and after resurfacing. Unfortunately, the addition of noise level sampling was incorporated into the app after the resurfacing works had completed and therefore could not be examined.

Resurfacing the road could be seen to reduce the magnitude of accelerations caused by the large pothole at 0.5 km along the test route (Figure 4.10). Peak pothole acceleration was reduced by 47%, although the majority of the route was unchanged and averaged to within 3.8% of previous recordings, suggesting that the app could detect surface changes appropriately.

The majority of the recorded acceleration was within 0.3 $m \cdot s^{-2}$ (R^2 = 0.84) when comparing before and after resurfacing, as shown in Figure 4.11. Each marker within the plot represents the r.m.s. magnitude, averaged over all journeys either before or after resurfacing, for each 10 m interval along the test route. The two greatest outliers in the data were due to the road improvements visible at 0.5 and 1.25 km. These locations, combined with how the unchanged portions of route registered closely matching vibrations, proved that the detection of the surface repairs did not occur by chance. Clearly showing improvements due to repair works suggested that any surface changes, including those that happen gradually, would be extractable from the recorded data.

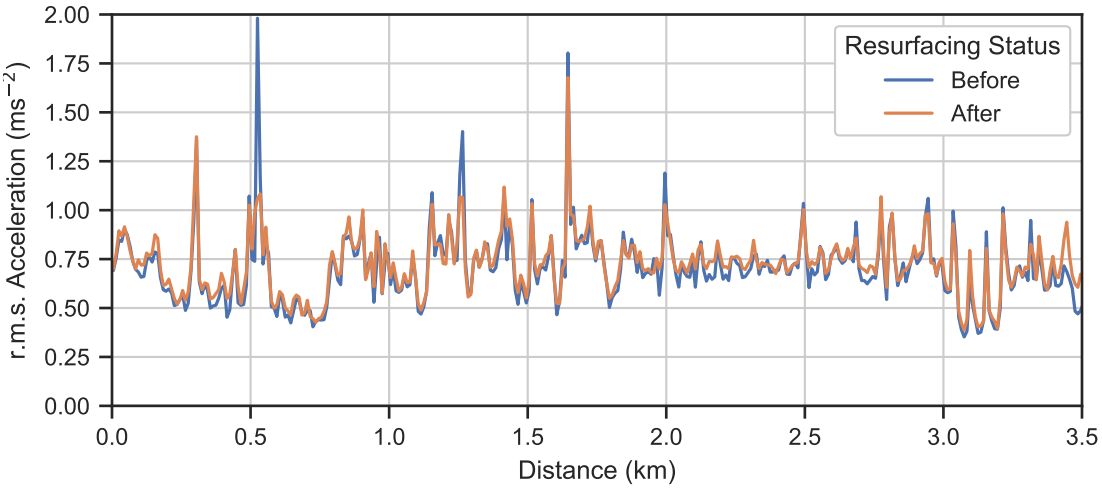

Figure 4.10: Comparison of 10 m resampled r.m.s. high-pass (0.5 Hz cut-off frequency) vertical acceleration recorded along the test route before and after resurfacing, sampling at 100 Hz, plotted against distance.

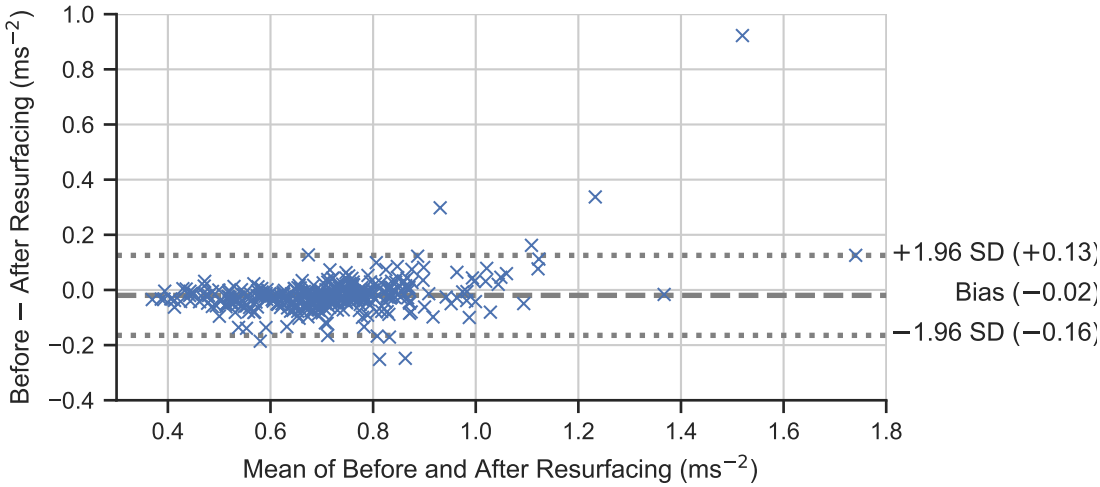

Figure 4.11: Bland-Altman plot comparing the average vertical accelerations recorded before and after resurfacing work along the test route.

4.6 Impact of Radio on Noise Levels

Ambulance staff need to converse during transfers, and clinical equipment will output some noise, but the overall level of noise will be dictated by the road. Ideally, the app would only register the noise due to the road itself as this can be affected by route choice. All additional noises, and their effects on the total magnitude, would need to be studied specifically to see if they should be reduced.

One of the problems with crowdsourcing data is the lack of control over environmental factors. Car drivers listen to music the majority of the time and will vary the volume depending on listening preferences. If app users were told they had to drive in silence, levels of participation will likely suffer, however it was highly likely that audio listening habits will affect recordings. The effect of supplementary music within vehicles could be assessed by comparing datasets with a known radio status.

4.6.1 Method

Two journeys along the test route were chosen where the only theoretical difference was whether the driver was listening to a podcast or silence. These journeys occurred within the space of a few days (14^{th} & 16^{th} August 2017) and were performed with similar speeds (average difference of 0.06%) and no roadworks or extreme weather conditions to possibly affect results. Although natural human variation in driving meant the exact position within the lane of the road may have varied, the resultant noise should have been similar as shown in Figure 3.12.

4.6.2 Results

Both datasets displayed similar trends along the test route, although the magnitudes were clearly different (Figure 4.12). The presence of radio entertainment resulted in an average increase of 3.9 dB with an interquartile range between 2.7 and 5.3 dB greater than that of the journey without. A few (< 2%) of the 10 m segment were louder during the journey with the radio off, but only by an average of < 1 dB. Overall, it appeared that the addition of supplementary audio amplified the noise level from the road.

Noise levels recorded with the radio on and off had a low correlation of $R^2 = 0.40$ (Figure 4.13) despite the waveforms looking visibly similar. This may be due to the particular audio being listened to in the journey with the radio on, or due to slightly different sections of road resulting in variations in noise input to the vehicle. Additionally, this could be due to the limitations of the resampling required for aligning signals

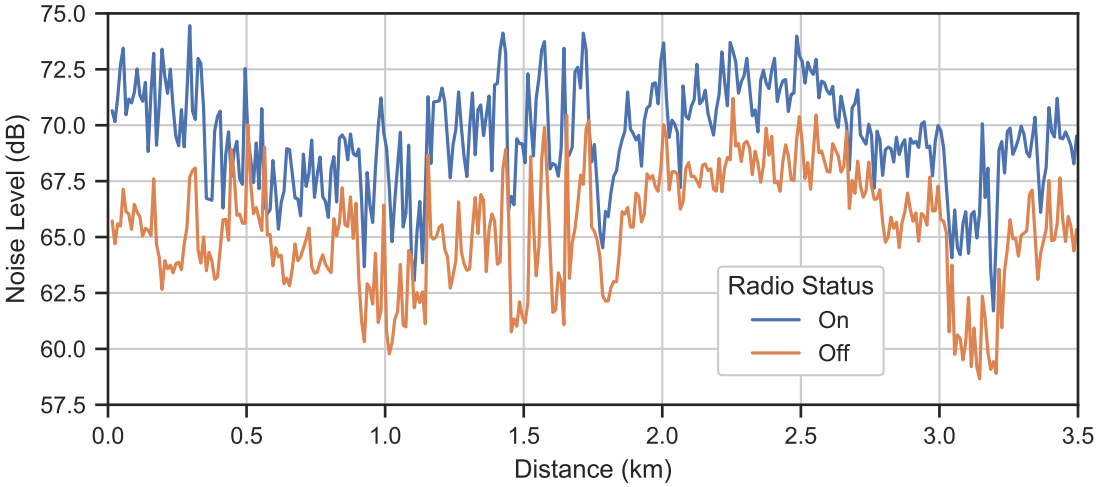

Figure 4.12: Unadjusted noise levels along the test route, recorded on consecutive days, showcasing the effect of radio.

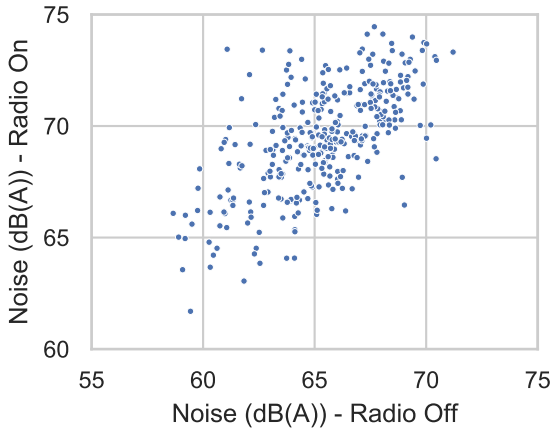

Figure 4.13: Distribution of noise levels with the radio on against the distribution of noise levels with the radio off.

from different journeys. Further work is required using controlled inputs along with a variety of different supplementary audio tracks to enable the assessment of the true impact that noise within the cabin has on the total magnitude.

4.7 Impact of Smartphone Selection

Another possible complication arising through crowdsourcing is the range of different smartphone manufacturers and models available which use the Android OS. These models can vary greatly in cost, with the quality of the built-in electronics typically increasing with cost. Recordings between different smartphones could therefore be affected by the sensors included due to different sampling frequencies, accuracies, and precisions. Restricting crowdsourcing to only a single device would have the benefit

of producing directly comparable data but would severely limit the number of eligible volunteers and therefore the amount of recordings which could be collected over a set time-frame. Instead, analysis would need to account for the differences between smartphones.

The manufacturer and model of an Android device are available programmatically via the built-in 'os.Build' class. This information could then be sent, along with the device ID, by the app to the server to aid the analysis of data. Calibration adjustments (such as noise level offset) could then be applied when the contents of each file are read, accessing the appropriate values based on the device ID associated with the recording. A specific model would only require calibration once as, theoretically, each of the same model of smartphone should contain near-identical sensors — subject to manufacturing tolerances — which could be assumed to have the same values. While only one of each smartphone model would need to be calibrated, this would still need to be performed in a lab environment.

An alternative method of calibration would be to compare the data from an unknown smartphone to that of a previously calibrated device. It has previously been shown that multiple recordings made by a single smartphone along the same route result in similar levels of both vibration (Figure 3.10) and noise (Figure 3.12). Thereby it follows that the data from a different device along that same route would be comparable. Calibration of this new device could then be performed by identifying the differences between the two sets of data.

4.7.1 Comparison between Moto E2 and Moto G5 smartphones

The phone used for road recordings was changed from the Moto E2 to a Moto G5 (Motorola, Chicago, IL, USA) in January 2018. This enabled data along the test route to be recorded by a new device while keeping the other variables (driver, car, listening habits) as constant as possible, facilitating direct comparison with the data recorded using the Moto E2. A second benefit of this upgrade was the availability of the Moto E2 for calibration in the lab as it was no longer required by the owner.

The Moto G5 was not available for the calibration of either the on-board IMU or microphone at the owner's request due to being their personal device. This represented a similar situation to crowdsourcing, where the owner would likely be unknown, with the need for calibration through comparison of devices.

Components inside the Moto G5 may have been improved compared to the Moto E2 due to being released 2 years later and being slightly more expensive. Although the

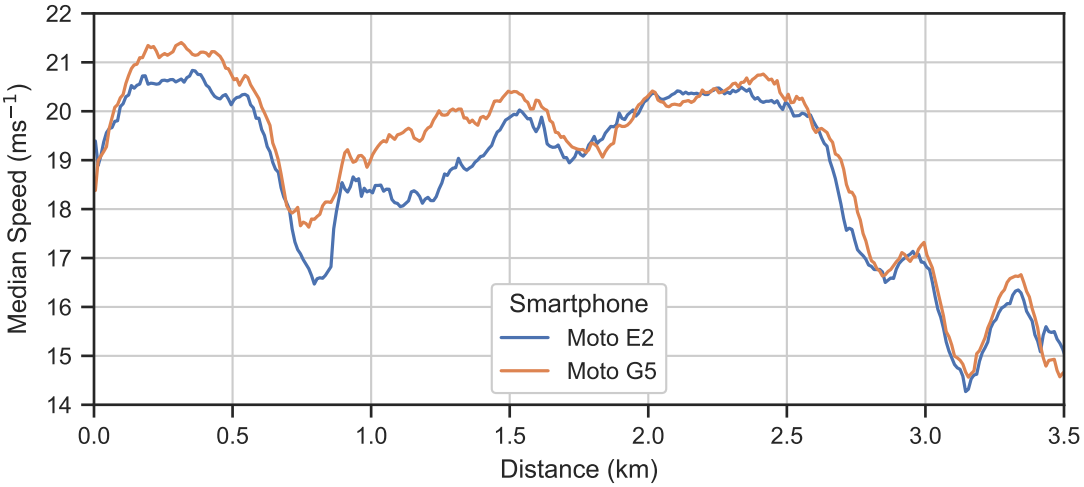

Figure 4.14: Comparison of 10 m resampled median vehicle speeds recorded along the test route using different smartphones, plotted against distance.

maximum IMU sampling frequency of 100 Hz was the same as that of the Moto E2, the precision may have differed. Similarly, the material composition of the body of the phones (aluminium/plastic for the G5, rubberised for the E2) may have affected the vibration transmission. No data was available on the microphones in either smartphone.

46 journeys along the test route were made using a Moto G5 between 29th January and 6th June 2018. These journeys could be compared against the data recorded by the Moto E2 after the resurfacing, as the values before were not directly comparable as shown in Section 4.5. A total of 90 journeys were recorded after the resurfacing, although only the final 57 of these included noise data due to microphone implementation being added later in development.

4.7.1.1 Vehicle Speed

Average vehicle speed recorded by the Moto G5 was largely repeatable, with half of all journeys within 2.1 m·s^{-1} (11.1%) of the average for each 10 m section of the test route. This was similar to the results from the Moto E2 journeys, as shown in Figure 3.7, which varied by 2.3 m·s^{-1} (12.7%).

Both smartphones registered similar speeds for the second half of the test route, while the journeys with the Moto G5 tended to be slower along the first 1.75 km (Figure 4.14). This resulted in an average 0.4 m·s^{-1} (2.1%) increase across the whole route compared to the Moto E2. The similarity of the speeds suggests the data recorded by both smartphones should be reasonably comparable.

Splitting the comparisons into both halves of the route gave contrasting increases of 0.7 m·s^{-1} (3.5%) and 0.1 m·s^{-1} (0.7%). Matching closely between smartphones in terms

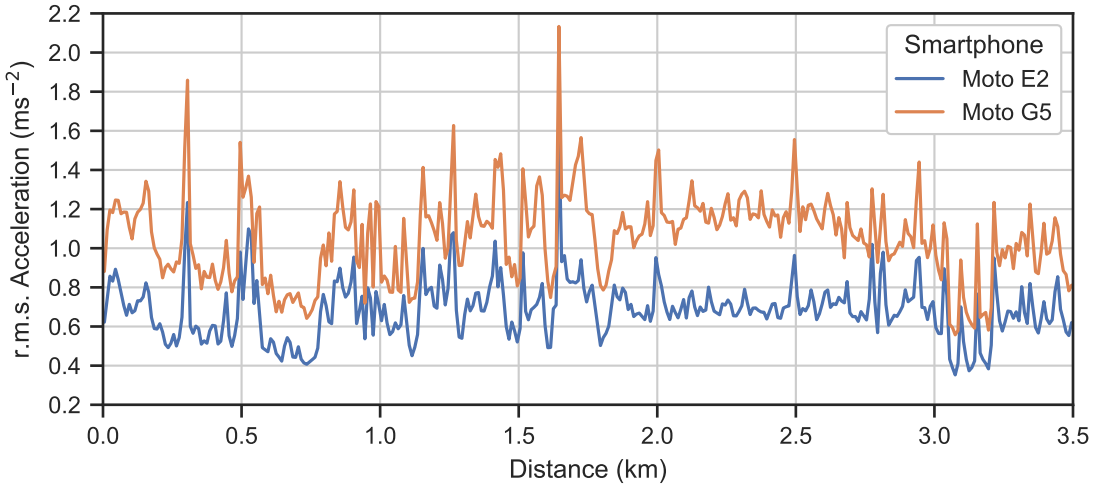

Figure 4.15: Comparison of 10 m resampled r.m.s. high-pass (0.5 Hz cut-off frequency) vertical acceleration recorded along the test route using different smartphones, sampling at 100 Hz, plotted against distance.

of speed, this suggests that the second half of the route may have produced more comparable sensor values.

4.7.1.2 Acceleration

Average vibration recorded by the Moto G5 along the test route varied by over twice as much as data from the Moto E2. Spread of magnitudes varied by 0.403 m·s^{-2} (39.1%) about the average, compared to 0.188 m·s^{-2} (27.0%) for the Moto E2. This may have been due to the difference in the built-in sensors or the transmissibility of the smartphone body materials.

Vibration along the test route followed the same trend for both smartphones, however was recorded at higher magnitudes using the Moto G5 (Figure 4.15). Magnitudes recorded by the Moto G5 were an average of 46.1% (0.329 m·s^{-2}) greater across all 10 m intervals. This suggests that there may have been some additional gain within either the Moto G5 or the interface between it and the car dashboard as the Moto E2 was earlier (Section 4.3.1.3) found to give an almost ideal response up to 40 Hz at the magnitudes experienced. The majority of vehicle vibration is concentrated below 20 Hz which suggests a more accurate IMU within the Moto G5 would not account for the differences.

No discernable difference was identified between the magnitude difference along the second half of the route compared to the first, despite the average speeds recorded by the smartphones matching more closely. Average increase for the first 1.75 km was slightly lower at 0.318 m·s^{-2} (44.7%) as opposed to 0.340 m·s^{-2} (47.4%). This shows that

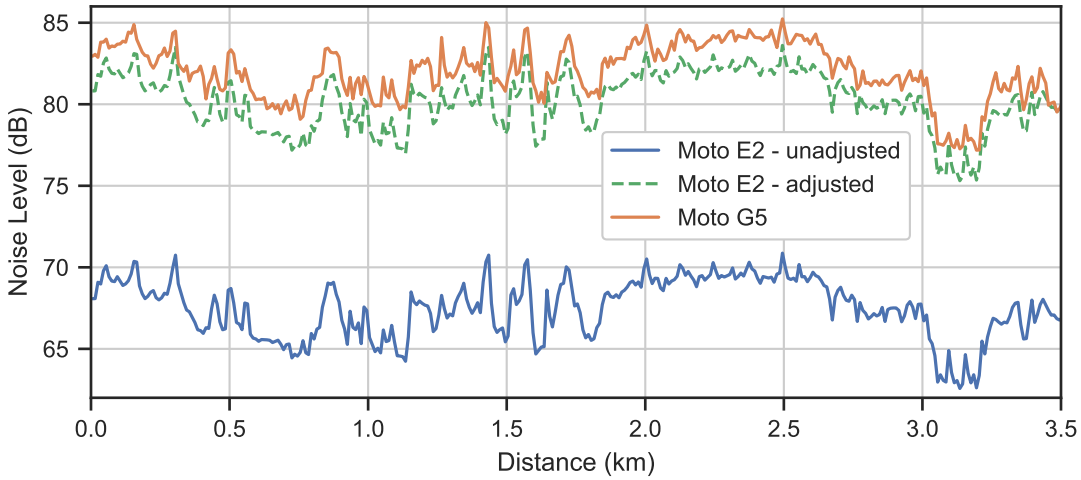

Figure 4.16: Comparison of 10 m resampled r.m.s. noise levels recorded along the test route using different smartphones, plotted against distance.

the slight variations in speed had no effect on the resultant vibration.

4.7.1.3 Noise

Noise levels for each 10 m interval along the test route varied from the average by greater magnitudes for the Moto G5 (IQR: -1.9 – +1.8 dB) compared to the Moto E2 (IQR: -1.4 – +1.5 dB). Although the variation was greater than the Moto E2 in terms of magnitude, the Moto G5 recorded higher overall levels resulting in similar values for percentage variation (±2.3% compared to ±2.1%).

Trends in noise level along the route showed high levels of similarity for both smartphones (Figure 4.16). Magnitudes recorded by the Moto G5 were 14.5±1.8 dB greater, on average, than the raw value from the Moto E2. Adjusting the Moto E2 values by the 12.7 dB offset identified during calibration (Section 4.4.2) reduced the average difference between the smartphones to 1.8 dB. This suggests that the Moto G5 needed to be reduced by 1.8 dB to match the adjusted levels of the Moto E2 — shown earlier to be equal to the levels registered by a professional sound meter — which could be assumed to be the true noise level.

Average difference between the smartphone noise levels differed by only 0.4 dB when comparing the two halves of the test route. This further suggests that the differences between the similarity in smartphone speeds of the two halves, as mentioned above (Section 4.7.1.1), had minimal effect on the sensor results.

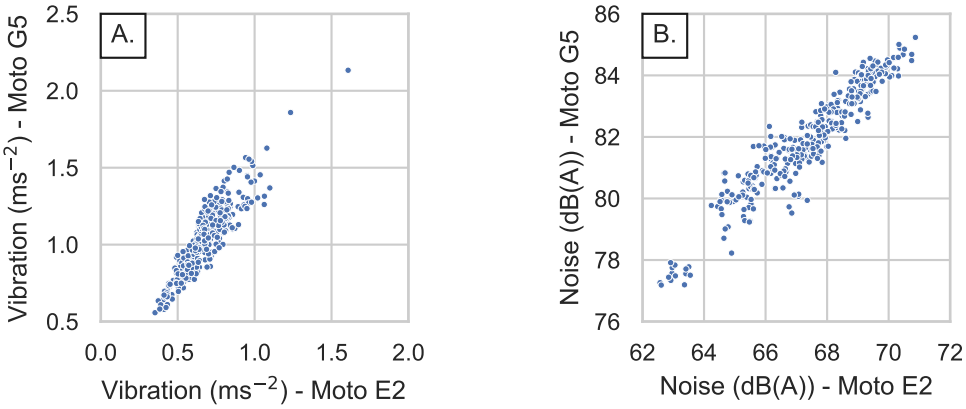

Figure 4.17: Distributions of vertical acceleration (A) and noise level (B) comparing data recorded by the Moto G5 with the Moto E2.

4.7.2 Inter-smartphone Calibration Uncertainties

Data recorded by the app was affected by the smartphone used, as was evident when comparing only two devices. The Moto G5 registered greater values in both accelerations and noise levels than the calibrated Moto E2. All devices on which the app is installed would require calibration to ensure all recorded data was both comparable and accurate. Comparability is required to enable the ranking of roads recorded by different devices, while accuracy would enable the effect on neonatal infants to be quantified.

Comparing two smartphones could provide calibration albeit at a lesser degree than when performed in a lab. Although offsets between device values could be calculated, the differences were not constant. Values of both noise and vibration were found to vary slightly between journeys for each device. Ideally, these variations would have averaged out and data from the two devices would be perfectly correlated. Instead, the results from the comparison of Moto G5 with Moto E2 gave reduced correlations, as shown in Figure 4.17, for both vibration (R^2 = 0.77) and noise (R^2 = 0.89). These imperfect correlations would introduce uncertainty in the calibration results.

Uncertainty errors were introduced in the smartphone-smartphone calibrations due to the lack of control. All variables which may affect calibration could be controlled inside a lab environment, however performing calibration in situ resulted in uncertainty due to the variations in input.

One of the problems with calibrating in situ was the ability of cars to move freely across roads. Without rigid location relative to the lanes of a road, tyres may travel over different areas of tarmac and produce different levels of noise and vibration. This means that two journeys along the same lane of the same road may not have travelled over

the same physical space. Therefore, it would be theoretically impossible to use the data from these journeys.

A combination of the time-based sampling of data and variations in vehicle speed would result in misalignment of journeys, even if the exact pieces of road were travelled over. In an ideal situation, sensor values would be registered at specific points along the route to ensure the vibration and noise inputs are identical. This could only be achieved if all recordings are started at the exact same location and the vehicle was driven at the exact same speed. Human error would render this impossible due to the need for incredibly high precision in both the starting of recordings and in driving, along with the need for control over other traffic on the roads to ensure all external factors were identical. Although rounding by distance, as outlined in Section 3.3.1, provided a reasonable method of comparing datasets along the same road, it cannot account for all variables. Rounding by distance was achieved by effectively summarising chunks of road, thereby reducing the quality of information available from any comparisons and introducing further uncertainties with regards to calibration.

Inter-smartphone calibration required the uncalibrated smartphone to record a journey along the same route as a calibrated smartphone. Without data for the same roads it is not possible to compare both, regardless of any uncertainties. Crowdsourcing would therefore be limited to devices in the surrounding area to the roads covered by the calibrated Moto E2. The area of crowdsourcing could increase as more smartphones get calibrated against the Moto E2, as there is a higher chance of similarity in road coverage with a greater number of calibrated devices. Although introducing further stages on the calibration chain would increase the crowdsourcing area, it would also greatly increase the uncertainty in values.

4.8 Conclusion

The purpose of this chapter was to determine how well the developed app fulfils the requirements for recording the in-vehicle environment during road journeys, along with assessing the feasibility of crowdsourcing this information.

The IMU accelerometer inside the Motorola Moto E2 smartphone was found to give an accurate response to sinusoidal vibration up to a frequency of 40 Hz. This suggests that the majority of accelerations sensed by the app reflect the true values closely.

Time intervals recorded by the app do fluctuate around the desired value, however these appeared not to affect results when assumed to be constant. It may be that the

IMU itself has a constant sampling frequency, with delays (and quick influxes of samples to make up for the delays) occurring in the reporting of values back to the app.

Noise levels recorded from the smartphone microphone using the built-in getMaxAmplitude method correlated perfectly with those of a precision sound meter when subjected to pink noise. These noise levels could then be adjusted using a constant offset in decibels to attain the true values.

Reductions in vibration magnitudes were clearly visible along the test route after the road was resurfaced. This shows that the changing state of roads was capable of being identified within data recorded by the app. Optimal routes can therefore be updated if and when a road is found to have degraded or been repaired.

Listening habits of users were shown to impact recorded noise levels as predicted. App-recorded noise was found to be at greater magnitudes while a podcast was played through speakers compared to the entertainment system being switched off. This suggests a constant auditory status is required to enable all recorded journeys to be comparable. Ideally, there would be no additional noise to provide a baseline to judge true noise levels of a road.

The addition of supplementary audio was found to keep a similar trend along the route, albeit at increased volumes. This suggests that all equipment within the ambulance during transfers should be optimised to reduce noise levels, along with reducing the input from the road, to provide an ideal environment.

Different smartphones result in varying levels of both vibration and noise in similar conditions. The Motorola Moto G5 gave greater values in both vibration and noise when compared to the vibration and both the raw and calibration-adjusted noise levels recorded by the Moto E2. This will mainly be due to the wide range of components available and built into the different devices. The physical composition of the smartphone body was probably a factor in the increase in vibration registered by the Moto G5 due to the rigid material not enabling a secure attachment to the dashboard. Every phone which is used for the recording of data using the app would need to be calibrated to ensure sensor values were both comparable and reflect the true conditions.

While calibration could be performed by comparing datasets from two smartphones which travelled along the same road, this leads to large amounts of inaccuracies due to the uncontrolled inputs and imperfect alignment. All calibrations should be performed within a laboratory environment as this enables all factors affecting data to be controlled. The need to calibrate all devices in a lab would severely reduce the effectiveness of crowdsourcing as volunteers would be limited in their participation by their

hardware.

Crowdsourcing the comfort of roads using the developed app does not appear feasible due to the requirement of reducing external factors. Although the app itself works as intended and the sensors built into smartphones are capable of recording accurate information, crowdsourcing would introduce too many unknowns which would render any analysis of the data impossible. Instead, a different approach was required in which all factors which may affect recordings are both known and as close to identical as possible.

CHAPTER 5

Neonatal Ambulance Data Collection

5.1 Introduction

Levels of noise and vibration recorded during data collection need to both accurately reflect the in-vehicle environment and be comparable between journeys to remove any biases. Testing found that both the app and the sensors within smartphones are capable of recording vibration and noise which correlate strongly with values recorded by professional equipment. However, collecting data via public crowdsourcing is not viable because the amount of additional factors (method and location of positioning the smartphone within the vehicle, suspension characteristics, amongst others) which can affect the data, along with the need to individually calibrate each unique smartphone, has shown that a more managed collection process is required. This chapter will outline the method used to record comparable data for the analysis of road comfort.

Mass data collection within a controlled environment was made possible through the participation of the local neonatal transport team, CenTre (CenTre Neonatal Transport) [144]. CenTre perform over 1,600 transfers per year by road ambulance [21], predominantly within the area defined by the East Midlands Neonatal Operational Delivery Network (EMNODN), meaning that a large number of journeys can be recorded. The information gathered during these journeys would be directly indicative of that experienced by a transferred baby due to being recorded within the same environment and the transport incubator being rigidly attached to the ambulance chassis. Recording the journeys conducted by CenTre would also provide a highly relevant set of data as they would focus on the roads of interest to the transfer service. Additionally, continued participation is far more likely than with crowdsourcing as the staff have a vested

interest in the collection. The standard layout of the vehicles used by CenTre would also provide an ideal opportunity to develop a protocol for collection ensuring comparability of data.

Before collection could be conducted, however, an Android device, with which to record the journeys using the app, needed to be provided to the transport team. Using a single smartphone model for all recordings would provide the highest level of consistency and comparability. Therefore, a device needed to be selected, and subsequently validated, for use.

Controlling the recording environment enabled both the method used and the data obtained to be focused. Steps were therefore taken to minimise variability between recordings by determining an available location for the device during journeys, along with a repeatable means of positioning. Furthermore, discussions were conducted with the transport staff which led to additional data being included in recordings to add to analysis.

5.2 Smartphone Choice

Controlling the recording enabled a smartphone model which fulfilled the project requirements to be selected, rather than using whichever device a crowdsourcing volunteer owned. These requirements related to the sampling frequency and jitter of the device Inertial Measurement Unit (IMU), the quality of the microphone, the size of the battery and the amount of storage.

An ideal device for data collection would possess a sampling frequency sufficient for assessing the harm of accelerations. As the effect of vibration on babies is currently unknown, the International Organisation for Standardisation (ISO) standard for perceived adult comfort of vibration [49] could be used to give an indication of the potential risk. Determined effects of vibration in the standard are derived from accelerations at frequencies up to 80 Hz, which therefore required an IMU sampling frequency of at least 160 Hz.

The Moto E2 (Motorola, Chicago, IL, USA) mentioned in previous chapters was found to have a rather high level of jitter in the IMU sampling (Section 4.3.2). Although the data appeared to be unaffected, a more consistent sampling rate would install greater confidence in recorded values.

The microphone built into the selected smartphone must be reasonably accurate at recording noise levels to enable comparisons with previous studies which have linked

aspects of sound disturbance to physiological effects.

Finally, the chosen smartphone needs to be capable of recording as many journeys as possible. This requires both a large battery, to provide life to record at least a return journey on a single charge, and sufficient internal non-volatile storage, to ensure that data is not lost due to a lack of space.

Both of the smartphone models used in Chapter 4 were limited to a maximum sample rate of 100 Hz and were therefore not suitable for data collection as they did not meet the minimum requirement of 160 Hz sampling. Instead, a different model smartphone was required.

5.2.1 Choosing a device

Smartphones are not marketed as research devices and as such the manufacturers provide little to no information regarding either the built-in IMU and microphone. This meant that only devices which belonged to people in the local area could be assessed for compliance with the requirements, as the only method of obtaining the required information was through the Android API (Application Programming Interface) and physical testing.

5.2.1.1 Sample rate

During development, it was observed that some Android devices exceeded the minimum sampling frequency of 160 Hz. A G4 smartphone (LG, Seoul, South Korea) was used during development alongside the Moto E2 to test out new features, or any code changes, and ensure operation was as intended before releasing for data collection along the test route. When examining the data collected by the G4, it was subsequently identified as having a maximum available sampling rate of 200 Hz.

Despite having a desirable IMU sample rate, the G4 smartphone was ultimately not suitable for use in data collection due to the shape of the construction. The unique curved design would result in problems regarding the attachment of the device for accurate vibration transmission. The fully-rounded back to the smartphone provided only a single contact point in the centre of the phone, while placing the device with the screen facing the surface would only result in small contact points in each corner. Although this would be sufficient for taping the smartphone in place, as in the shaker testing (Section 4.3.1), this method would not be practical for long-term collection.

An alternative smartphone model was required which also met the minimum IMU

sample rate requirement of 160 Hz. Ideally, this model would be completely flat, to enable easy fixation, and available at a reasonable price to facilitate future expansion of data collection.

Performing checks of sample frequencies on the smartphones belonging to both friends and colleagues, it was observed that maximum frequencies of 200 Hz or greater tended to be limited to higher end devices, although it is acknowledged that this was determined from a relatively small sample of models. One exception emerged in the form of the Redmi 4 (Xiaomi, Beijing, China; released May 2017) which fulfilled the minimum requirement but was priced more reasonably below £200 (in May 2018).

5.2.1.2 Microphone quality

There was no available data regarding the microphones inside any Android device examined. Due to requiring physical testing, it was also not possible to gather data to inform a decision. Instead, it was assumed that the microphones inside the majority of the latest smartphones would be high quality, due both to being core to the function of the device and being shown to be accurate (if offset) using the lower end Moto E2 (Section 4.4.2).

5.2.1.3 Sampling jitter

Similar to the microphone, it was not possible to ascertain information regarding the expected levels of jitter for different smartphones. Examining the specifications for the Moto E2, however, led to the hypothesis that the jitter experienced was potentially due to the system being unable to cope with the demands of recording. Therefore, prospective devices required a more powerful CPU (Central Processing Unit) and greater amount of RAM (Random Access Memory) than the Moto E2.

5.2.1.4 Storage and battery capacity

Unlike the above components, both storage size and battery life are reported for all smartphones. This is to enable consumers to make informed decisions reflecting the amount they interact with their device, and therefore could inform the choice of smartphone for data collection as well.

Table 5.1: Maximum sample frequencies of the Xiaomi Redmi 5 Plus smartphone.

Data Type	Sample Frequency (Hz)
Acceleration	200
Noise Level	50
Location	1

5.2.2 Final device choice

After identifying the required sampling frequency, along with the low cost and sufficient general specifications, the Redmi 4 was initially chosen as the device for data collection. Unfortunately, there was insufficient supply of the specific model due to being from a previous year. Instead, due to observing that the features within smartphones were never removed, and sometimes improved, with subsequent generations, the Redmi 5 Plus (Xiaomi, Beijing, China; released February 2018) was ultimately selected.

The Redmi 5 Plus was the next generation of the Redmi 4 and was therefore highly likely to sample accelerations at a minimum of 200 Hz. An improved CPU was also present on the Redmi 5 Plus compared to the Redmi 4, and provided twice the amount of processing cores, at higher frequencies, than the Moto E2 which could reduce jitter. Similarly, a minimum of 3 GB of RAM was available inside the Redmi smartphones, compared to only 1 GB within the Moto E2. Regarding storage, a minimum of 32 GB was built into the Redmi 5 Plus that, along with the 4000 mAh battery, would likely have resulted in largely unimpeded recordings.

Four Redmi 5 Plus smartphones were purchased for use in data collection — enough to enable each of the local neonatal transport group's 4 ambulances to be recorded simultaneously — based on the above information. The sensor sample rates used by the app on these smartphones are shown in Table 5.1.

The back of a Redmi 5 Plus smartphone is mostly flat apart from a circular extrusion for the camera lens. A silicone case is provided inside the box of each smartphone which is roughly the same thickness as the camera extrusion and results in a large, flat contact area for attachment.

Each of the 4 smartphones in turn was designated with an identifier from 1 to 4. These numbers were then printed onto the backs of the devices while both the identifier and the app device ID were noted. This would enable any similarities or differences to be observed between the presumed identical smartphones. Knowing which smartphone

corresponded to each app ID also enabled any problems in the field to be diagnosed and treated easier as the device in question could be recalled and examined without interrupting collection from the remaining devices.

5.3 Assessing the Redmi 5 Plus smartphones

Before any data could be collected, it was important for the sensors inside the new smartphones to be calibrated to ensure data collected reflect the true in-ambulance environment. This followed the same procedures as set out in Sections 4.3 and 4.4 for the accelerometer and microphone respectively.

Neonatal ambulances tend to make a large number of journeys, of varying duration, between hospitals. The life expectancy of the battery while recording with the app was checked to ensure there was minimal risk of the phone dying during a journey. Storage use by the app was also checked to ensure there was sufficient room for recordings on the occasions where wi-fi, and thus automatic deletion of files, was unavailable.

5.3.1 Accelerometer

5.3.1.1 Sampling Jitter

The Motorola Moto E2 was subject to considerable jitter between accelerometer samples, as shown in Figure 4.5. Although it did not appear to have any negative effect on the recorded data, this was still of concern. It was possible that the jitter was due to the hardware within the Moto E2 struggling to cope with the demands of multiple sensors at once. The 5 Plus was therefore subject to the same 90-minute stationary recording in an effort to quantify the levels of jitter.

The Xiaomi smartphones were found to vary minimally with respect to sample interval and produced only a single peak value (Figure 5.1). While the sample frequency did fluctuate, the standard deviation for the full 90 minutes was 0.42 ms (8.4% of the average interval of 5.00 ms). This suggested that the hardware within the 5 Plus was capable of recording data with acceptable sampling variance using the app.

5.3.1.2 Sinusoidal Frequency Analysis

Accelerations recorded vary between smartphone models, as shown in Section 4.7.1.2. Therefore each model requires calibration in order to accurately analyse the data collected. Calibration of the 5 Plus followed the same basic procedure as outlined in

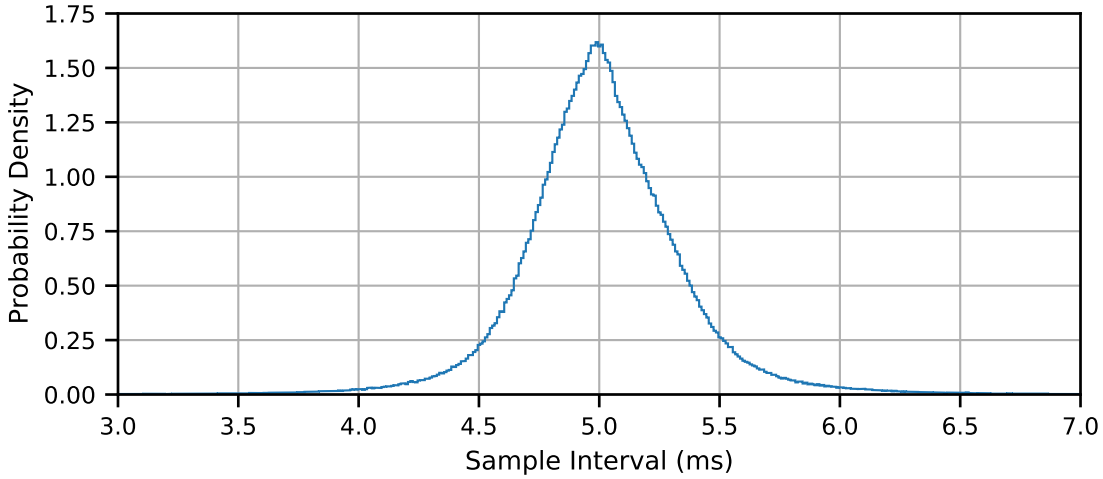

Figure 5.1: Distribution of accelerometer sample intervals over a 90-minute time period (bin size = 1×10^{-5} s).

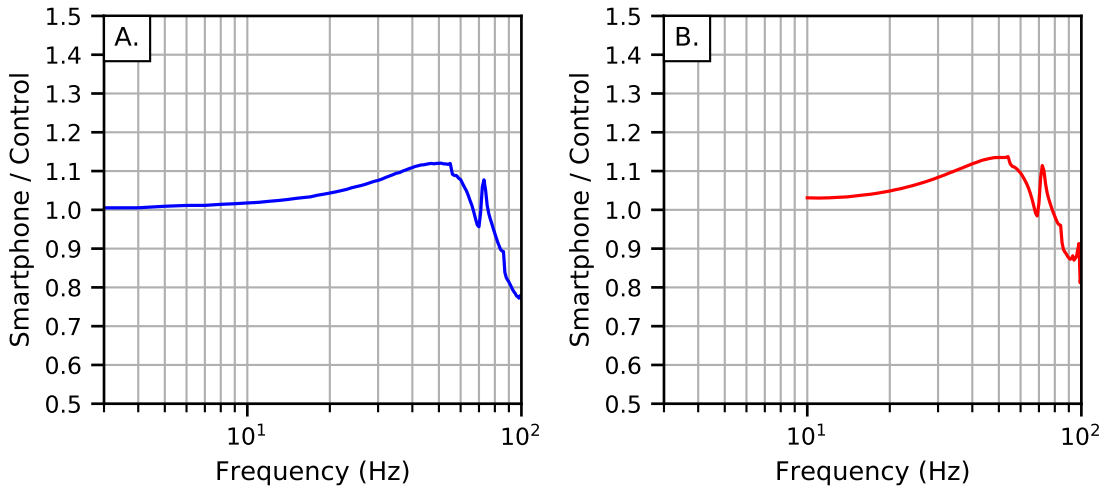

Figure 5.2: Sinusoidal frequency response of the Xiaomi Redmi 5 Plus, taped to a shaker, compared to the control accelerometer (PCB Piezotronics 352C65) at constant magnitudes of 1.00 m·s^{-2} (A) and 9.81 m·s^{-2} (B).

Section 4.3.1: controlling a shaker platform to output a constant r.m.s. magnitude (either 1.00 or 9.81 m·s^{-2}) at frequencies from 3 to 100 Hz and analysing the accelerations recorded by the smartphone attached.

Initial calibrations were performed on one of the Xiaomi smartphones taped directly onto the perspex platform. The r.m.s. response at a magnitude of 1.00 m·s^{-2} was almost identical to the response at 9.81 m·s^{-2} (Figure 5.2). Unlike when testing the Moto E2, the smartphone response increased to a >10% gain before dipping. This increase, along with the fluctuating response between 65 and 75 Hz, suggested an element of resonance within the setup. This may have been because of the interface between the smartphone and the shaker platform as the smartphone possessed a metal body compared to the rubber of the Moto E2.

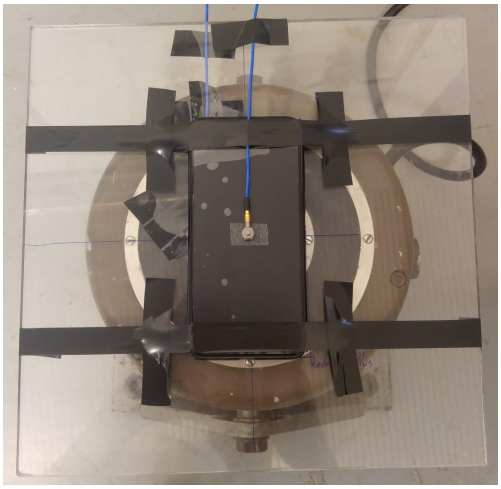

Figure 5.3: Photo of shaker setup, with a PCB Piezotronics 352C65 attached to the geometric centre of the smartphone.

A second single-axis accelerometer (352C65, PCB Piezotronics, Depew, NY, USA) was attached to the geometric centre of the smartphone screen using double-sided tape (Figure 5.3) to act as a reference. This reference accelerometer would enable assessment of the accelerations that the smartphone actually experienced, and therefore improve comparison with the expected input.

Double-sided tape was used to attach the reference sensor in lieu of a more rigid method to minimise damage to the screen. The screen of a smartphone is the main method of interfacing with the device and therefore any damage would be highly detrimental. Double-sided tape did not disrupt the operation of the smartphone or the app. The accelerometer did not appear to detach from the smartphone at any point during testing.

Compared against the screen-mounted reference accelerometer, the r.m.s. response of the smartphone matched to within 5% of the reference at both input magnitudes up until 54 and 55 Hz respectively (Figure 5.4). As the response vs the reference was flat, this further suggested there was some error occurring between the control accelerometer under the platform and the accelerations at the top. Although the response of the smartphone dropped off after 55 Hz, this may have been due to the imperfect interface between the screen and the reference accelerometer.

5.3.1.3 Improving smartphone-shaker interface

Although the smartphone accelerometer has been shown to match closely to a reference attached to the screen, it is important that the recorded values from the smartphone closely match the excitation of the surface to which it is attached. One cause

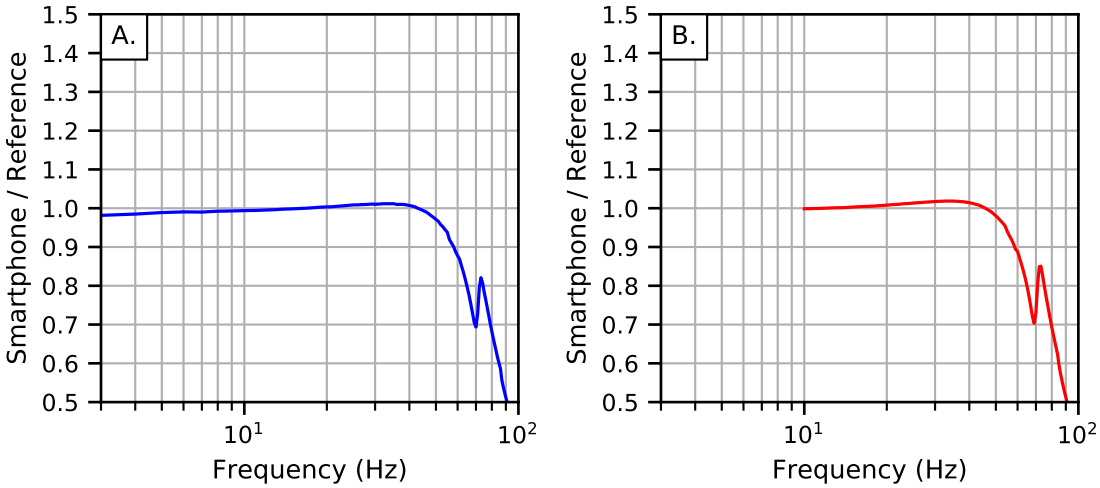

Figure 5.4: Sinusoidal frequency response of the Xiaomi Redmi 5 Plus, taped to a shaker, compared to the reference accelerometer (PCB Piezotronics 352C65) at constant input magnitudes of 1.00 $m \cdot s^{-2}$ (A) and 9.81 $m \cdot s^{-2}$ (B).

of problems between the control accelerometer values and the resultant smartphone vibrations may have been the poor interface between the 5 Plus and the shaker platform. Each Xiaomi smartphone came with a soft silicone case to protect from scratches. Protecting the smartphones from damage was imperative to ensure data collection was not interrupted, therefore use of the case may have been advantageous. This case also had the added benefit of providing a flat contact area on the back of the device where, without the case, the camera lens protruded slightly. The response of the smartphone inside the case was assessed to see if recorded data would be affected.

The addition of the silicone case appeared to improve the interface with the shaker platform, as shown in the differences between the responses with the case and without (Figure 5.5). Maximum gain between the smartphone and control accelerometer was reduced from 13.8% (at 45 Hz) to 10.7% (at 45 Hz). The overall response was significantly smoothed for the silicone case tests, with the sharp decrease above 70 Hz showing high levels of similarity with the approach towards the Nyquist frequency of the Moto E2 (Figure 4.4). Due to the combined benefits of improving acceleration transmission and also protecting the smartphones from damage, the silicone cases were used during data collection.

No reduction in response fluctuation was shown between the frequencies of 65 and 75 Hz. These fluctuations were not evident in the Moto E2 calibrations. It was not clear what caused the abnormal response but they were present in every test using the 5 Plus smartphones.

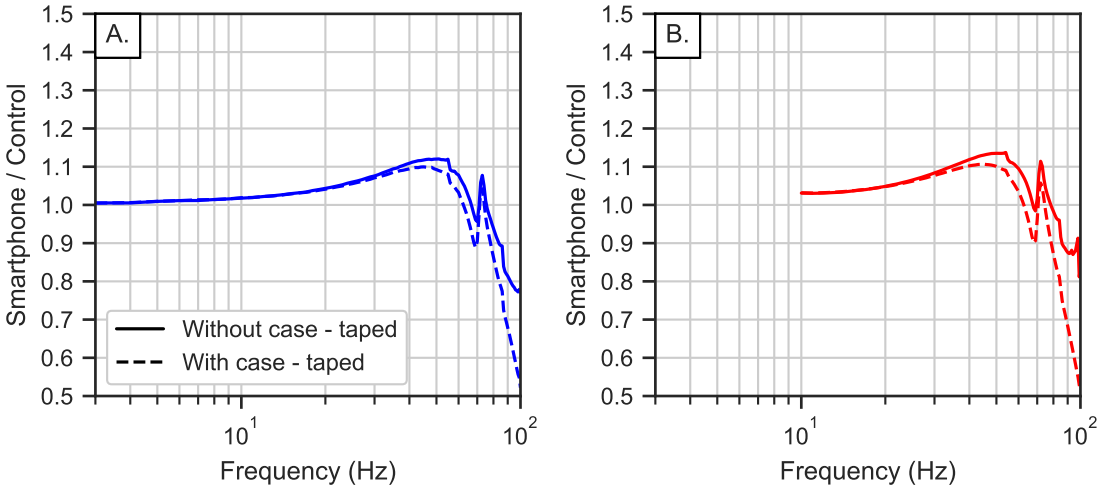

Figure 5.5: Sinusoidal frequency response of the Xiaomi Redmi 5 Plus taped to a shaker, with and without the silicon case, compared to the control accelerometer at constant input magnitudes of 1.00 $m \cdot s^{-2}$ (A) and 9.81 $m \cdot s^{-2}$ (B).

5.3.2 Microphone Noise Level Calibration

Documentation on the microphone built into the Redmi 5 Plus smartphones was not available[1], therefore it was imperative that the noise response was assessed to accurately quantify discomfort. Section 4.4.2 confirmed that the output of the 'getMaxAmplitude' correlated with the true noise level in A-weighted dB (dB(A)), however a decibel offset value was required for them to match. Each new smartphone model needs to be individually calibrated due to registered noise levels varying between devices (Section 4.7.1.3).

Noise levels recorded by the Xiaomi smartphones were calibrated against a professional sound meter, following the same methodology described in Section 4.4 where 5 second bursts of pink noise are played through a ring of loudspeakers in 5 dB intervals. An additional burst of pink noise at 80 dB was added to the calibration runs after finding the adjusted Moto E2 noise levels regularly reflecting this value (Figure 4.16).

Noise levels recorded by all four smartphones were found to match closely to the sound meter with minimal need for adjustment, unlike the Moto E2. Average smartphone r.m.s. noise levels were within 0.4 dB of the averages from the sound meter at each of the magnitudes tested. The magnitude of this difference is minimal and could be a result of slight misalignment between the microphones, therefore the decision was

[1]Xiaomi Redmi 5 Plus smartphones were supported up to Android 8.1 (API 27) at the time of writing. This meant that the microphone specifications could not be accessed through the 'MicrophoneInfo' class, which was introduced with Android 9 (API 28).

made not to apply any calibration offsets.

Average decibel offset increased slightly for the final burst of pink noise at 80 dB(A) compared to the offsets at reduced input levels. Although this may imply that the microphone working limit was being approached, the absolute difference was still minimal.

Values for the smartphones at each input level varied by less than 1 dB between devices. This minimal difference suggests that the noise recorded by the microphone in each smartphone is comparable as well as approximately equal to the true environment.

Considering the maximum output value of the 'getMaxAmplitude' method, the 5 Plus smartphones theoretically have a maximum noise level of 90.3 dB(A) which could be accurately registered. Any noise above 90.3 dB(A) would therefore be reported equally and could result in some environments being underestimated. This upper limit was smaller than that of the Moto E2 due to the lack of offset between the Redmi smartphones and the true noise levels.

5.3.3 Storage and Battery Usage

Use of the app requires both sufficient built-in non-volatile memory to store data along with a suitable battery which outlasts the longest possible journey duration. These were assessed by recording with the app for 5 hours while taking note of battery percentages and checking the size of the recorded file. The recording was conducted inside with the smartphone located next to the window to enable access to satellites for location data. All sensors were accessed and logged as they would be during normal operation. The "auto-stop" feature was disabled to ensure continuous recording despite the device being stationary.

The battery readout on the 5 Plus reduced by 16% of the maximum capacity over the course of the 5 hours, from 93 to 77%. This equates to an average consumption of 3.2% per hour, suggesting each smartphone would be capable of recording for up to 31 hours on a single charge, assuming a linear discharge during operation. The true recording life would vary depending on how often the smartphone was charged. Leaving the smartphone unplugged and on standby between journeys would gradually consume the battery (7% of the maximum capacity was consumed over an 18 hour period on standby, consuming at a rate of 0.39% per hour) and leave less for recording.

Recorded files were split every 30 minutes which equates to a logging rate of 20 MB per hour. The lowest specification 5 Plus smartphones have a total of 32 GB of built-in

flash memory. Only 22 GB (69%) of the stated 32 GB was available when examining the device, due to the operating system and stock applications. This still gave more than sufficient storage space for recordings, with capacity for 1,100 hours of data. It is unlikely that all of this space would be required due to the automatic deletion of successfully uploaded files.

5.3.4 Summary

The Xiaomi Redmi 5 Plus smartphones have been shown to be capable of accurately recording both vibration and noise. They also possess ample storage and should be capable of recording even the longest of journeys without the device battery going flat. Although the smartphones themselves are shown to work as required, the process of recording in the ambulances needs to be defined to ensure fully comparable data.

5.4 In-ambulance Collection

Data comparability was maximised by controlling as many aspects of the data collection as possible. The first stage of this was the selection of a single smartphone model for use in the ambulances, which was calibrated in the previous section. Other factors which would affect the collected data were the location of the device within the ambulance and when each recording should be started and stopped. This section will outline the processes of controlling these aspects.

Registered levels of vibration and noise would likely differ depending on the location of the smartphone within the ambulance. A single location was required for recording using the app which would remain constant between all journeys and would not interfere with patient care.

Smartphones were taped to the vehicle dashboard and to the shakers during development and validation. The use of tape did not facilitate accurate, repeatable positioning and also was relatively time-consuming to set up, although did result in good transmission of vibration. A method of locating each smartphone securely and in the correct place will therefore be investigated.

Several changes were made to the app before it was provided to the neonatal transport group. These were made both to simplify operation of the app and also to record additional data which were recommended by the clinical staff. Each change will be outlined in full along with the reasons for implementation.

Finally, the process of recording within the neonatal ambulances will be outlined in

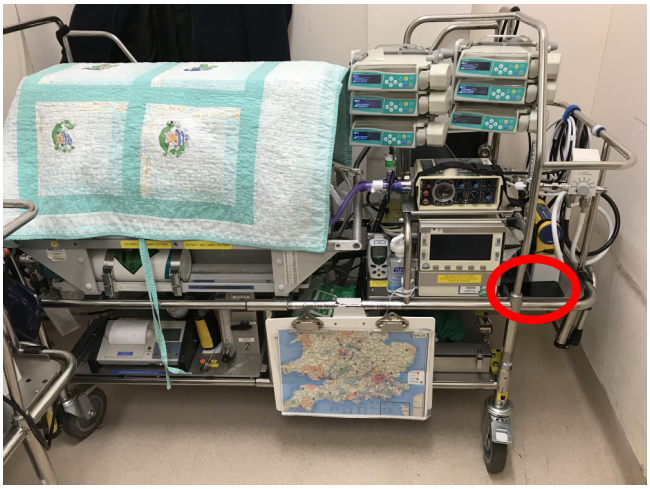

Figure 5.6: Photo of one of the neonatal transport trolleys, with a smartphone (circled) in the chosen location.

an attempt to maximise the amount of data collected and to ensure each journey was recorded in the same manner to achieve full comparability.

5.4.1 Positioning

There are two main aspects to positioning which are equally important for repeatable recordings: the location of the smartphone and the method of attachment. The location is important to ensure all recordings get exposed to the same environment. An ideal method of attachment will keep the smartphone secure while also transmitting accelerations. Discussions were conducted with senior members of CenTre during the decision process to ensure normal clinical practice was not obstructed.

As the reason for investigating route comfort is to reduce the risk of neonatal infant Intraventricular Haemorrhage (IVH), data collection should be as close to the baby as possible. Recording inside the incubator was not determined to be feasible as it would require a cleanable device as well as permission from parents. Instead, the rear left corner of the transport trolley (Figure 5.6), in terms of the ambulance axes, was chosen due to its close proximity. This location was flat, unobstructed by clinical equipment, and allowed the smartphone axes to be aligned with those of the ambulance: x-axis = sideways, y-axis = forwards/backwards, z-axis = vertical.

During testing, the smartphones were attached to both the vehicle dashboard and the shaker platform with insulation tape. While this gave a firm attachment and allowed full transmissibility of vibration, it did not promote repeatable positioning and had a limited life cycle. A new method was required to reliably locate the smartphone, minimising human error, while capable of being cleaned as per the transport team's

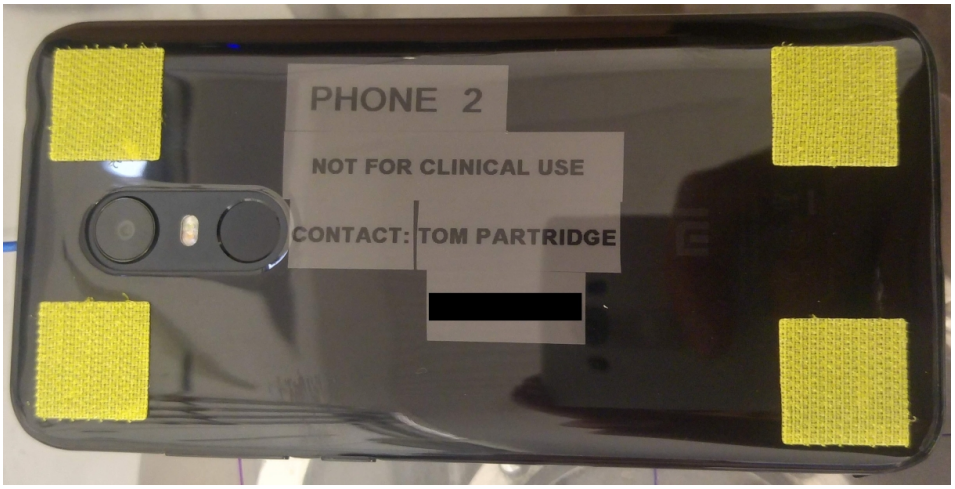

Figure 5.7: Photo of one of the smartphones with adhesive hook fasteners attached in each corner.

protocol.

5.4.1.1 Hook-and-loop fasteners

Hook-and-loop fasteners are a cheap method of attaching two objects, typically clothing, together. It has the advantages of being very cheap with a minimal reduction in adhesiveness over time. The effect on transmissibility was tested using adhesive hooked squares at each of the four corners of the smartphone case (Figure 5.7), with the corresponding looped squares on the shaker platform.

Subject to low magnitude vibration, the hook-and-loop fasteners gave the same transmissibility at 1 $m \cdot s^{-2}$ as directly taping the smartphone to the shaker up to 30 Hz (Figure 5.8A). Above 30 Hz, however, the amount of play between the hooks and loops contributed to excessive resonance. The resonant effects were even more prominent at magnitudes of 9.81 $m \cdot s^{-2}$ (Figure 5.8B) where the accelerations using hook-and-loop fasteners were greater than when taped at all low frequencies, with the sharp increase in r.m.s. occurring more than 10 Hz earlier than the 1 $m \cdot s^{-2}$ runs. This poor transmissibility, combined with the high likelihood of misalignment and tendency to collect dirt, discounted hook-and-loop fasteners from being used during data collection.

5.4.1.2 Magnets

Neodymium magnets are the strongest type of permanent magnet [145] and are used in a vast array of applications from simple fasteners to electric motors. The low-impact environment inside the ambulance, combined with the low mass of the smartphone

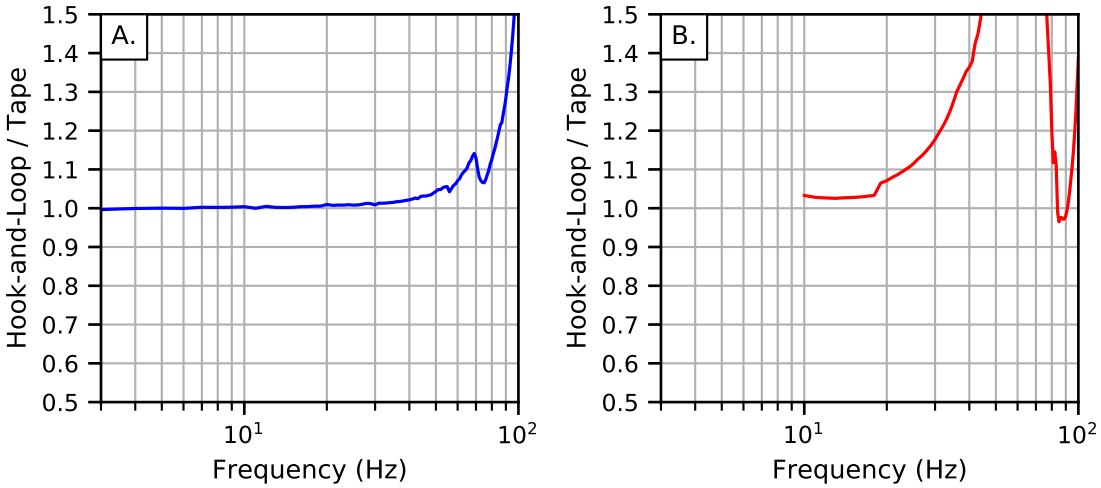

Figure 5.8: Ratio of r.m.s. magnitudes recorded by the Xiaomi Redmi 5 Plus smartphone attached to the shaker using hook-and-loop fasteners compared to using tape, at constant input magnitudes of 1.00 m·s^{-2} (A) and 9.81 m·s^{-2} (B).

(0.18 kg), meant that only small magnets would be required to affix to the trolley. Magnets were required on both the smartphone and trolley, due to its aluminium construction. Positioning was controlled by using magnets axially oriented through their thickness, with the trolley having a South contact face and the smartphones having a North contact face. To ensure secure mounting, magnets with a 0.58 kg pull were used (Adhesive 10mm dia x 1mm N42 Black Epoxy Magnets - 0.58kg Pull, Magnet Expert Ltd, Tuxford, Notts., UK).

A combination of 8 magnets were used to attach each smartphone to the trolley surface. The pull rating for each magnet was determined using a flat steel plate (>10 mm thick) free of dirt and using a steadily increasing perpendicular force [146]. Using two magnets to attract one another only results in a slight increase to the individual pull force [147] and so all calculations were based on the stated pull value. Therefore, forces of greater than 22.8 N (4 x 0.58 kg x 9.8065 m·s^{-2}) would be required to separate the smartphone from the trolley. This should be sufficient to keep the smartphone attached as it would require an acceleration of almost 13 g to provide enough force to overcome.

A laser-cut template was used to align the smartphone and trolley, with the magnets at the top of the smartphone spaced wider than those at the bottom to prevent misalignment. The template was dimensioned to match the shape of the 5 Plus (Figure 5.9) with a central slot aligning with the one in the silicone case for the camera and fingerprint sensor. Including the slot both helped with the directionality of the template, with the slot being towards the top of the device, and also when positioning the template on the back of the smartphone as both the outer edges and the inner edges of the template

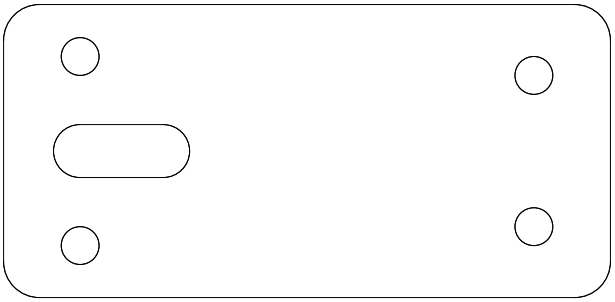

Figure 5.9: Laser-cut template used to position magnets (½ scale).

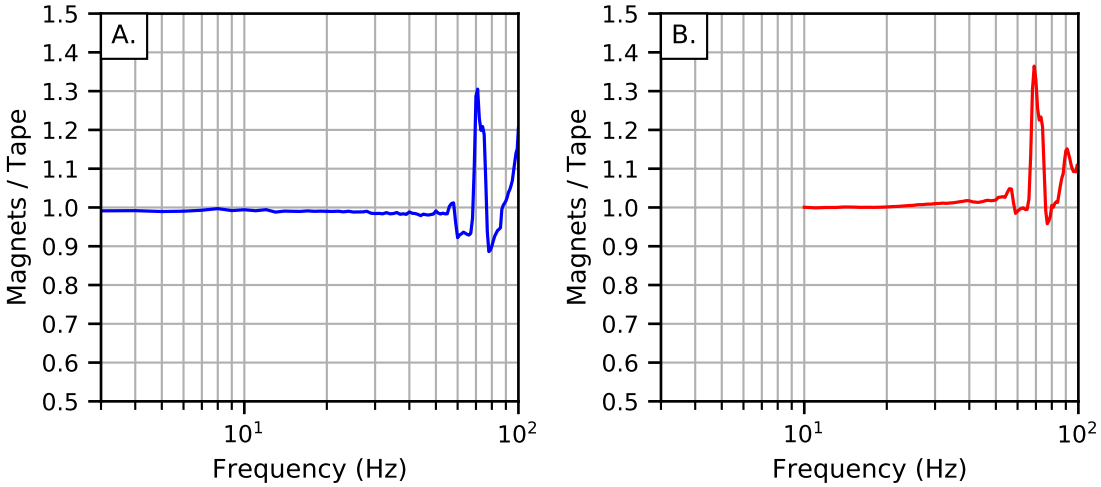

Figure 5.10: Ratio of r.m.s. magnitudes recorded by the Xiaomi Redmi 5 Plus smartphone attached to the shaker using magnets compared to using tape, at constant input magnitudes of 1.00 $m \cdot s^{-2}$ (A) and 9.81 $m \cdot s^{-2}$ (B).

could be aligned. Four 10 mm diameter holes for the magnets were positioned near the corners with each pair being placed equidistant from the centreline along the long edge of the device. The template was cut from cardboard and was therefore rigid enough for positioning while being flexible enough for the magnets to be pushed through the holes.

Compared to the use of electrical tape, affixing the smartphone using magnets resulted in recorded accelerations within 5% up to 59 and 66 Hz, at 1.00 and 9.81 $m \cdot s^{-2}$ respectively (Figure 5.10). At higher frequencies, especially between 68 and 80 Hz where resonance was evident within the shaker setup earlier (Section 5.3.1.2), the magnetic fastenings audibly separated during testing which subsequently resulted in the inputted excitations being amplified. This should not be a problem during data collection, as the vibration is composed of a range of frequencies with the most dominant being below 15 Hz.

Magnets were chosen for data collection because they gave a uniform transmission of acceleration in the frequency range at which the Xiaomi smartphones display an ideal

response. There was only one possible alignment due to the polarity of the magnets along with the use of the template, ensuring that accelerations from different journeys would always be comparable. Although the close proximity of the magnets to the smartphone corrupted the built-in magnetometer and removed the ability to transform the smartphone axes into world coordinates (North, East and down), this was not needed as it was aligned with the ambulance axes. Magnets also met the clinical requirement of being cleanable by having smooth non-porous surfaces.

5.4.2 App Modifications

Discussions were conducted with the clinical staff of the local neonatal transport group, CenTre, before data collection was undertaken. These conversations provided some extra insight into possible shortfalls of the use of smartphones as well as suggestions for additional information which could be recorded for each journey.

CenTre are one of 16 teams currently operating within the UK. Expanding collection to the whole of the UK would result in an almost ten-fold increase in the number of journeys which could be recorded [21], plus doubtlessly cover a greater selection of roads, compared to CenTre alone. All of the changes outlined below therefore consider the future expansion of data collection around the UK.

5.4.2.1 Ease of use

Many of the clinical staff who would be using each device owned Apple devices as their personal smartphones. Although the operation of both Apple and Android devices have become increasingly similar in recent years, there are still differences between them which could cause confusion. By removing the need to navigate to the app, operation of the devices could be confined to knowing how to turn the screen on and unlock it before entering the simple user interface of the app itself.

A launcher is the term for the home screen within Android, named as such due to its default use being the screen from which all other apps on the device can be selected. This home screen is the first thing a user will see once a device has been turned on and unlocked, with a dedicated system-wide button (or gesture, depending on user preference) to facilitate a return to it from within apps. Essentially, the launcher is started when the phone is first turned on and cannot be stopped by the user.

Setting the app to run as a launcher minimised the amount of screen interaction required and significantly reduced the chance of the app being lost to the user. The transport staff which would be controlling the recording of data using the app have a main

responsibility in caring for the patient being transferred. Therefore, by setting the app as a launcher, less time would be spent navigating the device and thus there would be minimal intrusion in the patient care.

Changing the default home screen to the app was only possible due to the use of dedicated smartphones for the purpose of recording the data within the ambulances. This would not be a viable option for crowdsourcing as that would have used the smartphones belonging to the volunteers. Those volunteers would doubtlessly use their smartphones and setting the app as the "launcher" would severely limit the functionality as no other apps could be opened intuitively.

The app was encoded to be able to be set as a launcher but it was not possible to actually set it on each device. Only the user can change the launcher of a device, which prevents an app from locking functionality of a device upon install when it may not be desired. Instead, the app was set up to check if it was set as the device launcher upon start-up. A prompt to set the app as the launcher was displayed if it was not already, providing a link directly to the required page in the device settings where the user could make the required change.

5.4.2.2 Additional information

The local neonatal transport team have four transport trolleys along with the four designated Fiat van-based ambulances. Recording equipment used for each journey would enable possible differences to be explored and recommendations made if a particular trolley or ambulance resulted in harsher levels of vibration or noise.

Including the clinical information attributed with each journey also enables possible trends within the recorded data to be explored as well as facilitating the future comparison of the recorded vibration and noise with the patient outcomes. All ambulance journeys would be recorded and therefore it was important to make a note of whether there was a patient on board or not as the driving style may be different, which may in turn affect the levels of vibration and noise. The reason for transfer was also recorded when a patient was on board, mainly as a clinical link to assess the scope of recording.

The final piece of information to be recorded about the journey was whether any emergency driving was used. Emergency driving occurs when there is a more urgent need for the ambulance to reach the destination. The ambulance drivers then have the authority to rush through traffic using blue lights and, in the worst cases, the siren to warn other motorists [148]. Emergency cases can occur either from the start of the journey or from midway to the destination if a complication arises with the patient. The

presence of emergency driving was recorded as a simple "yes/no" response at the end of the journey to ensure there was no interference with potential care, although this subsequently meant that the point at which emergency driving began is unknown. Examining the noise levels for an increase due to the siren may be a possible route to identifying periods of emergency driving, however the drivers for CenTre stated that siren use was minimal.

Each of the five pieces of information introduced above describes an aspect of the overall recording, and therefore will be encompassed in the term 'meta-data' for the remainder of the thesis.

Selection screens were added to the app to prompt the user to input the above information regarding the journey. Ambulance and trolley usage, along with the presence of a patient, were entered at the start of the recording as these were constants for the whole journey. The reason for transfer and occurrence of any emergency driving were entered at the end of the recording because they were susceptible to change as the journey progressed. In keeping with the rest of the app, the interface of these selections was kept simple with large, equally sized buttons so as not to subconsciously lead the user to select one option over another.

The ambulance codes designated to the neonatal transport group were pre-filled on the first screen, with the addition of being able to type in a different code (for example, if a replacement ambulance was used due to maintenance) (Figure 5.11A). The trolleys used for transfers were fixed, therefore the option of typing was replaced with a large button to return to the ambulance screen in case of an erroneous selection (Figure 5.11B). Each of the other options screens followed the same design principle, with either "Yes" or "No" options or the two reasons for transfer used by CenTre: "Repatriation or Capacity" (transferring patient closer to home when reduced care is required and transferring to reduce load on a ward, respectively) and "Uplift" (transferring to a hospital with a higher level of care than the birth hospital).

Values for each button on the selection screens were stored as arrays within the app structure rather than being hard-coded for each option. The default values were those which corresponded to the local neonatal transport group as they would be first to use the app. Ambulance and trolley identifiers corresponding to the other transport groups around the UK could be used instead by selecting the relevant group name within the app settings screen.

Where a device had previously been used to record a journey on the same calendar day, the user was provided with a "same as before" option when selecting the ambulance and trolley (Figure 5.11C). This was included as it was likely that each journey

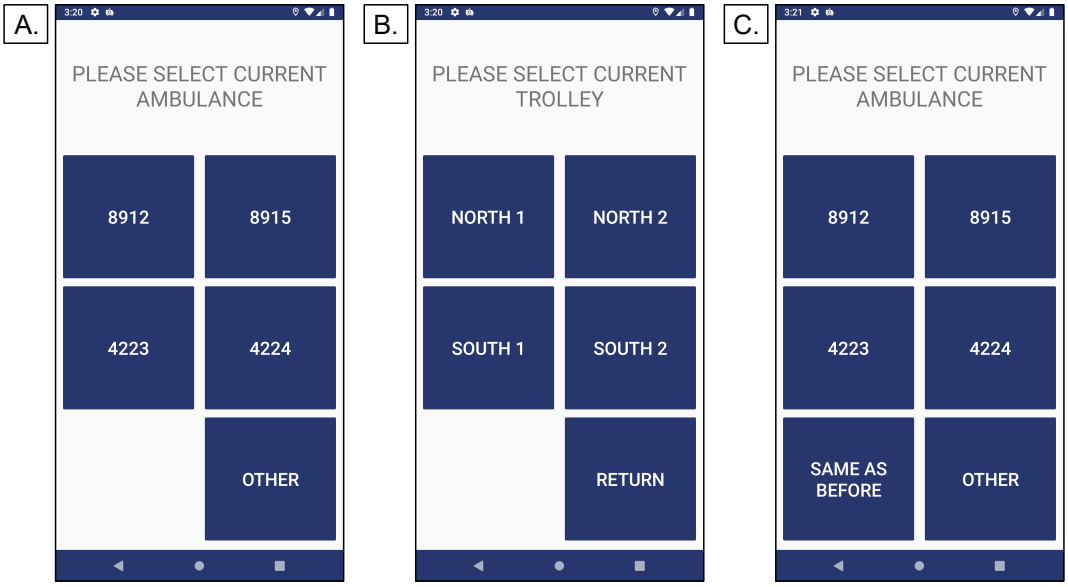

Figure 5.11: Emulated screenshots of the app showing the ambulance (A) and trolley (B) selection screens, along with the addition of a "same as before" button for same-day recordings (C).

```
Ambulance,8912,Trolley,NORTH 2,Patient,NO,Reason,N/A,
    Emergency,NO
```

Figure 5.12: Example string of ambulance options.

would be accompanied by a return trip. In these instances, neither the ambulance nor trolley would be changed. Including this option would further streamline the selection process, especially if a new ambulance code needed to be inputted.

The value selected for each of the five pieces of meta-data were stored in 'SharedPreferences' before being written to a file. SharedPreferences is an object within Android which enables simple reading and writing of key-value pairs to a file [149]. Storing within SharedPreferences allowed access to the information from all classes within the app and provided a resource in which to check the values to enter when the user selects the "same as before" option.

Meta-data was written to file as a single comma-separated line. This line alternated between the data identifier and the selected value (Figure 5.12) to enable the information to be human readable. A separate file was used to store the meta-data as it would need to be edited at both the start and end of the recording and the use of GZIP does not facilitate editing of rows without overwriting all other data.

The meta-data file was created at the start of the recording and then written over at the end to ensure that at least the journey constants (ambulance, trolley and patient presence) were attributed to a recording on the occasions that the smartphone ran out

of battery midway through a journey. Values for the reason for transfer and whether emergency driving was used were stored as "Unknown" until they are overwritten with the true data at the end of the recording. This also enabled instances where the recording stopped mid-journey to be easily identified and separated from those which were completed in full. After the journey had finished and the recording stopped, the meta-data file was inserted as the first row of the main data CSV file.

5.4.2.3 In-hospital recording

The final modification was made in response to the transport team's observations. In meetings before data collection began, the possibility of excessive vibration before reaching the ambulance was flagged. The transport trolleys were required to leave the hospital via the same doors as walk-in patients and the staff had observed babies visibly shaking when being pushed over the entrance matting designed for the removal of dirt from shoes. Although babies had been seen to shake, it was important to be able to quantify how much harm to which they may be exposed.

Accelerations and noise levels were therefore logged while the satellite receiver determined a location to enable the app to record data when inside buildings. This was simply a case of starting the IMU and microphone services directly after the "start recording" button was pressed, rather than the crowdsourcing setup of only recording the sensors once there was location data to which they could be attributed. The satellite timeout was also increased from 1 to 10 minutes to reflect the extra time which may be spent recording indoors before reaching the ambulance.

5.4.3 Method of collecting data

The four 5 Plus smartphones specified above, pre-loaded with the modified app, were provided to the local neonatal transport team with instructions to record all of their ambulance journeys. Two of the smartphones (IDs 1 & 2) were given to the staff based at Nottingham City Hospital (NCH), with smartphones 3 & 4 stationed with the staff at Leicester Royal Infirmary (LRI). To maximise data collection, journeys both with and without patients were to be recorded.

Interaction with the smartphones was limited to the start and end of the transfer process, allowing the user to focus on patient care. At the beginning of the transfer, the user started the recording and inputted the ambulance, trolley and patient options. The smartphone would then be attached to the transport trolley while still inside the hospital, enabling the in-building comfort levels to be recorded until the trolley was

loaded into the ambulance.

The smartphone was left untouched while the ambulance was travelling and recording continued until the transport trolley was unloaded and inside the destination hospital. This was to ensure that no manual touches of the smartphone affected any of the sensor measurements. Recording would subsequently be terminated via the emergency driving and, if transferring a patient, reason for transfer selection screens.

5.4.4 Data Retrieval

As outlined in Section 3.9, no user interaction was required to retrieve recorded files as they were automatically uploaded via wi-fi. Both of the transport group's base hospitals are affiliated with universities, therefore each of the smartphones was setup with a login to an international wi-fi network designed to provide network access at all educational institutions. This ensured that no further setup was required by the transport staff.

Organising of files once they had been downloaded could be made more intuitive as the device IDs were known. Instead of moving each file to a directory which had the device ID as the name, the directory was instead named after the physical smartphone identifier. Thus, recordings from each device were to be placed into directories labelled "Phone 1" through to "Phone 4". This was to enable the ambulance data to stand out from the other recordings and also be easily attributed back to a device without having to compare the device ID within the app.

5.5 Conclusion

This chapter described the steps taken towards the controlled mass data collection within neonatal ambulances, from choosing a recording device and assessing the onboard sensors, to specifying the protocol to be followed by the transport staff.

The Xiaomi Redmi 5 Plus smartphone was chosen as the recording device for the ambulances as it met the minimum IMU sampling frequency, as dictated by ISO 2631. This meant that the effect of vibration on the health and comfort of adults could be quantified, in lieu of a neonatal-specific index. Possessing an ideal sample rate while being priced below £200 would also facilitate easier uptake around the UK if data collection was expanded beyond the single transfer team.

Both the accelerometer and microphone registered similar levels compared to controlled inputs. The built-in accelerometer registered very little jitter between samples

and matched to within 5% of a reference accelerometer, up until 54 Hz, when subjected to sinusoidal excitation. Noise levels for the 4 smartphones varied slightly, but were all within 0.32 dB of the expected level when subjected to pink noise.

Battery life of each smartphone was more than sufficient to record the longest of journeys, with over 30 hours of recording available on a single charge. Similarly, on-board flash storage would be capable of storing over 1,000 hours worth of journey data. The large memory, along with the routine deletion of successfully uploaded files, means that the limit of storage will never be reached.

Neodymium magnets were used to affix the smartphones to the transport trolleys to ensure repeatable positioning between recordings. Levels of vibration when attached using the magnets were within 5% of those using the electrical tape, suggesting transmission was near-perfect. The use of magnets disrupted the on-board magnetometer of the smartphone, which meant that accelerations could not be converted from device to world coordinates as easily. However, the smartphones were aligned to match the axes of the ambulance, rendering conversion unnecessary.

Several modifications were made to the app due to the change in user base from crowdsourcing to the neonatal transport team. The app was set as the "home" screen on each of the smartphones — possible because the devices were dedicated for the app — to reduce the amount of interaction required with each device. Information about the transfer itself was recorded to enable future analysis into the possible effects of equipment and transport reasons on the recorded environment. Recording was enabled for up to 10 minutes without a satellite fix to enable analysis on the comfort when transitioning from inside the hospital to the ambulance. Staff were also instructed to continue recording until inside the destination hospital, to enable assessment of unloading the trolley from the ambulance.

After validating the sensor outputs and preparing the recording protocol, the next step towards data collection was to provide the transport team with the smartphones. These were split equally between their bases in Nottingham and Leicester. The devices were left with the team for a year, with the instructions to record all journeys made using the neonatal ambulances.

CHAPTER 6

Data Preparation

6.1 Introduction

Vibration, noise and location data was collected using the developed app during ambulance journeys made by CenTre (CenTre Neonatal Transport). This collection occurred over a 12 month period from 24th October 2018 to 14th October 2019, and was controlled in terms of the devices used and the protocol followed to ensure data was highly comparable.

A total of 1,890 recordings were made within the neonatal transport ambulances over the course of a year. This is 18% more than the approximately 1,600 transfers reported by CenTre [21], however this was expected as the staff were instructed to also record journeys without patients. Recordings were split fairly evenly between the teams at the two hospitals with only 30 more files (1.6% of the total) recorded by the smartphones based at Nottingham City Hospital (NCH) compared to those based at Leicester Royal Infirmary (LRI). There appeared to be a slight bias by the staff at NCH towards choosing Phone 2, however as the 4 devices used were shown to be highly comparable in Section 5.3, the initial collection could be determined as successful.

Storing data in Comma-Separated Value (CSV) files was adequate for logging purposes within the app, but did not facilitate easy analysis. Therefore, this chapter will outline the process of preparing the data collected for analysis of comfort by means of insertion into a database.

Problems with recordings were identified both during the collection period and when performing analyses. These problems will be explored and addressed with the aim of ensuring all resultant datasets followed the protocol set out in the previous chapter.

6.2 Using a database

Databases are an efficient method of storing large volumes of data. This is essential after collecting data from 1,890 recordings of varying length that totalled 193 GB in filesize (average of 105 MB per recording) and almost 2 billion rows of data. Reading in the data from each individual journey (an average of around 1 million rows per file) would be slow and require knowledge of the precise contents of each file to ensure the correct information is obtained. Storing the data within a database, however, would enable quick access to all data, assuming an appropriate configuration.

InfluxDB was used as it is an open-source purpose-designed time series database specified for high write and query loads [150]. Time series databases are optimised for the storage and processing of time-stamped data (aggregating data over a specified period of time, for example) such as that recorded by the app.

Data is structured within InfluxDB through a combination of timestamps, fields and tags. A field is defined as a key-value pair which can record either information about the recording or data for analysis [151]. Tags, on the other hand, may only contain meta-data because they are indexed within the database. Indexing the tag values enables faster queries [152] as well as helping identify a record. Each record in InfluxDB is uniquely identified by the combination of timestamp and tag values.

6.2.1 Structuring the data

Recorded data were separated into the components for analysis and those that facilitate searching or filtering the data. Noise and acceleration values would predominantly be used for analysis as these were believed to have an effect on the patients. Conversely, the meta-data do not describe the in-ambulance environment and could only be used when stratifying data. Both location and time data, however, could fall into either category as they could both contribute to the journey assessments (vehicle speeds, travelling durations) and are vital in identifying any repeatability (vibrations on a specific road, changes over the course of weeks or months). Therefore, the meta-data were set as the tags while the sensor data were set as fields.

Recording data with 4 smartphones led to the possibility of more than one journey being conducted at the same time. Timestamps were logged to millisecond precision and, with the highest sample rate being the 200 Hz belonging to the IMU, there was a high chance of two recordings registering the same timestamp. This, combined with the chance that staff may enter ambulance or trolley options incorrectly which could

Table 6.1: Example of the line protocol for writing points to InfluxDB.

Measurement	Tags	Fields	Timestamp
raw,	phone=1,ambulance=8912	x=0.2609,y=0.3256	1540371180

lead to duplicates which would not be physically possible, required a further, unique identifier. Therefore, the device which recorded the journey was added as another tag because it was set before the smartphones were provided to the transport team and could not be modified by the user.

There was no method of extracting a specific journey from the database unless the device and meta-data were known. This meant that if the sensor values from a certain recording were of interest, the CSV would have needed to be accessed first to identify the meta-data. Similarly, and more importantly, there was no straightforward method of identifying a recorded file on the occasion that a possible problem was found with data within the database. Therefore, a final tag was included which held a string of the start time (already used to name each recorded CSV file). This tag was termed the 'date' to aid differentiating from the timestamp.

6.2.2 Writing to InfluxDB

The default method of inserting data into an InfluxDB database is via line protocol. Each point — the terminology used by InfluxDB for a single record — requires the data to be set out in a string of text in a specific manner, making use of commas and blank spaces to differentiate between tags and fields (Table 6.1). All points must include a measurement and at least one field value. Although they are not mandatory when inserting data to InfluxDB, each point discussed here will include both tags and a timestamp.

A measurement in InfluxDB terminology is the location in which data is stored — similar to a table in more traditional databases. It was decided to separate recorded data between two main measurements to improve performance of queries and processing. The names of each measurement was kept short, maintaining clear descriptions of the contained data, to help reduce the length and complexity of queries. Unprocessed data from the IMU and microphone were inserted into a *raw* measurement while the location data were inserted into a *gps* measurement. Separating the data in this way improved queries of location data as it removed the need to exclude the 199 points produced due to the IMU for every point which included location data. Device information was

therefore included with the meta-data to be set as tags in all databases, ensuring data from the same journey could be easily connected.

Further measurements were required for data pertaining to before and after the ambulance transfer had been conducted. Earlier, the recording of data both inside the hospital and during loading and unloading of the ambulance was outlined (Section 5.4.2.3). Both this data and the data recorded during the journeys were independent of one another and therefore would be analysed separately. A *hosp* measurement contained the IMU and microphone data which was not recorded during the transfer, with a *gps_hosp* measurement containing the location data — to check for incorrect categorising. Both measurements included the meta-data as tags, as before, along with an additional *before* tag which contained either 'true' or 'false' to indicate whether the data was pre- or post-transfer.

Filtered accelerations were stored in additional measurements to reduce the need for repeated processing. The effect of vibration on the health and comfort of adults has been standardised in International Organisation for Standardisation (ISO) 2631 [49] and can give an indication as to what may be experienced by a neonate. This standard assesses the levels of acceleration after first applying frequency weighting factors. It was decided that the required memory needed to store the processed accelerations was insignificant compared to the time savings provided on each occasion assessment was conducted. These data were stored in measurements entitled *iso*, containing accelerations during transfer, and *iso_hosp*, containing those at the hospital.

6.3 Journey processing

Each journey file required interpreting in order to create any points for the database. This included the extraction of meta-data tag values, splitting data into transfer and non-transfer components, and separating the sensor data from the location information. The process of preparing each file for insertion is detailed below.

The first step in the preparation of data was to convert the filename into a timestamp. Data timestamps needed to be comparable between journeys and not simply relative to the start of each recording. Therefore, by converting the start time component of the filename (Table 3.2) into an absolute timestamp the journey timestamps could themselves be adjusted. Human-readable dates (in terms of years, months, days, etcetera) are traditionally converted into the elapsed number of seconds since 1st January 1970 [153] and termed, amongst other things, "unix" time. Converting filenames and sample timestamps to unix time provided each row of data with a unique times-

tamp, ensuring data would not be overwritten by the database.

Although it would be possible to extract data from a specific journey using a combination of the set of tag values and journey start time, this would require information about the journey to be known. Selecting a subset of a journey would also require the total journey duration to be known in order to filter the database by timestamp. Instead, the filenames were encoded as strings and included as tag values for each journey under the 'date' key. This enabled easy identification of distinct journeys — useful both in specifying data to query from the database and in correlating any problems back to the original CSV file — and further reduced the possibility of two entries possessing the same combination of timestamp and tag values.

Next, the indices of the CSV rows which pertained to the start and end of the transfer needed to be identified to ensure data was inserted into the relevant measurement. This involved analysis of the location data. The journey was deemed to have begun when the registered speed first breached 3 $m \cdot s^{-1}$ (7 mph) with a radial accuracy of less than 20 m, and therefore the index at which this occurred was stored for future use. End of ambulance transfer was identified by the last sample of location data matching the requirements for movement, with all data from this point onwards at the destination hospital.

Meta-data was read from the first line of the file to form the initial collection of tag key-value pairs in the form of a dictionary. The reason for transfer was removed from the dictionary for transfers that did not include a patient because it would not provide any useful function in querying the data. Tag key-value pairs were finalised by inserting both the string date and the device number (between 1 and 4). This dictionary of tags was used for every point in the journey.

6.3.1 Conversion of data

Before inserting into the database, the sampled data from each journey required preparation. The first step to this was reading in the data within the CSV columns that corresponded to the variables of interest — to easily associate values with the contributing sensor — and setting the data types, such as 64-bit integer for timestamps and 32-bit floats for the results of getMaxAmplitude. The data was read in chunks to ensure that processing was uninterrupted as the reading of longer recordings was identified as a bottleneck in the insertion process.

Before any data was read, empty dataframes were initialised to contain the values which would be inserted into the *gps* and *iso* measurements. Initialising dataframes

enabled the *gps* and *iso* data to be collated until the whole file had been read. Whereas the *raw* measurement would be filled before progressing to the next chunk of the file, this would be inefficient for the other measurements due to the significantly reduced sampling (*gps*) and the need for vibration filtering (*iso*). Collating the location data until the end of the file resulted in a greater number of points to be processed at once. Similarly, applying filters to the acceleration data was faster when conducted over the whole dataset than for each individual chunk.

Timestamps refer to a single point of data, for a specific set of tag values, and therefore must be unique within each recording. Multiple rows of data will occasionally — predominantly at the start of a recording — register identical timestamp values. These required altering to ensure all samples were included in analysis because duplicate timestamps would lead to points being overwritten. Therefore, the timestamps were inspected for each file chunk and any duplicate values were altered to assume linear intervals between the duplicates and the subsequent unique sample. The final timestamp of each chunk was stored to be included in the analysis of the next set of timestamps in case duplicates occurred across the chunks. Once uniqueness was ensured, the starting timestamp was then added to all values in preparation for insertion.

Noise levels were converted to the calibrated decibel values, in accordance with the results from Section 5.3.2, before the sensor data was split into the relevant dataframes. Location data was then moved to the prepared *gps* dataframe along with a copy of the timestamps while the *iso* dataframe was filled with copies of the timestamps and IMU accelerations in each axis. The data intended for the *raw* measurement was then converted to points and inserted, followed by the *gps* data. Accelerations were then filtered in each axis, as well as combined to produce a total vibration value for each timestamp, as stipulated by ISO 2631 [49] before being converted and inserted.

6.3.2 Inserting new recordings

The processing and insertion of recorded files into the database was run daily at 1am during the collection period. This enabled sufficient time for the files to be retrieved from the server (Section 3.9.3) which ensured all new data was processed.

Each of the directories which contained the recorded files from the separate smartphones were cycled through in turn. This helped to provide the device number for the tag value required to aid faster debugging. Files within each directory were accessed in alphabetical — and therefore chronological, due to the naming structure (Table 3.2) — order to determine whether processing should take place.

It was important to check journey data was not present in the database before performing any processing. Inserting data which already existed was a waste of time and duplicates were neither desired, as they could possibly skew any analysis, nor possible, due to the requirement for unique combinations of timestamps and tag values. A log of all inserted journeys was kept, in the form of a Python set of unique values, with the file represented by the formatted date string. The processing of a file was therefore skipped when the formatted date exists within the log.

Database contents were checked before a file was added to the insertion log, rather than after processing, because an error may have occurred when sending some points. The first check was to ensure the date string was present as a tag value in each of the database measurements, with a negative response indicating insertion was still required. After confirming the presence of the date string, the number of points in the *raw* and *iso* measurements were checked against each other. This involved the extraction of values from a single axis which, due to the way InfluxDB queries work, also included the timestamps for each point. The journey was determined to have been fully inserted if both sets of queried data had the same number of points and matching timestamps. The *gps* measurement was not checked due to the size difference compared to the acceleration measurements, under the assumption that *gps* would be present if both *raw* and *iso* data were inserted.

6.4 Data Cleansing

A number of problems within the recorded data became apparent both when looking at the CSV files as they were received along with when exploring the database and performing analysis. Some problems were identified during the data collection period, whereas others only became apparent at the end of the 12 months. These problems could be split into three different categories: incorrect durations, meta-data problems, and unwanted data. The descriptions, handling, and possible causes of these problems are outlined below.

6.4.1 Incorrect durations

Recorded files were expected to contain a short period of time where the device was inside the starting hospital, followed by an ambulance journey to a new hospital, and finally some data pertaining to the inside of this destination. Although the majority of recordings were as planned, there were many which required sorting.

6.4.1.1 Excess recording duration

The largest of the problems identified were recordings which continued long after arriving at the destination hospital. This would have been caused by the transport staff either forgetting to stop the recording, or having urgent clinical duties with which to deal, and the "auto-stop" feature of the app (Section 3.5.3) not activating. This was identified during the collection period and led to the inclusion of a timer to determine stationary duration — rather than simply counting location samples — and starting this timer when there was a lack of location samples, as well as when vehicle speed was below the threshold, to account for any loss of satellite signal. As this was only introduced partway through the collection period, all files needed to be checked for excessive duration and shortened accordingly.

Truncating files followed a similar process to that of the decision to end a recording with the auto-stop feature. The ultimate decision, as to which data to remove, was governed by a check for 20 minutes spent at speeds below 3 $m \cdot s^{-1}$ (6.7 mph) for which the device was termed as stationary. Unlike the app, this check only included location samples which reported a radial accuracy of less than 20 m. Examining the data suggested that 20 m was a suitable threshold as reported accuracies above this value tended to happen within buildings and produced clearly false information in terms of unrealistic speeds and large jumps in location. Using these conditions, the final moving sample could be identified and all data more than 20 minutes after this point could be removed.

Performing stationary analysis on the complete recording, rather than on-the-fly as within the app, enabled the end destination to be used to help determine when the transfer finishes on the occasions where a satellite fix had dropped. Locations which were within 20 m of the final recorded sample were selected as these could all be assumed to be at the same place. Timestamp differences between these location samples were then calculated. The recording was determined to have arrived if any two consecutive timestamps differed by greater than 5 minutes as it was highly likely that a building was blocking the signal. Any data after the first of these samples was designated as within the destination hospital and anything after 20 minutes could be removed as before.

Over 80% of the 84 recordings made by Phone 4 during the first 2 months were identified by the truncating algorithm as excessive. Removing the extra data resulted in a reduction in size of 58.6% (3.95 GB). The sheer number of identified recordings, combined with the amount of excess data, suggested there was a problem with the device

itself. This was furthered by the file size of most post-truncated files being around 34–38 MB in size and the files which did not require editing being at or below this value. Uncompressed data recorded by the 5 Plus smartphones tended to require over 80 MB per hour (Section 5.3.3), suggesting that the problem was occurring within the first 30 minutes.

Phone 4 was retrieved for examination in early December 2018. Testing revealed that the location receiver was being shut off by the operating system despite the wake lock (Section 3.4.2) and foreground declaration (Section 3.4.4). The only solution found was to manually "lock" the app to the "recents screen" [154]. Locking an app ensures that it remains when the user clears all other apps. Although there was no mention of any additional functionality, it also prevented the operating system from cancelling the location receiver. The phone was returned to staff and was recording ambulance journeys again from 23rd December 2018 with no further issues.

6.4.1.2 Multi-journey recordings

Several recordings, which were initially thought to be excessively long due to the size of the files, were found to contain more than one journey. This could have occurred when the transport staff failed to stop the recording on arrival and a second, possibly return, journey begun before the auto-stop feature kicked in. In these circumstances, the staff then decided not to intervene with the device until the end of the final journey, which suggests that this was done on purpose and the instructions for operation got confused. Each recording would ideally consist of a single journey with a unique start time identifier and precise meta-data, therefore these multi-journey files need to be separated.

Splitting up of files into the constituent journeys was not easily automated due to the variability regarding the length of time spent travelling, the time inside the hospital and the number of journeys recorded. Instead, each file was examined manually in turn. By comparing the location, accelerometer and time data the periods within hospitals should be visible.

The first step for each file was to add an additional entry for each location sample to describe whether the device was close to a hospital. This entry consisted of a descriptive string (for example, "Nottingham City") when the device location was in the surrounding area and a 'Nonetype' otherwise. The inclusion of the Nonetype was to enable quick exclusion of the data while travelling. A list of hospitals visited, in order, was then obtained by cycling through all of the descriptive strings which coincided with a

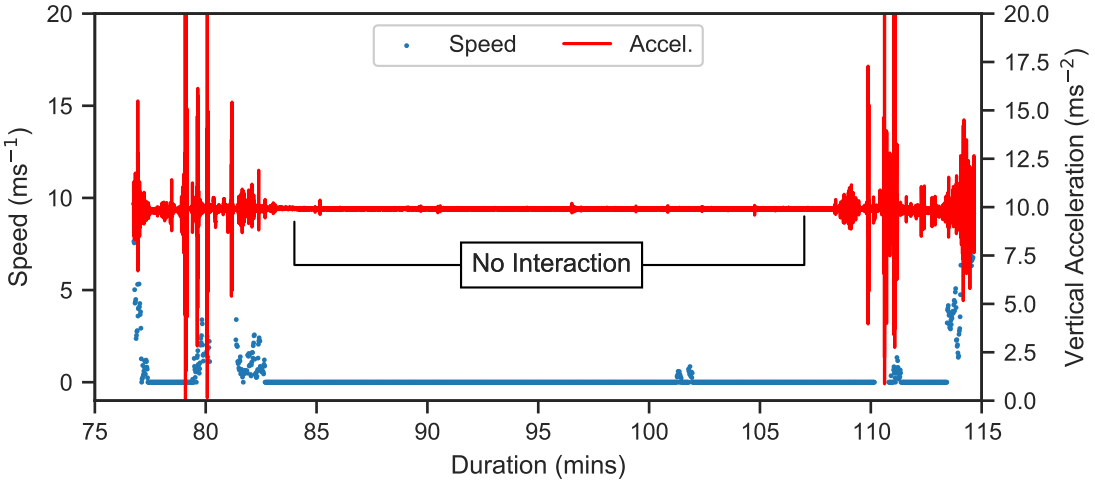

Figure 6.1: Example dataset showing the clear period without movement recorded between two journeys.

recorded speed less than 10 m·s^{-1} (22.4 mph) — ensuring the approach to the hospital was included, but excluding instances where the ambulance drove past a hospital.

Data at each identified destination hospital was then examined in turn to identify the row indices at which the journeys should be split. Index values were decided by looking at a combination of the recorded speed and vertical vibration. There were clear periods of time between ambulance journeys where the device registered minimal speed (or even no location data at all) and zero vibration (Figure 6.1). It was assumed that these periods had contained no interaction between the clinical staff and the transport trolley, therefore they provided an optimal point at which to separate the journeys.

New files were created to contain the individual journeys from each multi-recording. A function was created which required the original filename, start index and end index to be inputted. A new filename was created by calculating the adjusted journey start time from the original filename plus the elapsed time until the first index. For the first journey from a multi-recording, the start index was simply 0 and rows were simply copied directly from the original file until the end index was reached. All journeys which did not start with an index of 0 required the data of interest to be read and the time reset, so the first value was 0 ms, before writing to the new file. Resetting the start time aided the quick visual examination of the resultant CSV files by providing the true elapsed time.

Meta-data was copied from the original file to form the first line of the new file. The ambulance and trolley selections were assumed to be a constant for all journeys, as a different trolley cannot be used and it was unlikely the ambulance would be switched at a hospital. Presence of a patient on board was modified under the assumption that

each transfer was made for a single baby and the return journeys were empty. In reality, there were probably occasions where patients were transported between hospitals in consecutive journeys in order to increase the efficiency of the service. Unfortunately, both the reason for transfer and the use of emergency driving for each split journey had to be left blank because it was not possible to infer the information.

Performing separation of multiple journeys from within a single file resulted in 57 excess recordings being split into a total of 102 individual files, an increase of 79%. Two of the 57 recordings which were flagged as containing multiple journeys actually performed a slow circuit of the Nottingham QMC campus before continuing on to their final destination. These recordings were excluded from analysis because the detour would never be suggested as part of a route between hospitals and it would skew analysis of both journeys. A few of the other flagged files contained more than one journey, however the initial starting hospital was cut off. These recordings reduced the expected increase in the number of files as they only produced a single journey file.

6.4.1.3 Late location fix

Two hundred and thirty-seven (12.5%) of the original recordings did not record any location data until the journey had begun. Many of these appeared to be due to satellites being obstructed by tall buildings, typically in the more urban hospitals and surroundings, and thus the first fix being delayed. Comparing the duration of the recordings before the first satellite fix with the expected travel time from the nearest hospital suggested several of the late location data were caused by the clinical staff. In these cases, the overall recording was started while travelling, rather than inside the hospital as planned. Whichever the fault, a delayed start to the recording would result in the incorrect designation of in-hospital data as well as break up the journey, and therefore required correcting.

Each of the files designated as starting late were examined manually to assess how much of the journey would be usable. Thirty-five files (15.6% of all original recordings) started within very close proximity to a hospital and would be acceptable for analysis, however the timings suggested that no in-hospital data was present.

A further 2 recordings were found to match up with the truncated parts of files designated as containing multiple journeys. For these journeys the staff must have realised the app had not been restarted and quickly stopped the previous recording and started a new one. Each journey could then be rebuilt by combining the two file parts, with the correct meta-data being provided by the late-starting file and the timestamps of the

second file part adjusted to assume continual, uninterrupted sampling, although there was likely a delay of a second or two.

6.4.1.4 Premature finishing

Along with those that started late, a small number of recordings were found to have stopped before arriving at the destination hospital. Eleven (0.6% of the initial recordings) stopped due to a lack of registered movement. These were instances where the "auto-stop" feature worked as designed as the preceding data showed 20 minutes of very slow speeds, presumably due to excessively heavy traffic.

Some recordings appeared to have been interrupted by the smartphone running out of battery. These 16 (0.8%) files were identified via the final line of the file being incomplete. An incomplete line suggested the writing of the file was interrupted. Performing tests with the smartphones in a controlled setting confirmed that running out of battery frequently caused the final write values to be broken.

Twenty-two (1.2%) additional recordings stopped before the destination hospital and while travelling at speeds far above the stationary threshold. The final lines of each of these files were complete, which suggested the battery was not the cause of the early finish. This was further supported by further recordings being made by the same devices shortly afterwards, which would not have had enough time to adequately recharge. Ultimately, the reason for these shortened files remains unknown and could not be reproduced.

All recordings which ended early needed to be excluded from analysis, regardless of the reason for interruption. Overall journey values from these recordings would be affected by the truncated recording, while the automatically stopped recordings would not represent normal traffic conditions. Similarly, the end destination cannot be known for definite and therefore the journeys will not be easily used in comparisons. Portions of the route could have been included in the splitting of the roads into a network — discussed in full in the next chapter — however they consistently resulted in errors during the automation process. Additionally, no information would be known on the reasons for transfers or whether any emergency driving occurred and therefore further analysis would not be possible.

6.4.1.5 Developer error

Seven recordings (4 from Phone 2; 3 from Phone 4) were lost due to a problem introduced by an update to the app causing the data to end after 30 minutes. The error was

spotted within a couple of days and rectified by releasing a new update to all devices. No other recordings were attempted during this 2 day span.

6.4.2 Meta-data

A few files needed to be excluded due to the inputted meta-data being either missing or not making sense. A lot of these files had some overlap with those above which stopped before reaching the destination hospital. In the instances where this was caused by the phone running out of battery, the staff did not have the opportunity to input either the reason for transfer or whether any emergency driving was used. Similarly, the meta-data for journeys which were split from a multi-journey recording could not be ascertained and were excluded.

Several instances of incorrectly inputted information were present in the early recordings. These were identified as incorrect on the basis of the supposed trolley code not matching any of the 4 trolleys belonging to the transport group and set up with the magnets. The fault partially lies with the staff who inputted the trolley code, however it was only possible due to the original inclusion of an "Other" button on the trolley selection screen (Section 5.4.2.2). Incorrect trolley values appeared to be entered on occasions where the user realised they entered an incorrect ambulance identification code. No more selection errors were made after the "Other" button was removed and replaced with a clear "Return" button so the user could go back and change the ambulance if required.

Some of the recorded meta could be corrected by examining other recordings made around the same time as the file of interest. Occasionally, the device which recorded the confused data subsequently recorded correct meta-data for a second journey within an hour or two. The file of interest was then assumed to be using the same ambulance and trolley, due to probably conducting a return trip. Similarly, the trolley could be identified if other smartphones record a journey at the same time with the other trolley based at the starting hospital. There were only two trolleys at each of the main hospitals, therefore if one smartphone was using a certain trolley, the other smartphone would be attached to the second.

6.4.3 Unwanted data

The remaining files which were to be excluded from analysis all contained data which would not be relevant in the analysis of route comfort. These were split into three categories: those with no movement, small files, and non-hospital journeys.

Fifty-three recordings (2.8%) did not register any speeds above the stationary threshold of 2 $m \cdot s^{-1}$. These "journeys" were all contained to within a single hospital. These recordings were probably made with the intention of conducting a transfer but the actual journey was delayed for some reason. As no movement was registered during the recording, the data from the sensors would not have any use in the planned analyses.

A further 33 recordings (1.7%) were flagged as having extremely short durations. These also occurred within single hospitals and were likely conducted when introducing new staff members to the recording process. It was important that staff were capable of, and felt comfortable, using the app, and therefore permission was given to the transport groups to record these short recordings. These test files were easily identified and excluded by checking the size of the full recording.

Five journeys (0.3%) were excluded as the route was not relevant to inter-hospital transfers. Each of these recordings began at a hospital but then finished at a different establishment. It is unknown whether staff knew that they were not driving directly to a hospital, or if the decision was made mid-journey, but the routes driven to arrive at eateries were not of interest to this research. Similar to the journeys which were stopped prematurely, parts of the routes taken could have been used in the analysis of roads but resulted in errors in the automation process. While some roads may have been used during true transfers, the alternative destinations may have resulted in different lanes and speeds being used which could then affect the aggregated data.

6.4.4 Summary

As mentioned in the introduction, 1,890 files were recorded over the 12-month collection period. Removing all erroneous data resulted in a final total of 1,487 individual journeys which were successfully recorded, a loss of 19.4% (357). Data were removed due only to identified problems with either recording duration or locations. Neither levels of vibration nor noise constituted any part in the exclusion process and therefore no bias should have been introduced into the data. Although a relatively large number of journeys had to be discarded, this study has collected more than forty-two times the number of ambulance trips than any previous study (Table 6.2).

After accounting for errors in the recorded files, the final dataset of 1,487 journeys used a total of 129.8 GB — a reduction of 33% compared to the original files. Average filesize was also reduced, by 15%, to 89.4 MB, with an average of over 900 thousand lines of data per file.

Table 6.2: Summary of the ambulance journeys recorded here and by previous studies.

Study	Number of Journeys
This study	1,487 (653 patient transfers)
Blaxter et al. [44]	35 (12 patient transfers)
Hall [41]	7 patient transfers
Prehn et al. [47]	<50 minute-long runway trials
Karlsson et al. [40]	16 (combination of road and air)
Romano et al. [155]	4 45-mile circuits
Bouchut et al. [43]	10 hours-worth
Shah et al. [45]	10 (5 return trips)
Buckland et al. [73]	8 (>30 minutes, country roads)
Gajendragadkar et al. [48]	24 test runs

6.5 Expanded data overview

More in-depth analysis of the recorded journeys was possible using the database after cleansing the data. Before insertion, the number of journeys successfully recorded by each device was obtainable by counting the files which had not been moved due to problems, however this gave no indication of the true extent of the data collection. Performing simple queries on the measurements within the database provided a broader overview.

Inserting the cleansed data into the database resulted in over 946 million data points within the *raw* measurement alone. A further 946 million data points were entered into the *iso* measurement, as it contained filtered accelerations, while just under 5 million samples were inserted into *gps* (Table 6.3). The use of an optimised database enabled quick and easy access to all data, including the selection of specific data from multiple journeys, which would not have been possible to perform using the CSV files alone. Including the additional data within the hospital-specific measurements further showcases the need for a managed storage system.

A total of 36 different hospitals had at least one successful recording both as the origin and destination of journeys, with a further 6 hospitals only present in one direction. The majority of journeys occurred between hospitals within the transfer team's network — covering a portion of central England [156] — although a few journeys were found to travel significantly beyond these bounds (Figure 6.2).

Travelling components of recordings totalled 81,901 km (50,891 miles) of driving over

Table 6.3: Number of unique data points within the database, separated into the 6 distinct measurements.

Measurement	Number of samples	Number of variables
raw	946,880,064	4
iso	946,880,064	4
gps	4,730,813	7
hosp	393,116,256	4
iso_hosp	393,116,256	4
gps_hosp	1,229,342	7

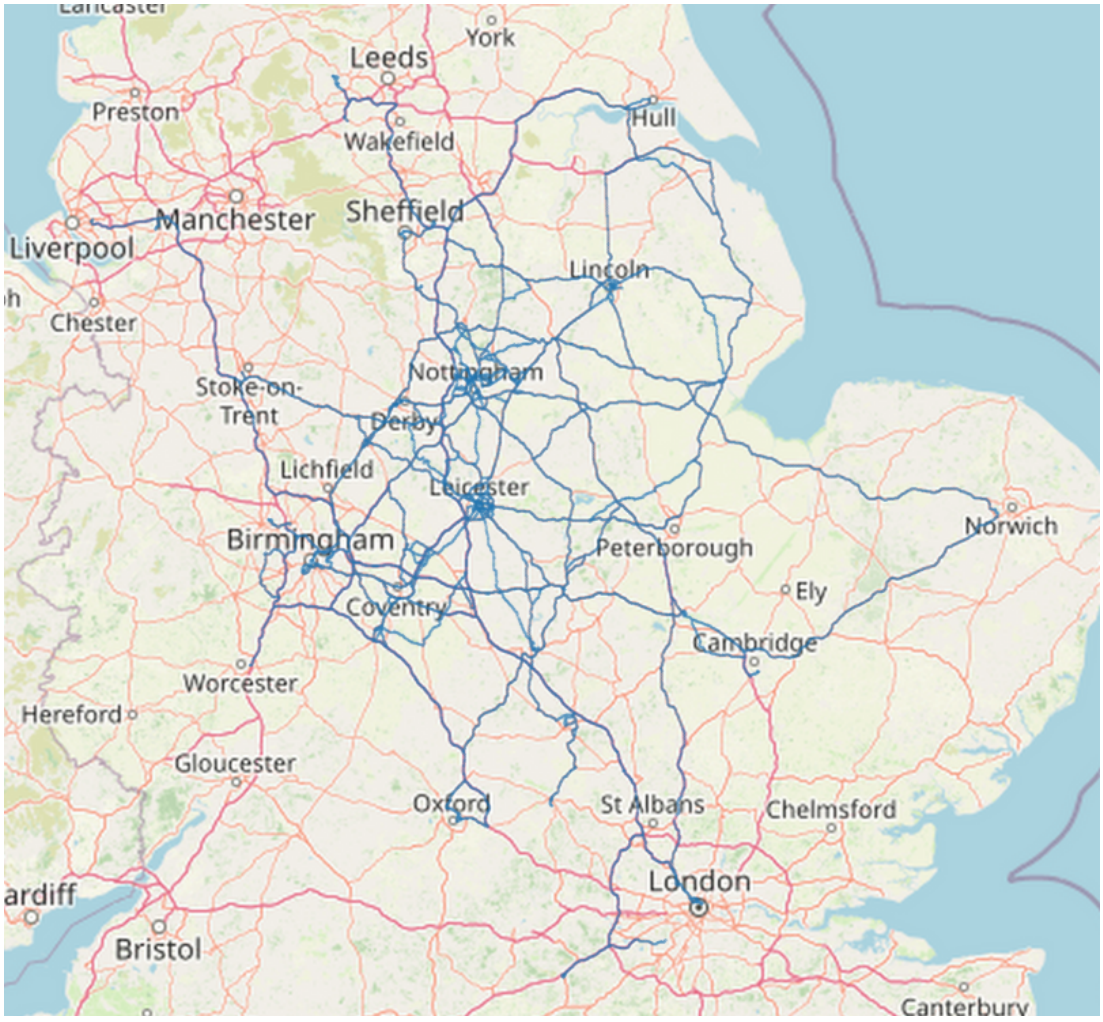

Figure 6.2: Map of all roads successfully recorded in the neonatal ambulances between 24th October 2018 and 14th October 2019 (© OpenStreetMap contributors).

the course of 1,318 hours. This resulted in an average duration of 53.2 minutes, covering 55.1 km (34.2 miles) between hospitals. Inner-city transfers consisting of a maximum of 10 km in distance were found to account for 268 journeys (18.0%). Journeys between cities were found to vary greatly, with the furthest journey being almost 290 km (180 miles) and the longest duration over 3.5 hours.

6.5.1 Meta-data analysis

The 4 main ambulances which were assigned to the neonatal transfer team were used in 1,425 (95.8%) of all journeys. Twelve additional ambulance codes were entered by staff for the remaining 62 journeys.

Recordings were generally well distributed between the available 4 transport trolleys. Only 63 journeys separate the most commonly used trolleys from the least used. The two transport trolleys based at NCH conducted 53.7% of journeys, with an even distribution between them. A 2% preference was seen between the two trolleys from the south base.

Patients were transferred during 653 (44%) of the final journeys, with the remainder of the 1,487 journeys either travelling to collect a patient or returning to one of the base hospitals. Although the number of patient transfers equates to less than 41% of the expected 1,600 performed by CenTre over a typical 12 month period [21], it remains far greater than any previous study.

Fifty-one percent (333) patient transfers were recorded with a reason of either repatriation or capacity (moving patient either nearer to the birth hospital when stable or to increase space on the ward, respectively), with the remainder as uplift (transferring to a hospital with a higher level of care than the birth hospital). Use of the t-test[1] suggests that the data collected was representative of all transfers as the proportions matched those reported for both 2018 ($t(8,245) = 0.02$, $p = .98$) and 2019 ($t(8,424) = –0.19$, $p = .85$) [21].

Thirty-seven of the final patient transfers were reported as involving an aspect of emergency driving, matching closely to the expected value of 40 time-critical transfers [21], although at a significantly greater proportion compared to the first 6 months of 2019

[1]The t-test compares two sets of data for statistical similarity. The test statistic, t, is a standardised value that indicates how closely data matches a reference set, with the value inside brackets displaying the degrees of freedom for the calculation. The p value provides the statistical probability of the two sets being similar, with values below .05 (<5% chance of similarity) typically referred to as an indication of significant differences. As with all statistics, the results of these tests are only used to illuminate potential dissimilarities and not to prove an idea.

($t(1{,}456) = 3.12$, $p < .01$). Some discrepancy was observed when comparing the data recorded solely during the same 6 month period, with 28 transfers with emergency driving indicated compared to the 20 time-critical transfers officially reported. It is possible that this difference could have stemmed from alternative definitions.

The majority (33) of the emergency driving occurred during uplift transfers, comprising 10% of all uplift transfers. This was slightly greater ($t(4{,}157) = 2.27$, $p < .05$) than the expected 7% [21].

The general similarity between the recorded meta-data and the officially reported values install confidence in both the entered values and the overall use of the app by the neonatal transfer team. There did not appear to be any bias in the decision of whether a journey was recorded or not and therefore data should be fit for analysis.

6.6 Conclusion

Over the course of 12 months, 1,890 CSV files were recorded using the developed app, containing vibration, noise and location data within neonatal ambulances. The CSV format was adequate for logging the data, however a more comprehensive system was required to facilitate analysis and comparison between journeys.

Recordings were inserted into the InfluxDB database which was specifically designed for storing time series data. Sensor data were separated into three main measurements within the database to both improve query times and reduce repetitive computation: *raw* containing unprocessed noise and vibration, *gps* containing location data, and *iso* containing vibration filtered in accordance with ISO 2631. Data were further separated based on whether they were recorded during transfer or not, with the latter being denoted *hosp*, *gps_hosp*, and *iso_hosp* respective of the main measurements. meta-data and device information, along with the start time of each journey, were included as tags for each sample to both enable filtered analysis and link data directly to the original CSV files.

Several problems were apparent with the data when examining the CSV files and the database. These yielded durations and meta-data which did not match the expected protocol. Rigorous cleansing of the data was performed with the objective of remaining journeys consisting of a full hospital-transfer-hospital dataset with unique sets of tag values.

A total of 1,487 separate journeys remained after the data cleansing process. While some data was discarded, this total remains far greater than any other in literature with

the next highest being Blaxter et al. [44] with forty-two journeys recorded. In terms of overall quantity, a total of over 81,000 km of driving were recorded over the course of 1,318 hours.

The main aim of this work is to assess the health and comfort of roads with a view to offering an optimal route between hospitals. After ensuring the database only contained information pertaining to ambulance journeys between hospitals, steps were required to optimise analysis. Comparing overall journey values from each recording would provide an insight into the expected values of ambulance transfers, however comparisons would be limited to the journeys which drove between the same hospitals. Due to both the scale and spread of the data recorded, it is inevitable that some journeys would drive along some of the same roads even if they had different destinations or starting points. Therefore, there is a need to aggregate all data which has travelled along each section of a route to maximise the use of the data.

CHAPTER 7

Splitting Recorded Road Network by Decision Points

7.1 Introduction

One of the main aims of this work is to analyse how comfort levels inside ambulances may vary with route choice. Assessments could be performed individually for each of the 1,487 journeys recorded in the neonatal ambulances, however this would limit comparisons to only the journeys between the same hospitals. Instead, a method was required to aggregate data for analysis. Large amounts of data were collected to improve the confidence in analyses by averaging values of acceleration and noise, for example, to minimise the effect of anomalous events. A potential cause for anomalous events would be abnormally heavy traffic which would significantly reduce the vehicle speed and subsequently the values from the sensors. The likelihood of average assessments reflecting the true conditions would increase with the number of journeys as multiple journeys along the same stretch of road were earlier shown to result in similar values, under the same conditions, of both noise and vibration (Figures 3.10 & 3.12). As traffic conditions can not be controlled, the preferred analysis would be the expected comfort on a normal day.

Averaging all data which travels along the same roads, in the same direction, would increase the confidence in the assessed values by reducing the impact of any outlying data (for example, caused by a traffic jam). Before performing any aggregation, the data needs to be grouped by road. One method of grouping the data would be to first group the recordings which started and ended at the same hospitals. The different routes could then be found by checking if the same roads were taken. The comfort would then be assessed on a route-by-route basis. While this is a relatively simple

approach, it effectively assumes that each route was on completely unique roads. The problem with this approach is that it will ignore the majority of data collected and thereby invalidate the need for mass data collection. Instead, a method of grouping the data which accounts for the true nature of roads was required.

In reality, routes between different places can overlap. The clearest example of this is in hospital grounds where there may be only one or two roads available to travel away from the neonatal ward. This will mean all routes which depart from the hospital will have these roads in common. Recorded data needs to be split into unique segments to take advantage of these common sections of road, wherein all the data within a segment would travel along the same path from beginning to end. These segments would be split up by locations at which a decision was made, for example, a junction where more than one exit was taken. This would then maximise the number of journeys aggregated during analysis.

A method of identifying the unique segments was required before any analysis could be performed. Initially, the aim was to analyse routes while data collection was ongoing. This meant that there was potential for a new journey to travel along previously unused roads, which would alter the network of segments. Therefore, the method of dividing up the data needed to be programmatic and quick to enable analysis of routes to include newly recorded data with minimal delay.

Open source road networks are readily available online, with the level of detail varying with region. For the UK, highly detailed networks are available from OpenStreetMap data. These networks are split into many segments, each containing information about the road. Some of the details, such as speed limit and travel direction, could be useful, however the majority of them are irrelevant for the purposes of splitting and analysing the recorded data. Taking into account all the extra details, such as whether the road is in a tunnel or on a bridge, can result in a single road being split up multiple times over a relatively short span. This would introduce far more separate segments than required, with the possibility of some segments not having enough data samples to accurately analyse.

One of the drawbacks of using an established road network is needing to download the data before it can be used in analysis. When the roads of interest are known, this is straightforward. As this study was recording real-world ambulance journeys, it was impossible to predict which roads would be used and how large an area would be covered. Typically, neonatal ambulances keep to main roads during transfers. While this suggested that the minor classifications could be discarded, the roads around hospitals and within urban areas tend to be smaller. Needing to include all road classifications

along with a large geographical area, to try and ensure all future recordings would be within the road network, would result in a massive reference dataset. Recorded data would also need to be matched to true road coordinates, which would be further complicated by the sheer number of possible roads.

Road networks are continually evolving as infrastructure projects add, modify, and remove roads. As these changes would undoubtedly occur during the data collection period, the reference network would need to be updated to account for any changes which the ambulances may come across. This meant that updates would need to be continually downloaded and implemented, assuming that the online data was not delayed.

For the purposes of this study, only data on the roads which were used by the ambulances were required as the comfort of a road cannot be assessed if no data was collected on it. Given the number of roads which would not be required, downloading and updating an online UK network did not seem sensible. Ultimately, either the whole road network dataset would be used for identifying segments in the recorded journeys, or a new method was required.

This chapter outlines the splitting up of recorded roads using only the recorded coordinates, along with the feasibility of the resultant planar networks.

7.2 Location representation

The first thing required in order to be able to split up roads into unique sections, without any external information, was a method of identifying neighbouring locations. It should thereafter be straightforward to group coordinates together which are from the same route. Similarly, the coordinates which do not match with another group would be identifiable and thus assigned to a different segment.

The raw coordinates, in the form of latitude and longitude, could be used to identify whether one location was in close proximity to another. Calculating the distance between two sets of coordinates and comparing the distance to a specified threshold (25 metres, for example) would determine whether or not to class the coordinates as neighbours. Distance between two points on the surface of the Earth are calculated by multiplying the average radius of the Earth by the surface angle between locations (Equation 7.2.1). Different methods of acquiring small surface angles are available, with the Pythagorean Theorem (Equation 7.2.2) providing an approximate value or the haversine formula (as shown earlier in Equation 3.3.1) calculating a more accurate re-

sult [127]. Both of these methods require multiple calculations and need to be repeated for each location comparison. While slightly reducing the number of calculations is possible by using a surface angle as a threshold instead of distance, converting latitude and longitude to a single location identifier enables further reductions.

$$\text{distance} = \text{surface angle} \cdot \text{Earth's radius} \tag{7.2.1}$$

$$\text{surface angle} = \sqrt{(\delta\ \text{lat})^2 + (\delta\ \text{lon})^2} \tag{7.2.2}$$

Multiple systems of representing coordinates as a single code or string exist, ranging from algorithmic combinations of numbers and letters (for example, geohashes [157]) to words and phrases (for example, what3words [158]). These systems all work by dividing the Earth's surface into a grid, with each code representing an area (or cell) within the grid. Instead of calculating distances, it should be possible to identify neighbouring locations by determining if they occurred in the same grid cell. Proximity of neighbouring locations would depend on the size of the grid used. By using a system with a variable resolution, the grid area can be optimised.

Locations in close proximity but in different cells of the grid would be ignored if neighbouring locations were restricted to those within the same area. Identifying neighbouring cells would enable more locations within close proximity to be found and also a potential method of identifying when paths diverge.

Open location code [159], developed by Google, was chosen as the base for creating a network of roads from the recorded data because it can encode locations in a range of resolutions (Table 7.1) and these are easily incremental in each of the cardinal directions. Open location code works by converting latitude and longitude values to a string of 2–15 digits — depending on the requested resolution — referred to as a "plus code". Development use is aided by the availability of libraries, in multiple programming languages, for encoding and decoding plus codes, along with their implementation within Google Maps which enabled quick visualisation and sanity checks.

7.2.1 Incrementing plus codes

Much of this chapter is reliant on the incrementing of plus codes to find neighbouring codes. Before codes can be incremented, it is important to understand what plus codes represent. At lengths of 10 digits and below, plus codes are formed of pairs of digits which represent the latitudinal and longitudinal positions within the larger grid. The open location code system uses 20 distinct digits (Table 7.2) each of which has an as-

Table 7.1: Valid open location code lengths and the resultant resolutions. Reproduced from the Open Location Code specification.

Code Length	Resolution			
	Lat. (°)	Lon. (°)	Lat. (m)	Lon. (m)
2	20	20	2,226,000	2,226,000
4	1	1	111,321	111,321
6	1/20	1/20	5,566	5,566
8	1/400	1/400	278	278
10	1/8,000	1/8,000	14	14
11	1/40,000	1/32,000	3	4

Table 7.2: Plus code digits along with their numerical values. Reproduced from the Open Location Code specification.

Symbol	2	3	4	5	6	7	8	9	C	F
Value	0	1	2	3	4	5	6	7	8	9
Symbol	G	H	J	M	P	Q	R	V	W	X
Value	10	11	12	13	14	15	16	17	18	19

signed value. Neighbouring plus codes can be found by incrementing the latitude or longitude digit, depending if interested in north/south or east/west, starting with the digits at the right of the code. Depending on the direction of the neighbour of interest, the digit either gets incremented towards 'X' (north or west) or towards '2' (south or east).

Computing the plus code to the east would require incrementing the longitude digit furthest to the right towards 'X' or to a larger value. This can be shown by using the 8-digit plus code for the centre of the Nottingham City Hospital (NCH) campus, '9C4WXRRV+' (Figure 7.1A). In this case the furthest digit to the right which corresponds to the longitudinal direction is the 8^{th} digit (highlighted with a single underline). Incrementing eastwards therefore results in the 'V' being changed to a 'W' to result in '9C4WXRRW+', as shown in Figure 7.1B.

On the occasions where the digit to be incremented is at the end of the list, it needs to be reset and the next corresponding digit to the left is incremented. This continues until a digit can be successfully incremented. Continuing east from the previous plus

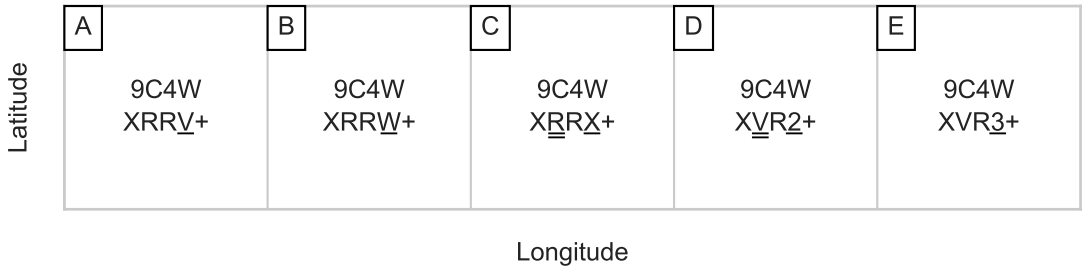

Figure 7.1: Example of neighbouring plus codes. Single underlines highlight the main longitudinal digit that requires incrementing when moving between the codes. Double underlines highlight the second longitudinal digit that also requires incrementing due to the main digit being reset.

code, it can be seen that the 8th digit gets changed to 'X' (Figure 7.1C). As the digit is at the end of the list, incrementing east a second time would mean the 8th digit is reset to '2' and the focus would shift to the next longitudinal digit. This corresponds to the 6th digit, 'R', as indicated by the double underline, which gets changed to 'V'. Therefore, incrementing east from '9C4WXRRX+' results in '9C4WXVR2+' (Figure 7.1D).

Further incrementing east would then shift focus back to the 8th digit (Figure 7.1E), as before.

7.3 Forming a Planar Network using plus codes

Recorded journeys needed to be converted into a road network to enable the greatest amount of data were used in analysis. Splitting up the roads used during the collection period into unique segments — termed 'edges' in network vernacular [160] — would enable similar data to be aggregated. The resultant values for comfort and health for each edge could then be combined to produce an expected value for a full route.

This section outlines the processes required to generate a planar — because there is no altitude component — network using plus codes. First, the impact of plus code precision is explored, before the full algorithm for the network is detailed.

7.3.1 Deciding on plus code precision

All coordinates for each dataset needed to be converted to plus codes before they could be used to form a network by grouping similar journey segments. The size of the area represented by a plus code depends on the precision, as mentioned earlier, and can be seen in Figure 7.2. Depending on the required resolution of the network, different precisions could yield better results. Therefore, the next step was to decide on the code

Figure 7.2: Example of the reduction in area size represented by plus codes with precisions of 4 (A), 6 (B) and 8 (C) digits, with subsequent precision levels outlined in black.

length to use.

The main deciding factor in choosing a plus code precision was finding a balance between resolution and the number of locations recorded within a specific area. Increasing the plus code precision would enable a higher resolution to the network which would then enable more roads to be distinguished from one another. This can be seen by comparing the number of roads within Figure 7.2B, at 6-digit precision, and Figure 7.2C, at 8-digit precision. Road resolution would significantly increase with each precision level, however this would also result in fewer location samples. The number of location samples which may occur inside a particular plus code can be estimated using the plus code dimensions, the average distance travelled through the area and the expected vehicle speed.

First, the average distance between two points inside a unit square was calculated. Once this was obtained, the average distances within each plus code were calculated by multiplying the value for the unit square by the known resolution in metres from Table 7.1.

Assuming a straight line for reasons of simplification, the average distance between two points on the perimeter of a unit square was approximated by a simple program that simulated potential paths. This program split each side of the square into equal sized divisions and calculated the distance between each point using the Pythagorean theorem. Taking the mean of these values resulted in the average distance of 0.735 units.

Table 7.3: Average expected location samples inside a single plus code for a single journey, at standard UK speed limits of 30, 50 and 70 mph (13.4, 22.4 and 31.3 $m \cdot s^{-1}$)

Code Length	Average Number of Location Samples		
	30 mph	50 mph	70 mph
2	12,201	7,321	5,229
4	6,102	3,661	2,615
6	305	183	131
8	15	9	7
10	<1	<1	<1

Multiplying this value by the plus code resolution then provided the average distance travelled for each code length.

The expected number of location samples were then calculated for the standard UK speed limits of 30, 50 and 70 mph (13.4, 22.4 and 31.3 $m \cdot s^{-1}$ respectively) for the first five plus code precision levels (Table 7.3). As the location data was sampled at 1 Hz, this calculation required dividing the average distance by the vehicle speed. Although 10-digit precision plus codes would give great road resolution, the small area covered by each code means that not all journeys which travel through at 30 mph would register any location data inside. These odds reduced further at higher speeds and therefore 10-digit precision was not suitable for use in forming the network. Instead, 8-digit plus codes were chosen as they possess the resolution to distinguish between major roads and, at an average of 6.5 location samples at the maximum UK speed limit, are large enough to ensure that journeys would not miss them.

Once a precision has been set, in this case 8-digits, the coordinates for all journeys need to be converted. This conversion could either be performed directly before generating a network or when inserting the location data into a database. Converting coordinates to plus codes before inserting would speed up the network generation by reducing the number of computations required. It is recommended to insert plus codes at a higher resolution than determined to enable sensitivity analyses to be performed as lower resolution plus codes can be formed by the simple truncating of strings, whereas higher resolutions would require full computation from the coordinates.

Table 7.4: ID and description for each of the three different sets formed when comparing a journey with an edge.

ID	Contents
in_both	plus codes which occur in both the journey and the edge
in_edge	plus codes which occur only in the edge
in_journey	plus codes which occur only in the journey

7.3.2 Splitting algorithm

The splitting of road data into a network of unique edges consists of multiple interlinked stages. These will be set out below.

The first stage would be to gather all plus codes, at the required resolution, to be analysed and group them by recording. Duplicate plus codes should be removed on a journey-by-journey basis to form a type of collection termed a set. This is because the algorithm to be introduced involves checking for the presence of specific codes and having more than one of each plus code would only serve to increase computation time due to extending the length of the collections to be searched. The use of sets both reduced the amount of data to be compared — as each contained entry must be unique — and sped up computation due to being designed partly for comparison, with multiple implemented methods [161].

Edges of the network can start to be identified once each journey is in the form of a set of unique plus codes. This algorithm works by iterating through the journeys and adding new splits where required. Therefore, with the generation of a new planar network, the first journey becomes the first edge as it is a constant, unbroken set of coordinates. This edge will likely be split up as more journeys are included.

The next step towards building the network is comparing the next journey with the identified edge. When comparing plus codes of a journey with those of an edge there are five possible outcomes, as shown in Figure 7.3. Three new sets (Table 7.4) are created to handle the different outcomes with the processing dependent on the contents of each.

7.3.2.1 Outcomes of Journey comparison with a Network consisting of a single edge

The simplest outcome is when all of the journey plus codes match those of the edge (Figure 7.3A). In this case, only the *in_both* set will be filled, with *in_edge* and *in_journey* both being empty. Network analysis can progress to the next journey as no change has

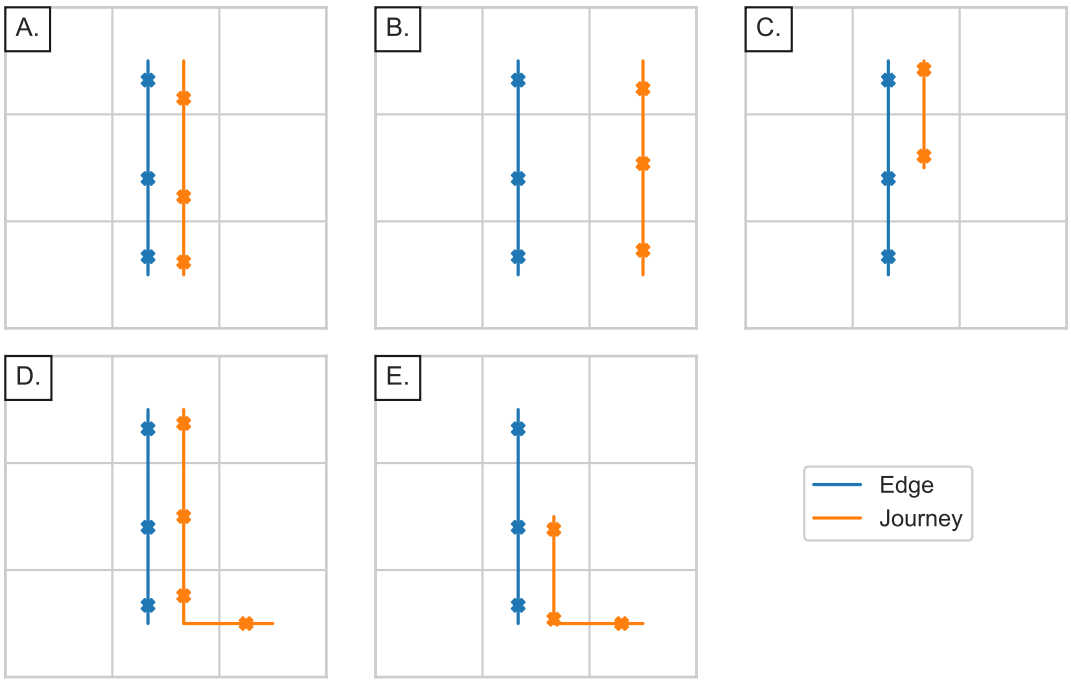

Figure 7.3: Simplified illustrations of the types of result when comparing a set of journey plus codes and an edge, with each grid representing a separate plus code area.
A: All plus codes in both edge and journey; the journey codes match the route of the edge.
B: No plus codes in both edge and journey; the journey codes take a different route to the edge.
C: Journey plus codes inside edge, but some edge codes remain; journey misses a plus code of the edge, for example, due to travelling through final area before a location sample is received.
D: Edge plus codes inside journey, but some journey codes remain; journey continues after travelling along edge.
E: Some plus codes in both edge and journey, some codes of each remain; combination of situations C & D.

been made to the list of edges.

Next simplest is when there are no shared plus codes between the journey and the edge (Figure 7.3B). This results in *in_both* being empty, leaving both the *in_edge* and *in_journey* sets containing separate journeys. The *in_journey* set is then added to the network, resulting in a total of two edges, before progressing to the next journey.

Complications arise when there are plus codes in *in_both* as well as at least one of *in_edge* and *in_journey*. These fall into three outcomes: the journey is part of the edge but some edge plus codes remain (Figure 7.3C), the edge is part of the journey but some journey plus codes remain (Figure 7.3D), or the edge and journey share some plus codes but codes remain in each (Figure 7.3E). Each of these outcomes will result in an increase in the number of edges, with the number dependent on the continuity of each set.

As the number of journeys increases, so does the likelihood of the network consisting of multiple edges. This therefore requires comparison with each edge before finalising the network.

7.3.2.2 Outcomes of Journey comparison with a Network consisting of multiple edges

Previously, when comparing a journey with a single-edge network, the resultant sets would be added straight to the network. Where more than one edge already exists in the network, remaining journey plus codes need to be checked against each pre-existing edge. This does not affect the possible outcomes of comparing the journey and edge plus codes, but rather the handling of the resultant sets.

The *in_both* and *in_edge* sets get added to the network as before, unless they are empty, as they are known to contain unique plus codes which are not present in any other edge. This is because the remaining edges of the network would have already been through the process of comparison against these codes.

Any codes remaining in *in_journey* are carried over to the next edge comparison regardless of whether any codes were in *in_both*, as the results of each comparison are independent. Each subsequent edge of the network is then compared with the remaining journey codes, as above, until either there are no codes or no edges remaining. Where no pre-existing edges remain in the network, but *in_journey* is not empty, the remaining codes become a final edge of the new network due to being unique.

Before the network analysis can progress to the next journey, each edge needs to be checked for continuity to ensure that each contains an unbroken chain of locations.

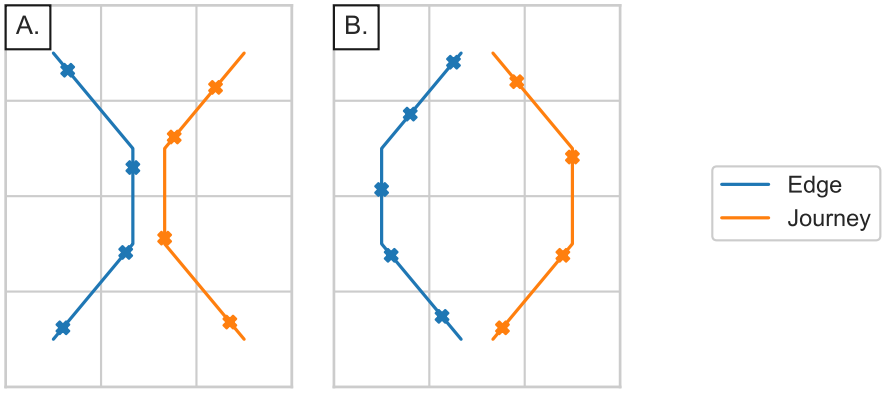

Figure 7.4: Simplified illustrations of 2 ways comparing a journey with an edge could result in continuity splits within the resultant sets.
A: *in_both* is continuous, *in_edge* and *in_journey* are both split.
B: *in_edge* and *in_journey* are both continuous, *in_both* is split.

Otherwise, assessment of each edge could provide incorrect values. This could occur in the calculation of duration, for example, where the difference between the first and last timestamp is found without realising there is a gap in the middle, resulting in greater durations than expected when edges are combined to assess routes.

7.3.2.3 Continuity checking of plus code sets

When an edge and journey share some, but not all, plus codes with each other then each set needs to be checked for continuity. This is required to ensure that each new edge is an uninterrupted section of road. Continuity can be interrupted by the edge and journey converging before diverging (Figure 7.4A), which interrupts both *in_edge* and *in_journey*, or by diverging before converging (Figure 7.4B), which causes *in_both* to be discontinuous. Due to the unordered property of sets, it is impossible to establish which are continuous and which are not without checking.

The process of checking for continuity involves iterating through the plus codes contained in each set and forming new sets based on the presence of neighbouring codes, where each new set will be termed a 'group' for clarity. This then results in either a single group — in which the plus codes are all adjacent and, therefore, continuous — or multiple groups.

To start with, the first plus code within the original set is put into a new group as there are no existing groups with which to compare against.

Neighbouring codes, in each of the cardinal and ordinal directions (Figure 7.5A) (as outlined in Section 7.2.1), are calculated for every subsequent plus code in the set in turn. These are then compared against every existing group for similarity. Where one

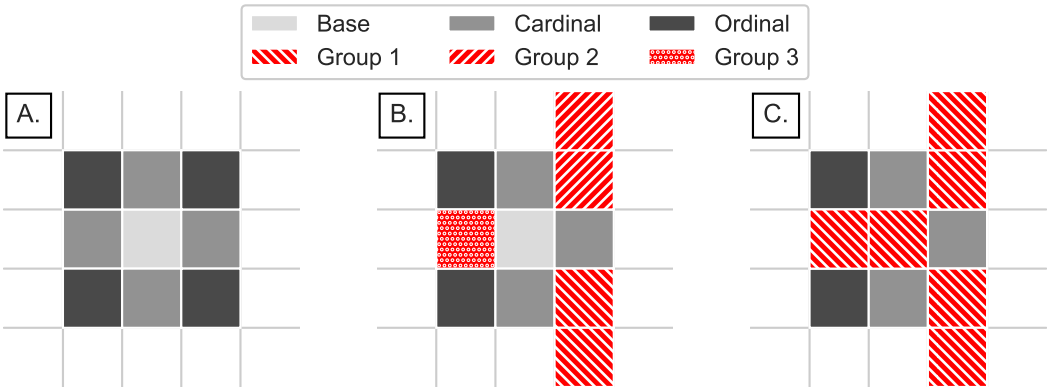

Figure 7.5: Simplified illustration of a base plus code and its neighbouring plus codes in the cardinal and ordinal directions (A), along with an example of how three separate groups (B) can be joined together by the base plus code to become one (C).

of the neighbours is identified as being present in a group, the new plus code is added to the rest of the values in the group. Otherwise, a new group is formed containing solely the new plus code.

Presence of plus code neighbours needs to be checked in every group which has been identified. When there are multiple groups, the neighbours can either be found in a single group, in multiple groups, or not in any group. The cases where either a single or no group matches any neighbours is handled as before, with either the plus code being inserted into the matching group or added as its own group. Multiple groups can occur when there are two or more groups separated by a single plus code (Figure 7.5B). These can then be combined to form a single, unbroken group (Figure 7.5C) by including this plus code.

Once all plus codes have been checked and added to the groups, the input set has been split up into its continuous segments which can then be added into the network as edges. The next step is to verify if the small edges should truly be separate.

7.3.2.4 Handling small edges

While small edges — in terms of the number of plus codes — may reflect a short, unique section of road, it was found that the nature of the algorithm occasionally created a small edge which ran parallel to a larger one. Vehicle speed, GPS drift, and the time-based sampling of location data can all contribute to the formation of these anomalous edges. Figure 7.6 highlights the effect of time-based sampling with two journeys travelling at the same speed but recording location at different points, identified using triangle and circle markers respectively, showing that a small trip through a

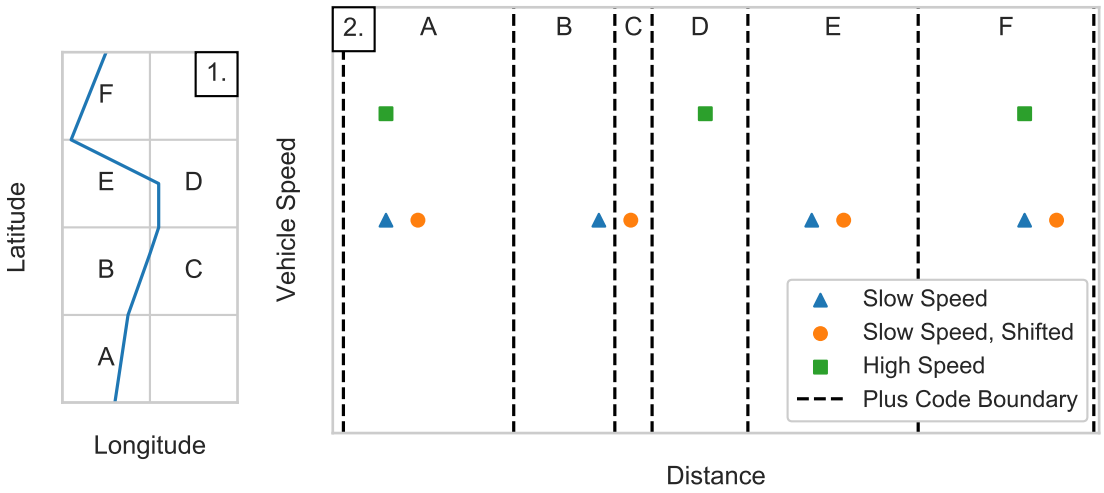

Figure 7.6: Illustration of how different journeys along the same route (1) may not register the same plus codes due to the difference in sample time and vehicle speed (2).

plus code (C) is picked up only by the delayed journey and B is missed. Similarly, the effect of vehicle speed is shown between the triangle and square markers, assuming the location samples started at the same position. In this case, the faster vehicle registers three plus codes (A, D & F) while the slower vehicle records four (A, B, E & F). GPS drift can also have the effect of registering a plus code adjacent to a road, although this is less common. These factors can all result in different journeys not having the same plus codes despite travelling along the same route, and thus mistakenly produce multiple edges.

Initially, it was thought that only the edges which contained a single plus code should be absorbed into other edges, and therefore the term 'singleton' was used — in mathematics singleton refers to a set which contains only a single element [162]. The possible anomalous small edges were then found to exist in various sizes, leading to 'singleton' becoming unfit. Instead, the term 'multiton' was coined — derived from the previous, singleton, term — to represent the small edges requiring checking. This term is used henceforth to clearly distinguish from larger edges of the network.

The assessment of multitons needs to be conducted after each journey has undergone network analysis. This assessment is again performed through the use of plus code neighbours. First, all edges with a content length at or below a specified threshold are separated from the network. These multitons are then iterated through, with their neighbours computed and checked against all other remaining edges. If an edge contains a neighbouring plus code of a multiton, this multiton is then absorbed into the edge.

It is up to the user to determine the maximum size threshold for which an edge becomes a multiton, similar to the choice of plus code precision in Section 7.3.1. Setting the threshold value too small could result in roads being broken into far too many sections, whereas a threshold value which is too large could absorb edges which belong on their own. Ultimately, it is dependent on the resolution of the roads of interest.

7.3.3 Summary

A method of creating a network of roads from recorded routes was devised. This took the form of comparing locations, in the form of plus codes, from a journey and comparing them with edges already identified. Depending on the outcome of each journey-edge comparison, the resultant network could contain either unchanged, split or brand new edges or a combination of the three.

Two variables can affect the outcome of the network: the plus code precision and the multiton threshold. Careful consideration is required for both values as they could result in the final network consisting of fewer or greater number of edges than expected. The 8-digit plus code precision was recommended after comparing the average number of location samples one could expect to fall within each plus code area. A multiton threshold would need to be derived for each use as it would be highly dependent on the size of the roads to be analysed.

After designing a method of creating a network, the next step was to apply it to some recorded data.

7.4 Road selection for network assessment

The aim of this chapter is to generate a dynamic method of creating a road network to aid the analysis of route comfort. Aggregating data from each journey which travels over each unique section would give greater confidence in the computed values. It is important that the roads are divided up correctly to ensure the analysed data is comparable. Therefore, it makes sense to test how well the roads are split up between two destinations.

Recorded journeys were analysed to decide upon a route to use to test the method of network creation. Two main requirements were chosen for the route selection, both to provide a reasonable test of the splitting algorithm and to provide a large amount of data for the route assessment later: there needed to be multiple possible road combinations between the start and end hospitals, and ambulances needed to have recorded

Table 7.5: Top six most frequently recorded journeys made by CenTre neonatal transport between 24th October 2018 and 14th October 2019.

Start Hospital	End Hospital	Number of Journeys
NCH	QMC	85
LRI	LGH	64
QMC	NCH	55
LRI	UHC	51
NCH	LRI	39
LGH	LRI	33

along at least two of them. The number of journeys which began and finished at certain hospitals were able to be aggregated by identifying the plus codes surrounding each neonatal ward and comparing these to the journey plus codes.

Four of the six most frequently recorded routes (Table 7.5) were discounted from the network generation. This was due to the number of road combinations being limited, with the journeys being between hospitals within the same city (NCH & Queen's Medical Centre (QMC); Leicester Royal Infirmary (LRI) & Leicester General Hospital (LGH)). Instead, routes travelling from NCH to LRI were chosen, after journeys from LRI to University Hospital Coventry (UHC) were excluded due to being a direct route consisting of one main road. Both destination hospitals also serve as bases for the CenTre neonatal transport team, which should mean that the surrounding roads were covered in the majority of recordings by ambulances either leaving or returning.

The 39 recordings which travelled from NCH to LRI used a wide range of routes over multiple different road sections (Figure 7.7) which should give a good insight into the performance of the network. Examining the recordings manually showed that 13 different combinations of roads were used resulting in 38 unique sections of road, in a range of sizes, based on where recordings overlapped. Choices during journeys varied from large decisions, such as whether to use the motorway along the west edge of the map or the A-road along the east side, to small decisions, such as which direction to loop around the LRI maternity building. Using these routes should provide a good test of the network generation due to the number of road combinations and the spread of section sizes.

Network generation using the full set of recordings (Figure 7.7B) was assessed along with those which travelled directly from NCH to LRI. The purpose of recording such a high volume of data was to increase the number of recordings which could be included for analysis of a specific road. Limiting analysis to solely the NCH to LRI record-

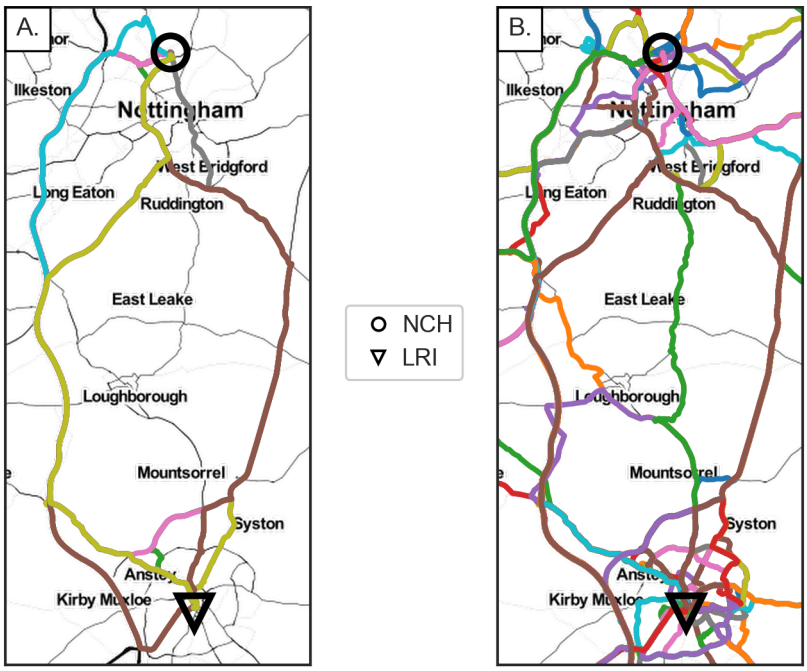

Figure 7.7: Coordinates of all recorded journeys which travel from NCH to LRI (A) along with all recorded coordinates in the surrounding area (B) overlaid on a map (© OpenStreetMap contributors).

ings would result in data being ignored from 550 other recordings which were manually identified to travel over parts of the roads of interest. These additional recordings would result in the 38 road sections being divided up by a factor of 3.5 to form 133 new sections under visual analysis. After generating the NCH to LRI network, the full dataset was split into unique sections and the new splits to the NCH to LRI roads will be examined to ensure network generation can cope with large datasets.

The choice of routes travelling from NCH to LRI are significantly varied as to test the splitting algorithm at both large and small scales. Including all available recordings in the network analysis of these routes would further examine the robustness of the method by exposing it to a significantly greater amount of data in an uncontrolled manner. The next step, as the journeys had been identified, was to perform the network analysis.

7.5 Network generation using 8-digit plus code precision

The generation of a network using plus codes requires both the plus code resolution and the multiton threshold to be assigned, along with a set of journeys to be analysed. A plus code resolution of 8 digits was selected in Section 7.3.1 while the journeys travelling from NCH to LRI were chosen in the previous section to provide a basis for

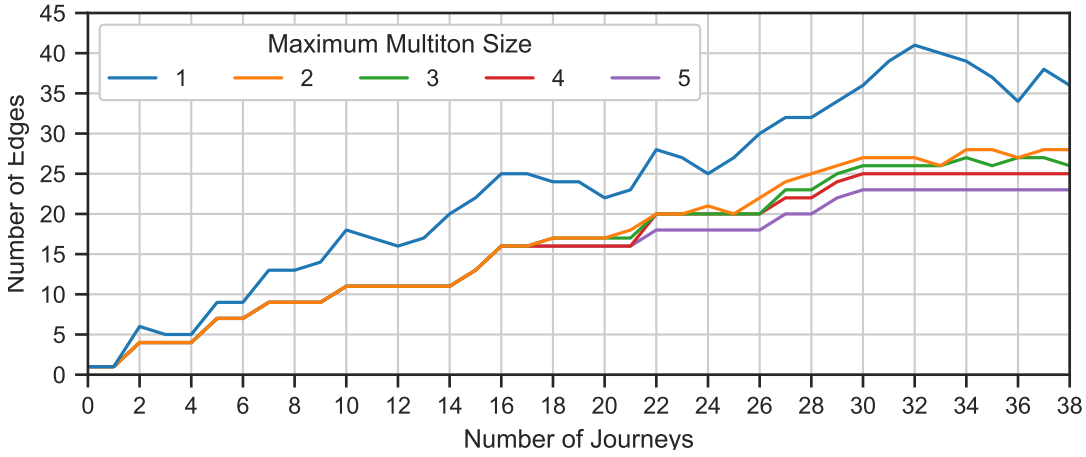

Figure 7.8: Impact of the multiton threshold on the number of identified edges during network generation.

analysis. This section will therefore explore the possible effects of multiton threshold first, followed by the application of the algorithm to roads travelled by the NCH to LRI recordings, first in terms of the 39 journeys which used these routes and then including all 1,448 remaining journeys.

7.5.1 Multiton sensitivity analysis

Before assessing the resultant network using the 8-digit plus codes, the maximum size of multiton needed to be determined. The multiton threshold would affect any generated network as it specifies the minimum number of plus codes in an edge, with any smaller edges being absorbed into a larger edge containing at least one neighbouring code (Section 7.3.2.4). Ideally, the threshold would be small enough to enable short road sections to be kept while still absorbing incorrectly split plus codes. Threshold values were therefore explored using a sensitivity analysis.

It can be seen in Figure 7.8 that a multiton threshold of one would produce poor networks due to the number of edges fluctuating. The number of edges in a network should only increase with new journeys as decision points can be created but not removed. A decision point, by definition, is a location where two or more journeys took different roads and, therefore, can only be removed by removing one of the journeys which took a different route. As well as fluctuating, a threshold of 1 frequently resulted in incorrect edges so is not usable.

Multiton thresholds between 2 and 5 produced an identical number of edges for the first 18 journeys from NCH to LRI (Figure 7.8). At and above 18 journeys, thresholds of 2 and 3 produced a higher number of edges and also resulted in fluctuations due

to new plus codes parallel to existing edges becoming new edges and later absorbed into the original edge. These fluctuations suggest the networks cannot be trusted, even if the final edges produced were correct, and they were rejected. Although both final thresholds produced stable networks after each journey, the lower threshold of 4 was chosen to assess network generation as it would allow for a greater section resolution.

7.5.2 NCH to LRI journeys

Figure 7.9 displays the progression of the network generated from the 39 NCH to LRI journeys using 8-digit plus codes and a maximum multiton of 4, with each change of colour representing a new edge. Overall, the algorithm appeared to work, with decision points being identified as more recordings were analysed. Although 25 different sections of road were identified, this fell short of the expected number of 38.

Lack of resolution was one of the reasons for the reduced number of generated edges. Eight-digit precision was chosen because 10-digit precision had an average chance of less than 80% of a location sample falling within its area (as shown earlier in Table 7.3), with the odds decreasing as vehicle speeds increased. This meant that multiple small roads could fall within the same plus code. This is mostly a problem in urban areas due to the density of the roads, which is evident when in close proximity to the hospitals. Figure 7.10 highlights this problem by showing the edges generated in Leicester city centre with multiple routes converging towards LRI. Neonatal care at LRI occurs in the building at the centre of the orange circle on the map. The network algorithm resulted in the blue edge encompassing the hospital grounds. In reality, there were three routes around the neonatal building, which suggested a smaller area plus code is required.

The relatively low resolution plus codes also caused the handling of the multitons (Section 7.3.2.4) to yield inaccurate results. The checks mainly worked as expected in rural areas, where the generated multitons tended to be parallel to long edges, but could not cope with the intricate roads in urban environments. This is visible when examining the green and orange edges in Figure 7.10. It is clear from looking at the pale orange roads on the map that journeys which travelled along the green and orange edges have the two left-most orange plus codes in common. These two plus codes cover a very small section of road, however they were absorbed into the rest of the orange edge because of the singleton threshold. More plus codes would cover this road section if the resolution was increased, which therefore should result in it becoming separated.

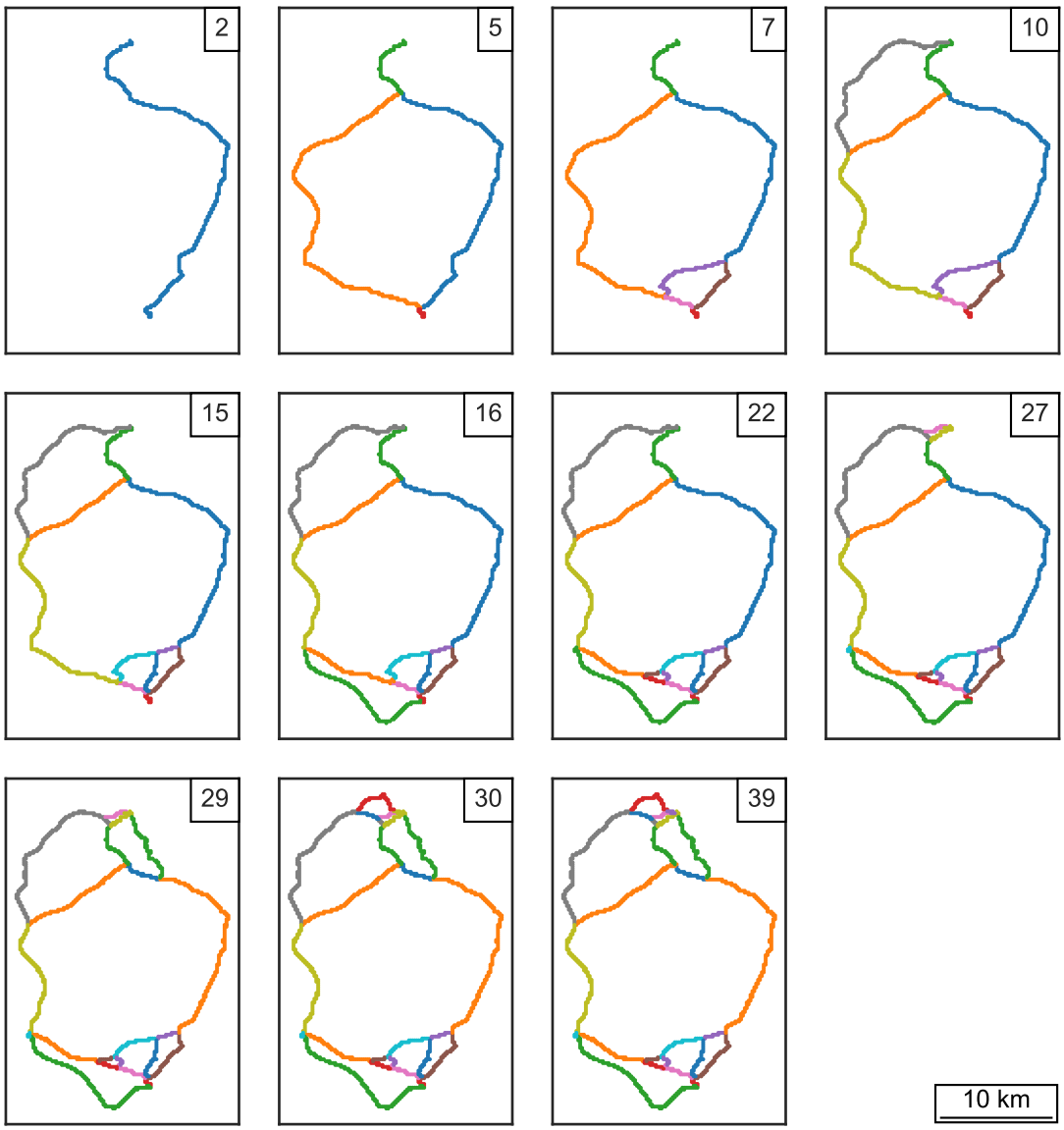

Figure 7.9: Development of the network using 8-digit plus codes with a multiton size of 4 (number of journeys processed shown in the top right corner).

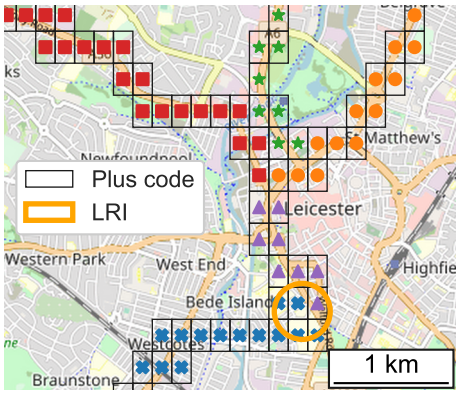

Figure 7.10: Leicester city centre overlaid with the edges found using 8-digit plus codes (© OpenStreetMap contributors).

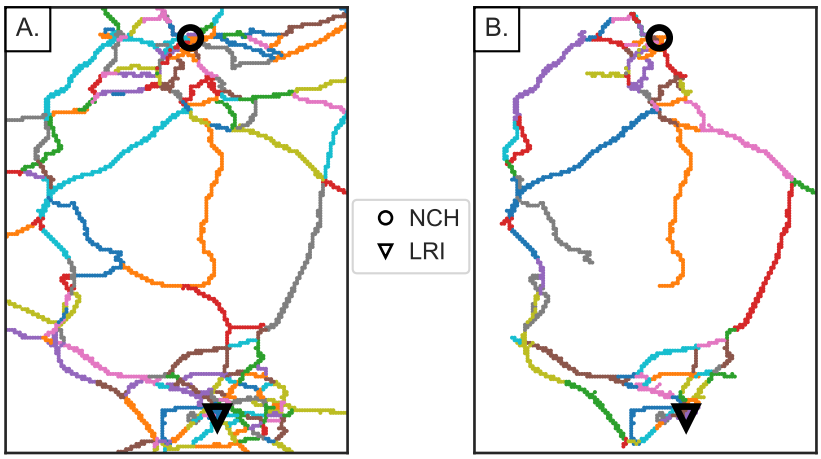

Figure 7.11: All edges, computed using 8-digit precision plus codes and a multiton threshold of 4, in the area surrounding the NCH to LRI routes (A) and only those which contain plus codes found in the NCH to LRI routes (B) where incorrect edges are evident from roads travelling to dead ends.

7.5.3 Full Network Edges

Before trying to improve the plus code resolution, it was worth looking at the edges computed using the full set of recordings as this would be the resultant network when splitting the roads dynamically. Visually inspecting the area surrounding the NCH to LRI journeys (Figure 7.11A) it can be seen that there are a large number (632) of edges which cover a large proportion of the map. As with the NCH to LRI route network, the edges became more condensed as the roads got closer to urban areas. This was in part due to multiple neonatal facilities being present, with Nottingham having a second hospital (Queen's Medical Centre) and Leicester having two others hospitals in close proximity (Leicester General Hospital and Glenfield Hospital). Journeys either leaving or accessing these hospitals could introduce further splits to roads both due to the different locations and the presence of one-way systems. Directionality of roads is also problematic because edges would need to be checked for directionality for extracting and processing.

Removing all network components apart from those which had at least one plus code in common with the NCH to LRI recordings resulted in a slightly more manageable 89 edges (Figure 7.11B). This network was 44 short of the true number of splits identified in Section 7.4, although was better than 632.

One thing which was more clearly evident when using the full collected dataset, rather than solely using the NCH to LRI recordings, was the presence of incorrectly split roads which were caused by the creation and handling of multitons. The inclusion of recordings which travelled opposite to the direction of interest could cause multitons to be

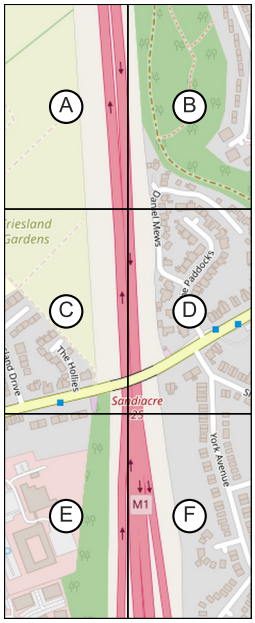

Figure 7.12: Example of how journeys travelling opposite directions along the same road could result in different plus codes. Each grid section identifies a separate 8-digit plus code.

created as the road width could cause the lanes to be inside different plus codes. This can be understood by examining Figure 7.12, where the southbound lane of the M1 goes through the plus codes represented by A, D & F, while a northbound slip road falls into E and the northbound lane itself resides in C. Additionally, journeys which used other roads may join the NCH to LRI routes at new junctions and therefore should result in split edges, however they could also produce multitons around the decision points. Multitons are currently combined with the first edge which is identified to contain a neighbouring plus code, however some method of deciding on the most suitable edge may result in a better network.

Simple dynamic network generation using plus codes did not appear possible due to the many complications of including every recording. There were many things which would need consideration and result in the algorithm becoming more and more convoluted. While dynamic networks may not be feasible, reducing the size of the plus codes may enable network generation when focusing on specific routes.

7.6 Dividing plus codes into smaller divisions

Reducing the size of the plus codes may improve network generation, however the next available resolution is too small. As shown in Section 7.3.1, 8-digit plus codes contain an average of 6.5 location samples per journey at 70 mph compared to only 0.3 samples

Table 7.6: Area identifiers for 4 (A), 9 (B) and 16 (C) extra divisions.

A.

2	3
0	1

B.

6	7	8
3	4	5
0	1	2

C.

12	13	14	15
8	9	10	11
4	5	6	7
0	1	2	3

Table 7.7: Resolutions of the plus codes with extra divisions, compared with the standard plus codes.

Code Length	Divisions	Resolution(°)	Resolution (m)
8	0	1/400	278
	4	1/800	139
	9	1/1,200	93
	16	1/1,600	70
10	0	1/8,000	14

at 10-digit resolution. One possibility was to create an intermediary resolution which would divide up the 8-digit plus codes.

Plus code precisions from 2 to 10 digits divide up the Earth's surface into smaller and smaller 20×20 grids. Introducing a method of creating a variable number of divisions could result in an optimum area size for network generation. The simplest way of dividing up an area would be to use square numbers of divisions, with each division represented by an integer (Table 7.6). This was built upon the standard open location code algorithm and employed similar practices.

Encoding plus codes with extra divisions first required the grid size to be computed, as this varied with the number of divisions. The length of each side of the grid would be the root of the prescribed divisions when using square numbers. Resolution of the new grid was then calculated by dividing the resolution of the standard plus code precision being used by this length (Table 7.7).

For each pair of latitude and longitude coordinates the standard plus code was computed to act as a base. The bottom left (southwest) corner of this plus code was then decoded as a datum for the new grid. Latitude and longitude offsets within the grid were calculated by subtracting the true coordinates by the datum coordinates, before being divided through by the previously computed resolution to attain rough X & Y positions. The final X & Y positions were found by taking the minimum of the rough position rounded down to the next integer and the grid length minus one, to ensure

positions fell within the true grid (Equation 7.6.1). These were combined as in Equation 7.6.2 to compute the grid ID which was then appended to the standard plus code. For example, the centre of the NCH campus was identified earlier as '9C4WXRRV+' when using 8-digit plus codes. Increasing the resolution using 9 extra divisions, this would become '9C4WXRRV+4', comprised of the 8-digit code and the numeric identifier for the centre of a 9 division grid ('4').

$$\text{position} = \min\left(\text{floor}\left(\frac{\text{position}}{\text{grid resolution}}\right),\ \text{grid length} - 1\right) \tag{7.6.1}$$

$$\text{grid ID} = \text{position}_X + (\text{grid length} \cdot \text{position}_Y) \tag{7.6.2}$$

7.6.1 Incrementing to find neighbouring codes

Identifying neighbouring locations was fundamental to the network analysis described in Section 7.3.2. Therefore, after encoding all locations to be in the form of a plus code with extra divisions, a method of finding the neighbours to these new-format codes was required. Neighbouring codes were found in a similar method to that of the standard plus codes as described in Section 7.2.1. First, the code to be incremented needed to be split into the standard plus code and grid ID, before then checking if the ID was at a grid boundary.

Boundary checks altered the method of incrementing the code as when the ID was at a boundary the standard plus code would also require incrementing. Methods of checking the boundary condition differ for each of the cardinal directions. The difference in ID values between the top and bottom rows (δh) was calculated as in Equation 7.6.3 before the grid ID was checked against it to assess if it could be incremented northwards (Equation 7.6.4). While the ID being in the top row of the grid meant that it was not at the south boundary, not being in the top row did not guarantee the ID is in the bottom row unless 4 divisions were being used. To identify whether the ID was in the bottom row, it needed to be checked against the grid length (Equation 7.6.5). Longitudinal boundaries were assessed by checking if the modulo of the grid ID (west) or (grid ID + 1) (east) divided by the grid length had a remainder greater than zero (Equations 7.6.6 and 7.6.7). These boundary verdicts were then passed on to the incrementing functions.

$$\delta h = \text{divisions} - \text{grid length} \tag{7.6.3}$$

$$\text{boundary}_{\text{NORTH}} = \text{grid ID} \geq \delta h \tag{7.6.4}$$

$$\text{boundary}_{\text{SOUTH}} = \text{grid ID} < \text{grid length} \tag{7.6.5}$$

$$\text{boundary}_{\text{WEST}} = \text{mod}\left(\text{grid ID},\ \text{grid length}\right) > 0 \tag{7.6.6}$$

$$\text{boundary}_{\text{EAST}} = \text{mod}\left(\left(\text{grid ID} + 1\right),\ \text{grid length}\right) > 0 \tag{7.6.7}$$

Incrementing the codes required adding or subtracting values to/from the grid ID, along with the plus code incrementing when the boundary condition was met. Latitudinally, the grid length was either added to (north) or subtracted from (south) the grid ID when the boundary condition was false. When the boundary conditions were met the plus code got incremented and then the arithmetic signs flipped to result in δh either being subtracted from (north) or added to (south) the grid ID.

Similarly, incrementing longitudinally required a single unit to be combined with the grid ID along with the grid length. In all cases, a value of 1 was either added to (east) or subtracted from (west) the grid ID. When the boundary condition was false, these were the final IDs, however when the boundary condition was met then the grid length needed to be either subtracted (east) or added (west) as well as the base plus code being incremented.

7.6.2 Interpolating location samples

Due to the small area of the divided plus codes the decision was made to interpolate the location samples. It was thought that increasing resolution might result in spatial gaps between subsequent codes as vehicles travelling on fast rural roads could travel over the defined area size in less than the time between location samples. The problem was actually found to occur in urban areas before gaps were visible on motorways. Several of the recordings which travelled from NCH to LRI resulted in a gap in Leicester city centre, as shown in Figure 7.13A, which was caused by an underpass. Gaps were minimised by linearly interpolating location data for every IMU sample, resulting in an effective 200 Hz location sampling, before encoding. If there was a long period without a location sample, linear interpolation could potentially bisect roads and cause an incorrect plus code to be identified. However, with the NCH to LRI dataset this method successfully ensured the codes for each recording were continuous (Figure 7.13B), although the odds of discontinuity remaining increase with further plus code resolution.

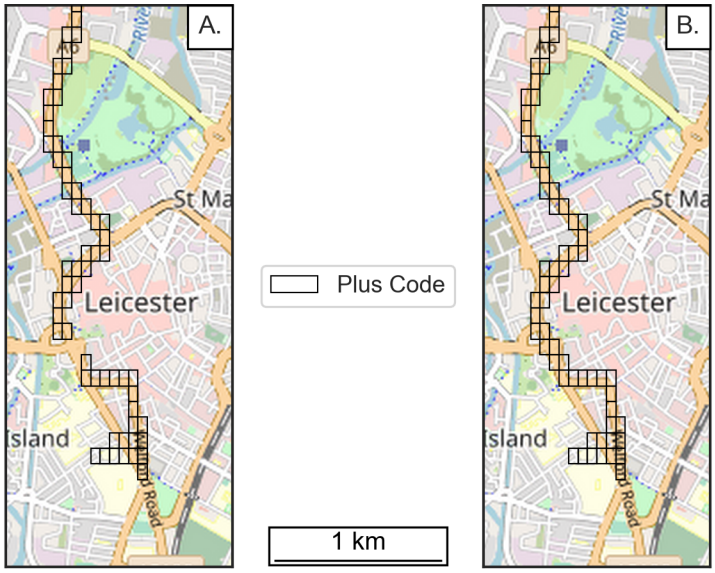

Figure 7.13: 8-digit plus codes with 9 divisions for a single recording inside Leicester city centre using raw (A) and interpolated (B) location samples (© OpenStreetMap contributors).

7.6.3 Results using 8-digit plus codes with extra divisions

Three versions of the NCH to LRI network were generated using 8-digit plus code with an extra 4, 9 and 16 divisions, as defined in Table 7.6, and a multiton threshold of 4. Extra divisions were used after problems were found with the base 8-digit plus codes when examining the splits around Leicester city centre in Figure 7.10. While the use of 4 extra divisions clearly increased the resolution there was ultimately no change in the generated edges (Figure 7.14A), with roads not being split where required. Increasing the number of divisions to 9 and 16 (Figure 7.14B & C) successfully separated the orange edge, however both resulted in additional incorrect splits. These incorrect splits appeared to be due to locations occurring around the edge of the plus code area and thus a small fluctuation creating a multiton.

Although some of the erroneous splits could be absorbed by increasing the multiton threshold this would then lose some true splits, and the intricate roads around LRI were still not identified. This can be seen in Figure 7.15A where a multiton threshold of 5 results in the surrounding edges merging while the blue and brown edges remain incorrectly split. Further increasing the multiton threshold (Figure 7.15B & C) still did not correct the incorrect splits and, in the case of a threshold of 7, resulted in the grey edge being wrongly absorbed into the orange. Ultimately, the non-linear, complex nature of road networks were not conducive with simple planar networks and a more intensive process was required.

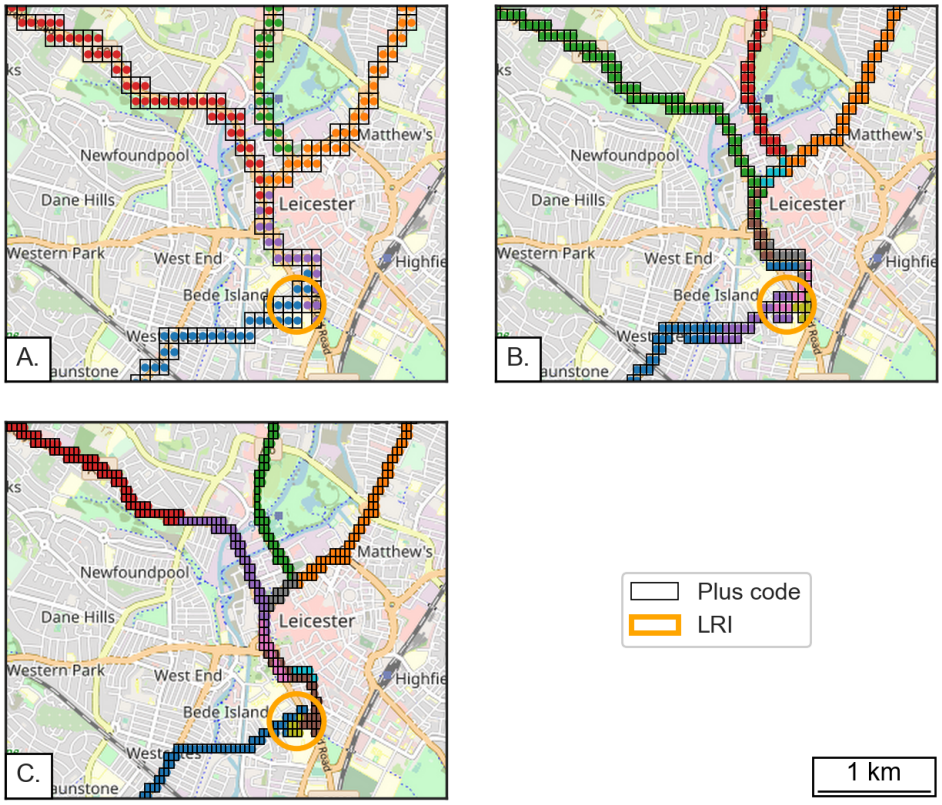

Figure 7.14: Comparison of the edges around Leicester city centre as generated with 4 (A), 9 (B) and 16 (C) divisions and a multiton threshold of 4. Each separate colour indicates a unique edge to the network (© OpenStreetMap contributors).

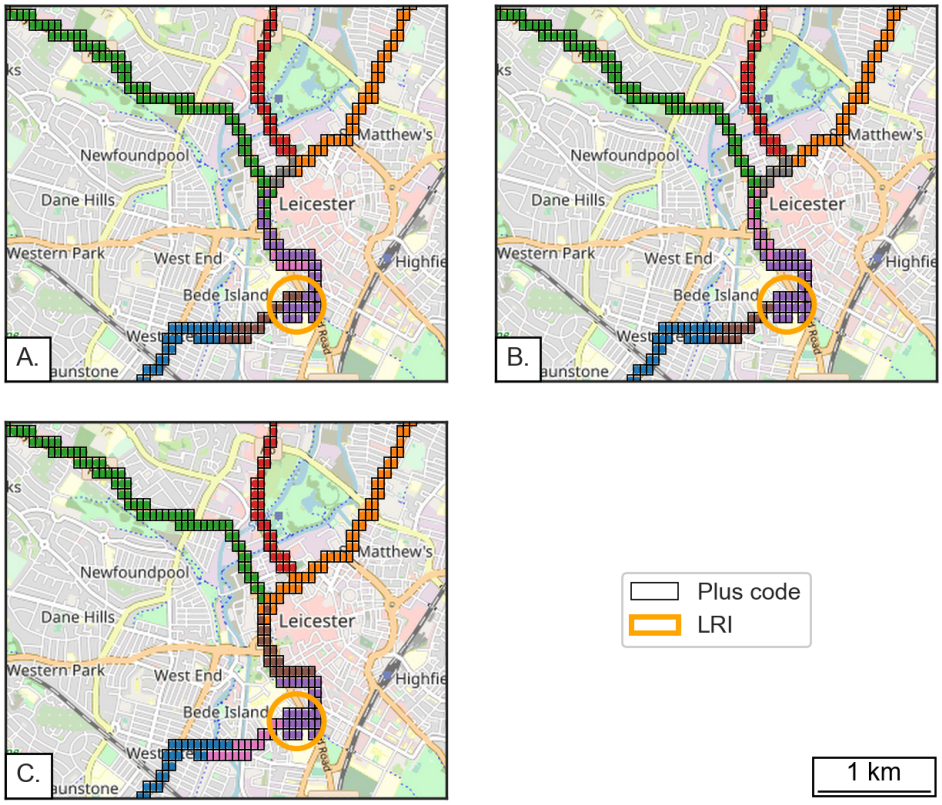

Figure 7.15: Comparison of the edges around Leicester city centre as generated with 9 divisions and a maximum multiton threshold of 5 (A), 6 (B) and 7 (C). Each separate colour indicates an unique edge to the network (© OpenStreetMap contributors).

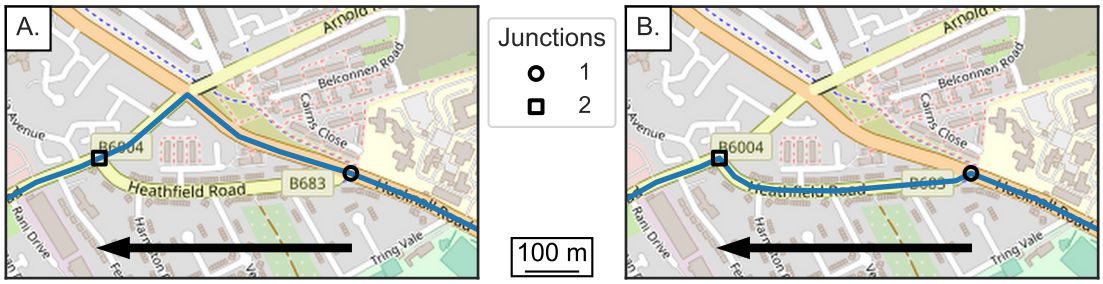

Figure 7.16: Example of road junctions which were used during some of the recordings travelling from NCH to LRI, with the arrow indicating the direction of travel (© OpenStreetMap contributors).

7.7 Manual identification of decision locations

Automatic road network analysis could not distinguish small edges while keeping large edges whole, mainly due to the density of urban roads. The roads in close proximity to cities tend to have many junctions within short distances of one another, which would result in the formation of multitons. Separating roads into unique sections should therefore be conducted by hand, considering how planar networks cannot comprehend two roads crossing over each other (but not physically intersecting) or single direction roads.

Through consideration and further visual inspection it was realised that decision points are often not at the location where routes diverge — the principle used with the plus code algorithm — as drivers prepare for a junction before it is reached. For example, a driver may stay at a steady speed when continuing straight or slow down in order to take a junction. This is best understood by looking at the routes and junctions in Figure 7.16. Although the indicator for junction 1 is where two roads meet, route A will continue straight at the speed limit while route B would be required to slow down to take the turning. The actual split point for separating these two roads should be where a driver would start to slow down for the junction, which would be a reasonable distance before. Similarly, although the routes meet at junction 2, vehicles which travel along route A should be up to speed whereas vehicles along route B would join the junction from a standstill. The true split location would actually be some distance after the junction to account for route A vehicles accelerating. It was therefore imperative that routes are examined closely and the splits were chosen accurately so as to ensure analysis would reflect the true routes.

7.7.1 Generation of NCH to LRI network

The first network to be created manually involved the 39 journeys which travelled from NCH to LRI. This would involve close examination of the roads used, identification of the decision points (or nodes) and then the generation of the edges which connected these nodes. By creating the network manually, there should be zero incorrectly split or combined sections of road.

First, the lines of coordinates from each of the recordings were plotted concurrently over a map to enable visual analysis of decision points. Locations at which routes diverged or converged could then be identified by inspecting all overlaps. Each location was examined in depth using a combination of Google Maps and Street View. These online tools enabled precise identification of the coordinates at which the routes could reasonably be judged to be split.

A directional graph was created — using the 'NetworkX' package [163] — using each of the decision locations as nodes and creating edges between nodes which belonged to the same route. This resulted in 27 nodes, of which 3 nodes were solely created because of a driver excursion on one journey. Decision points were placed at the entrance and exit of a motorway slip road to enable this diversion to and from a service station, also included as a decision point, to be excluded. The network of interest would therefore contain 26 nodes and 36 edges, which would effectively be 24 nodes and 34 edges, when the service station node has been removed. These edges could be combined to form 31 different routes, over twice the number of road combinations used.

Recordings were assigned to the edges of the network by comparing journey coordinates to those of the nodes. Nodes were checked in order of travel on a journey by journey basis, starting with the NCH node (*start_node*). Proximity to the node coordinates was checked using the haversine formula (Equation 3.3.1) until a pair of recorded coordinates were within 25 metres and their corresponding index value was logged as *start_index*. Analysis then progressed to the adjacent nodes (*check_nodes*) and their coordinates (*check_coords*).

Iterating through journey coordinates continued while checking against the new set of *check_nodes*. The only difference in the process was that each coordinate needed to be compared to multiple coordinates. This continued until the journey was within one set of *check_coords*, at which point the values from the *start_index* up to the current index were assigned to the edge between the *start_node* and the corresponding node in *check_nodes*. The *start_index* and *start_node* were then updated along with the next set of nodes and coordinates stored in *check_nodes* and *check_coords* respectively. The

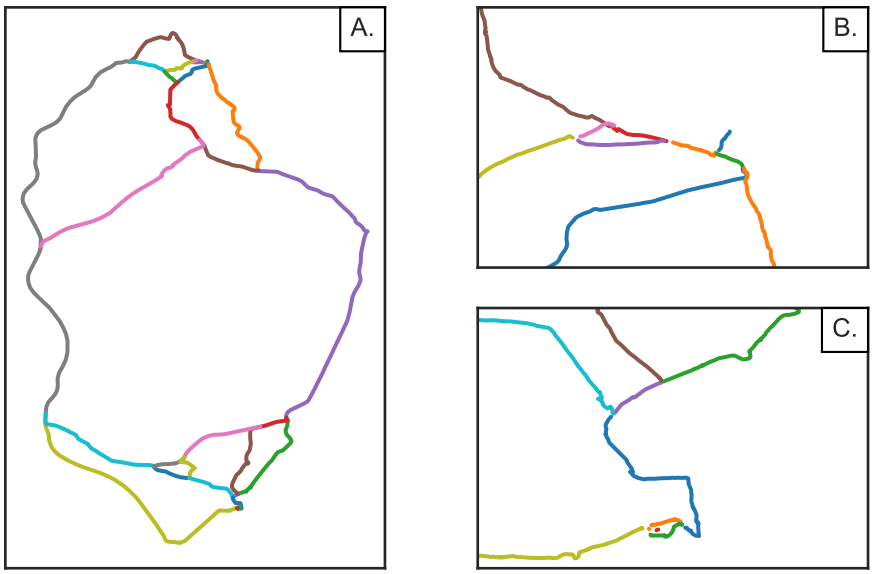

Figure 7.17: Plots of the manually identified network of the NCH to LRI routes showing the full area (A), roads surrounding NCH (B), and roads surrounding LRI (C).

process of checking journey coordinates against *check_coords* continued until either the *start_node* was LRI or the coordinates ran out, at which point the journey was completed.

Coordinates of each of the NCH to LRI edges are plotted in Figure 7.17A. The increased intricacy of the manual edges can be seen by focusing on the areas surrounding both hospitals, in subplots B & C.

The use of manually-identified decision points with the NCH to LRI journey routes provided an accurate means of separating coordinates into unique stretches of road. This should result in precise assessments of the comfort of routes due to the data for each edge was collected along the same stretches of road and will therefore be highly comparable when aggregating. However, before any analysis of routes could be conducted, it was important to be able to include in the computations all other recorded data which travelled along each of the roads of interest.

7.7.2 Inclusion of full set of recordings

In order to increase the confidence of the future metric computation, analysis of the NCH to LRI routes was expanded to all recordings which traversed the roads following a similar process to the simple visual analysis described above. First, the NCH to LRI roads were plotted over the map using thick black lines before the coordinates of all surrounding recordings were overlaid. Having the simple network in bold enabled

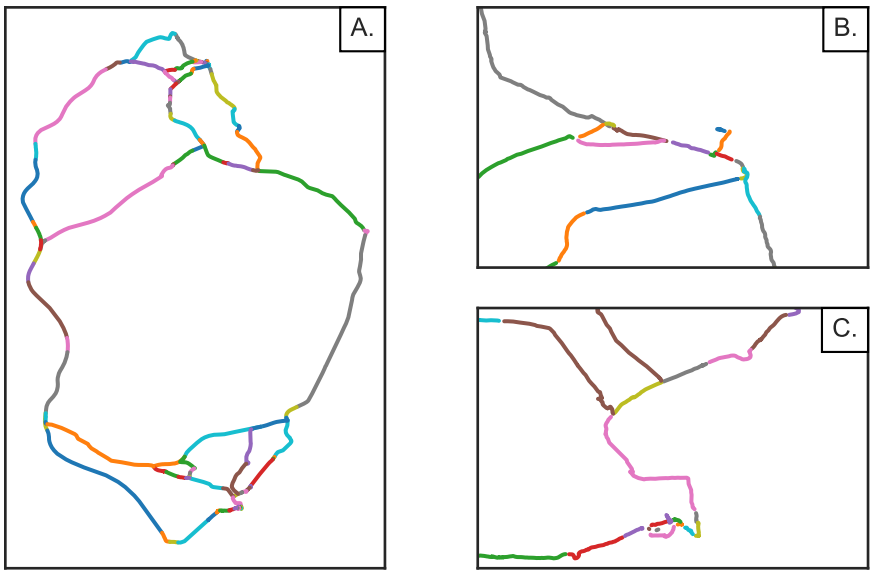

Figure 7.18: Plots of the manually identified network of the full split NCH to LRI roads showing the full area (A), roads surrounding NCH (B), and roads surrounding LRI (C).

visual analysis of the new splits to focus solely on the roads of interest, identifying locations where new roads converged, diverged or overlapped.

Coordinates of each new split were found and combined to create an expanded directional graph, with the majority of the previous edges being further segmented. A total of 114 nodes were identified where routes differed, which was a fourfold increase on the simple network. Nine of these locations were not on the roads of interest, but were required to exclude recordings which departed from the routes before rejoining again. Three nodes were required to exclude each unwanted location: one on the route before the journey departed, one which identified the unwanted location itself, and one back on the route where the journey rejoined. Including all journeys from the surrounding area resulted in the NCH to LRI routes being split into over three times as many unique sections (115 vs 34, excluding the edges in the graph used for exclusion purposes), as highlighted in Figure 7.18.

Manually dividing up the road network enabled all possible splits within roads to be identified, including those which needed to be included for exclusion purposes. A further benefit was the ability to ignore any presumed decision points where, on closer examination using photographic data, it was only relevant to traffic travelling in the opposite direction to that of interest. Although time consuming, and not able to be conducted until data collection was completed, the generation of the network by manual means would enable the assessment of routes to be conducted with the greatest amount of data, and therefore with the highest levels of confidence in the aggregated

metric values.

7.8 Final network

The ultimate aim of this chapter was to produce a road network with which roads can be assessed, in terms of comfort and health, to see if route choice may affect the outcomes of transferred neonatal infants. First, an algorithm was devised with the intention of being able to split up roads — by the decision points used by drivers — programmatically, enabling the network to be updated as new data became available. This algorithm had shortcomings which led to the manual separation of roads, of which an overview of the method and the resultant network was provided. This network will now be detailed more finely to act as a reference for the subsequent route analysis.

Roads were divided up based on the coordinates from 39 journeys which travelled directly from NCH to LRI between 24th October 2018 and 14th October 2019. Twenty-four nodes were identified where these routes diverged from one another. Each node in the network was assigned a unique numeric ID, starting at 0 for NCH and increasing to 23 for LRI. These IDs are mapped out in Figure 7.19 and also given in Table 7.8, along with their corresponding coordinates to 4 d.p., to provide a full understanding as to their locations. The proximity of the nodes around the hospitals can be seen to be orders of magnitude greater than in the rural areas, showing how complex a splitting algorithm would be required to work perfectly.

The purpose of the network was to provide edges along which there was no possible deviation in route. Thirty-three such edges were identified during the manual network generation, which correspond to the nodes as shown in Table 7.9. It is the data collected along each of these edges which are converted into metric values, and therefore it is imperative that the result can be linked back to the relevant road.

Five-hundred and fifty further journeys recorded inside the neonatal ambulances were found to travel over parts of the identified NCH to LRI network of roads. Due to the different start and end destinations of these journeys, additional junctions along the routes were taken. This resulted in the 34 edges of interest being broken down to 115 smaller edges. Analysis will only be shown for the each of the 34 original edges from which different routes can be formed, however the metric values for these will be combined from each constituent edge of the 115.

Ultimately, it is the potential difference in metric value between routes which is of interest. The NCH to LRI edges can combine to form 31 distinct routes. Each route

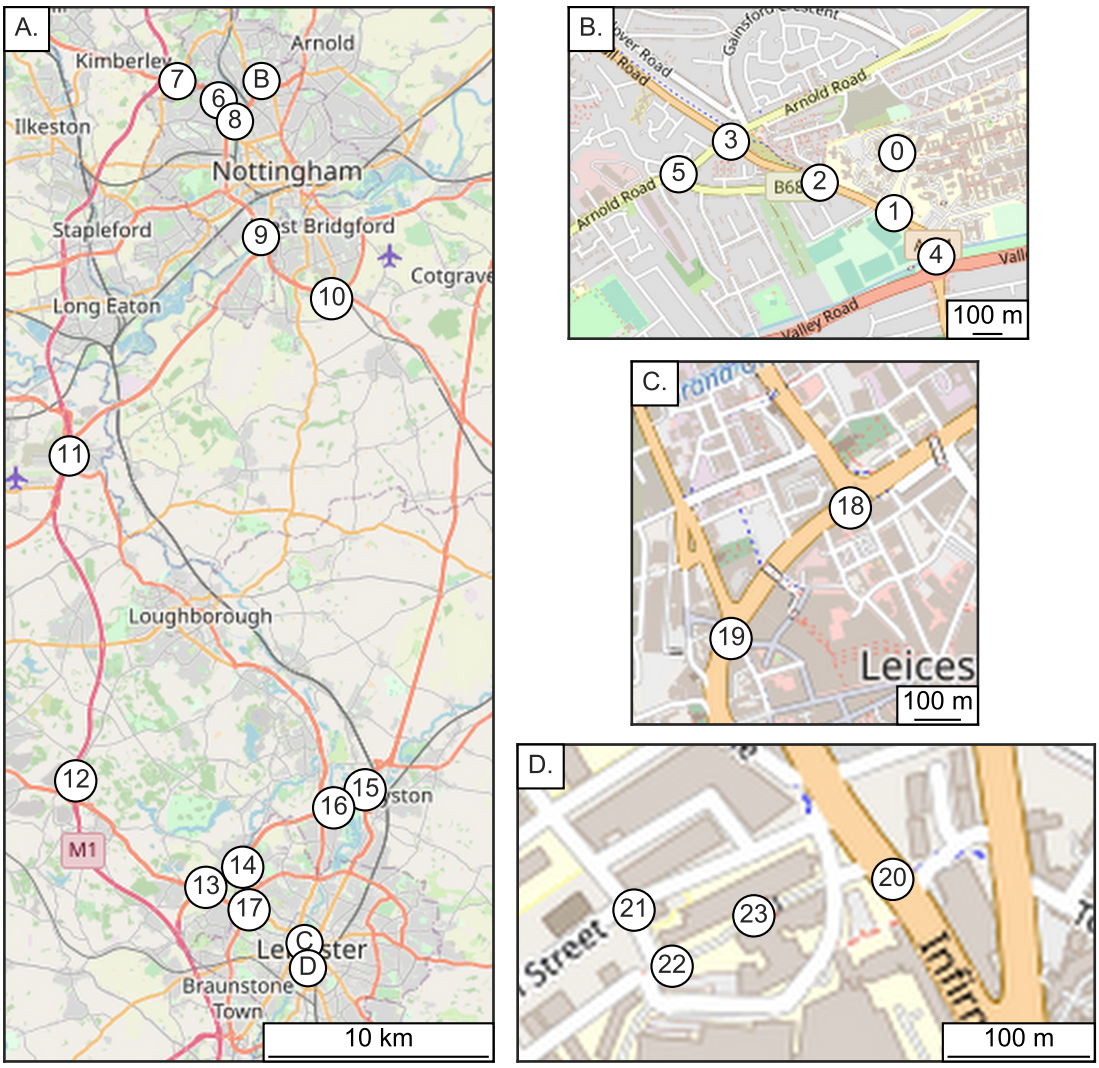

Figure 7.19: Locations of the final nodes (route splits) of the NCH to LRI route network, with enhanced resolution of densely-packed nodes around NCH (B), Leicester city centre (C) and LRI (D) (© OpenStreetMap contributors).

Table 7.8: Latitude and longitude coordinates for each of the final nodes of the NCH to LRI route network, rounded to 4 significant figures.

Node ID	Latitude	Longitude	Node ID	Latitude	Longitude
0	52.9906	-1.1615	12	52.7048	-1.2919
1	52.9888	-1.1617	13	52.6613	-1.2041
2	52.9897	-1.1654	14	52.6695	-1.1792
3	52.9909	-1.1699	15	52.7012	-1.0966
4	52.9875	-1.1596	16	52.6938	-1.1178
5	52.9900	-1.1725	17	52.6521	-1.1749
6	52.9822	-1.1954	18	52.6392	-1.1361
7	52.9899	-1.2233	19	52.6367	-1.1398
8	52.9736	-1.1844	20	52.6283	-1.1343
9	52.9266	-1.1669	21	52.6282	-1.1370
10	52.9015	-1.1192	22	52.6278	-1.1366
11	52.8374	-1.2963	23	52.6281	-1.1357

Table 7.9: Start and end nodes for each edge of the NCH to LRI network.

Edge ID	Nodes	Edge ID	Nodes	Edge ID	Nodes
0	0-1	12	8-6	24	18-19
1	1-2	13	8-9	25	16-18
2	1-4	14	10-15	26	16-14
3	2-3	15	9-10	27	14-13
4	2-5	16	9-11	28	14-17
5	3-7	17	11-12	29	17-19
6	3-5	18	12-21	30	19-20
7	7-11	19	12-13	31	20-21
8	5-6	20	13-17	32	20-22
9	6-7	21	21-22	33	22-23
10	4-8	22	15-18		
11	4-10	23	15-16		

was assigned an alphabetical identifier, as outlined in Table 7.10, with which to use in the results comparison. It would then be possible to identify exactly which roads were travelled along on the map to give each metric value.

After generating a network and defining both the edges and the possible routes, analysis of the discomfort is needed. It is with this analysis that the potential for improving outcomes by directing ambulances along different roads can be assessed.

Table 7.10: Edge combinations for the possible routes between Nottingham City Hospital and Leicester Royal Infirmary as identified in the final NCH to LRI network.

Route ID	Combination of Edges
A	0-1-3-5-7-17-18-21-33
B	0-1-3-5-7-17-19-20-31-21-33
C	0-1-3-5-7-17-19-20-32-33
D	0-1-3-6-9-7-17-18-21-33
E	0-1-3-6-9-7-17-19-20-31-21-33
F	0-1-3-6-9-7-17-19-20-32-33
G	0-1-4-8-9-7-17-18-21-33
H	0-1-4-8-9-7-17-19-20-31-21-33
I	0-1-4-8-9-7-17-19-20-32-33
J	0-2-10-12-9-7-17-18-21-33
K	0-2-10-12-9-7-17-19-20-31-21-33
L	0-2-10-12-9-7-17-19-20-32-33-
M	0-2-10-13-15-14-22-24-30-31-21-33
N	0-2-10-13-15-14-22-24-30-32-33
O	0-2-10-13-15-14-22-23-25-24-30-31-21-33
P	0-2-10-13-15-14-22-23-25-24-30-32-33
Q	0-2-10-13-15-14-23-26-27-20-29-30-31-21-33
R	0-2-10-13-15-14-23-26-27-20-29-30-32-33
S	0-2-10-13-15-14-23-26-28-29-30-31-21-33
T	0-2-10-13-15-14-23-26-28-29-30-32-33
U	0-2-10-13-16-17-18-21-33
V	0-2-10-13-16-17-19-20-29-30-31-21-33
W	0-2-10-13-16-17-19-20-29-30-32-33
X	0-2-11-14-22-24-30-31-21-33
Y	0-2-11-14-22-24-30-32-33
Z	0-2-11-14-23-25-24-30-31-21-33
AA	0-2-11-14-23-25-24-30-32-33
BB	0-2-11-14-23-26-27-29-30-31-21-33
CC	0-2-11-14-23-26-27-29-30-32-33
DD	0-2-11-14-23-26-28-29-30-31-21-33
EE	0-2-11-14-23-26-28-29-30-32-33

CHAPTER 8

Improving Comfort of Ambulance Journeys through Routing

8.1 Introduction

Transferring neonatal infants by ambulance within the first 28 days of birth has been shown to lead to worse outcomes compared to those born at the required facilities [22–24]. Multiple papers have found the presence of both excessive noise [73] and excessive vibration in relation to adult comfort [43, 44, 47] during these transfers. One study investigated the noise and vibration inside ambulances at the same time as measuring the biological effect, but the results were largely inconclusive due to the low population involved [40]. Despite the lack of evidence, it is still believed that the noise and vibration levels could be a contributing factor in the outcomes.

Ambulance drivers typically follow routes which are determined by satellite navigation systems, along with their local knowledge (from personal communication with the drivers for CenTre (CenTre Neonatal Transport)). These systems are programmed to output the fastest route by default, which may be over rougher or noisier roads. It is plausible that driving along a different set of roads could yield lower levels of vibration and/or noise. Driving along a more comfortable route may then improve outcomes for transfers. Determining which roads will reduce discomfort requires comfort to be assessed for multiple routes to the same destination.

Data on the in-ambulance environment during transfers are required to be able to assess route comfort. Previous chapters have shown the development (Chapter 3), validation (Chapter 4) and use (Chapter 5) of a smartphone application for recording acceleration, noise and location data on a neonatal transport trolley. Twelve months of recording data during ambulance journeys conducted by CenTre resulted in 1,487 dif-

ferent journey recordings, after cleansing the data (Chapter 6), which travelled almost 82,000 km. With data collected, the next step was to compute comfort levels.

Roads between Nottingham City Hospital (NCH) and Leicester Royal Infirmary (LRI) were divided into a network of unique sections in Chapter 7, within which all journeys took the same route from beginning to end. Identifying unique sections, or edges, is vital to the computation of comfort as it enables the largest amount of data to assess each stretch of road. Comfort values can be calculated using the data recorded over the roads on a journey-by-journey basis for each edge. The values from each journey can then be averaged to attain the edge comfort, with route comforts being the combination of the comfort of each constituent edge. By calculating the comfort levels of different routes between the same hospitals the existence of a most comfortable route can be determined.

This chapter will examine the comfort computation for the different road combinations travelling from NCH to LRI. First, the identification and then extraction of the relevant data was required before any assessment could be performed. A range of metrics will then be defined, along with the methods for calculating values on a journey, edge and, finally, route scale. Feasibility of improving comfort levels during inter-hospital transfers will then be assessed by comparing the calculated comfort of each edge and how these then combine for each route.

8.2 Identification of route data

Knowledge on the data required for analysis was required before any extraction and processing could be conducted. The routes travelling from NCH to LRI and the constituent edges were identified in Section 7.7 to create a network. This network needed to be passed back to the recordings to pinpoint which data samples belonged in each edge, for each journey.

The use of unique identifiers for each edge, along with the choice of database, would facilitate quick extraction of relevant data by searching for the desired ID. This required analysis of which edges occurred in which journeys so that the edge IDs could be assigned appropriately. Performing this analysis was only required once and then resulted in speedier extraction by minimising the amount of data. The method used for determining the presence of any edges of the NCH to LRI network inside a single recorded journey is set out below.

8.2.1 Attributing journey data to network edges

Coordinates of each node were checked against the first latitude and longitude pair of the journey to identify a starting point. A node qualified as 'matching' the journey coordinates if the haversine distance (Equation 3.3.1) between them was within 25 metres. Nodes were checked in turn, starting at NCH and working along the routes towards LRI to ensure journeys which did not start at NCH, but still used parts of the routes of interest, were included. Successive pairs of journey coordinates were checked until one matches a node.

Upon matching with the coordinates of a node, the row of the journey pair was logged and a new variable (*check_nodes*) was created to contain the nodes which succeed. For example, if the matched node was at the entrance to the NCH campus, *check_nodes* would contain the nodes which are found by driving either to the right (north) or to the left (south). Successive journey coordinates were then compared to the nodes in *check_nodes* until a new match was found. When coordinates matched with one of *check_nodes* the ID for this edge was inserted for all rows from the original match up to the preceding timestamp to this new match. The process continued through the network with updated *check_nodes* until either LRI or the end of the journey was reached.

All matched data were then reinserted into a new database with their assigned edge IDs for future extraction. Raw and ISO-weighted values with the timestamps of the matched data were then extracted and reinserted into the new database with the edge IDs ready to be analysed.

8.2.2 Interpolation of recorded coordinates

Location data was the only way of identifying the segments in a journey because an edge was only defined by the start and end coordinates. Matching parts of a journey to an edge using the raw 1 Hz sampled location data could have resulted in both data incorrectly being missed from an edge and data being incorrectly attributed. This can be shown by the simplified illustration in Figure 8.1, where the threshold for matching recorded coordinates to an edge occurs between two blue location samples and therefore several samples of data from the Inertial Measurement Unit (IMU) — which possessed the greatest sampling rate of all sensors accessed — would not be included in the analysis. Linearly interpolating the coordinates with respect to the number of IMU samples would ensure that all data which was on the roads of interest was attributed to the edge, assuming a constant vehicle speed between location samples.

The maximum amount of time which could be incorrectly missed, or attributed, from

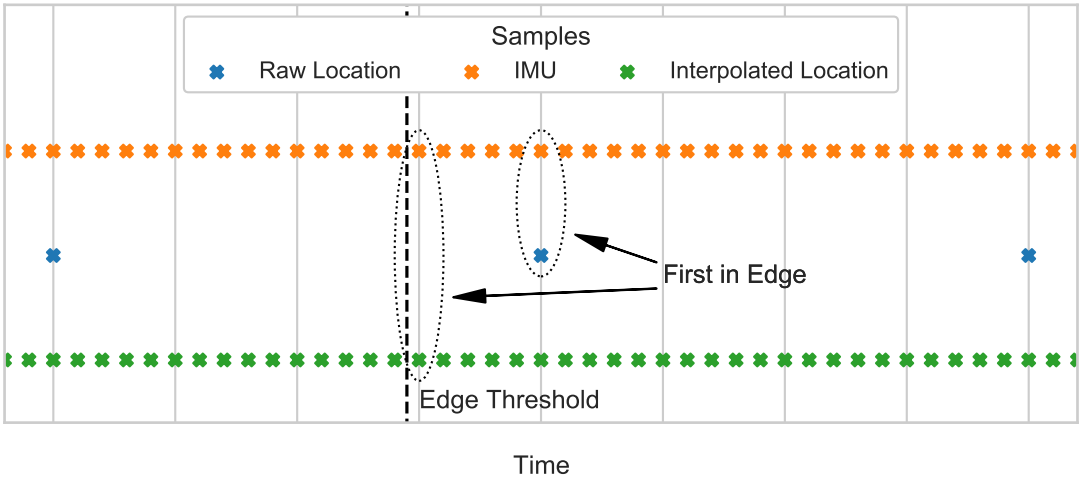

Figure 8.1: Illustration of the method and benefit of interpolating location samples with respect to samples from the IMU.

an edge was around 1 second when using the sampled location data due to this being the rate at which the satellite receiver was accessed. Some segments were shorter than 50 metres due to the complexity of urban roads, especially surrounding LRI. Almost 9 metres worth of data could then be affected around these roads which predominantly impose a 20 mph speed limit. Although this is not much compared to a full route, it is over 75% of the shortest edge and could significantly alter the analysis. By contrast, interpolating the location data would result in an effective rate of 200 Hz, reducing the amount of data affected to less than 50 mm or less than 0.4% of the shortest edge.

8.3 Extraction of data for edge analysis

The steps outlined below explain the procedure of extracting and preparing the data for a single edge of the network for comfort analysis.

8.3.1 Extraction of location data

All location fields were first extracted from the database by selecting the specific edge identifier. These were grouped by recording to ensure that each journey stayed separated. Analysis of the edge comfort could then be conducted by processing each journey in turn.

8.3.2 Calculation of distance travelled

Cumulative distance which was travelled between location data was calculated using the haversine method (Equation 3.3.1). Calculating the distance travelled enabled a check to see if all journeys covered the same amount of road. Comparing the distances travelled with that provided by Google Maps routing algorithms also provided a confirmation that the data had been correctly split up and extracted.

8.3.3 Alignment of journeys

Journeys within an edge could either be compared by generating a single summary value or by resampling each dataset in accordance with distance, as described in Section 3.3.1. Comparing journeys along the edge could aid the inspection of differences as summary values often do not tell the whole story. Coordinates therefore needed to be aligned between journeys to ensure that the same distance was reached at the same location along the edge. Once aligned, the waveforms of the acceleration should look similar for each journey, for example. This could also help the comparison with Maps as aligning the journeys can shift the calculated total distance.

The process of alignment was skipped when only a single journey had been recorded along the edge as there were no comparisons to be made.

First, the alignment coordinates needed to be determined with the help of the bearings recorded from the satellite receiver. Bearings were required as each edge was a collection of road sections which could travel in any direction. The bearing of alignment was determined using the median of the first sampled bearing from each journey. The starting coordinates furthest in the direction of this bearing were chosen to ensure all journeys covered the same roads. These alignment coordinates were identified from the first sampled coordinates of each journey by finding either the maximum or minimum of either latitude or longitude, depending on the value of the alignment bearing.

Distances between the alignment coordinates and the start of each journey were calculated as the final alignment process. These offset values were then subtracted from the journeys to generate comparable distances along the edge, with distances starting at zero in line with the alignment coordinates. All data attributed to coordinates before the zero point were discarded because they were classed as occurring before the other journeys in the edge.

8.3.4 Extraction of acceleration and noise data

After aligning the coordinates of each journey, the data from the other sensors could be included. Databases were created to contain all sensor data which corresponded to the roads in the NCH to LRI network. These contained both the weighted and unweighted acceleration samples in the x-, y- & z-axes, in $m \cdot s^{-2}$, along with the noise levels in A-weighted dB (dB(A)). Extraction of this data was then conducted by querying with the specific edge identifier, as with the location data. This sensor data was then combined with the location data for each journey, discarding any rows whose timestamps were removed due to alignment.

8.3.5 Extraction of meta-data

Finally, information about each recording was extracted from the database to enable extra analysis of the data in the future. This data encompassed the additional data inputted at the time of recording by the user, such as which ambulance was used, as well as which of the four smartphones was used. The meta-data was stored in each row of the database, but only one row per journey was required to be extracted due to staying constant.

Once all journey data along an edge had been extracted and aligned, then the comfort could be calculated.

8.4 Definition of comfort

A neonatal comfort index is yet to be determined, however studies have suggested that both vibration and noise may be factors [43, 44, 47, 73]. A range of metrics will be used to compare routes in lieu of a true index derived from neonatal responses to stimuli. The metrics defined below may or may not influence the outcomes of infants undergoing transfer, but it will not be known until further studies have been conducted. These metrics should give an indication of the effect of route choice, and serve as a feasibility study which can be used once a true comfort index to be optimised has been identified.

Methods for obtaining the value of a metric for a single journey along an edge will be described alongside the metric itself. Calculations were conducted after extracting the relevant data as described in Section 8.3.

8.4.1 Duration

Routes cannot be assessed using different comfort metrics without accounting for journey duration. It was important to check how comfort compares to the travelling time to see if any of the metrics result in a different route to the quickest option, which would be provided by traditional satellite navigation systems. If there was a direct correlation, the advice would be to follow the route already provided. Similarly, it may be found in the future that returning infants to the hospital environment as soon as possible would be most beneficial.

Journey duration was calculated by subtracting the final timestamp by the first. The value was converted into minutes as the length of the scale of the journeys are too long to be easily interpreted in the S.I. unit of seconds.

8.4.2 Vehicle speed

Some metrics may be correlated to the average vehicle speed as well as to surfaces of the driven roads. This could be assessed at the most basic level by averaging the speeds recorded by each journey and comparing them to the overall comfort. Average vehicle speed was typically dependent on the classification of the roads, compared to the duration which was a combination of vehicle speed and route length.

The average vehicle speed was taken as the median of all recorded vehicle speeds in the data, rather than the mean, to reduce the influence of abnormally slow speeds that may be caused by traffic jams, for example.

8.4.3 Unweighted Vibration

It is unknown exactly how neonatal infants respond to vibrational stimuli. While a standard exists for evaluating the health and comfort of Whole-Body Vibration (WBV) (International Organisation for Standardisation (ISO) 2631 [49]), this is specific to adults. This therefore cannot be directly attributed to neonatal infants due to their lack of development. ISO 2631 does, however, recommend reporting the average vibratory power — under a section on the perception of comfort — presumably to compare with the results of the weightings so any adverse events may be highlighted.

Examining the raw vibration, in all recorded frequencies and in each axis, was conducted to indicate the expected input to the transport incubator along each route. This would enable any differences between road types to be inspected. Performing this analysis would also provide a useful reference for future work to try and identify neona-

tal response by comparing transport outcomes with the recorded vibration, as well as quickly determine optimal roads to use when the response is understood.

Analysis was performed on both the total vibration recorded and on a per-frequency basis.

8.4.3.1 Average power

A summary of the average vibration recorded in each axis could serve as a simple indicator as to the comfort of a road. This was calculated by computing the r.m.s. (Root-Mean-Square) of the unweighted vibration at all frequencies (Equation 8.4.1) in each axis.

$$a_{\text{raw}} = \sqrt{\frac{1}{n}\sum_{i}^{n} a_i^2} \tag{8.4.1}$$

a_i : instantaneous unfiltered acceleration ($\text{m}\cdot\text{s}^{-2}$)

8.4.3.2 Shocks

Along with high magnitudes of sustained vibration, abrupt events have been shown to cause considerable injury to humans. One study, by Blaxter et al. [44], examined the occurrence of sudden shocks within neonatal ambulances. This has not been standardised, however events with a magnitude of 2 g were reported. Both 20 & 100 ms events were shown, however it was decided to solely focus on the 20 ms events due to the lack of 100 ms events in either the previous study or the results here. On a similar note, additional thresholds of ½ g, 1 g and 3 g were assessed to provide more clarity.

$$\text{number of shocks} = \sum_{ms}^{T} S(a_{ms}) \tag{8.4.2}$$

a_{ms} : 20 ms r.m.s. acceleration ($\text{m}\cdot\text{s}^{-2}$)

$$S(a_{ms}) = \begin{cases} 1 & \text{if } a_{ms} \geq a_{\text{threshold}} \\ 0 & \text{otherwise} \end{cases}$$

While computing both shocks and the average power would provide an overview of the vibration along a road, another measure was required to give greater detail.

8.4.3.3 Frequency bands

Vibration at all frequencies should be considered as neonatal infants may each respond differently due to varying lengths of gestation and body weight. Investigating the

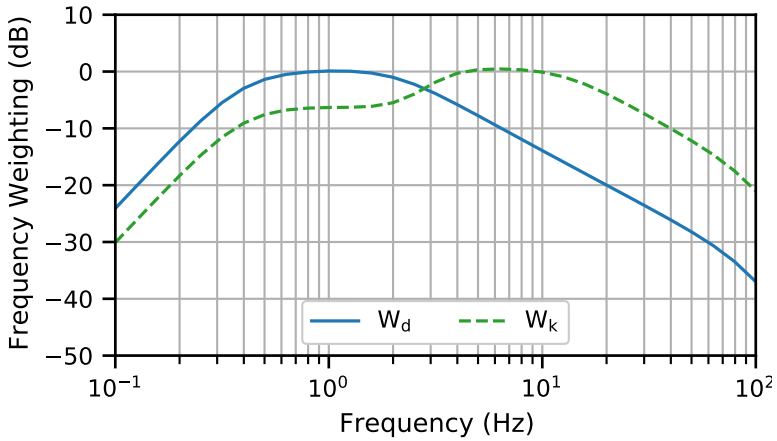

Figure 8.2: Frequency weighting curves for the W_d and W_k weightings from ISO 2631 [49].

spread of frequencies between routes would provide a quick reference if an individual frequency was identified as most significant. Therefore, the power of different frequencies was assessed by calculating the r.m.s. of the vibration in frequency bands of 0.5 Hz around every 0.5 Hz interval (0.25–0.75, 0.75–1.25, etc.) using the formula in Equation 8.4.3.

$$a_f = \sqrt{\frac{1}{n}\sum_{i}^{n} a_{fi}^2} \tag{8.4.3}$$

a_{fi} : instantaneous bandpass-filtered acceleration ($m \cdot s^{-2}$)

8.4.4 Frequency weighted vibration

As mentioned above, the effect of vibration on adults has been well documented and is subject to an international standard [49]. This standard specifies a range of frequency weightings to be applied to recorded acceleration which focus on the most harmful frequencies, depending on the circumstances. Figure 8.2 shows the weighting curves used to examine comfort of a recumbent body, selected as the patients are laid down during transfer, where the W_d weighting is applied to horizontal accelerations and W_k is applied to vertical accelerations. These weighting functions are specified for the analysis of both health and comfort for recumbent persons, and are predominantly related to vibrations through the pelvis. A further weighting function, W_j, is available for analysing the comfort of vibration under the head, however there is currently no available guidance on results.

Different measures of vibratory impact can then be assessed using the weighted accelerations. These measures could be determined for each individual axis or as a total

value by combining the axes using the following formula:

$$v_w = \sqrt{k_x^2\, a_{wx}^2 + k_y^2\, a_{wy}^2 + k_z^2\, a_{wz}^2} \tag{8.4.4}$$

a_{wx}, a_{wy}, a_{wz} : frequency weighted vibration in the x-, y- and z-axes ($m{\cdot}s^{-2}$)

k_x, k_y, k_z : multiplying factors for the x-, y- and z-axes

v_w : total frequency-weighted vibration ($m{\cdot}s^{-2}$)

Each measure, which will be outlined in the next sections, could assess the impact on health — in terms of WBV and Vibration Dose Value (VDV) — or the perception of comfort. Thresholds for each measure were provided in either the ISO 2631 standard itself, or in an affiliated regulatory document to enable analysis of human response as well as comparisons between roads.

8.4.4.1 Health impact

Analysis on the effects of WBV on health was conducted using a reference time-period of 8 hours as defined in ISO 2631. This is defined in the standard as the assessment of vibration exposure, and is usually denoted A(8). Each axis was assessed individually, with ISO 2631 stipulating that the horizontal axes are also multiplied by a factor of 1.4, however the reasoning behind this multiplication is not explained and the effects of vibration on health are only provided for seated occupants (rather than the recumbent in this scenario). These values could then be compared to the exposure action and limit values (0.5 and 1.15 $m{\cdot}s^{-2}$, respectively) specified by an EU Directive [50] and implemented by the UK Health and Safety Executive [51].

The formula for computing A(8), as defined in ISO 2631, is shown below in Equation 8.4.5.

$$A(8) = k\sqrt{\frac{1}{T_0}\sum_i a_{wi}^2 T_i} \tag{8.4.5}$$

a_{wi} : instantaneous frequency-weighted acceleration ($m{\cdot}s^{-2}$)

$$k : \begin{cases} 1.4 & \text{x- or y-axis} \\ 1.0 & \text{z-axis} \end{cases}$$

T_0 : reference duration of 8 hours (s)

T_i : period of acceleration (s)

This could be simplified for the current dataset as the accelerations recorded during a

journey had a sampling interval which could be assumed constant:

$$A(8) = k\sqrt{\frac{T}{T_0}\sum_i a_{wi}^2} \tag{8.4.6}$$

Vibration exposure assessment was to be conducted independently for each axis, however the total vibration value could be used to give a further indication of potential neonatal discomfort. Total vibration exposure was calculated from the individual axis exposures to reduce the amount of computation. Rearranging Equation 8.4.4 shows how combining the exposure values would produce the same result as combining accelerations at each sample and then assessing. The first step was to look at the total exposure at each sample:

$$v_{wi} = \sqrt{k_x^2\, a_{wxi}^2 + k_y^2\, a_{wyi}^2 + k_z^2\, a_{wzi}^2} \tag{8.4.7}$$

$a_{wxi}, a_{wyi}, a_{wzi}$: instantaneous frequency-weighted vibration in x-, y- and z-axes ($m{\cdot}s^{-2}$)

k_x, k_y, k_z : multiplying factors for the x-, y- and z-axes

v_{wi} : instantanteous total frequency-weighted vibration ($m{\cdot}s^{-2}$)

Next, the vibration exposure needed to be rearranged with respect to the square of the weighted vibration:

$$a_{wi}^2 = \frac{T_0}{T_i}\left(\frac{A(8)}{k}\right)^2 \tag{8.4.8}$$

Before substituting in to the original equation (Equation 8.4.7):

$$\sqrt{\frac{T_0}{T_i}\left(\frac{A_{vi}(8)}{k_v}\right)^2} = \sqrt{k_x^2\,\frac{T_0}{T_i}\left(\frac{A_{xi}(8)}{k_x}\right)^2 + k_y^2\,\frac{T_0}{T_i}\left(\frac{A_{yi}(8)}{k_y}\right)^2 + k_z^2\,\frac{T_0}{T_i}\left(\frac{A_{zi}(8)}{k_z}\right)^2} \tag{8.4.9}$$

$A_{xi}(8), A_{yi}(8), A_{zi}(8)$: instantaneous vibration exposure in x, y & z axes ($m{\cdot}s^{-2}$ $A(8)$)

$A_{vi}(8)$: total instantaneous vibration exposure ($m{\cdot}s^{-2}$ $A(8)$)

k_v : multiplying factor for combined axes $= 1$ as undefined

Next, the equation needed to be further rearranged and constants needed to be cancelled out:

$$A_{vi}(8)\sqrt{\frac{T_0}{T_i}} = \sqrt{\frac{T_0}{T_i}\left(k_x^2\left(\frac{A_{xi}(8)}{k_x}\right)^2 + k_y^2\left(\frac{A_{yi}(8)}{k_y}\right)^2 + k_z^2\left(\frac{A_{zi}(8)}{k_z}\right)^2\right)} \tag{8.4.10}$$

$$A_{vi}(8)\sqrt{\frac{T_0}{T_i}} = \sqrt{A_{xi}(8)^2 + A_{yi}(8)^2 + A_{zi}(8)^2}\sqrt{\frac{T_0}{T_i}} \tag{8.4.11}$$

$$A_{vi}(8) = \sqrt{A_{xi}(8)^2 + A_{yi}(8)^2 + A_{zi}(8)^2} \tag{8.4.12}$$

Which could then be generalised to give the total vibration exposure as a product of the exposure calculated for each axis:

$$A_v(8) = \sqrt{A_x(8)^2 + A_y(8)^2 + A_z(8)^2} \tag{8.4.13}$$

8.4.4.2 Vibration dose value

ISO 2631 defines an additional method of evaluating the impact of vibration where shocks may be underestimated when using the r.m.s. method. The VDV is defined to be more sensitive to high magnitude spikes in vibration through the use of the fourth power. Thresholds for VDV have been defined in ISO 2631, as a "caution zone" existing between 8.5 and 17 $m{\cdot}s^{-1.75}$.

The equation for calculating the VDV is defined as below:

$$VDV = \left[\int_0^T \left(a_w\left(t\right)\right)^4 \mathrm{d}t\right]^{\frac{1}{4}} \tag{8.4.14}$$

This could then be simplified, as with Equation 8.4.20, to account for the constant rate of sampling acceleration:

$$VDV = \sqrt[4]{t \sum_i^n a_{wi}^4} \tag{8.4.15}$$

t : sampling interval (s)

Before finally adjusting the weighted vibration to account for the impact of different axes on health:

$$VDV = \sqrt[4]{t \sum_i^n \left(k\, a_{wi}\right)^4} = k \sqrt[4]{t \sum_i^n a_{wi}^4} \tag{8.4.16}$$

$$k : \begin{cases} 1.4 & \text{x- or y-axis} \\ 1.0 & \text{z-axis} \end{cases}$$

Action and limit values of VDV have been defined by the EU Directive [50] at magnitudes of 9.1 and 21 $m{\cdot}s^{-1.75}$ respectively. These are the equivalents of the WBV exposure limits (Section 8.4.4.1) which could be substituted for a_w inside the equation for estimated vibration dose value:

$$eVDV = 1.4\, a_w\, T^{\frac{1}{4}} \tag{8.4.17}$$

Vibration dose values reflecting the total acceleration in each axis could be calculated in a similar method to exposure (8.4.13), taking into account that VDV employs the 4^{th} power:

$$VDV_v = \sqrt[4]{VDV_x^4 + VDV_y^4 + VDV_z^4} \tag{8.4.18}$$

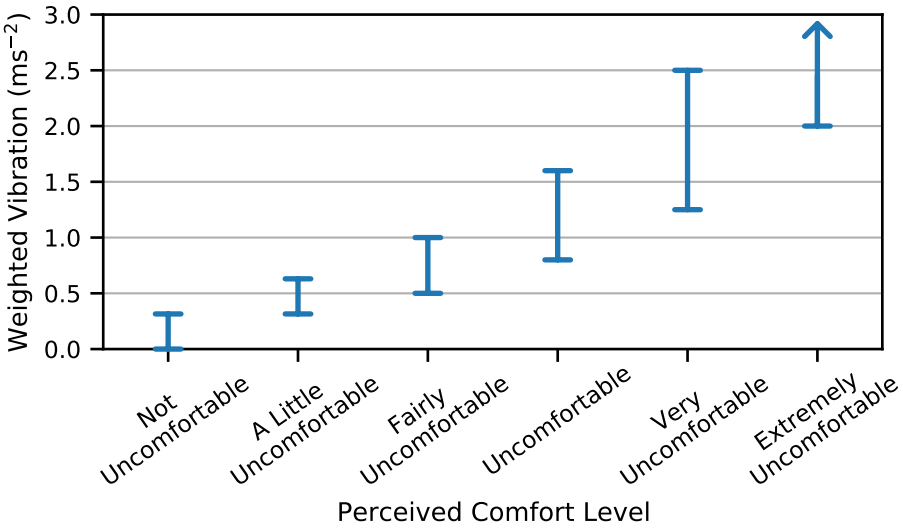

Figure 8.3: Perceived comfort levels and the vibration ranges at which they occur, as defined in the ISO 2631 standard [49].

8.4.4.3 Perceived comfort levels

Reducing the levels of discomfort experienced by neonatal infants may reduce the stress during transfers, which could in turn have health benefits. Perceived comfort of adults has been defined to range from "not uncomfortable" through to "extremely uncomfortable" [49]. Figure 8.3 shows how each comfort level refers to a range in which the r.m.s. of the weighted vibration may fall. These levels are not exact and are known to vary with the subject's environment. It is also worth noting that, aside from the "not uncomfortable" level, this variation resulted in the determined vibration ranges overlapping. It was therefore important to examine whether the vibration result was in the middle of a bound or approaching the upper limit, to provide a greater resolution to the comparison between roads.

Calculating the r.m.s. of the weighted vibration in a single axis is defined in the standard as below:

$$a_w = \left[\frac{1}{T}\int_0^T a_w^2(t)\,\mathrm{d}t\right]^{\frac{1}{2}} \tag{8.4.19}$$

$a_w(t)$: instantaneous frequency-weighted acceleration (m·s^{-2})

T : duration of measurement (s)

The time component could be removed from the calculation as the accelerations were

sampled at a constant rate. This then meant that the r.m.s. equation can be simplified:

$$a_w = \sqrt{\frac{1}{n}\sum_{i}^{n} a_{wi}^2} \tag{8.4.20}$$

n : total number of samples

All axes should be assessed both independently as well as combined to get the overall vibration value. The base equation for combining axes (8.4.4) could be simplified as the multiplying factors for recumbent comfort equal 1 for each axis and could therefore be ignored:

$$v_w = \sqrt{a_{wx}^2 + a_{wy}^2 + a_{wz}^2} \tag{8.4.21}$$

The typical use of the specified levels is to aggregate vibration values and report the overall comfort. It is worth exploring the effect of routing on the individual vibration levels as future development of a neonatal index may identify specific magnitudes as being most harmful. Therefore, the overall vibration values will be supplemented in the analysis here by calculating and comparing the proportion of travel time which was spent with vibration at each level (Equation 8.4.22).

$$\text{time}_{\text{comfort}} = \frac{T}{n}\sum_{i}^{n} C(v_{wi}) \tag{8.4.22}$$

$$C(\text{comfort}) = \begin{cases} 1 & \text{if level}_{\text{lower}} \leq \text{comfort} < \text{level}_{\text{upper}} \\ 0 & \text{otherwise} \end{cases}$$

v_{wi} : total instantaneous frequency-weighted vibration ($\text{m}\cdot\text{s}^{-2}$)

8.4.4.4 Peak WBV

Although peak values are not quantified by ISO 2631, they have been reported by studies investigating neonatal ambulance vibration and are therefore included here for comparison.

The first metric, implemented by Bouchut et al. [43], was a count of the number of 1 s averaged vibrations which breached the lower limit of the "extremely uncomfortable" classification (Equation 8.4.23). Additional metrics were computed similarly but with the lower limit of each of the remaining classifications, excluding "not uncomfortable"

which would always return a result equivalent to the total duration.

$$\text{number of shocks} = \sum_{s}^{T} B(a_s) \tag{8.4.23}$$

$$a_s : 1 \text{ s r.m.s. weighted acceleration } (\text{m}\cdot\text{s}^{-2})$$

$$B(a_s) = \begin{cases} 1 & \text{if } a > 2 \text{ m}\cdot\text{s}^{-2} \\ 0 & \text{otherwise} \end{cases}$$

Supplementary to the above, the overall maximum values were reported for comparison with Karlsson et al. [40]. It was not mentioned in the study what time period these maximum values were calculated for, only that the sample rate was 200 Hz and weightings were applied in accordance with ISO 2631. Therefore, to try and ensure comparability, as well as hopefully determine whether a time period was used, both the sampled maximum values and the maximum 1 s r.m.s. values were identified. One second averages were implemented due to equipment often employed by medical professionals reporting processed values in set time periods, with 1 s commonly selected (as in Bouchut et al. [43]).

8.4.5 Noise

Studies [67, 68] have shown that noise levels could cause adverse health effects in neonatal infants, however the methods of assessment varied. Until more thorough investigations are conducted, with large sample sizes, a range of different noise-related metrics should be analysed. All metrics will use A-weighted sound levels which are designed to reflect how the human ear responds to sound stimuli of different frequencies, as used in all previous studies despite some infants' ears still developing.

Assessment of the noise recorded along each road will be conducted in a similar manner to the comfort-weighted vibration earlier, with the calculation of both an overall value plus the proportion of time spent at certain levels being performed.

8.4.5.1 Average power

Average noise level is often reported when looking into effects of noise on infants in the neonatal intensive care unit and during transport. This noise level is the average power over the course of the recording and is computed by taking the r.m.s. of all samples. A maximum level of 45 dB(A) has been recommended for NICU wards [62], however this was based on a previous suggestion to enable uninterrupted communication between staff and was not medically derived [64]. Similarly, a threshold of 60 dB(A) has been

standardised for the transport equipment [72], although no explanation for the choice of value is provided.

All noise values in the data were copied to a separate variable while removing all blank entries caused by sampling at a lower rate than the IMU. Multiple methods of calculating the average power were then implemented in order to ensure levels could be compared with previous studies.

True average noise levels, typically termed as the equivalent continuous levels, are calculated by taking the r.m.s. of all values in the form of Sound Pressure Level (SPL) [61], before converting to dB(A). This therefore requires noise levels to first be converted into sound levels (Equation 8.4.24) that are then combined as shown in Equation 8.4.25.

$$\mathrm{SPL}_i = 10^{\frac{1}{20}\mathrm{noise}_i} \tag{8.4.24}$$

noise_i : instantaneous noise level (dB(A))

SPL_i : instantaneous sound pressure level (Pa)

$$\mathrm{noise} = 20\log_{10}\left(\sqrt{\frac{1}{n}\sum_i^n \mathrm{SPL}_i^2}\right) \tag{8.4.25}$$

Although the above is the established method of calculating the average noise level, several studies [40, 42, 43, 47, 73] present values which are reported as the "mean". The noise values by these studies tended to be recorded by purpose-built sound meters which would apply Equation 8.4.25 to sampled SPLs and output summaries in set time periods. The use of the term "mean" suggests that the authors of these studies did not process their results correctly and therefore the values would not be directly comparable to the results here. Therefore, three additional summaries were also calculated.

The first of the additional summaries calculated was the mean, which was expected to provide comparable results to the previous studies:

$$\mathrm{noise}_{\mathrm{mean,dB}} = \frac{1}{n}\sum_i^n \mathrm{noise}_i \tag{8.4.26}$$

Additionally, the r.m.s. of the decibel noise levels were calculated. These were calculated to explore the differences in results in case studies which realised the need to use r.m.s. but did not convert to SPLs:

$$\mathrm{noise}_{\mathrm{r.m.s.,dB}} = \sqrt{\frac{1}{n}\sum_i^n \mathrm{noise}_i^2} \tag{8.4.27}$$

A final averaging method was conducted by converting the noise levels to SPLs before calculating the mean value, and then converting back to decibels. This was performed

to address any studies that may have recognised the decibel scale as being logarithmic but did not know that the r.m.s. method was to be used when averaging [61].

$$\text{noise}_{\text{mean,SPL}} = 20 \log_{10} \left(\frac{1}{n} \sum_{i}^{n} \text{SPL}_i \right) \tag{8.4.28}$$

8.4.5.2 Sudden changes

Sudden increases in sound may have more of an impact on health than the sustained levels. Studies have shown that an increase in noise can result in negative effects such as an increased heart rate [67, 68] and a reduced rate of respiration [68], although the classification of events varied. Similarly, studies also looked at the presence of peak noise events during transfers [42, 43, 47, 73]. There is not a standardised measure for sudden noises, and therefore roads were assessed in accordance with the previous studies.

Kuhn et al. [68] produced the only comparison between quantifiable changes in noise level and the resultant physiological effects, therefore the method used in the study took precedent here. Noise levels were reported as 1 second averages and the events were classified as being either 5–10 or 10–15 dB greater than the value for the previous second. This therefore required the sampled noise to be aggregated for each second using Equation 8.4.25 before the differences between subsequent values assessed. These differences were then summarised into counts of the number of times peak events occurred. The two classifications used by Kuhn et al. were supplemented by reporting events of greater than 15 dB, to avoid loss of data.

$$\text{number of events} = \sum_{s}^{T} K\left(\text{noise}_s\right) \tag{8.4.29}$$

$$K(\text{noise}_s) = \begin{cases} 1 & \text{if } \text{noise}_{\text{lower}} < \text{noise}_s \leq \text{noise}_{\text{upper}} \\ 0 & \text{otherwise} \end{cases}$$

$$\text{noise}_s : \text{one second averaged noise level (dB(A))}$$

One second averaged noise levels were also used in a metric employed by Bouchut et al. [43]. Instead of comparing subsequent values, the number of spikes over 85 dB(A) were reported (Equation 8.4.30). Additional thresholds of 75 & 80 dB(A) were also used to further comparison options.

$$\text{number of spikes} = \sum_{s}^{T} B(\text{noise}_s) \tag{8.4.30}$$

$$B(\text{noise}_s) = \begin{cases} 1 & \text{if noise} > 85\text{dB(A)} \\ 0 & \text{otherwise} \end{cases}$$

Finally, the overall maximum noise values were computed to compare the various road types against previous findings from inside ambulances [42, 47, 73].

8.4.5.3 Repetitive exposure

Effects of specific sound levels on the health of neonatal infants are yet to be assessed. Studies have not investigated the effect of specific sound levels, despite looking into both the average levels and shocks. Average levels during most studies varied by only a few dB and a large proportion of shock studies introduced a loud sound for very brief periods. Future research may find that it is exposure to certain levels of noise which causes the most distress, therefore the time spent at different sound levels over the course of each route will be computed.

Time spent at different noise levels was calculated in 2 dB windows from 40 to 90 dB to encompass all expected noise levels. This should give a suitable resolution to enable variation between routes to be visible. Values below 40 dB and above 90 dB were contained in larger windows as they were not expected during transfers. Each window was assessed by dividing the length of the data between the upper and lower limits by the total number of samples, resulting in the percentage of the journey, and then multiplying by the journey duration:

$$\text{time}_{\text{window}} = \frac{T}{n}\sum_{i}^{n} W(\text{noise}_i) \tag{8.4.31}$$

$$W(\text{noise}) = \begin{cases} 1 & \text{if } \text{noise}_{\text{lower}} \leq \text{noise} < \text{noise}_{\text{upper}} \\ 0 & \text{otherwise} \end{cases}$$

8.4.6 Summary

Five core metrics for the assessment of routes have been defined thus far. These ranged from the standard of duration to the use of the noise and vibration recorded by the smartphone sensors. Neither the response of neonatal infants to vibration nor noise are understood, and therefore a range of comparisons will be performed with each of these metrics to form a feasibility study as to the validity of routing towards improving journey comfort. This will be performed by comparing results between roads and against standardised measures, as well as assessing any connection between metric levels and either duration or speed.

This section outlined how to compute metric values for a collection of data from a single journey, however to be able to compare routes the data along an edge needed to be aggregated before these edge values are combined to form an unbroken chain

between destinations.

8.5 Aggregating comfort metrics

Computation of comfort metrics consisted of the assessment and aggregation of values for 3 stages: a single journey segment, an edge, and a full route. After extracting the relevant data and applying the metrics for each journey along an edge, the resultant values needed to be averaged to attain the overall edge assessment. Route comfort was then calculated by combining the values of each constituent edge.

8.5.1 Edge assessment

Obtaining the metrics for an edge was achieved by averaging the calculated values for each journey which travelled along it. Averaging most metrics, apart from the times spent in different comfort levels or at specific noise magnitudes, could be conducted by taking the median of all journey values as each of these metrics was a summary of the same collection of roads. The use of median averaging reduced the impact of any outlying data (for example, extra long duration due to a traffic jam) as opposed to using the mean.

Calculating the average of the times spent at the comfort and noise levels required some adjustment to reflect the variation in vehicle speed between journeys. All journeys along an edge naturally differed in terms of vehicle speed, and therefore duration, which was a problem due to these metrics effectively being a proportion of the travel time. Instead, each metric value could be converted to the distance spent at a specific vibration/noise level, as the edge was a constant distance. These distances were then converted back to a comparable time using the average duration and distance recorded for the edge, with the formula for the full conversion given in Equation 8.5.1. Theoretically, the metric values could be adjusted by the ratio of average to journey duration. The inverse edge distance ratio was included to increase the numerical precision after finding that the recorded distances actually varied slightly due to factors such as (minor) satellite drift and the use of different lanes.

$$t_{ai} = \frac{T_{av}}{d_{av}} \cdot \frac{d_i}{T_i} \cdot t_i \tag{8.5.1}$$

d_{av}, d_i : distance along the edge (edge average, journey recorded)

t_{ai}, t_i : time spent within metric-defined levels (journey adjusted, journey recorded)

T_{av}, T_i : total duration of edge (edge average, journey recorded)

Calculating the median values of each metric would summarise an edge but the variation between journeys will be lost. Therefore, the 5^{th}, 25^{th}, 75^{th} and 95^{th} percentile values were also computed from the journey results for each metric, enabling the spread of data to be analysed.

8.5.2 Route assessment

Route metric values with units of time (duration, time spent at vibration/noise level) and count (number of shocks/sudden noise events) were obtained by totalling the individual values from each constituent edge. This was because the values were simple linear summaries which did not affect one another. Conversely, the vibration exposure, Vibration Dose Value (VDV) and multiple r.m.s. metrics required a more mathematical approach.

Combining metric values in units of vibration, or powers of, followed similar methods to the original calculations shown in Equations 8.4.6, 8.4.15 and 8.4.20. The difference in the combined calculations was the inclusion of varying time periods, as opposed to the constant sampling intervals used previously.

8.5.2.1 Root-Mean-Square (r.m.s.) metrics

Each of the metrics which were calculated using a r.m.s. method could be combined by accounting for the durations of each edge, as was suggested in the original definition of comfort-related vibration (Equation 8.4.19). This could then be generalised to obtain route values for each metric:

$$rms_r = \sqrt{\frac{1}{T_r} \sum_e rms_e^2 \, T_e} \tag{8.5.2}$$

rms_e : edge value of r.m.s. metric

rms_r : route value of r.m.s. metric

T_e : edge duration (s)

T_r : total route duration (s)

One exception to the use of Equation 8.5.2 is the true average noise level. As discussed in Section 8.7.4.1, the true noise level requires conversion to SPL before applying the r.m.s. formula. Therefore, aggregating the route noise was performed in a similar man-

ner:

$$\text{noise}_r = 20 \log_{10} \left(\sqrt{\frac{1}{T_r} \sum_e \text{SPL}_e^2 \; T_e} \right) \tag{8.5.3}$$

$$\text{SPL}_e = 10^{\frac{1}{20}\text{noise}_e}$$

8.5.2.2 Vibration Exposure

General calculation of vibration exposure, given in Equation 8.4.5, allowed the input of accelerations recorded over different periods. Expanding this equation could provide a simple method of combining previously calculated exposures.

First, it was assumed that there were two exposures which each had an overall weighted acceleration and time period:

$$A(8) = k \sqrt{\frac{1}{T_0} \left(a_{wi}^2 \; T_i + a_{wj}^2 \; T_j \right)} \tag{8.5.4}$$

a_{wi}, a_{wj} : weighted acceleration for time period T_i and T_j respectively ($\text{m}\cdot\text{s}^{-2}$)

This could then be rearranged algebraically to get the following:

$$A(8) = k \sqrt{\frac{1}{T_0} \; a_{wi}^2 \; T_i + \frac{1}{T_0} \; a_{wj}^2 \; T_j} \tag{8.5.5}$$

$$= \sqrt{k^2 \; \frac{1}{T_0} \; a_{wi}^2 \; T_i + k^2 \; \frac{1}{T_0} \; a_{wj}^2 \; T_j} \tag{8.5.6}$$

$$= \sqrt{\left(k \sqrt{\frac{1}{T_0} \; a_{wi}^2 \; T_i} \right)^2 + \left(k \sqrt{\frac{1}{T_0} \; a_{wj}^2 \; T_j} \right)^2} \tag{8.5.7}$$

Substituting the original formula for vibration exposure back in resulted in a simple expression with no extra variables:

$$A(8) = \sqrt{A_i(8)^2 + A_j(8)^2} \tag{8.5.8}$$

$A_i(8), A_j(8)$: vibration exposure for periods i and j ($\text{m}\cdot\text{s}^{-2}$ $A(8)$)

This could then be generalised to obtain a simple formula for combining the vibration exposures of each edge to calculate the overall route exposure:

$$A_r(8) = \sqrt{\sum_e A_e(8)^2} \tag{8.5.9}$$

$A_e(8)$: edge vibration exposure ($\text{m}\cdot\text{s}^{-2}$ $A(8)$)

$A_r(8)$: route vibration exposure ($\text{m}\cdot\text{s}^{-2}$ $A(8)$)

These final route values could then be both compared against the action and limit values of 0.5 and 1.15 $m{\cdot}s^{-2}$ as well as against other routes to determine whether road choice may have an effect on health.

8.5.2.3 Vibration Dose Value

Combining vibration dose values for multiple periods of time was outlined in the documentation for ISO 2631 as the following:

$$VDV_{total} = \sqrt[4]{\sum_i VDV_i^4} \tag{8.5.10}$$

VDV_i : individual vibration dose values ($m{\cdot}s^{-1.75}$)

This could be proved by expanding on the standard VDV equation (8.4.14) in a similar method to the vibration exposure above. First, it was assumed that there were two measurements which were recorded over different time periods:

$$VDV = \sqrt[4]{a_{wi}^4\, t_i + a_{wj}^4\, t_j} \tag{8.5.11}$$

Substituting Equation 8.4.14 back in resulted in Equation 8.5.10 as specified:

$$VDV = \sqrt[4]{VDV_i^4 + VDV_j^4} \tag{8.5.12}$$

Therefore, the vibration dose value for each route was calculated with the expression in Equation 8.5.13 before being assessed against the action and limit values of 9.1 and 21 $m{\cdot}s^{-1.75}$.

$$VDV_r = \sqrt[4]{\sum_e VDV_e^4} \tag{8.5.13}$$

VDV_e, VDV_r : vibration dose values of an edge and route, respectively ($m{\cdot}s^{-1.75}$)

8.5.3 Summary

The aggregation of metric values for a single edge from all journeys which travelled along the roads was straightforward after the data was aligned (Section 8.3.3). Each of the journeys was ensured to have used the same roads for the same distance and therefore data were directly comparable. Median and percentile values were computed for each metric to both average the edge and to illustrate the spread.

Computing overall route values from the constituent edges required different processes depending on the metric. The metrics which had units in terms of either distance or

time could be totalled quite. Those which involved powers of either vibration or noise appeared complex at first, however the process for each could be simplified through the manipulation of equations.

With the procedures in place to first assess metrics along each edge and then combine them to from route assessments, the next stage was to implement them.

8.6 Factors affecting metric values

This section will detail the trends which were identifiable between the journey components and the resultant metric values of the NCH to LRI network.

Summary values were computed for data recorded by each journey along each of the constituent edges which combined to form a complete edge of the final NCH to LRI network (as explained in Section 7.7.2). This enabled a far greater amount of data to be analysed than performing with the final NCH to LRI edges, as the latter would only be possible using the averaged metrics. The 114 edges of the fully split NCH to LRI network resulted in a total of 7,092 separate sets of metrics, with a mean of 62 journeys contributing towards the analysis of a specific edge (minimum of 1 journey; maximum of 221 journeys).

Edges could be classified by the types of roads within them. These could be urban roads, which are typically slower with a high number of traffic lights, A roads, which are predominantly dual carriageways split up by roundabouts, and motorways, which have multiple lanes and are designed for travelling at higher speeds. Manually inspecting the edges resulted in the classifications shown in Table 8.1, where some edges consisted of more than one road type. Although every effort was made to accurately classify the edges, it is acknowledged that the assessments could be disputed. For example, some road types were ambiguous, such as the Nottingham ring road that is in a rather urban location and possesses multiple roundabouts and traffic light intersections, however also has a relatively high speed limit of 40 mph (18 $m \cdot s^{-1}$).

The road classification for each edge will be clearly indicated on all figures within this section by means of colour, and outlined in embedded legends correlating the colours to the individual IDs from Table 8.9.

Three main factors were assessed for their influence on metric results: road classification, duration spent along the edge, and the average vehicle speed, along with an examination of the potential noise averaging methods. These were explored through examination of the formulae for computing the metrics, the application of statistical

Table 8.1: Amount of data available for each road type.

Road Classification (ID)	Number of Edges	Number of Samples
Urban (U)	24	1,433
A roads (A)	32	1,364
A & Urban (AU)	42	3,187
Motorway (M)	16	1,108

tests and by considering the physical components. Tests performed were either parametric — such as the *t*-test and Analysis of Variance (ANOVA) — or non-parametric — such as Mann-Whitney and Kruskal-Wallis tests — depending on the properties of the data to be compared. It is important to note that the scale of the data could lead to small differences being declared as significant and therefore the ultimate effect of each factor needed to quantified before any association was determined.

Analysis of trends will progress through the potential factors in order, from road classification to vehicle speed and then duration, before finishing with the noise variations. First, however, the general comparability of the dataset will be assessed.

8.6.1 Similarities between patient transfers and empty journeys

Gathering a large amount of data in order to increase the confidence in the determined metric values for each road requires all journeys to be comparable. Blaxter et al. [44] reported decreased vibration levels during ambulance journeys with patients on board versus without, potentially due to an effort (either conscious or unconscious) by the drivers to drive more smoothly. In Section 6.5.1, it was shown that only 44% of successfully recorded journeys were true neonatal transfers. A total of 23% (1,612) of the data along the 7,092 road segments were recorded during these patient transfers, meaning the majority of data was from empty ambulances. Focusing the route analysis solely on data from patient transfers would significantly diminish the overall confidence in results, as well as potentially reduce the number of unique road segments, and therefore tests were required to see whether all data could be included in the final analysis.

Ideally, data recorded during patient transfers would be comparable to those where the incubator was empty. To check this, tests were performed to compare the means of both sets of data. Four of the predefined metrics were assessed for similarity: average unweighted vibration in each of the three axes, and the average noise level. Each of the other metrics were deemed to be a subset of the four chosen metrics, and therefore

additional tests were not performed.

Data were first grouped together by the specific road segment as each segment could vary in terms of road classification, length, state of repair, etcetera, before being split into those transfers with and those without patients on board. It was not possible to perform an ANOVA of all data together using the two factors of edge and patient presence due to violating the requirement for homogeneity[1] ($F(214, 6{,}877) = 4.7$, $p < .001$). This violation was largely expected due to the vast range of sample sizes for each edge (varying from 1–22) and therefore an alternative approach was needed.

t-tests were performed separately for each edge that registered at least two journeys both with and without patients. Although this resulted in the exclusion of 22 (19%) of the total segments, at least 96% of the remaining edges (by distance) now possessed homogenous variances between the two groups in each metric. The *t*-tests for each of the metrics were computed for each edge and then combined to calculate the proportions of similarity for both edges and the cumulative edge distance.

Similarity between the resultant metrics during patient and non-patient transfers were found to vary depending on the edge. Each of the metrics produced some values which were either indicated by *t*-tests as significantly different between the journey data along a single edge (noise: $t(112) = 5.2$, $p < .001$; x-axis acc.: $t(191) = –1.3$, $p < .001$; y-axis acc.: $t(163) = –4.6$, $p < .001$; z-axis acc.: $t(190) = –4.0$, $p < .001$) or highly similar for a single edge (noise: $t(119) = 0.0$, $p = 1$; x-axis acc.: $t(56) = 1.0$, $p = .93$; y-axis acc.: $t(26) = 0.0$, $p = .98$; z-axis acc.: $t(53) = 0.0$, $p = .99$), where a p value below .05 indicates a less than 5% chance of similarity and a value of 1 strongly suggesting no differences are evident.

Aggregating the probabilities for each edge provided an overall estimate for the general influence of patient presence. Average noise, y-axis and z-axis accelerations had no significant differences for over 83% of the total distance covered by the various edges (Table 8.2). The similarity for x-axis accelerations was reduced at 74%, although the average difference between the journeys with patients and without was less than 13%. Average significant differences were equivalent across the vibration axes, while average noise level was relatively similar despite the statistical result.

Due to the majority of edges registering similar levels of both average noise and average vibration, regardless of whether a patient was being transferred, the removal of

[1]The *F*-test is used in ANOVA to indicate how closely the variances of multiple sets of data match one another. Two values are provided in brackets: the degrees of freedom between the sets (number of sets -1) and the degrees of freedom within the sets (number of data points $-$ number of sets). Probability values, p, below .05 indicate dissimilarity between the sets and therefore suggest that ANOVA cannot be accurately implemented.

Table 8.2: Proportion of edges which were not significantly different in metric value.

Metric		Similarity Proportion		Dissimilarity of
		Number of Edges	Distance	Significantly Different
Average Noise		87.0%	83.7%	2.6%
Vibration	X-axis	70.7%	73.7%	12.7%
	Y-axis	80.4%	87.3%	13.1%
	Z-axis	79.3%	92.8%	11.4%

the patient factor was judged as acceptable. Therefore, no further considerations were made regarding neonatal presence during analysis of other factors.

8.6.2 Effect of road classification on metric values

By their nature, road classifications indicate differences in the quality of road and the expected driving conditions. Motorways, for example, are designed for safe travel at speeds of 70 mph (31 $m \cdot s^{-1}$) over long distances. A roads, on the other hand, can vary in speed limit between 40 (18 $m \cdot s^{-1}$) & 70 mph and have multiple junctions and roundabouts which require vehicles to slow down, resulting in a less consistent experience. Finally, roads within urban environments tend to be far slower with an increased number of junctions and traffic lights. These differences, along with the varying standards at which the types of road are maintained, were likely to influence resultant metric values.

Assessments of the influence of road classification on expected noise and vibration levels were performed using average metric values for each of the 114 edges. This decision was made over the use of all 7,092 constituent journey segments due to the numerous data along the more frequently used edges skewing results to their averages.

The average noise and vibration levels in each axis were used in the road comparisons, as in the patient presence earlier, with noise expected to be greater along motorways compared to urban roads and vibration expected to behave oppositely. These metrics were supplemented with metrics that looked at the the frequency of peak events, included due to the increased likelihood of lower class roads having sudden road artefacts which probably would not be present along motorways, and the potential influencing factors of duration and vehicle speed. Absolute peak values were considered but found to follow the same trends as the averages.

The focus of this section will be on direct links between road classification and the resultant metric value, however later sections will explore additional influences in more

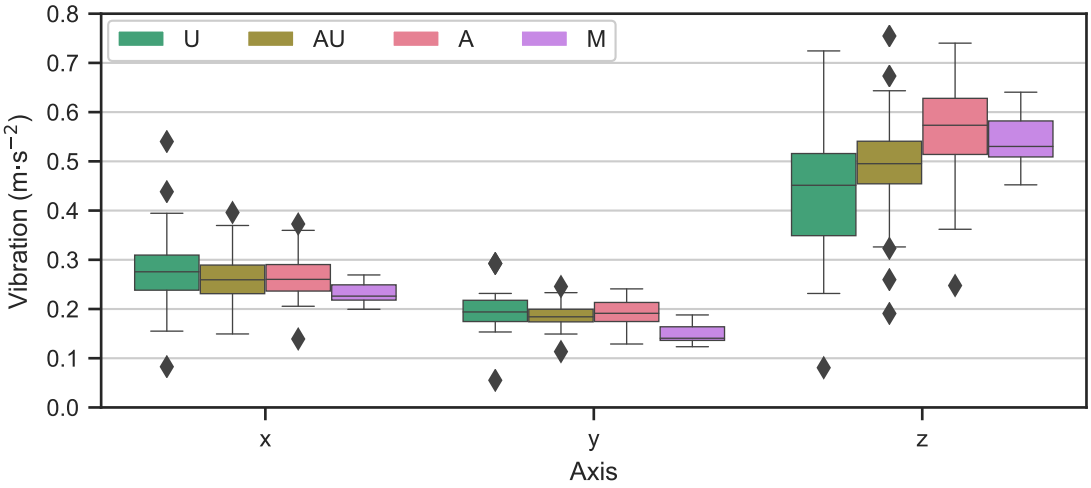

Figure 8.4: Difference in average vibration in each axis for the 4 road classifications.

detail. Of the assessed metrics, only the average edge durations ($F(3, 110) = 2.3$, $p = .08$) and the rate of shocks over ½ g ($F(3, 110) = .0$, $p = .98$) were determined to be homoscedastic (all sets of data had equal variances [164]) when grouped by road classification and were therefore suitable for analysis using ANOVA. All others were found to vary too greatly between road classifications, leading to the use of non-parametric tests.

8.6.2.1 Average vibration

Road classification was found to significantly affect average vibration in the horizontal axes (x-axis: $H(3) = 11.8$, $p < .01$; y-axis: $H(3) = 24.8$, $p < .001$). Upon closer examination using Figure 8.4, it appeared that only the use of a motorway resulted in a clear reduction in either the x- or y-axes. This was confirmed by Mann-Whitney tests, performed for each of the 6 possible pairings due to the plotted overlaps with the significances adjusted by applying a Bonferroni correction, where comparisons with motorway values were significantly different (x-axis: $U = 294$–497, $p < .05$; y-axis: $U = 339$–592, $p < .001$). No differences were evident between urban, urban/A mixed or A-roads.

Clear contrasts were evident in the average z-axis vibration ($H(3) = 19.1$, $p < .001$) and appeared to differ between all road classifications (Figure 8.4). Again, overlaps between groups was prominent and therefore all 6 pairwise comparisons were performed. Clear similarities were shown between urban and mixed urban/A roads ($U = 382$, $p = .63$) and between A-roads and motorways ($U = 321$, $p = .95$), while no significant difference was found between mixed urban/A and motorways ($U = 216$, $p = .23$). It can, however, be concluded that urban roads result in reduced vibration compared to both A-roads and motorways (U vs. A: $U = 191$, $p < .01$; U vs. M: $U = 93$, $p < .05$), with

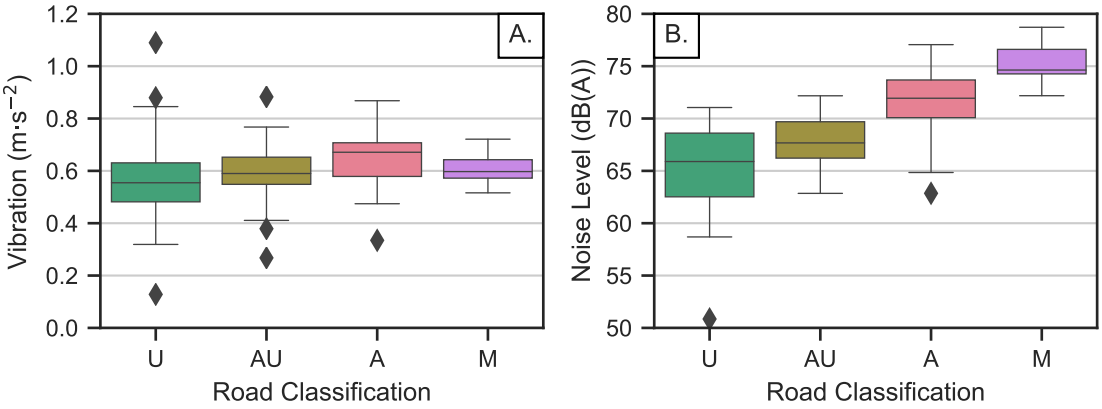

Figure 8.5: Variation in total vibration (A) and average noise level (B) for each road classification.

A-roads also resulting in greater vibration than mixed urban/A ($U = 364$, $p < .01$).

All roads registered the greatest vibrations in the vertical axis, meaning that the total combined vibrations were minimally influenced by the horizontal accelerations. This resulted in the total vibrations largely following the same trend as the vertical values above with slightly increased magnitudes (Figure 8.5A).

8.6.2.2 Average noise

Road classifications were found to have a clear influence on the expected noise levels ($H(3) = 68.1$, $p < .001$) and were visibly different for each, as shown in Figure 8.5B. Subsequent Mann-Whitney tests, performed stepwise through the classifications and therefore requiring a Bonferroni correction of 3, confirmed that noise levels were significantly different (U vs. AU: $U = 309$, $p < .05$; AU vs. A: $U = 189$, $p < .001$; A vs. M: $U = 68$, $p < .001$) albeit by only 2–4 dB in each case.

8.6.2.3 Peak events

Unweighted vibrational shocks of ½ g were found to rarely occur on any road classification and therefore resulted in minimal differences ($F(3, 110) = 0.1$, $p = .97$). Comparing values for 1 s shocks deemed as at least uncomfortable, on the other hand, were found to occur at different frequencies depending on the road classification ($H(3) = 16.8$, $p < .001$). Similar to the average vibration in the x- & y-axes reported above, insignificant differences were noted between the average rates along urban, urban/A mixed, and A-roads (Figure 8.6A), although the variation did increase at lower classifications. Motorways, however, were found to result in a significantly reduced rate of shocks (vs. A: $U = 80$, $p < .001$).

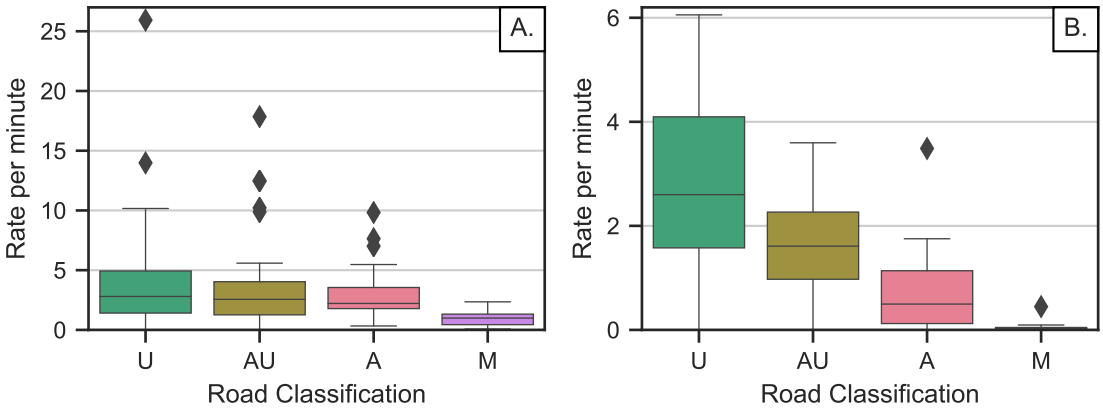

Figure 8.6: Rate of uncomfortable (over 0.8 $m \cdot s^{-2}$, weighted in accordance with ISO 2631) shocks (A) and 5–10 dB increases in noise (B) for different roads.

Noise stability was also identified as varying with road classification ($H(3) = 58.0$, $p < .001$), with rates of 5–10 dB increases in 1 s noise level appearing to be greater at lower classifications (Figure 8.6B). This trend was confirmed with all classifications being reduced compared to the previous, with all significances adjusted by applying a Bonferroni correction (U vs. AU: $U = 731.5$, $p < .05$; AU vs. A: $U = 1063.5$, $p < .001$; A vs. M: $U = 431.5$, $p < .001$).

8.6.2.4 Duration and vehicle speed

There are clear differences between the expected metric values for each road classification. What is not clear is if these differences are due to the roads themselves or a feature of the road classifications. For example, while the average noise levels increase as roads progress from urban to A-road to motorway, so do the respective speed limits. Similarly, the levels of horizontal vibration and the rate of fluctuations in noise level were more prominent in classifications which have more frequent junctions and roundabouts, and therefore may be correlated with segment duration with the segments being separated by these junctions. Analysis was therefore performed on both duration and vehicle speed to see how they varied with road classification.

As noted earlier, average durations were found to be homoscedastic between the road classifications and thus were suitable for ANOVA. Variations were subsequently identified as significantly different depending on whether urban, A, or motorway roads were used ($F(3, 110) = 3.1$, $p < .05$). Examining the spread of values themselves, Figure 8.7A shows that all classifications tend to vary in length apart from urban. Excluding urban roads from the ANOVA indicated that the remaining classifications were highly similar ($F(2, 87) = 0.2$, $p = .84$). This is to be expected as urban roads are closer together and have more frequent junctions than the other classifications. Although it may be ex-

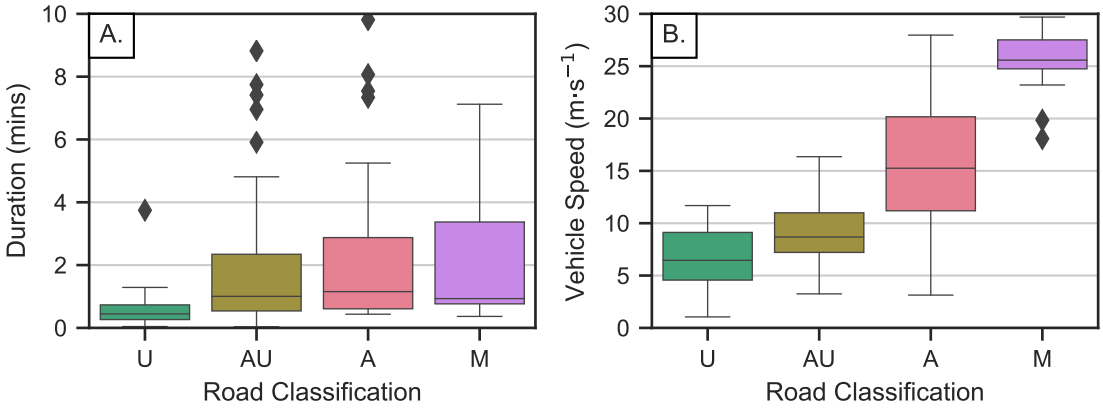

Figure 8.7: Variation in both average edge duration and vehicle speed with respect to road classification.

pected that motorways would mainly comprise long, uninterrupted stretches, multiple small segments were formed around the junctions to enable network analysis.

Average vehicle speeds were expected to differ greatly between road classifications, due to the variation in speed limits, and this was confirmed by the Kruskal-Wallis test ($H(3) = 66.0$, $p < .001$). Interquartile Ranges (IQRs) were only shown to overlap for the urban and mixed urban/A-roads (Figure 8.7B), with the motorway group clearly different from all others. Therefore, only two Mann-Whitney tests were performed, with the results concluding that there are significant differences between the remaining groups (U vs. AU: $U = 284$, $p < .01$; AU vs. A: $U = 235$, $p < .001$).

The trend between road classification and expected vehicle speed appears to be similar to those of some metrics while being the inverse of others. This suggested that part of — if not all — of the identified connections between metrics and road groups could be due to the vehicle speed.

8.6.3 Effect of vehicle speed on metric values

The general relationship between vehicle speed and resultant metric values was first explored by examining the full dataset of 7,092 journey components. Combining all values for speed from each edge resulted in a set of results which did not follow a normal distribution ($D(7{,}092) = 0.10$, $p < .001$) due to multiple prominent peaks, likely influenced by the different speed limits on the various road types, and therefore only non-parametric tests could be used.

Analysis was furthered by studying trends within each road classification and then within each edge, to see how speed may affect results over the same road surfaces. While trends could be assessed, it was not possible to investigate how much of each

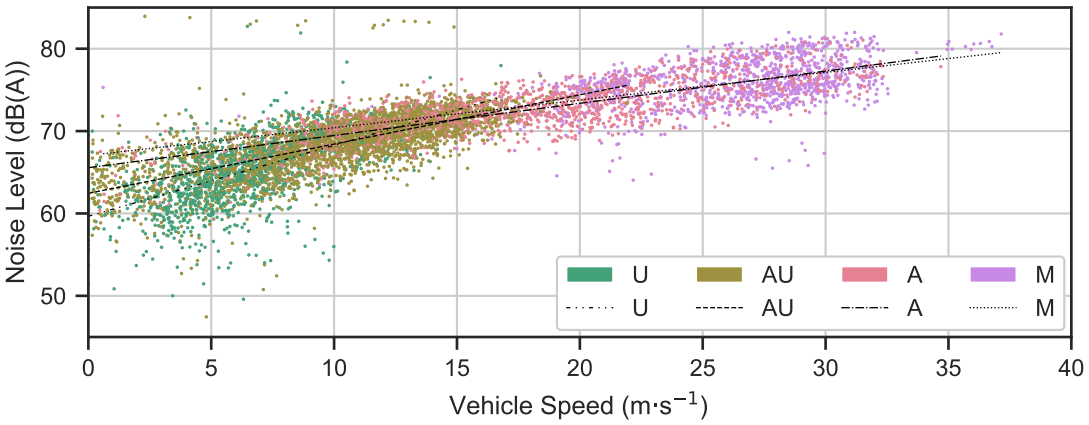

Figure 8.8: Distribution of resultant noise levels compared to the vehicle speeds at which they occurred.

metric may have been influenced by vehicle speed and how much by the road classification. Although methods such as Analysis of Covariance (ANCOVA) exist, they are designed to improve the testing of an independent factor (in this case, the road class) and cannot be used to control for additional factors [165]. Due to the clear and fundamental link between road classification and expected speed, ANCOVA was not suitable and could not be implemented regardless of whether the data met the assumptions of normality and homoscedasticity.

8.6.3.1 Average noise

As expected, average noise level appeared to increase with speed and produced the strongest visible correlation of all metrics when all data was plotted in Figure 8.8. Performing a Spearman's rank-order correlation between both variables confirmed the general relationship was significant ($r_s(7{,}090) = .88$, $p < .001$). The expected connection between road classifications and recorded noise levels were also evident, with urban roads tending to be quietest, followed by urban/A mixed and then A-roads and motorways, although some of the correlations are very poor and differences are likely due to the vehicle speeds themselves increasing as shown earlier (Figure 8.7B).

Performing linear regressions on the results showed clear differences between each road classification. Intercept values increased as classifications progressed, without overlap, highlighting how the higher classification roads tended to be louder (Table 8.3). Conversely, increases in vehicle speeds were found to have the greatest influence on noise levels along urban roads, whereas the smallest influence was on motorways.

Table 8.3: Intercept and gradient results (with 95% Confidence Interval (CI)) from linear regressions performed between average noise levels and speeds for each road classification.

Road Type	Intercept (dB(A))	Gradient	R^2
U	59.7 (59.3–60.1)	0.86 (0.81–0.92)	.39
AU	62.4 (62.2–62.7)	0.60 (0.58–0.62)	.52
A	65.6 (65.3–65.9)	0.39 (0.38–0.41)	.64
M	67.0 (66.1–68.0)	0.34 (0.30–0.37)	.25

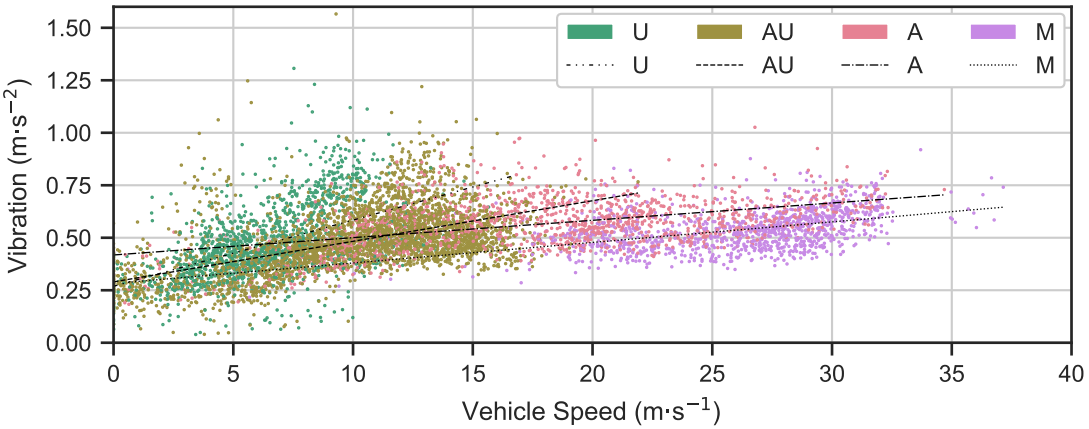

Figure 8.9: Distribution of resultant average vertical vibrations compared to the vehicle speeds at which they occurred.

8.6.3.2 Average vibration

Vertical vibration was similarly found to suggest an increase with speed ($r_s(7{,}090) = .51$, $p < .001$), albeit at a reduced correlation. One reason for this reduced correlation could be the differing relationships depending on the road classification. Examining the data visually (Figure 8.9), it appears that urban roads may have a stronger relationship between speed and vibration than the other classifications.

Sideways motions, on the other hand, were found to reduce slightly as vehicle speed increased, although there was a large spread of data ($r_s(7{,}090) = -.29$, $p < .001$), which followed the hypothesis of decreased junctions and roundabouts. The x-axis, however, appeared to be greatly dependent on the road classification and subsequently did not significantly determine any correlation for the full dataset ($r_s(7{,}090) = .02$, $p = .07$).

Examining the regressions for the range of road classifications showed how the trends differed for each axis (Table 8.4). The x-axis vibration on smaller urban roads, where junctions are frequent, were found to have a larger dependency on vehicle speed than the A-roads and motorways where vehicle speed is largely consistent. The design of motorways, with long sweeping curves and slip roads instead of roundabouts, was also

Table 8.4: Intercept and gradient results (with 95% CI) for vibration versus speed, in each axis, depending on road classification.

Axis	Road Type	Intercept ($m \cdot s^{-2}$)	Gradient	R^2
x	U	0.212 (0.201–0.224)	0.015 (0.013–0.016)	.19
	AU	0.189 (0.183–0.195)	0.007 (0.006–0.007)	.14
	A	0.263 (0.253–0.272)	0.000 (0.000–0.001)	.00
	M	0.154 (0.137–0.170)	0.003 (0.002–0.004)	.07
y	U	0.185 (0.177–0.192)	0.003 (0.002–0.004)	.03
	AU	0.182 (0.178–0.187)	0.000 (0.000–0.001)	.00
	A	0.225 (0.218–0.232)	−0.002 (−0.003–−0.002)	.10
	M	0.141 (0.126–0.157)	0.000 (0.000-0.001)	.00
z	U	0.269 (0.250–0.288)	0.032 (0.029–0.034)	.29
	AU	0.290 (0.279–0.301)	0.019 (0.018–0.020)	.30
	A	0.418 (0.402–0.434)	0.008 (0.007-0.009)	.22
	M	0.282 (0.252–0.313)	0.010 (0.009–0.011)	.20

found to produce the least sideways vibration with minimal change with speed. An increase in vehicle speed along A-roads was found to reduce the sideways vibration, however, likely due to junctions being more frequent on lower speed stretches of road. While the data largely agrees with expected results, it must be noted that the variations between recordings led to very low coefficients of determination.

Assuming that equivalent speeds are attainable on all recorded road types, motorways would result in the least amount of vertical vibration. In reality, it is not possible (legally, and, likely, physically) for ambulances to achieve a speed of 70 mph on the urban roads and thus it cannot be determined whether the obtained gradient would remain valid. It can be determined from the overarching trend relationship between vehicle speed and vertical vibration, however, that motorways would result in similar or reduced levels to those on an urban roads in the event of a traffic jam with greatly reduced speeds. Similarly, it can be determined that the recorded A-roads produced larger magnitude vibrations compared to the motorway, with multiple data at overlapping speeds. Analysis may be furthered by incorporating the assigned speed limit for each road along with the road classification, although several edges would consist of multiple stretches at different speeds to complicate matters.

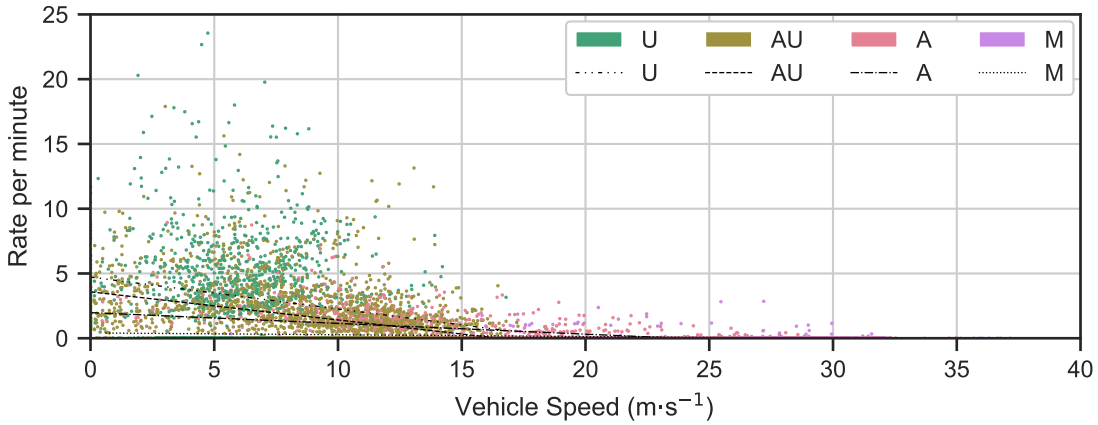

Figure 8.10: Distribution of frequency of 5–10 dB noise level increases compared to vehicle speeds (8 results omitted between rates of 25 and 55, for clarity).

Table 8.5: Intercept and gradient results (with 95% CI) for the rate of 5–10 dB increases versus speed depending on road classification.

Road Type	Intercept	Gradient	R^2
U	4.7 (4.3–5.2)	−0.25 (−0.31–−0.18)	.04
AU	3.6 (3.4–3.8)	−0.22 (−0.24–−0.20)	.13
A	2.0 (1.8–2.1)	−0.08 (−0.09–−0.08)	.29
M	0.4 (0.3–0.5)	−0.01 (−0.02–−0.01)	.06

8.6.3.3 Peak events

The rates of peak events occurring were earlier shown to be lowest along motorways. Significant differences were found within the 5–10 dB increases in noise level as road class increased in scale, suggesting that vehicle speed may be a factor. Examining the spread of the rates of noise increases against the average vehicle speeds (Figure 8.10) suggests a reasonable negative correlation, confirmed by performing a rank analysis ($r_s = -.56$, $p < .001$).

Separating the results based on the road classification enabled further trends to be observed. Linear regressions were performed upon the data in each group, with the results shown in Table 8.5. It can be observed that the greatest negative relationships occur within the urban and mixed urban/A road groups which encompass the bulk of the high event frequencies. It is likely that these increases are connected to the number of ambulance accelerations and decelerations in accordance with junctions, as it was earlier shown that noise increases with speed.

Although motorways were found to have a reduced rate of fairly uncomfortable (in accordance with ISO 2631) shocks, there was minimal difference between the other three road classifications. Unlike sudden noises above, there was no discernable correlation

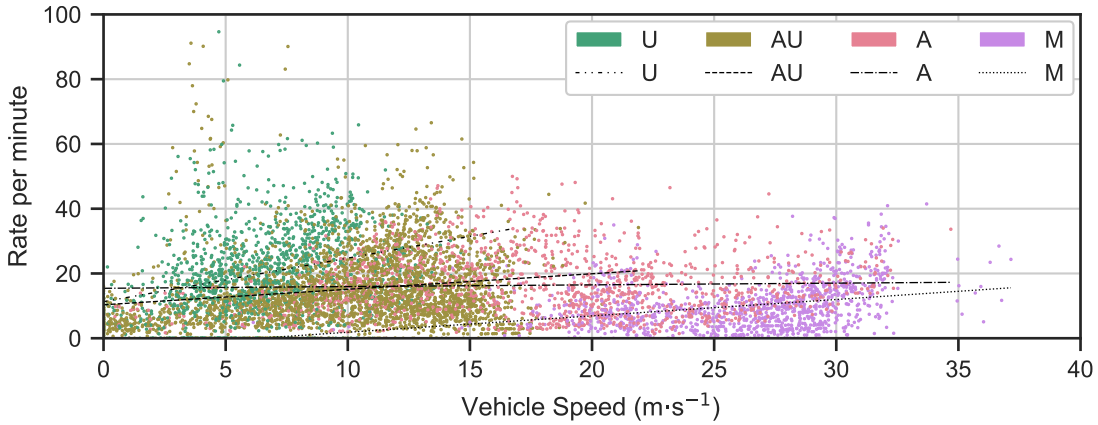

Figure 8.11: Distribution of frequency of 1 s fairly uncomfortable shocks compared to vehicle speeds (9 results omitted above a rate of 100, for clarity).

Table 8.6: Intercept and gradient results (with 95% CI) for the rate of 1 s uncomfortable shocks versus speed depending on road classification.

Road Type	Intercept	Gradient	R^2
U	11.3 (9.4–13.2)	1.3 (1.1–1.6)	.07
AU	10.3 (9.2–11.4)	0.5 (0.4–0.6)	.03
A	15.4 (14.1–16.8)	0.1 (0.0–0.1)	.00
M	−3.2 (−5.8–−0.6)	0.5 (0.4–0.6)	.09

between vehicle speed and the resultant rate of shocks (r_s = .00, p = .92). The extent of the apparent randomness can be seen in Figure 8.11 where the spread of rates increase up to around 10 m·s^{-1} before reducing afterwards.

Applying linear regression to the data suggested that increased vehicle speed increased the rate of shocks for all road classifications bar A-roads (Table 8.6). This was especially prominent for urban roads, although the correlation for each classification is minimal. The increase in rate with speed may in fact be the result of a decrease in duration. It may be that specific roads result in a set number of shock events and driving over them at a greater speed solely reduces the time interacting with the edge, supported by an average edge gradient of 1.3 (IQR: 0.6–2.3).

The number of shock events were initially converted into rates to enable analysis between different size segments, however the results suggest that there were no trends. Comparing the true count, instead of the rate, to vehicle speed suggested there was a slight positive relationship, however there was no correlation (R^2 = .00, .02, .07 & .06 for U, AU, A & M classifications respectively). The lack of correlations suggested that speed had little effect on the number of shocks and therefore the focus progressed to duration.

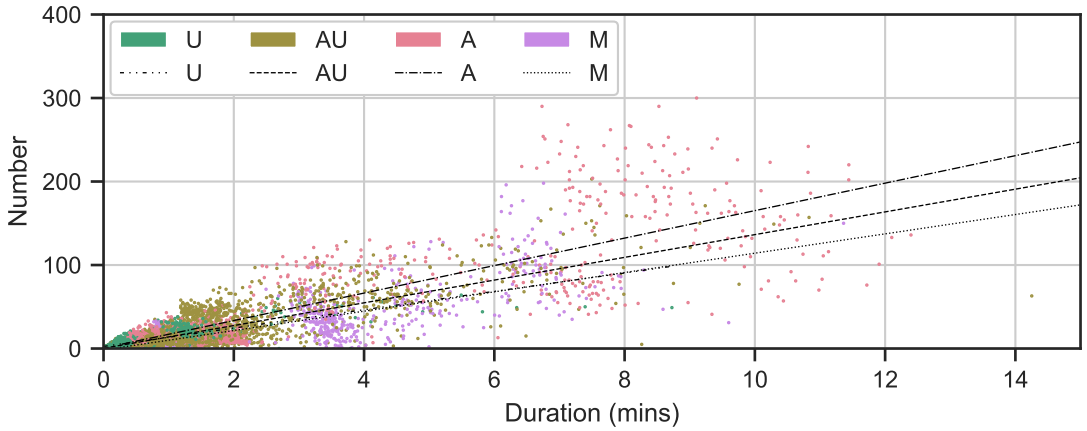

Figure 8.12: Distribution of 1 s fairly uncomfortable shocks compared to duration (5 results omitted above a duration of 15, for clarity).

Table 8.7: Intercept and gradient results (with 95% CI) for the number of 1 s uncomfortable shocks versus duration depending on road classification.

Road Type	Intercept	Gradient	R^2
U	2.6 (2.2–2.9)	11.0 (10.6–11.4)	.65
AU	0.3 (−0.3–0.9)	13.6 (13.3–13.9)	.67
A	0.6 (−1.6–2.8)	16.5 (15.9–17.0)	.69
M	−1.6 (−3.2–0.1)	11.5 (11.0–12.1)	.61

8.6.4 Effect of duration on metric values

Duration was not expected to influence many of the metrics, due to the majority either assessing average levels or sudden shocks. These were expected to be influenced by the road and/or the vehicle speed, with any connection with duration being due to the segment length combined with the speed. There were, however, a few metrics in which duration was integral to the formula. First, the peak vibrations from the previous section were assessed.

8.6.4.1 Vibratory shocks

Comfort-weighted 1 s shocks of at least 0.5 $m \cdot s^{-2}$ were found to increase in number with duration (Figure 8.12). Although this may appear obvious, it was not expected that all road segments would result in similar rates due to the different surface conditions and driving patterns. There was a reasonably strong correlation within each of the road classifications (Table 8.7), with A-roads influencing the expected total by the greatest amount.

Table 8.8: Intercept and gradient results (with 95% CI) for vibration exposure against the root of duration in each axis, depending on road classification.

Axis	Road Type	Intercept ($m \cdot s^{-2}$ $A(8)$)	Gradient	R^2
x	U	0.002 (0.002–0.002)	0.008 (0.008–0.009)	.62
	AU	0.001 (0.000–0.001)	0.007 (0.007–0.008)	.75
	A	0.001 (0.000–0.001)	0.008 (0.007–0.008)	.87
	M	0.000 (−0.001–0.000)	0.007 (0.007–0.007)	.87
y	U	0.000 (0.000–0.000)	0.009 (0.009–0.009)	.65
	AU	0.000 (−0.001–0.000)	0.008 (0.008–0.008)	.72
	A	0.002 (0.002–0.002)	0.004 (0.004–0.005)	.45
	M	0.000 (0.000–0.000)	0.003 (0.003-0.003)	.56
z	U	0.002 (0.001–0.002)	0.012 (0.011–0.012)	.55
	AU	0.000 (−0.001–0.000)	0.016 (0.016–0.016)	.70
	A	−0.002 (−0.002–−0.001)	0.020 (0.019-0.020)	.87
	M	−0.002 (−0.002–−0.001)	0.018 (0.018–0.019)	.87

8.6.4.2 Vibration exposure

Examining the formula for vibration exposure (8.4.6), it can be seen that resultant values are proportional to the root of both the duration and sum of accelerations. It was therefore unsurprising that all axes increased with duration (r_s = .84, .65 & .88 for the x-, y- & z-axes respectively, $p < .001$ for all).

Performing linear regressions on the data recorded along each of the road classifications resulted in gradients — presented in Table 8.8 — that reflected the differences found within the average vibration in Section 8.6.2.1. Again, this was to be expected due to the formula for calculation.

Similar to the vibration exposure, VDV is also calculated using a factor of duration. In this case, results are linearly correlated with the 4^{th}root of duration (Equation 8.4.15).

Before progressing on to the analysis of the final edges of the NCH to LRI edges, it was important to address the impact of choosing the correct method of calculating metric values.

8.6.5 Influence of processing method on results

Studies which investigated the extent of vibration and noise during ambulance journeys, along with the potential correlations with infant response, are inconsistent in their reporting. Noise levels often appear to be reported incorrectly, while the full de-

tails of gathering and processing vibrations tend to be missing entirely. Therefore, this section will investigate both the effect that different processing methods can have on results, both to guide future research and to enable the results here to be compared with previous work.

8.6.5.1 Methods of averaging noise

Most studies which reported noise levels in relation to neonatal infants or ambulance journeys used the term "mean" when summarising values. Although taking the mean of the noise levels would be applicable in cases where multiple journeys which were recorded along the same stretch of road were to be compared, averaging noise levels recorded during a single journey requires calculation of the r.m.s. of SPLs (Equation 8.4.25). Studies, however, appeared to compute the means of all reported noise levels — often 1 s averages — regardless of whether or not they were recorded consecutively. Computation of average noise levels using various methods were therefore performed for each edge both to identify the effect on the resultant values and to enable comparison with previous work.

The first method compared with the SPL r.m.s. was calculating the r.m.s. of the noise levels in decibels. Despite using the r.m.s. to account for the variations, decibel averaging was found to underestimate the true noise level by an average of 2.0 dB, as shown in Figure 8.13A. This was inevitable due to the logarithmic nature where an increase of 10 dB results in the equivalent SPL increasing by a factor of $\sqrt{10}$. Therefore, performing averaging in decibels applies reduced weighting to higher levels, leading to an underestimate.

Computing the mean of the decibel values was also found to underestimate average noise levels (Figure 8.13B). By squaring all inputs, r.m.s. places a greater significance on the larger values, as opposed to the mean in which all values have equal significance. Combining the reduction in weighting caused by the choice of the mean with the reduction from the use of decibel values therefore result in both a larger underestimate (mean: -2.1 dB) and a greater spread of data compared to the r.m.s. of decibel values.

The final incorrect averaging method investigated involved converting noise levels to SPL before calculating the mean value. Figure 8.13C shows this process resulted in a much closer approximation of the true noise level than either of the methods that computed averages using decibel values, although there remained a 1.0 dB underestimate.

All alternative methods of averaging noise levels resulted in a lower reported noise

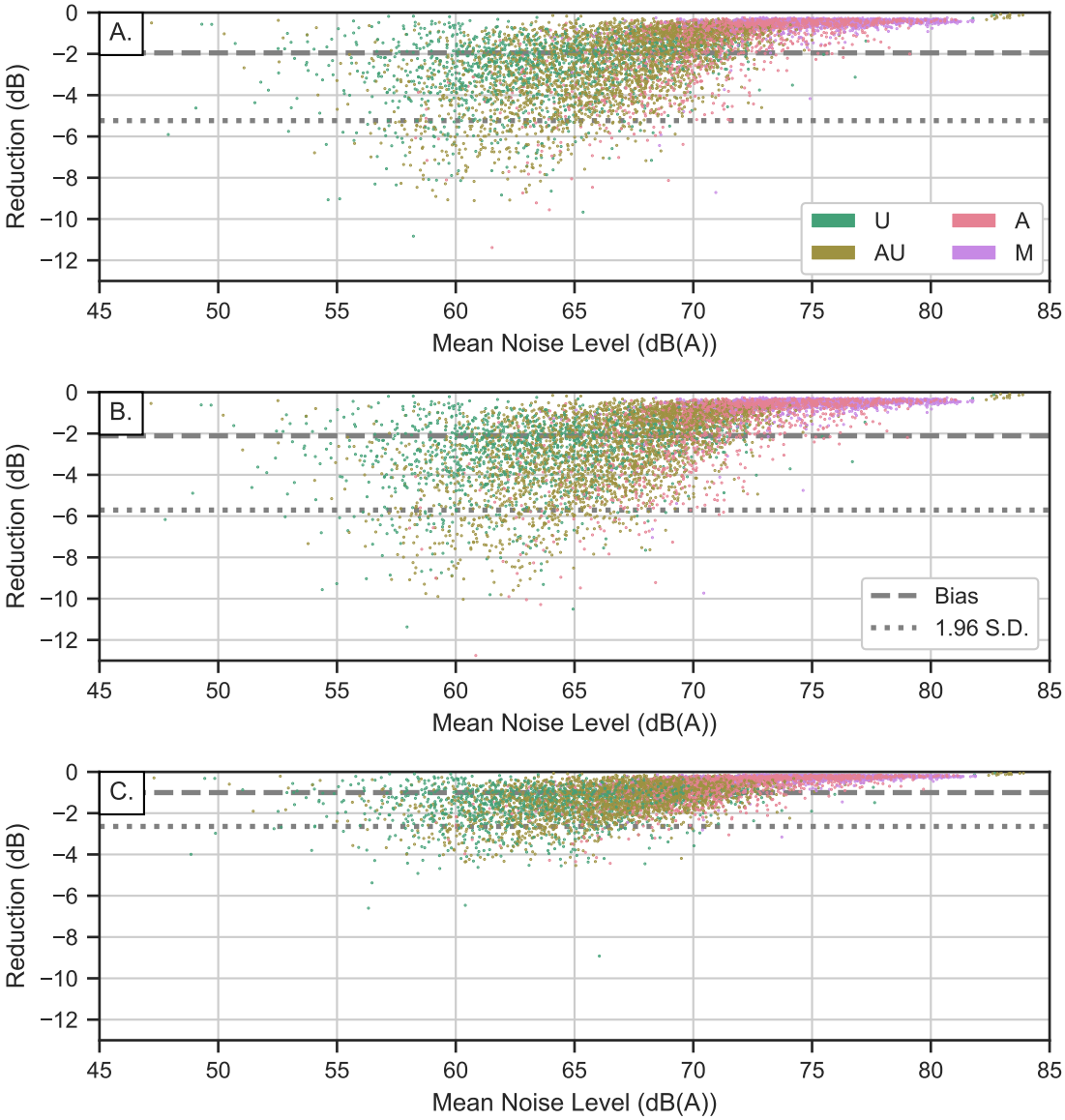

Figure 8.13: Comparisons calculating average noise level by taking the r.m.s. of dB values (A), mean of dB values (B) and mean of SPL values (C) against the correct r.m.s. of SPL values.

level compared to the true value. The worst performing method, with the greatest discrepancy, was the mean using decibels. Unfortunately, it is believed that this is the method implemented by previous studies and therefore these values will also be reported in the remainder of the results. While the reference to "mean" noise levels in other studies could theoretically refer to the computation of mean values using SPLs, and therefore a smaller underestimate, it is highly unlikely that the researchers would have recognised the need for conversion to a linear scale while not knowing that averaging by r.m.s. should be performed.

8.6.5.2 Averaging period choice

While Karlsson et al. [40] reported that vibrations were processed according to ISO 2631, the authors did not outline whether the reported peak values were as sampled or time-averaged. It was assumed that 1 s averaging was in place during their research due to both the magnitude of the values reported and observing medical researchers tendency to use averaged values for simplicity. Therefore, both as-sampled and time-averaged data recorded during the data collection here were compared.

Plotting all recorded peak values for each journey segment shows that averaging vibrations over a second results in an average reduction of 36% (Figure 8.14). Comparing these values to the maximum reported peak by Karlsson et al., 2.9 $m \cdot s^{-2}$, suggests that a 1 s averaging was indeed implemented during the study. Forty-two percent of the sampled vibrations from all journey segments were found to exceed the Karlsson threshold, far greater than the 3 edges that registered averaged vibrations above. Although it is possible that the journeys recorded by Karlsson et al. were along smoother roads than those here, the fact that 18% of sampled peaks also exceed the maximum shock the authors registered during loading of the ambulance suggests that the roads were not the reason for discounting sampled vibrations. Additionally, Karlsson et al. recorded vibrations within the incubator that has previously been shown to lead to greater magnitudes compared to the trolley [44] — where the smartphones were recording data — which may explain why the 1 s values here were all lower.

8.7 Edge Comparisons

After the trends within the data were explored, the edges themselves were compared. While understanding what may cause higher, and therefore worse, values for metrics is useful in providing generalised recommendations, it is important to determine if any

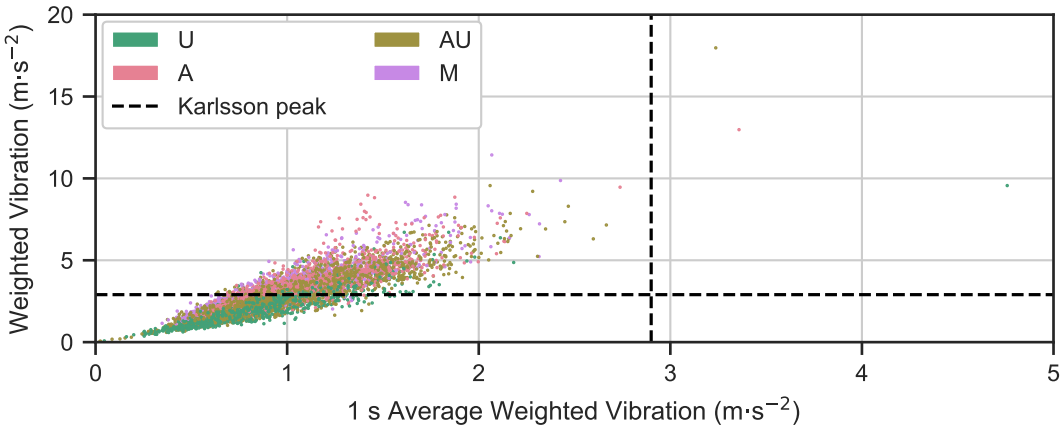

Figure 8.14: Comparison of peak comfort-weighted vibrations as sampled vs. averaged over one second.

specific edges should be avoided.

Thirty-nine separate journeys were recorded travelling directly from NCH to LRI between 24th October 2018 and 14th October 2019, travelling along 13 different routes as identified in Chapter 7. These routes comprised of 34 unique sections of road which could then be combined to generate 31 routes (an increase of 2.4× the recorded number). These sections, or edges in graph terminology, varied in both length and road type.

The reliability of the edge assessments was maximised by including data from all available recordings, resulting in a fifteen-fold increase in the total number of contributing journeys to 588. These additional datasets caused the original edges to be divided up into the 114 unique sections of road used in the previous section. The average number of journeys which contributed towards the analysis of each edge increased five-fold from 9 (IQR: 3.5–18) to 50 (IQR: 17.5–91) through inclusion of all available recordings.

As with the previous section, edges were grouped by road classification. This required manually assessing the roads used and resulted in the classifications shown in Table 8.9. A fifth classification, mixed urban/motorway, was required due to the lack of travelled options between junction 21 of the M1 motorway and LRI. All figures within this section will again distinguish between the types of road through use of colour.

Edges will be indicated in-text with a prefixed "*E*" to differentiate from regular numbers.

All possible routes would travel along either *E14* or *E17*, evenly split with 16 and 15 routes respectively. These edges are markedly different, consisting of mostly concrete A-road and multi-lane motorway respectively. Comparing the metric values along these edges should give an indication of the effect of road surface and type on

Table 8.9: Type of road forming at least 75% of each edge

Road Classification (ID)	Edge Numbers
Urban (U)	0, 1, 2, 3, 4, 6, 8, 12, 21, 24, 30, 31, 32, 33
A roads (A)	14, 15, 16, 19, 20, 23, 26, 27, 28
A & Urban (AU)	5, 9, 10, 11, 13, 22, 25, 29
Motorway (M)	7, 17
Motorway & Urban (MU)	18

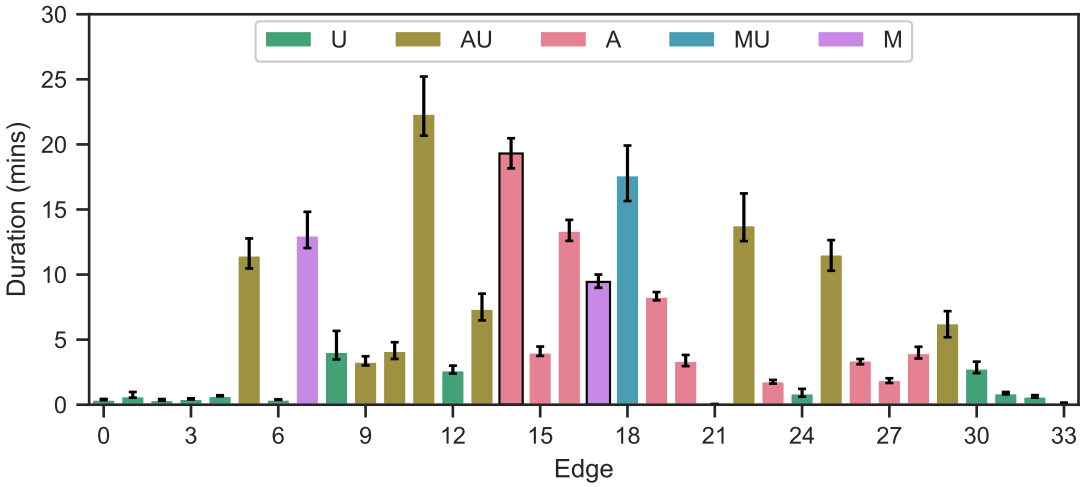

Figure 8.15: Average duration for each individual edge between NCH and LRI.

comfort. Each of these edges will be emphasised within the below figures by a bold outline.

8.7.1 Duration and Vehicle Speed

Time spent travelling along edges varied greatly from less than 1 to over 22 minutes (Figure 8.15). The edges with shorter durations tended to consist of urban roads in close proximity to the hospitals, with the longer edges being in rural areas.

Twice as much time was spent travelling along edge *E14* (19.3 minutes) than along *E17* (9.4 minutes) and thus any further comparison needed to be time-independent.

Vehicle speeds varied as much as 24 $m{\cdot}s^{-1}$ (54 mph) between edges (Figure 8.16). As with the durations, the lowest average speeds occurred along the small edges close to the hospitals and the higher speeds were further into the countryside. Travelling along roads with a lower speed limit resulted in a reduced average in *E14* (23 $m{\cdot}s^{-1}$ / 52 mph) compared to *E17* (27 $m{\cdot}s^{-1}$ / 61 mph). This difference in speed needs to be accounted for in the assessment of any variation in metric values between these edges.

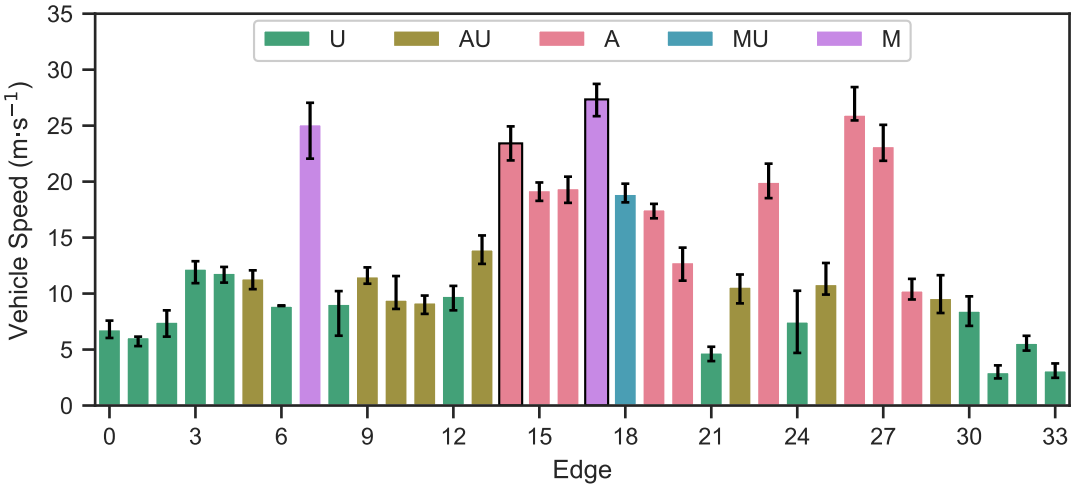

Figure 8.16: Average vehicle speed for each edge.

Average speeds were found to decrease significantly with road classification, as shown earlier. Fastest edges were those which travelled along a motorway, resulting in an overall average speed of 26 $m \cdot s^{-1}$ (59 mph). Next fastest were the A roads, at 19 $m \cdot s^{-1}$ (43 mph), followed by the urban edges, 7 $m \cdot s^{-1}$ (16 mph).

8.7.2 Unweighted Vibration

Assessment of the vibration as-recorded would provide an indication as to the severity along each edge. This was performed both in terms of the average power of the vibration and the frequencies at which it was most prevalent.

8.7.2.1 Average Power

Raw unfiltered vibration was greatest in the z-axis for all edges, followed by the x- and then y-axes. Average z-axis vibration (0.520 $m \cdot s^{-2}$) was just shy of 2 and 3 times as large as the x- & y-axes (0.269 and 0.183 $m \cdot s^{-2}$ respectively) as shown in Figure 8.17. Edge values for vibration varied hugely for each axis, with ranges in the x-, y- and z-axes of 0.142, 0.099 and 0.382 $m \cdot s^{-2}$ (53, 54 & 73% of the mean). No correlation was found between the axis values of each edge as coefficients were either very low (x-y: $R^2 = .36$) or around zero (x-z: $R^2 = .07$, y-z: $R^2 = .00$).

Despite travelling 4 $m \cdot s^{-1}$ slower on average, *E14* resulted in greater magnitude vibrations than *E17*. Each of the axes along the concrete route were higher than the motorway (x: 26%, y: 32%, z: 19%) which combined to give a 21% increase in the total vibration.

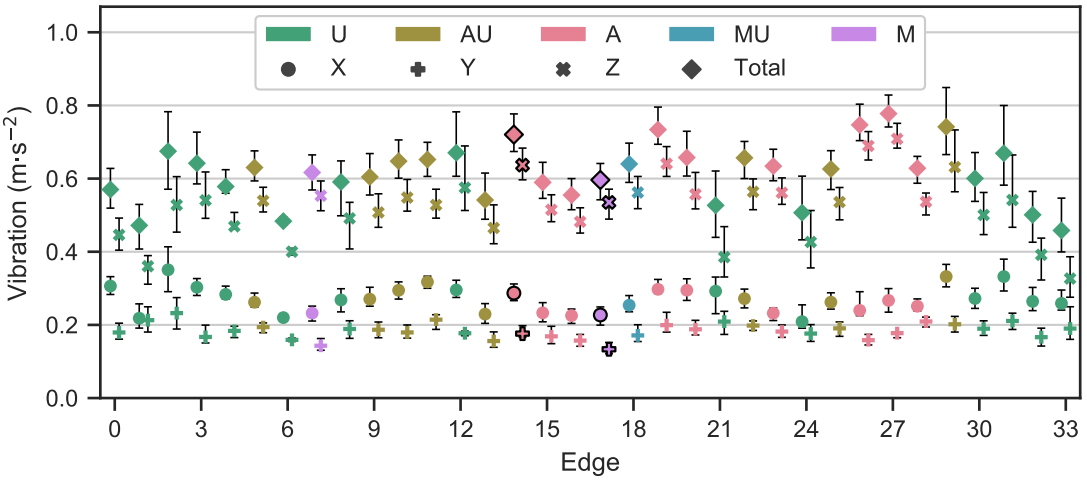

Figure 8.17: Unfiltered average vibration in each axis for each edge.

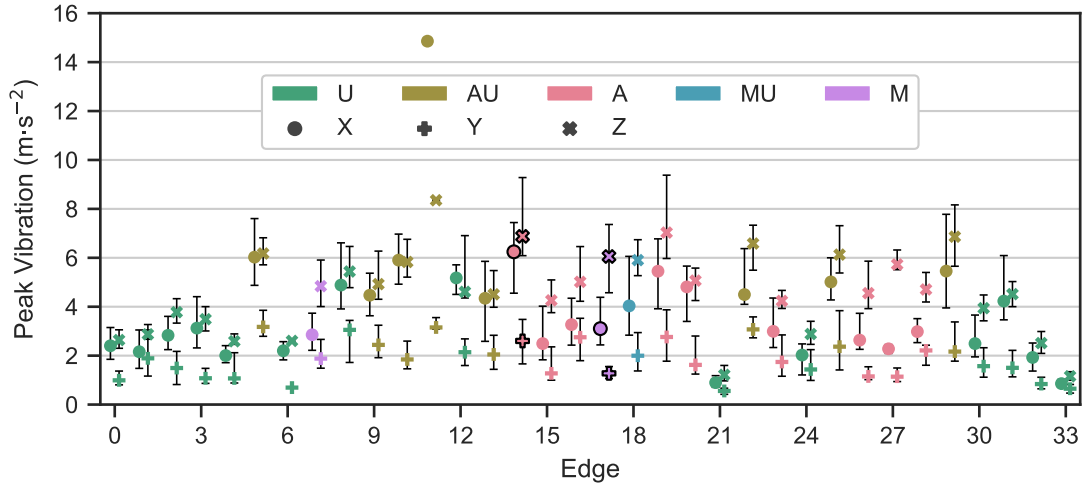

Figure 8.18: Unfiltered peak vibration in each axis for each edge of the NCH to LRI network.

8.7.2.2 Shocks

There were no recorded accelerations which breached a magnitude of 1 g for 20 ms, in any axis on any edge. Instantaneous accelerations rarely exceeded 1 g either, with only the worst 5% of journeys along 4 edges found to recorded sufficiently large z-axis values. This increased to 6 edges (*E27*, *E11*, *E29*, *E19*, *E14* & *E22*, in descending order of rate per minute) when using the ½ g threshold, suggesting that peak vibrations were minimal throughout. The highest peaks tended to be found in the vertical axis, with greater magnitudes than all y-axis peaks and all but 3 x-axis peaks (Figure 8.18). *E11* registered the greatest values for peak acceleration both vertically and parallel to the direction of traffic, although the anomalous x-axis peak appears to have been influenced by a section of road that only a single journey used.

Unfiltered accelerations fared worse along the concrete road compared to the motor-

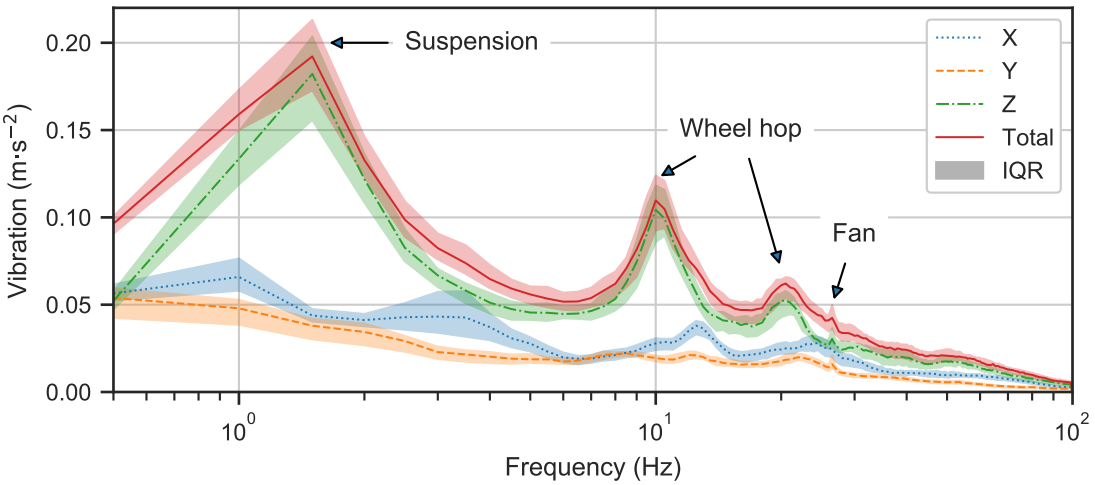

Figure 8.19: Comparison of the spread of edge vibration magnitudes in each axis, calculated from the median value of each edge in 0.5 Hz intervals.

way. Previously, the average vibrations in each axis was greater for *E14* than the counterparts on *E17*, and the same is true for the instantaneous shocks. The maximum shock magnitude in the x-axis for *E14* was even as large as, if not slightly greater than, the z-axis for *E17*, suggesting that a more stable vehicle speed was maintained along the motorway.

8.7.2.3 Frequency Bands

The majority of vibration in all frequencies occurred in the vertical axis (Figure 8.19). Most of the z-axis vibration was likely caused by the natural frequency of the vehicle suspension (1.5 Hz) and wheel hop (10.0 Hz), with over 29 and 14% of the combined vibration occurring at or around these frequencies. A further peak was evident around 20.0 Hz, and while the cause was unclear it is possibly a harmonic of the wheel hop. Two frequencies which had the highest levels of correlation against vehicle speed were 1.5 Hz (R^2 = .54) and 21.0 Hz (R^2 = .50) with all frequencies trending upwards in magnitude as speeds increased.

Peaks in the x-axis vibrations occurred at 1, 3.5 and 12.5 Hz frequencies. Lower frequencies (<9.0 Hz) were most susceptible to changes in vehicle speed and increased as speed reduced, having a maximum correlation of R^2 = .62 at 4.5 Hz, possibly caused by the increased proportion of accelerating and decelerating events. Vibrations above 9.0 Hz tended to increase with vehicle speed.

Slower vehicle speeds caused magnitudes in the y-axis to increase below 10.0 Hz but decrease at higher frequencies, following a similar trend to the x-axis. A high correlation with speed at low frequencies (R^2 = .82 at 2.5 Hz) was likely due to sharper

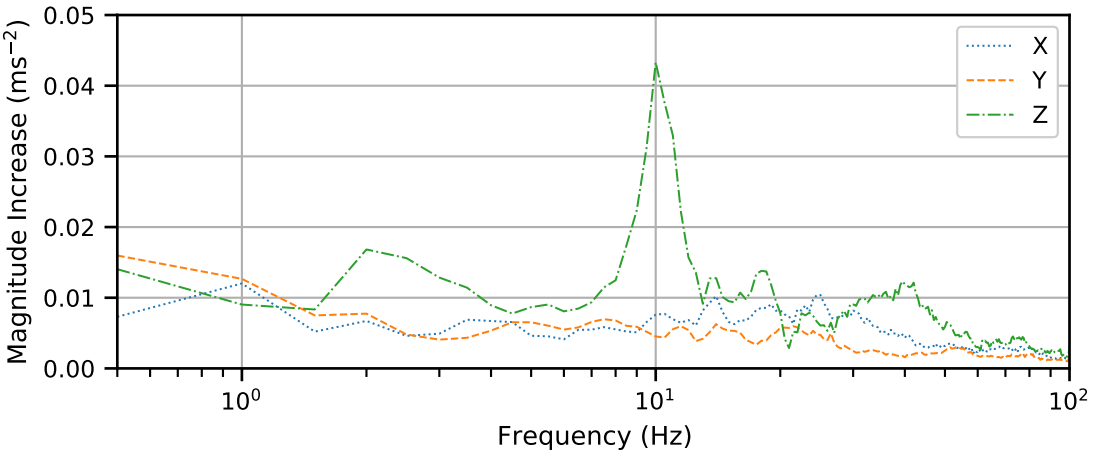

Figure 8.20: Increase in vibrational magnitude along a concrete portion of road (*E14*) compared to a motorway (*E17*), calculated for each axis in 0.5 Hz intervals.

cornering in urban areas. Aside from this correlation, few features were evident in the side-to-side vibration.

All edges and axes registered sharp spikes of vibration at 26.5 Hz, which has previously been suggested to be caused by a fan on the neonatal trolley [44]. Being caused by transport equipment may explain the presence of this spike in all axes, and at such a precise frequency, however it was unable to be replicated during tests run in the lab using a standard transport incubator. Also, no equipment which possessed a fan that spun at 1,590 rpm was evident when investigating the components used during neonatal transport.

Journeys along the concrete-surfaced roads registered higher magnitudes at all frequencies and in all axes. The largest increase occurred in the z-axis at 10 Hz, around the frequency of wheel hop (Figure 8.20). Increase in wheel hop, along with both x- & z-axis values being larger for the slower edge, indicate that the vibrational differences between these edges was predominantly due to the road surfaces.

8.7.3 Frequency weighted vibration

While the raw vibration reports the true input from the road, it does not indicate what effect may be had on neonatal infants. The ISO 2631 standard for the evaluation of human exposure to WBV was therefore used to assess the impact of the recorded vibration on adults as a substitute for a neonatal index. The standard required applying different frequency weightings depending on the axis, and was then assessed in terms of the impact on both health and comfort.

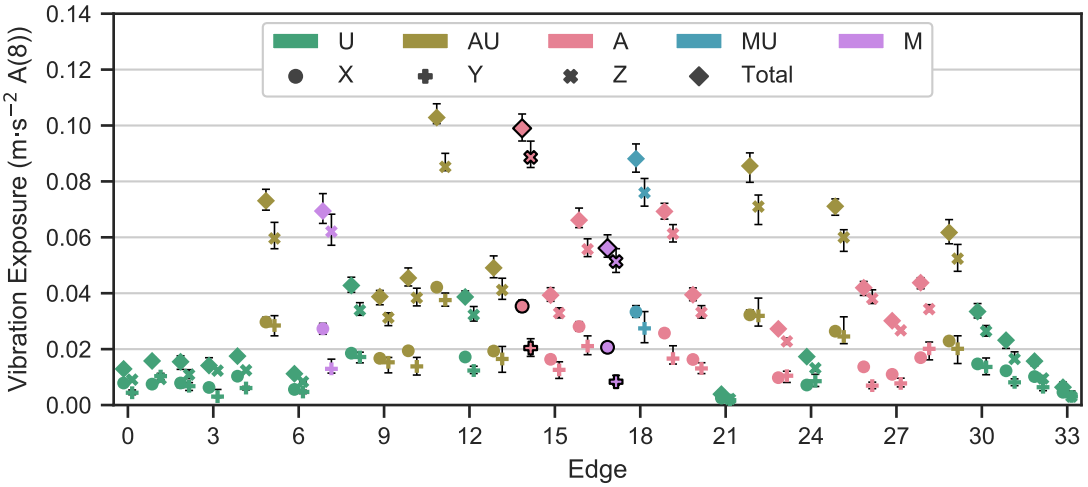

Figure 8.21: Average vibration exposure in each axis for each edge.

8.7.3.1 Health impact

Edge values for vibration exposure varied massively, by over twice the mean, for each axis (Figure 8.21), due to the range of edge durations. No values were close to the action limit of 0.5 m·s^{-2} A(8), however this was expected as the longest edge duration was less than 5% of the reference duration of 8 hours.

Vibration exposure was 6% greater than the fitted relationship along edge *E14* while 14% less than the expected value along *E17*. This would mean that only 8.2 hours of driving along the concrete surface would be required to reach the action value, compared to 12.5 hours of driving along the motorway.

8.7.3.2 Vibration dose value

Vertical accelerations resulted in vibration dose values almost twice as large as the horizontal axes on average (Figure 8.22). One edge (*E33*) registered VDVs larger in both x- & y-axes than in the vertical, while two additional edges resulted in the vertical value being less than either the x-axis (*E21*) or y-axis (*E1*). Edge *E11* recorded the highest total VDV but was only 51% of the action value. Average edge VDV was 2.57 m·s$^{-1.75}$.

Driving along the concrete road surface of *E14* would reach the VDV action value in half the time (6.7 hours) as driving along the motorway in *E17* (13.6 hours). Comparing with the earlier results, it can be seen that *E14* is dominated by sharp spikes of vibration, as it required more time to reach the action limit for vibration exposure, whereas the motorway was at a constant level with few peaks.

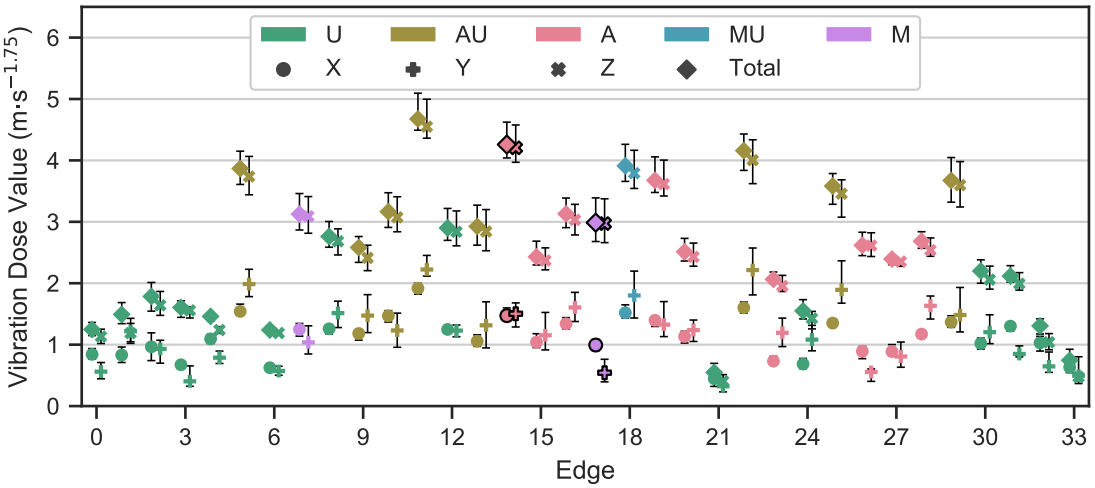

Figure 8.22: Vibration dose values in each axis for each edge.

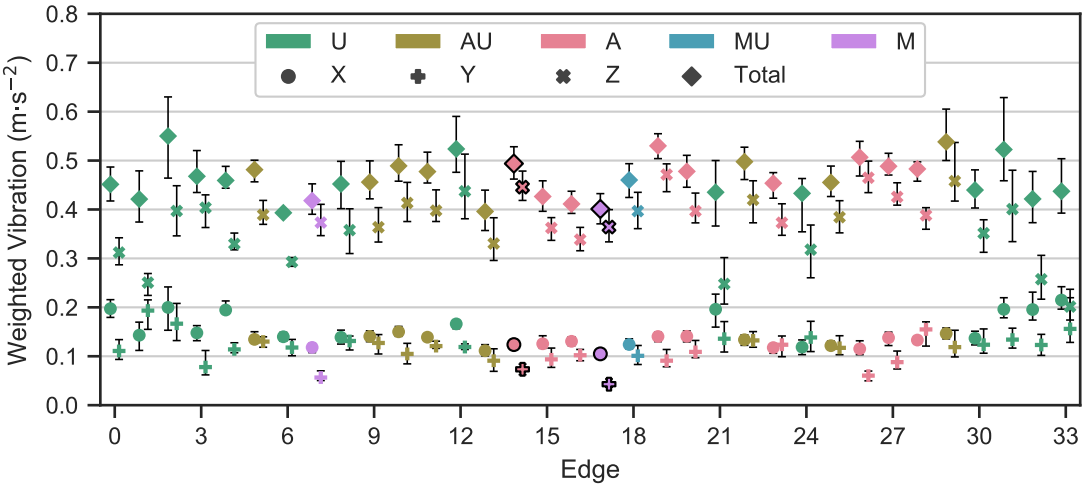

Figure 8.23: Comfort-weighted vibration in each axis for each edge.

8.7.3.3 Perceived comfort levels

Comfort-weighted vibrations were greatest in the vertical axis compared to the horizontal axes (Figure 8.23). Horizontal vibrations had less than half (x: 42%, y: 33%) the magnitudes of those in the z-axis. All edges were classed as "not uncomfortable" in both the x- & y-axes. Average vertical vibration, along with 28 (82%) of the edges, fell into the "a little uncomfortable" category. Combining all axes resulted in at least "a little uncomfortable" vibration for all edges, with 6 (18%) edges within the "fairly uncomfortable" classification.

Although they were both at "a little uncomfortable" levels, total comfort-weighted vibration was worse along *E14* than *E17*. Vibrations caused by the concrete surface were 23% greater than along the motorway and were only 1% (0.006 m·s^{-2}) less than the starting threshold of "fairly uncomfortable".

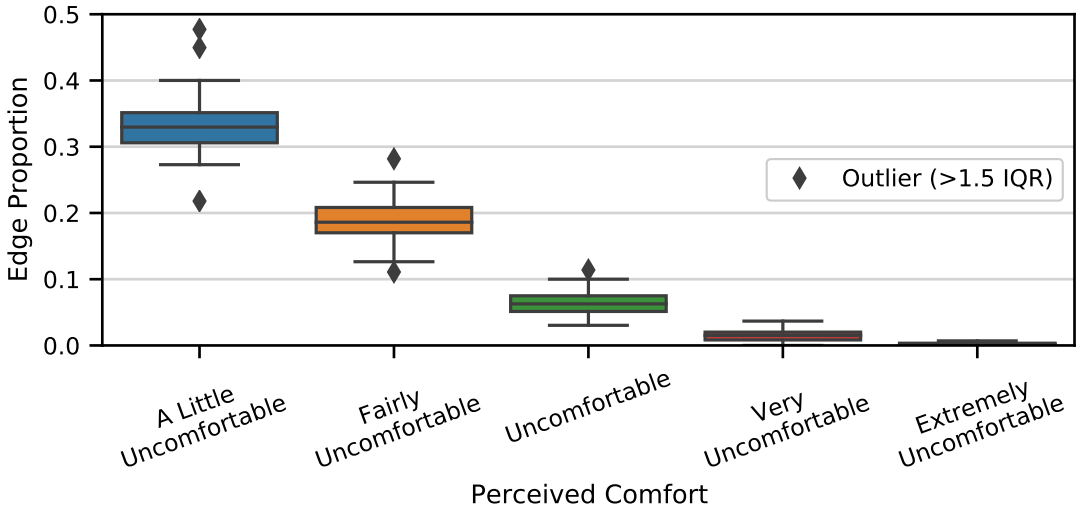

Figure 8.24: Spread of average edge proportion, in terms of duration, spent with total combined vibration within each of the defined comfort levels.

Most edges spent similar proportions within each of the levels of discomfort (Figure 8.24). The majority of discomfort was classed as "a little uncomfortable", as expected from the overall average values. Proportions rapidly decreased with discomfort which suggested higher magnitudes are caused by localised events and road features, as opposed to surface quality. Outliers were caused by edges with shorter durations, where a few shocks will skew proportions.

Stark increases in vibration along *E14*, compared to *E17*, were evident when examining the average proportions. Both edges spent approximately similar amounts of time at "a little uncomfortable" (*E14* 36% vs *E17* 31%), however they were indicated by the *t*-test as significantly different ($t(212) = 6.7$, $p < .001$). The differences between the two edges also increased with vibration magnitude. Increasing proportions at "fairly uncomfortable" (22 vs 14%; $t(212) = 7.0$, $p < .001$), "uncomfortable" (8 vs 4%; $t(212) = 3.7$, $p < .001$) and "very uncomfortable" (2 vs 1%; $t(212) = 2.9$, $p < .01$) emphasised how much worse the concrete road was compared to the motorway. Differences could also be seen in the proportion of time spent at vibration classed as "not uncomfortable", with 63% of the motorway edge not affecting adults compared to less than half (49%) of the concrete edge ($t(212) = –7.4$, $p < .001$).

Trends displayed for each edge were highly similar, with the majority of vibration occurring below 0.4 $m{\cdot}s^{-2}$ (Figure 8.25). The most commonly detected weighted vibrations for 18 (53%) of the 34 edges registered magnitudes between 0.20 & 0.25 $m{\cdot}s^{-2}$, with a further 12 edges (35%) within 0.05 $m{\cdot}s^{-2}$ of this bound. Edges which exhibited a greater spread of vibration across magnitudes tended to be those which had short durations in close proximity to hospitals (*E0*, *E21*, *E32* & *E33*).

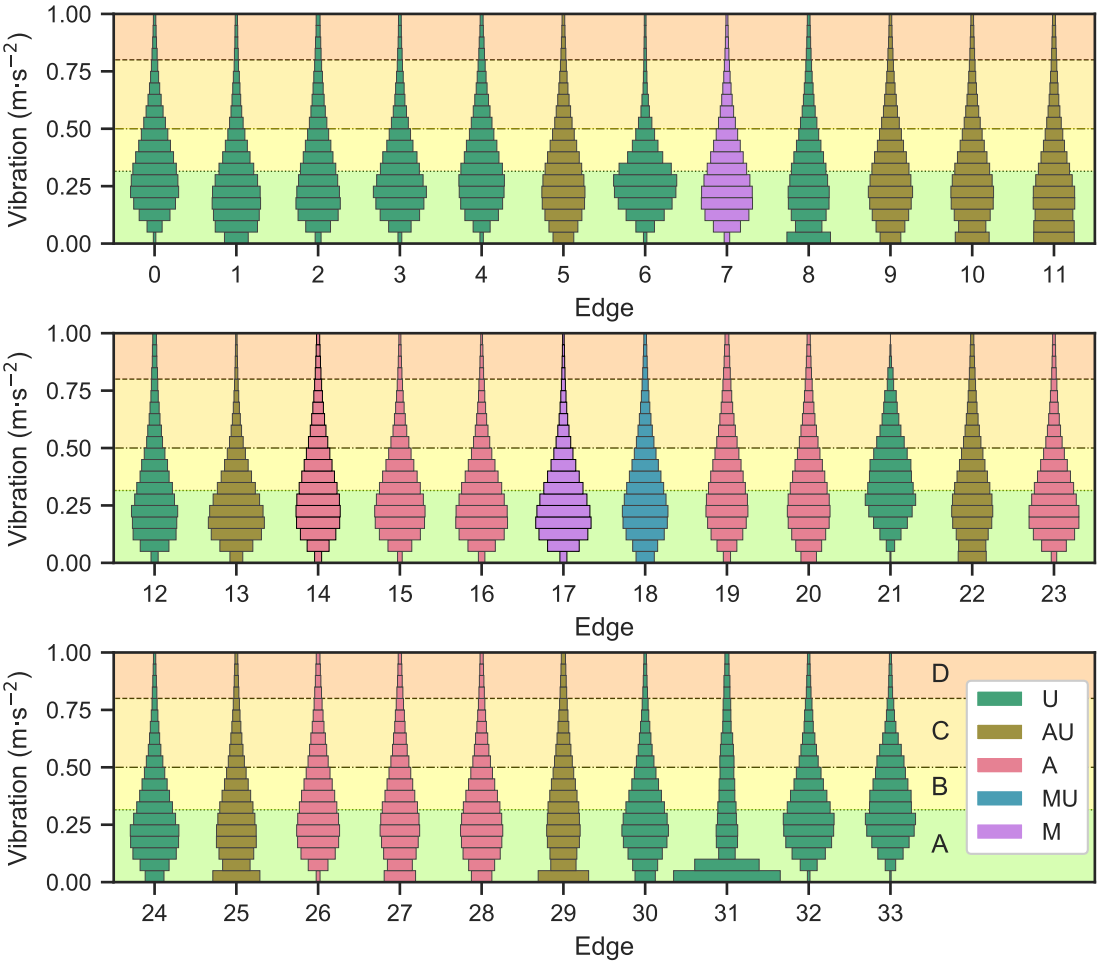

Figure 8.25: Proportion of time spent by each edge at different levels of weighted vibration, in accordance with ISO 2631, and how they compare to the "not uncomfortable" defined level (A) and lower limits of the "a little uncomfortable" (B), "fairly uncomfortable" (C) and "uncomfortable" (D) defined levels.

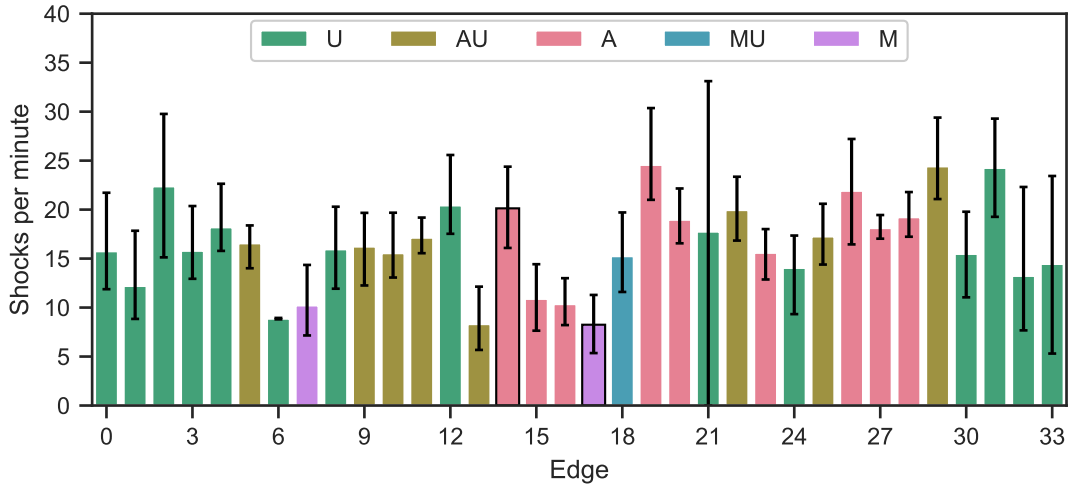

Figure 8.26: Rate of comfort-weighted shocks above a magnitude of 0.5 m·s^{-2} for each individual edge of the NCH to LRI network.

Comparing the spread of magnitudes along the motorway and the concrete A-road provided clear differences. *E17* recorded 84% of weighted vibrations below the "fairly uncomfortable" lower threshold of 0.5 m·s^{-2}, compared to only 74% of vibrations along *E14*. All magnitudes between 0.05 & 0.35 m·s^{-2} were more prominent along the motorway, with the greatest difference (3%) occurring in the 0.10–0.15 m·s^{-2} bound. The A-road registered an average of 2.3× greater a proportion of time than *E17* for each bound above 0.35 m·s^{-2}.

8.7.3.4 Peak events

A total of seven edges recorded at least 1 s of WBV above the 2 m·s^{-2} set by Bouchut et al. [43] and these were only during the worst 5% of journeys. Instead, the spread of shocks at a minimum of "fairly uncomfortable" magnitude were examined (Figure 8.26) after observing proportions varying in Figure 8.25. Most edges tended to produce an average rate of 16 shocks per minute, rising to 28 in the worst cases.

Figure 8.27 shows the distribution of maximum 1 s WBV values for each edge, where the worst edge recorded a peak of 1.6 m·s^{-2}. The majority of edges only registered levels of WBV around 1.2 m·s^{-2}, suggesting that either the ambulance or road conditions were worse for Bouchut et al. than in this study, or the lower levels found here were due to being at the trolley surface instead of inside the incubator.

The concrete road surface of *E14* resulted in a greater number of shocks than the motorway of *E17*, registering magnitudes above 0.5 m·s^{-2} at over twice the rate. Similarly, overall peaks were far worse along the A-road with the lower quartile of *E14* covering similar levels to the upper quartile of *E17*. Peak 1 s WBV along *E14* were also within

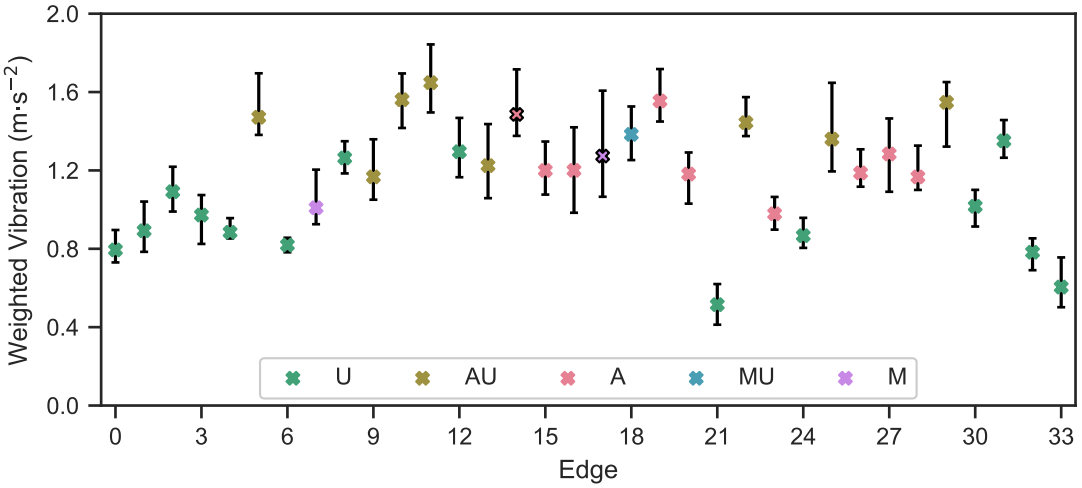

Figure 8.27: Maximum ISO-weighted vibration, averaged over 1 s intervals, for each individual edge of the NCH to LRI network.

the range of "very uncomfortable", compared to only "uncomfortable" for *E17*, although the guidance is unclear on the instantaneous comfort.

8.7.4 Noise

Noise levels were also sampled during journey recordings, however there is no standard — either for adults or infants — which specifies an expected response. Instead, both the average and the spread of the noises were calculated and compared, in preparation for when the effect on a neonate is known.

8.7.4.1 Average Power

Average noise levels varied by over 15 dB between edges, as shown in Figure 8.28. The majority of noise levels breached the 60 dB(A) equipment recommendation for transport, with all journeys having an equivalent level above the NICU recommendation. The mean of all edges was at 70 dB(A), which is commonly compared to the expected noise alongside traffic, while the extremes ranged from normal conversation levels to around the level of a vacuum cleaner.

Motorway noise was found to occur at a higher level than the concrete stretch of road, however the differences were minimal (< 1 dB) and did not warrant further discussion.

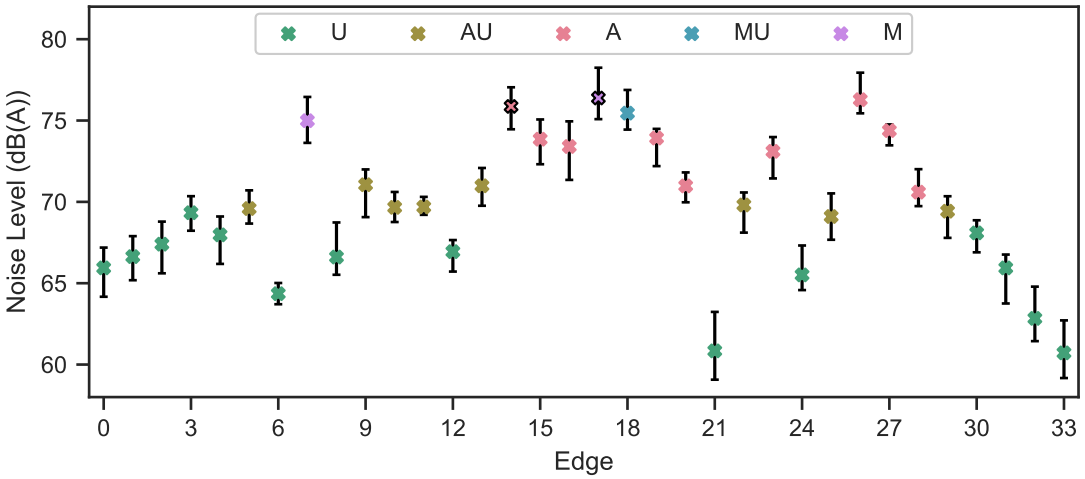

Figure 8.28: Average noise level for each individual edge between NCH and LRI.

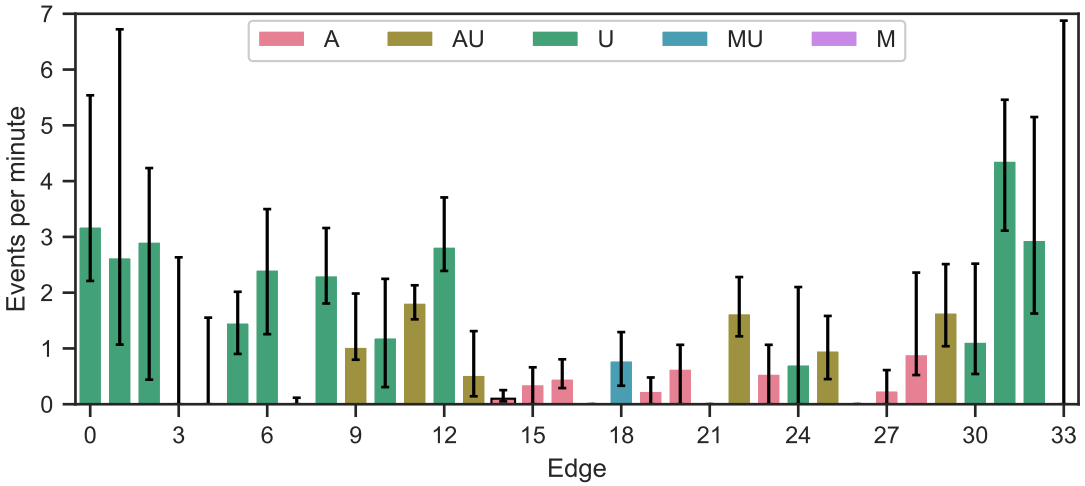

Figure 8.29: Rate of 5–10 dB increases in 1 s averaged noise levels, shown for each edge of the NCH to LRI network.

8.7.4.2 Spikes

Abrupt increases in 1 s average noise level were found to occur along all edges, although at a much higher rate in urban environments, as shown in Figure 8.29. Most frequently occurring spikes in noise were 5–10 dB louder than the ambient (mean: 1.2 per minute), but significantly reduced in frequency for the 10–15 dB range (mean: 0.1 per minute). Increases of greater than 15 dB rarely occurred and were only present when considering the worst 25th percent of journeys, of which 7 (21%) of edges experienced some events.

There were no journeys, along any of the edges, that recorded any noise levels over the 85 dB(A) threshold used by Bouchut et al. [43] when averaging over the course of a second. The lack of sustained levels above 85 dB(A) can be further seen when examining the instantaneous maximums in Figure 8.30, where the average of the maximums

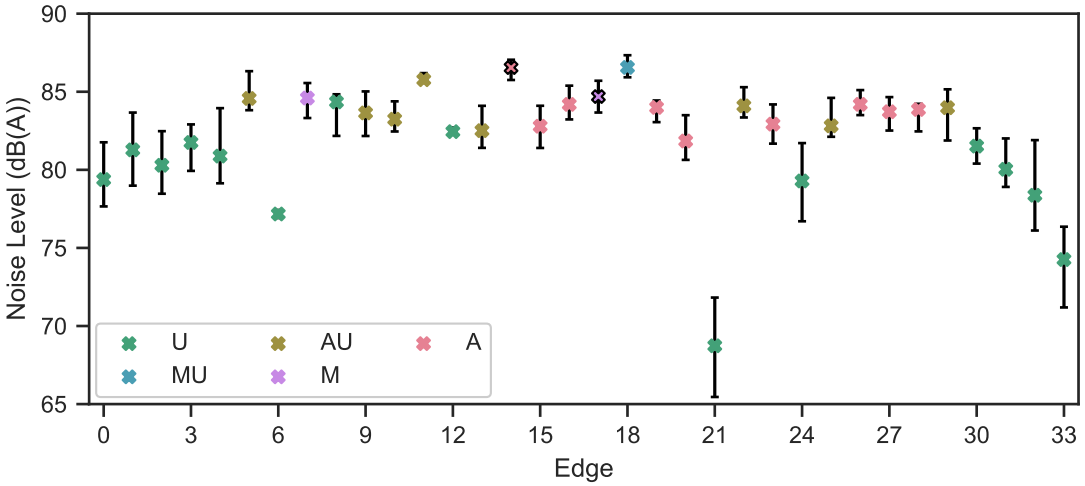

Figure 8.30: Maximum sampled noise level for each individual edge between NCH and LRI.

recorded along each edge was 83 dB(A) and most edges did not reach the threshold. Only 3 edges registered a peak value greater than 85 dB(A), which rose to 11 when considering the 75th percentile of journeys along each edge. Reducing the threshold to 75 dB(A) resulted in several edges registering values, however these were the edges with the highest vehicle speed — following the relationship identified in Section 8.6.3.1. *E21* registered significantly lower maximum noise levels than other edges, likely reflecting the combination of short duration and low speed.

The motorway edge of *E17* recorded zero spikes in noise level over 5 dB, accounting for 95% of the journeys which traversed it. On the other hand, around two spikes of 5–10 dB could be expected during the average journey along *E14*. Instantaneous peak noise levels were also 1 dB lower for *E17* compared to *E14*. Although this shows that the motorway noise is more stable which could be beneficial, the differences appear minimal and any benefit may be offset by the slightly greater average levels.

8.7.4.3 Repetitive Exposure

Each edge displayed a different trend in the registered noise levels (Figure 8.31). Most common noise levels ranged from 46–48 dB(A) to 78–80 dB(A), with an average of 68 dB(A). The edges which resulted in the highest average speeds (*E17*, *E26*, *E7* and *E14*, in descending order) recorded the majority of noises at levels above 70 dB(A). Faster edges also tended slightly to be more focused at individual noise levels, with the maximum proportion at a single 2 dB interval increasing with speed ($R^2 = .45$, $p < .001$).

Journeys along the motorway had a more prolonged exposure to loud noises than the concrete road. *E17* registered almost 100% of noises at levels of 70 dB(A) and above,

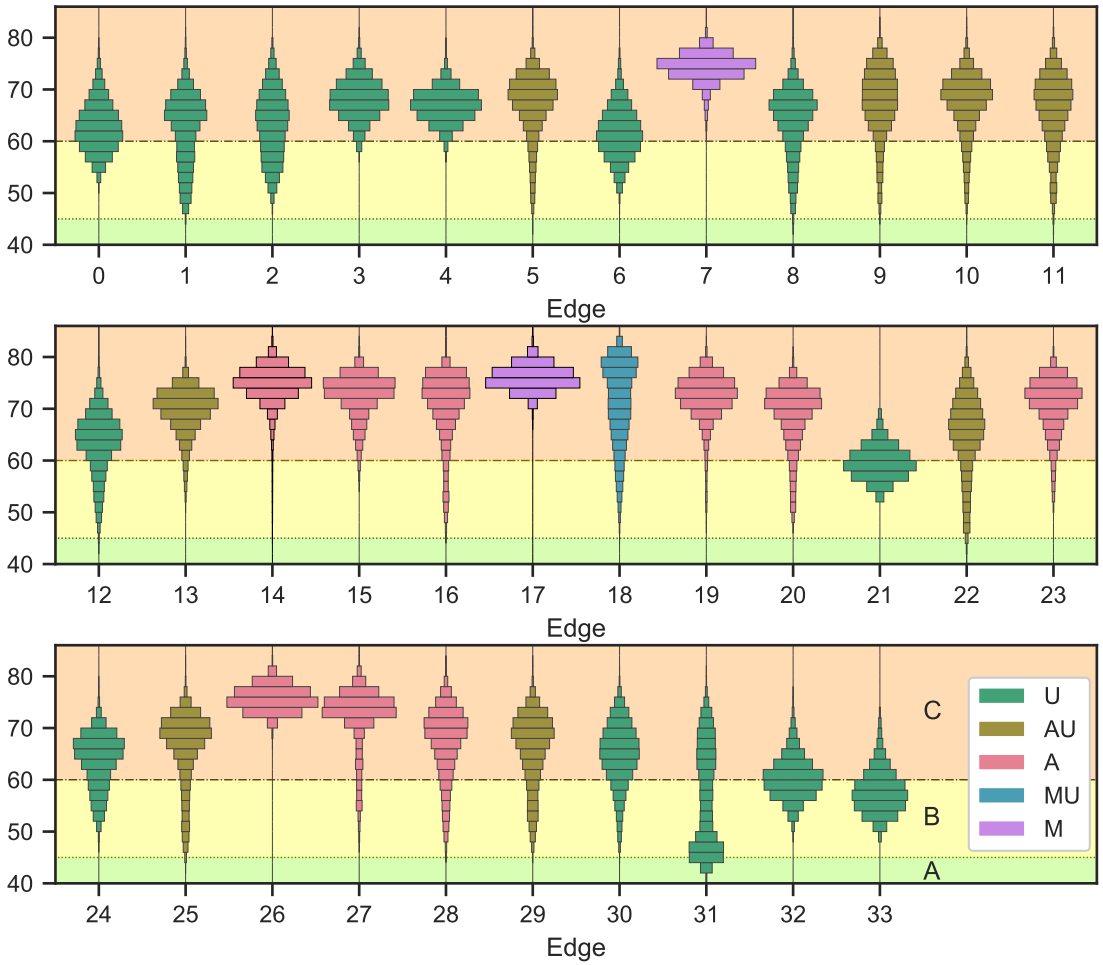

Figure 8.31: Proportion of time spent by each edge at different noise levels, calculated in 2 dB intervals. Background colours indicate whether the noise level is deemed adequate for NICUs (A), exceeds recommended NICU levels (B), or also exceeds the recommended transport equipment limit (C).

Table 8.10: Summary of the routes driven from NCH to LRI.

Route ID	C	D	G	I	J	M	N	P	R	T	V	W	AA
Number of Journeys	1	1	3	3	1	2	2	8	1	3	2	11	1

whereas *E14* only registered 92%. The largest differences were in the ranges of 74–76, 76–78 and 78–80 dB(A) for which a higher percentage of *E17* occurred (6, 6 & 4% respectively).

8.8 Route Comparisons

The overall aim of this project was to examine whether any significant differences in either vibration or noise could result from ambulances driving along a specified route. This therefore required the edges of the NCH to LRI network, and their assessed metric values, to be combined to form unbroken sequences connecting the two destination hospitals.

The 34 identified edges were able to be combined to form 31 routes, each designated by either a single or double capital letter (Table 7.10). These routes varied in distance and consisted of multiple road types, from small urban roads to large motorways. Metrics values for all routes were calculated from the constituent edges as described in Section 8.5.2.

Only 13 (42%) of the 31 possible combinations of edges were used by ambulances driving directly from NCH to LRI, with *W* being the most frequently chosen route (Table 8.10). Examining the number of times each route was chosen and comparing to the value of each metric would provide a true indication as to whether improvements in routing could reduce discomfort. These frequencies will be shown on all figures in this section using a colour gradient, where white indicates a route that was not used.

As with the comparison of edges, routes will be assessed using metrics of time, vibration and noise.

8.8.1 Route duration and vehicle speed

Average route duration varied from 49 to 63 minutes, with a median of 55 minutes. Figure 8.32 displays the variation within each route along with how they compared to one another. Route *G* was found to be the quickest and *BB* the slowest of the route combinations, with *G* also given as the quickest route by Google Maps at the time of

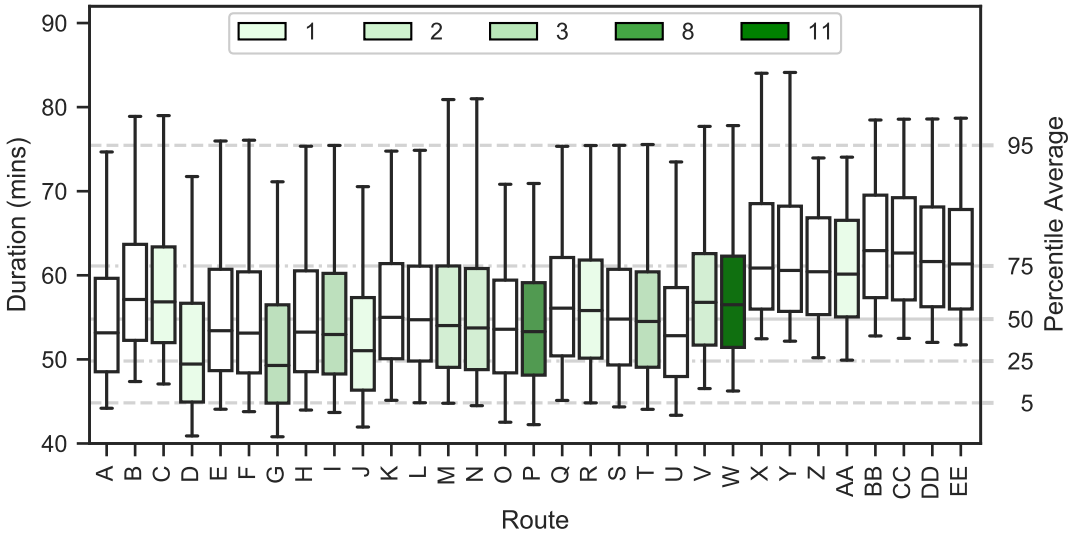

Figure 8.32: Comparison of the combined durations for each route.

writing (June 2021). Half of all durations for each route were within 8 minutes of the route average. 5% of route durations tended to be over 20 minutes slower than usual, whereas the fastest 5% were at least 10 minutes quicker.

The longest durations were all found to have edge *E11* in common, suggesting this was the confounding factor. Similarly, the shortest durations were along the routes which used all 3 edges which consisted of motorway sections (*E7*, *E17* & *E18*).

Only 8% of the journeys which actually travelled from NCH to LRI used the fastest route (Figure 8.32), however this was likely due to a large portion of motorway in *E7* undergoing maintenance and improvements during the earlier months of the data collection period. The most frequently used route was not the slowest but was slower than the mean duration of 56 minutes.

Interestingly, the routes with the shortest durations travelled the furthest distances (Figure 8.33) and vice versa. Although this is initially counter-intuitive, the shorter routes spent more time along slow urban roads whereas the longer routes consisted of large sections of motorway.

The use of motorway sections was further evident when looking at the vehicle speed for each route, which followed a similar trend to the distance travelled (Figure 8.34), varying between 14 and 21 $m \cdot s^{-1}$ on average (31 and 46 mph respectively). The highest average speed was associated with route *G*, which also had the shortest duration, however the lowest average speed occurred along route *X*. Route speeds varied by less than 3 $m \cdot s^{-1}$ (7 mph) for 50% of recordings. The slowest journeys differed from the average by almost twice the fastest journeys by 3 and 2 $m \cdot s^{-1}$ respectively (7 and 4 mph).

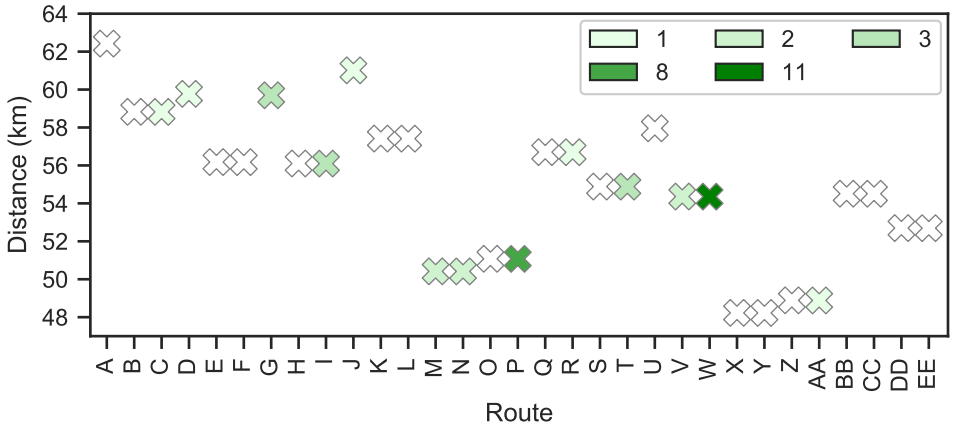

Figure 8.33: Comparison of the combined distance travelled for each route.

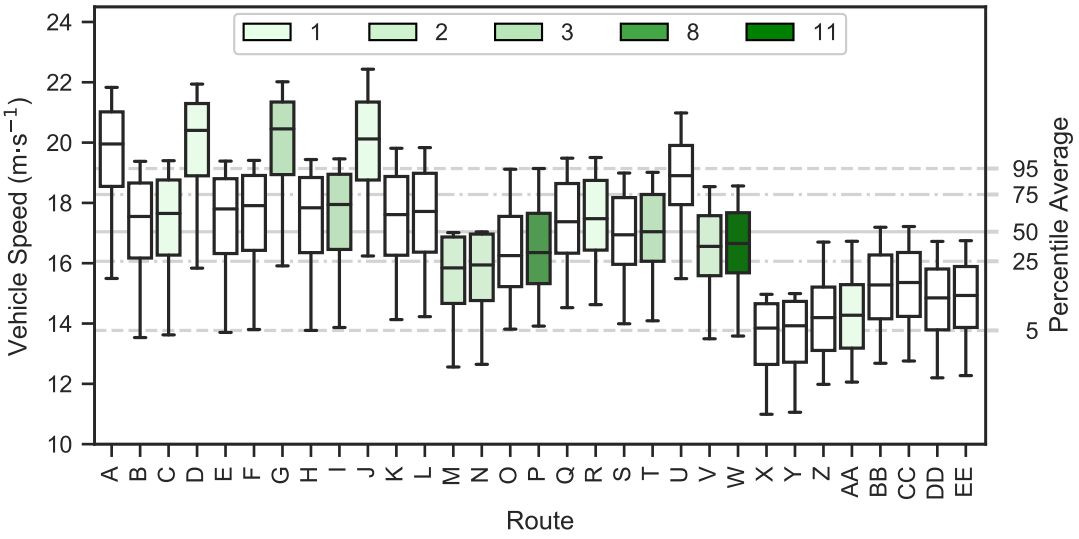

Figure 8.34: Comparison of the average vehicle speeds for each route.

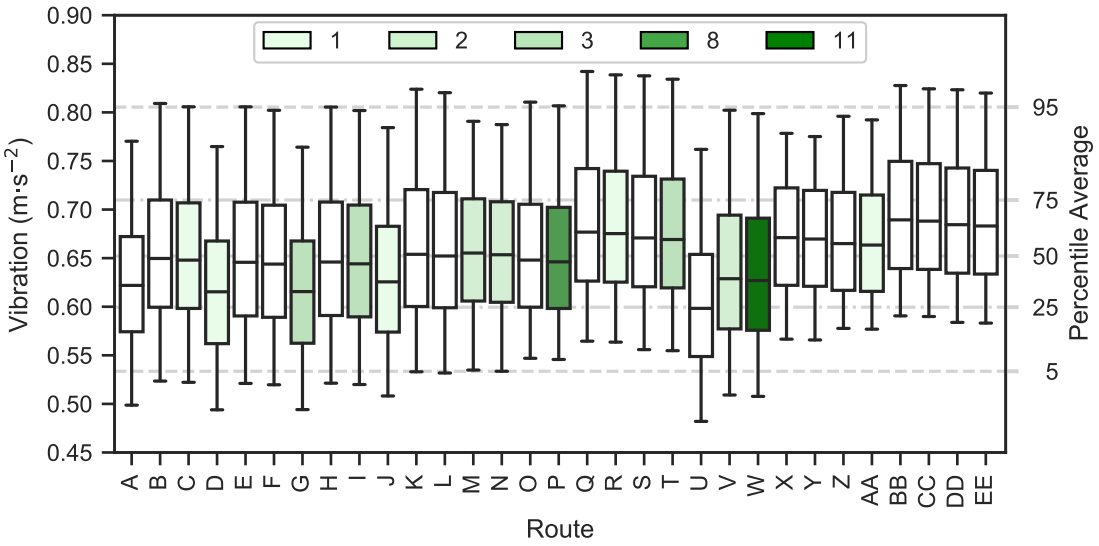

Figure 8.35: Comparison of the total combined unweighted vibration for each route.

The fastest average speeds were along the routes which used all 3 motorway edges (A, D, G & J). Slowest average speeds were in routes which used *E11*, skewed by a combination of the longest duration at a low speed of 9 $m \cdot s^{-1}$ (21 mph).

8.8.2 Unweighted Vibration

Raw vibration provides a base level comparison which encompasses all frequencies without any weightings applied. This was assessed here in terms of the r.m.s. magnitude, both in individual axes and combined, and by looking at the frequencies present between routes.

8.8.2.1 Average power

Half of all journeys resulted in a total vibration value within 9% of the average for the route (Figure 8.35).

Lowest total vibration was associated with routes which travelled along two edges consisting of stretches of motorway. The optimal route, *U*, only used 2 motorway edges (*E17* & *E18*), however the next 4 routes travelled along all *E7* as well. Levels of vibration were worst for the routes which travelled along Leicester's Western Bypass (*E23*, *E26* & *E27*), with the worst of these routes also having used *E11* through the centre of Nottingham.

There was minimal difference in the average vibration of a route depending on its speed (Figure 8.36), although there was a slight negative trend (R^2 = .54, p < .001). This appears to be due to the variation in edge values cancelling out when combined

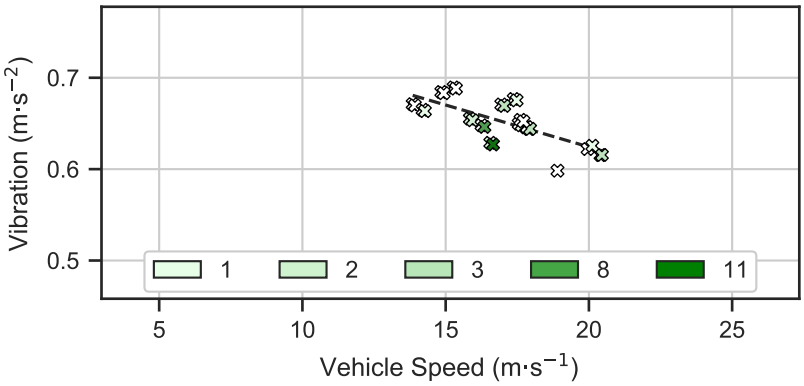

Figure 8.36: Comparison of the average unfiltered vibration and vehicle speed for each route, with the axis ranges set to the minimum and maximum of the edge averages.

into a full route. One of the biggest outliers was route *U* which was 8% slower than the fastest route but resulted in a 3% decrease in the total vibration.

8.8.2.2 Shocks

As discussed during earlier, there were no shocks of at least 1 g and only 6 edges which registered shocks of at least ½ g. These 6 edges, however, comprised sections of 25 of the 31 routes with only the routes which travel directly from the motorway to LRI (*E18*) not being exposed to any shocks. Overall, an average of 3 shocks could be expected along any of the routes, although this was heavily skewed by the routes which travel directly through the centre of Nottingham (*E11*).

8.8.2.3 Frequency bands

Average route frequencies followed the same trend as seen in Section 8.7.2.3, but with reduced variance (Figure 8.37). The majority of the recorded vibration occurred below 30 Hz, with peaks in the vertical axis centred around 1.5, 10.0 and 20 Hz, while a sudden spike was evident in all axes. A large proportion of the x-axis vibration occurred between 10 and 14 Hz — the cause of this was unknown.

Certain frequencies were found to vary significantly between routes, which could lead to differing comfort levels. In terms of vibration magnitude, noticeable variations between routes can be seen in Figure 8.38 around frequencies of 1.5, 10.5, 19.0, 23.0 and 40.5 Hz (differing by 0.02, 0.03, 0.01, 0.01 and 0.01 $m \cdot s^{-2}$, respectively). Vibration between 40.0 and 43.0 Hz fluctuated the most of these frequencies when also accounting for the average vibration, varying by 28–29%.

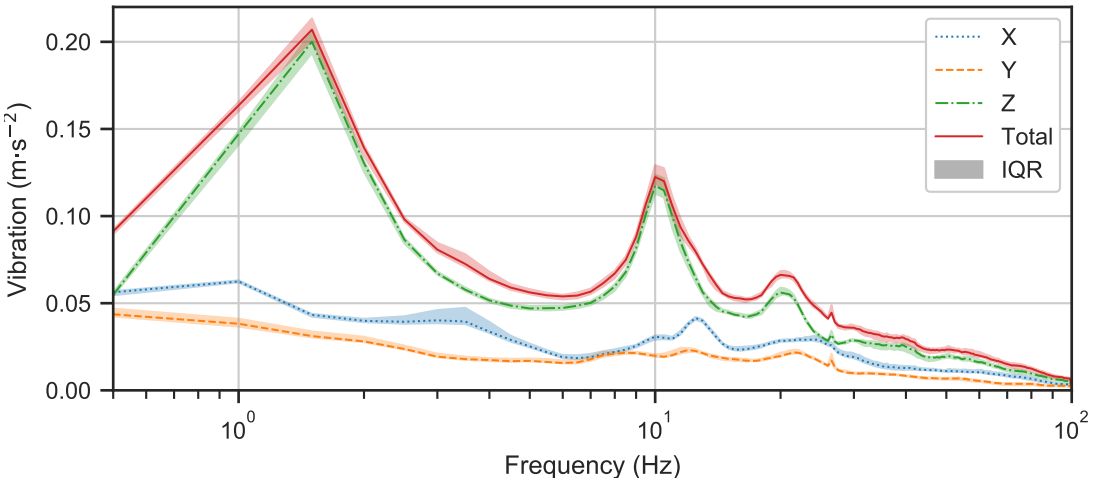

Figure 8.37: Average route vibration magnitudes as calculated in 0.5 Hz intervals for each axis.

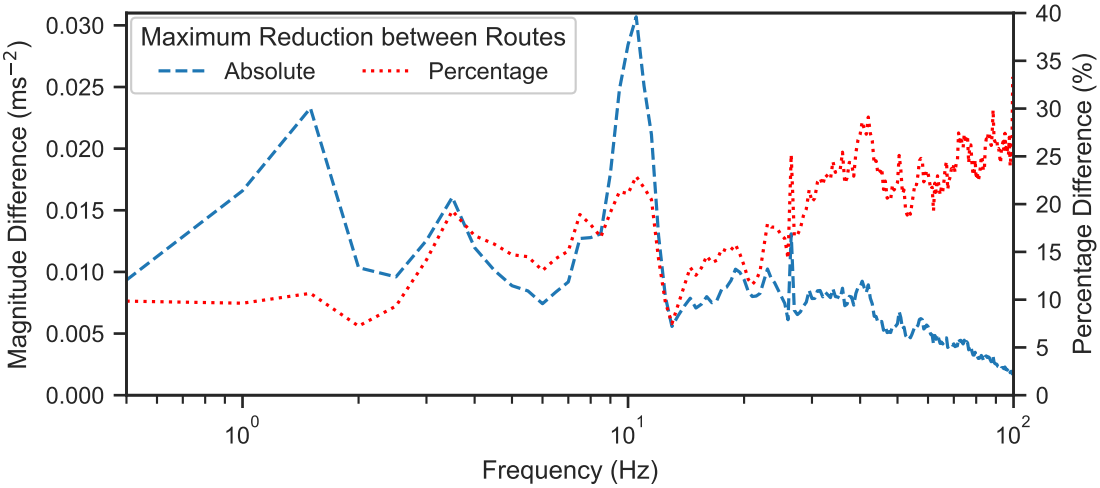

Figure 8.38: Range of 0.5 Hz interval total unfiltered vibration for all routes.

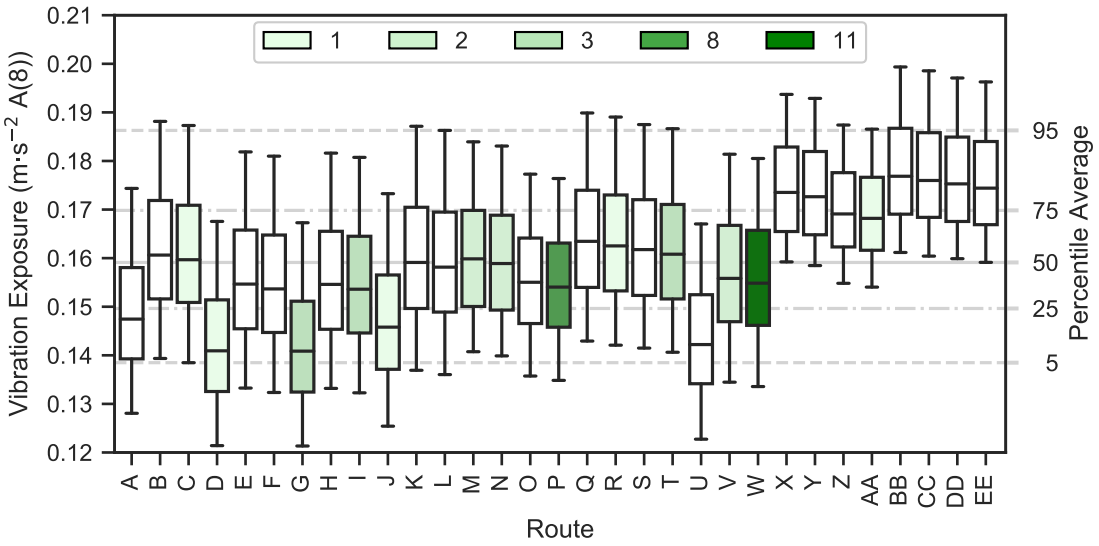

Figure 8.39: Comparison of the total vibration exposure for each route.

8.8.3 Frequency weighted vibration

Filtering the vibrations in each axis using the weightings designated by ISO 2631 enabled the expected effect on adults to be quantified, as a surrogate for neonatal-specific weightings. This was assessed in terms of the impact on health as well as comfort.

8.8.3.1 Health impact

Route exposures followed the same trends discussed in the edge analysis, with the the z-axis contributing the most towards the total, followed by the x- & y-axes respectively. Only the total values will therefore be discussed here.

Minimum and maximum exposure levels were found to occur along the fastest and slowest routes, *G* and *BB*, as shown in Figure 8.39. Choosing the optimum route would give a 12% decrease in exposure compared to the average of all routes (0.14 vs 0.16 $m{\cdot}s^{-2}$ $A(8)$) and a 20% reduction compared to the worst route (0.18 $m{\cdot}s^{-2}$ $A(8)$). No routes had a total vibration exposure within 50% of the action value, including the 95^{th} percentile extremes, suggesting that these levels would therefore not cause any long-term direct damage to adults (and likely not to neonates, although this remains unknown). Exposure again showed the strong correlation with the root of duration (R^2 = .89) but with a gradient that was twice as steep.

Low exposure values were associated with the motorway sections, earlier identified as leading to shorter route durations, with the lowest of these also having the lowest raw vibrations. All routes which used edge *E11* resulted in the highest exposure values.

The most frequently used routes resulted in slightly below-average vibration exposure,

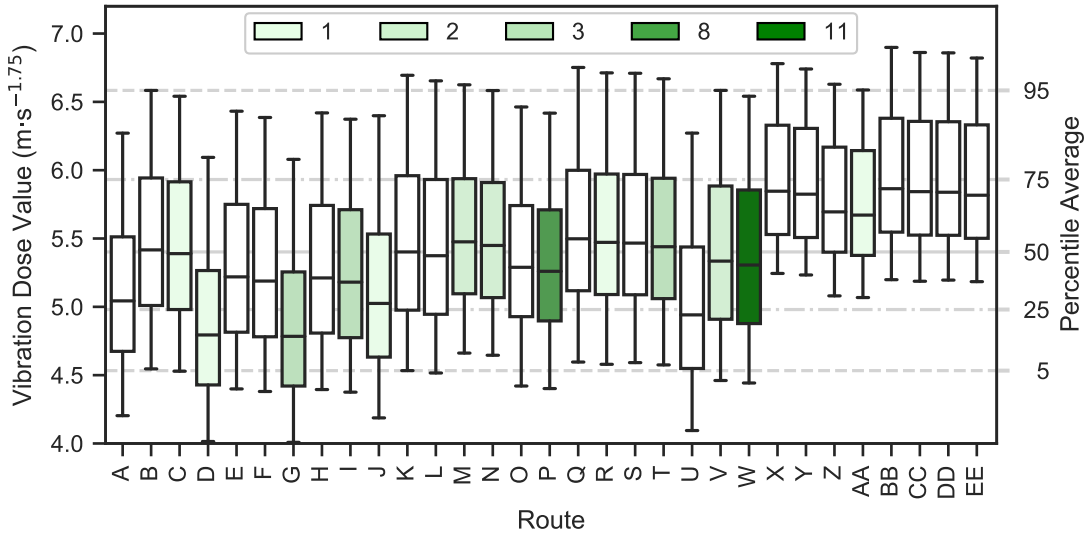

Figure 8.40: Comparison of the total vibration dose value for each route.

although the average was skewed by the few routes which involved edge *E11* (Figure 8.39). Most routes produced similar values, with only 4 (10%) of the 39 journeys driven from NCH to LRI found to have used the roads with the lowest exposure.

8.8.3.2 Vibration dose value

As with the total exposure, VDVs for each route were dominant in the z-axis. The y-axis contributed slightly more than the x-axis in 94% of routes, converse to the exposure relationship, suggesting that the y-axis vibration was more random.

It can be seen in Figure 8.40 that no routes recorded a total vibration dose value which reached the action value of 9.1 m·s$^{-1.75}$. Average VDVs for each route varied from 53 to 64% of the action value, with 5% of the journeys along the worst route at 76%, however these remain acceptable for adults. The best (*G*) and worst (*BB*) routes were the same as for vibration exposure, resulting in a 23% increase from 4.8 to 5.9 m·s$^{-1.75}$. As expected from the formula, VDV was strongly correlated (R^2 = .88) with the 4th root of duration.

As with vibration exposure, the lowest values for VDV were associated with motorway edges and the highest values occurred in routes which used *E11*.

VDV followed a similar pattern to vibration exposure and the most frequently used route between NCH and LRI was around the mean dose value of 5.4 m·s$^{-1.75}$ (Figure 8.40). Choosing the optimum route would result in an 11% (0.5 m·s$^{-1.75}$) reduction in total vibration dose.

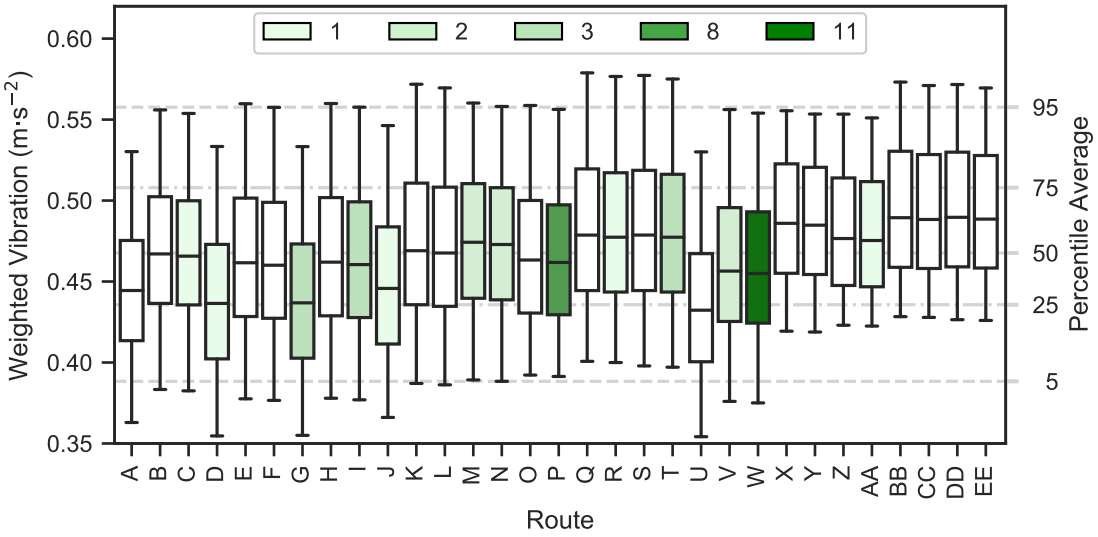

Figure 8.41: Comparison of the total combined weighted vibration for each route.

8.8.3.3 Perceived comfort levels

Examining total comfort-weighted vibration in Figure 8.41 shows that over 25% of all journeys between NCH and LRI occurred at "fairly uncomfortable" levels, as defined for adults. It can also be seen that at least 95% of journeys along all routes were at least "a little uncomfortable" as all whiskers were above the 0.315 $m \cdot s^{-2}$ threshold. There was a slight negative correlation with average vehicle speed (R^2 = .69), however the differences between routes were not significant.

Comfort-weighted vibration followed similar trends to the unfiltered vibration, with regards to the edges of causality. Minimum values occur along the routes which use sections of motorway (*E7*, *E17* & *E18*) while the worst values are along routes which travel through the centre of Nottingham (*E11*) and around Western Bypass (*E23*, *E26* & *E27*).

Of the routes which were used from NCH to LRI, the most frequently used route had slightly below average comfort-weighted vibration (Figure 8.41). Only one journey used the most comfortable route which would have provided a theoretical discomfort reduction of 6%.

As anticipated, higher unweighted vibration tended to lead to a higher comfort-weighted value (R^2 = .96, Figure 8.42). The optimum route for minimising comfort-weighted vibration was *U*, reducing the levels by 8% to 0.43 $m \cdot s^{-2}$ from an average of 0.47 $m \cdot s^{-2}$. Maximum vibration levels occurred along routes *DD* and *BB*, which both gave averages of 0.49 $m \cdot s^{-2}$ to 2 decimal places.

The average duration spent at each comfort level were plotted in Figure 8.43), with

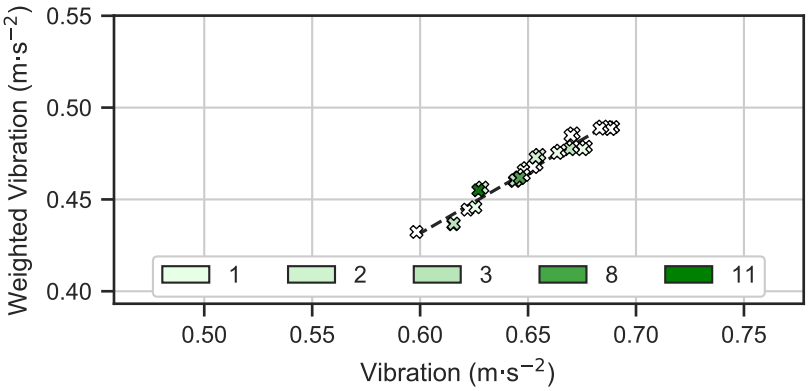

Figure 8.42: Comparison of the total comfort-weighted vibration and total unweighted vibration for each route.

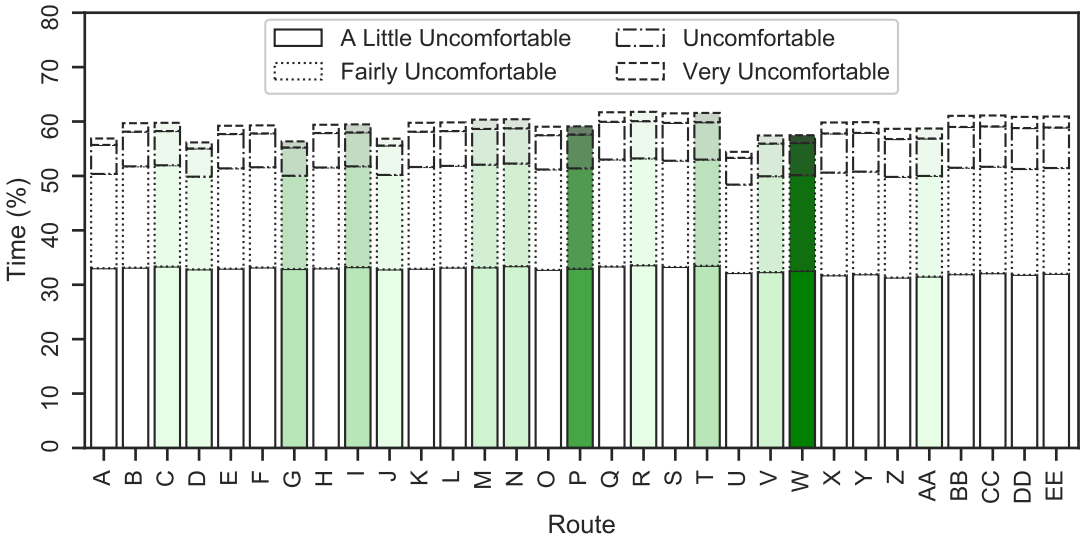

Figure 8.43: Percentage of route duration spent at different comfort levels.

times displayed as percentages to account for the different total route durations. Percentage distributions were highly similar for all routes, especially in the "a little uncomfortable" range. This was then reflected in the actual durations with the time spent at each individual comfort level being highly proportionate to the total duration (resulting in coefficients of determination of 0.93, 0.87, 0.88 and 0.88 as the scale increases).

Average times spent at "a little uncomfortable" vibration varied between routes with an average time of 18 minutes (Figure 8.44). Routes *G* and *D* resulted in the least amount of "a little uncomfortable" time at 16 minutes, an 11% reduction on the mean, while routes *CC* and *BB* performed the worst at 20 minutes. The minimum, average and maximum durations for each category of perceived comfort are summarised in Table 8.11 along with the IDs of the contributing routes.

Although the comfort classifications are derived from adult perception, it is possible that the extended periods of time at levels of vibration that are at least "a little un-

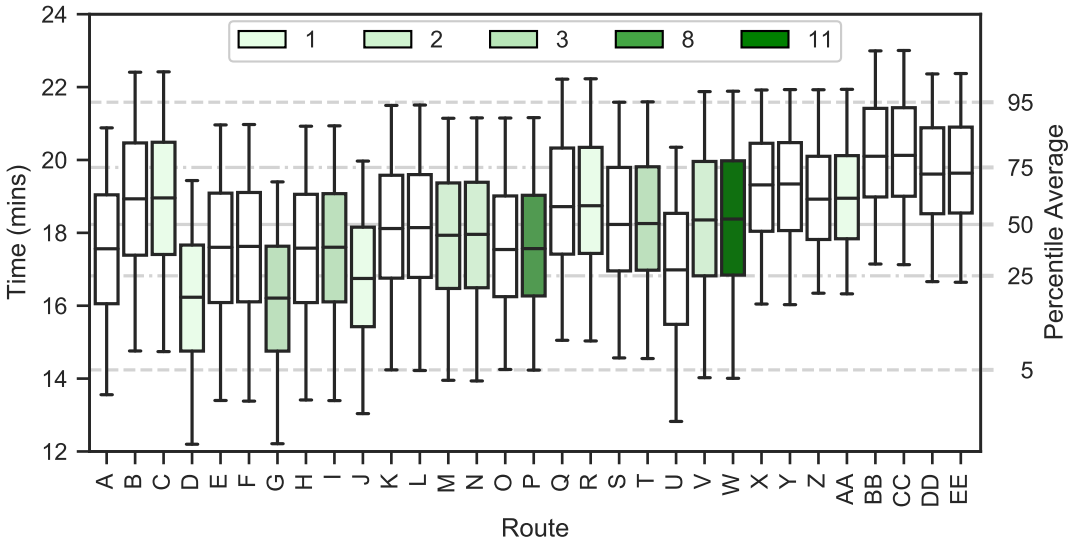

Figure 8.44: Comparison of the time spent at vibration classed as "a little uncomfortable" for each route.

Table 8.11: Minimum, maximum and average durations spent by NCH to LRI routes at the perceived comfort levels.

Perceived Comfort Level	**Route Duration (mins) / Route**				
	Minimum		Average	Maximum	
A Little Uncomfortable	16.21	*G*	18.26	20.13	*CC*
Fairly Uncomfortable	8.42	*D*	10.37	12.31	*BB*
Uncomfortable	2.56	*D*	3.62	4.72	*BB*
Very Uncomfortable	0.55	*G*	0.91	1.29	*BB*
Extremely Uncomfortable	0.07	*G*	0.14	0.21	*X*

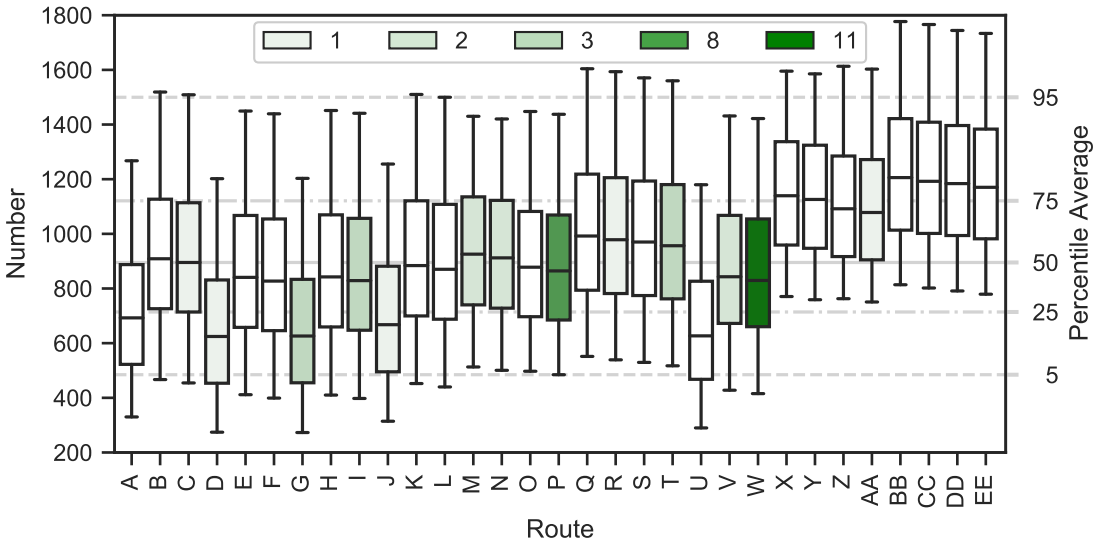

Figure 8.45: Comparison of the number of 1 s accelerations of at least 0.5 m·s^{-2} for each route.

comfortable" are a cause of the distress experienced by neonates during ambulance transport.

8.8.3.4 Peak events

The number of 1 s shocks that breached the 0.5 m·s^{-2} lower limit of the acceleration range perceived as "fairly uncomfortable" varied by almost a factor of 2 between routes (Figure 8.45). The lowest frequencies of occurrence corresponded with the routes that used large sections of motorway, earlier shown to result in some of the lowest number of shocks per minute travelled. The use of the motorway edges also resulted in the quickest times and contributed to the number of shocks being proportional to duration, as identified in Section 8.6.4.1. Similarly, the routes which travel through the centre of Nottingham resulted in the highest number of 1 s shocks.

The most frequently used route driven by the ambulances directly from NCH to LRI had a shock frequency within the 25th percentile of all road combinations. Choosing the route with the fewest number of shocks, which was only driven on a single occasion, would provide a 25% reduction over the most frequent choice.

8.8.4 Noise

It has been suggested that noise could be a factor which affects infants undergoing ambulance transfer [40, 43]. It is unknown whether this is due to the overall impact of the noises or the presence of specific levels, and therefore both were examined.

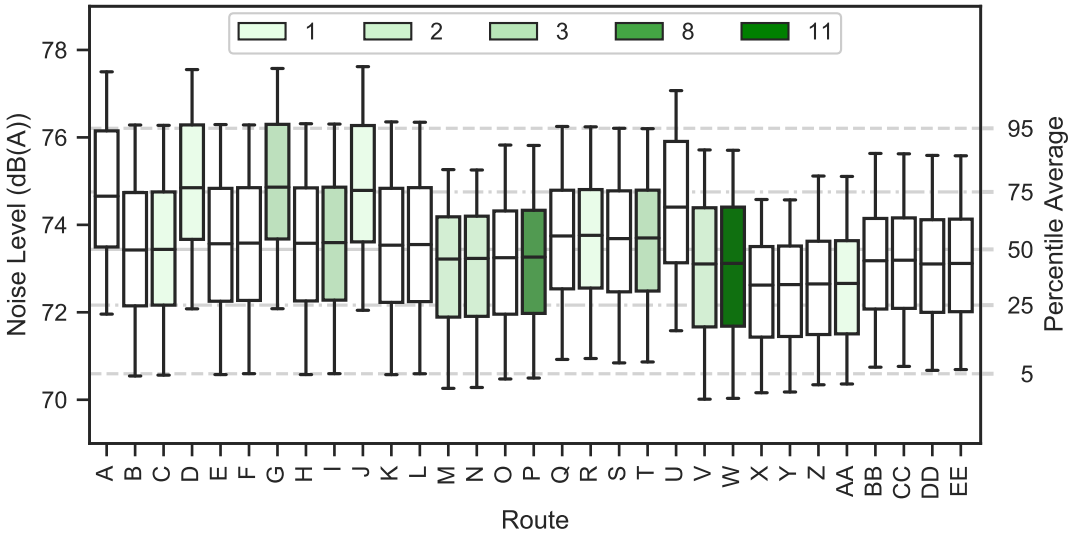

Figure 8.46: Comparison of the average noise level for each route.

8.8.4.1 Average power

Average noise levels varied minimally between routes (Figure 8.46). The lowest levels, at below 73 dB(A), were found along route *X*. These resulted in reductions of 1 and 2 dB(A) compared to the route average and the maximum level, along route *G*, respectively.

All noise levels were more than 20 dB above the recommended 45 dB(A) limit for neonatal intensive care units and averages for each route were between 12 & 15 dB greater than the stipulated maximum equipment level of 60 dB(A). The average noise level of 73 dB(A) is similar to that of a washing machine spin cycle, which could be classed as irritating but not painful. It is possible that these levels cause stress in the transported infants, however work is required to understand any link involved.

The most frequently used route resulted in below-average noise levels, although there was effectively no difference between all results. The route which produced the lowest noise levels was found to result in the highest vibration exposure (Figure 8.39).

Lowest average noise levels were associated with routes that used *E11* through the centre of Nottingham. Using at least two motorway edges in a route resulted in the loudest noise levels due to the amount of time spent at high speeds.

8.8.4.2 Spikes

At least 30 occurrences of a 5–10 dB increase in 1 s noise level can be expected during an average journey along each of the 31 routes (Figure 8.47). The average for all routes was slightly greater at 44 events (median of 38), however the edge that travels directly

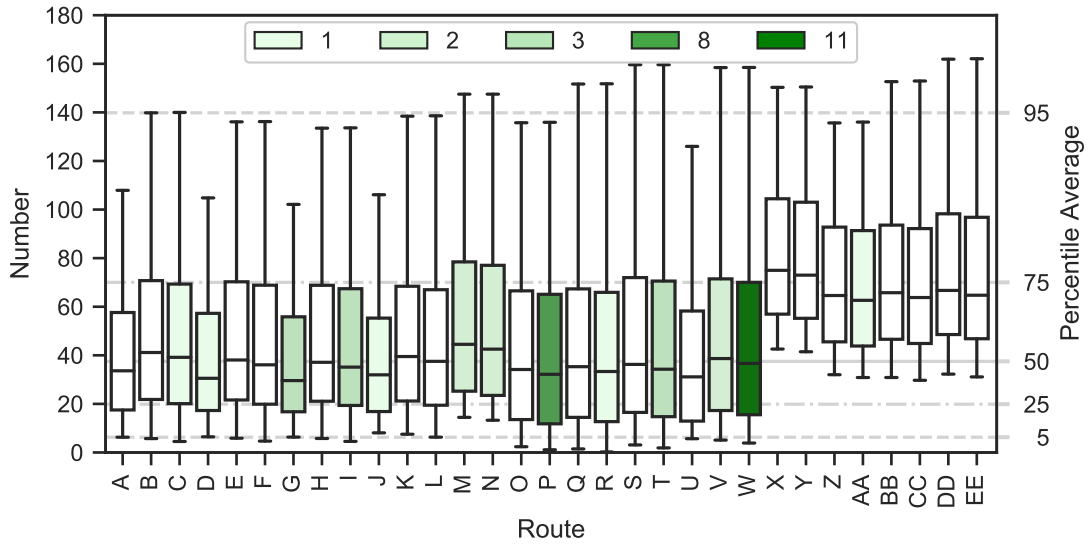

Figure 8.47: Number of 5–10 dB increases in 1 s averaged noise levels for each route.

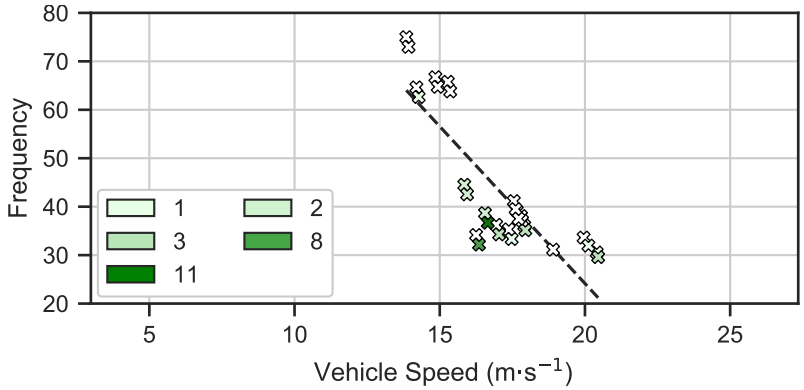

Figure 8.48: Average frequency of 5–10 dB noise spikes versus average vehicle speed, for each route.

through the centre of Nottingham (*E11*) increases the average by 50% to 67. Each route varied by an average of 49 occurrences for half of the recordings, with the worst 5% of journeys experiencing at least 102 events regardless of the route.

Kuhn et al. [68] demonstrated in the Neonatal Intensive Care Unit (NICU) that sudden spikes in noise level contribute to an increase in heart rate. It is therefore likely that the multiple occurrences found during ambulance transfer have a similar impact on the infant and should be minimised.

Figure 8.48 illustrates how there remained a negative correlation with vehicle speed in the number of events (R^2 = .71), although the routes which use *E11* heavily skew the gradient. This is thought to be due to lower speeds being associated with increased driving in urban environments, that are in turn associated with stop-start traffic causing speeds (and therefore noise levels) to fluctuate.

The majority of routes taken when driving directly from NCH to LRI resulted in highly

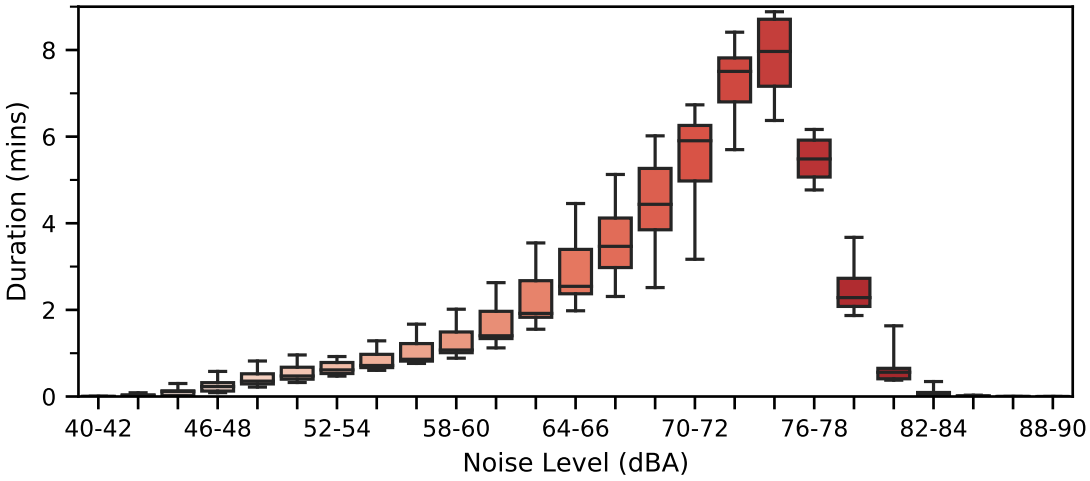

Figure 8.49: Spread of time spent at different noise levels by all routes from NCH to LRI.

similar frequencies of noise spikes and therefore there is not much to gain in differentiating between them. It is clear, however, that the use of *E11* should be avoided.

The number of 1 s noise levels over a 75 dB(A) threshold was earlier identified to be strongly connected to the use of high speed road segments such as motorways. This was reflected in the route analysis, where the values differed in proportion to the number of fast segments used. Therefore, if the amount of time over set levels are to be avoided, journeys should be driven at reduced speeds.

8.8.4.3 Repetitive exposure

More journey time was spent at noise levels in the 74–76 dB(A) range, equivalent to using a vacuum cleaner, than any other 2 dB(A) interval despite the average noise levels being below 71 dB(A) (Figure 8.49).

The 5 routes which registered the greatest amount of noise above 78 dB(A) (*A*, *J*, *D*, *G* & *U* in descending order) also resulted in the highest average speeds after travelling along large sections of motorway. Spending the majority of the journey time at high speeds, these routes also registered the fewest minutes between 60 & 72 dB(A) leading to the large ranges shown.

8.9 Conclusion

The aim of this chapter was to use the data recorded within the ambulances of CenTre, using the purpose-developed and validated app, to determine if routing could affect,

and hopefully improve, the outcomes of neonatal infant transfers. Roads taken by ambulances travelling from NCH to LRI were previously chosen as an ideal network to test the hypothesis due to both the large number of route combinations and the amount of data collected along these roads.

Sections of 588 separate journeys needed to be assigned to the 114 edges of the expanded NCH to LRI network to provide the greatest amount of data for metric computation, which would increase the confidence in the average values. Every recorded journey went through a process of checking coordinates against the network nodes and those that corresponded with an edge were assigned the numeric value for that edge to facilitate easy extraction from the database. Coordinates were only sampled at 1 Hz, compared to the 200 Hz of the IMU, and were therefore interpolated linearly to minimise the amount of sensor data incorrectly assigned to or missed from an edge.

Coordinates were extracted for each edge in turn and the journeys were aligned to ensure the same sections of road were travelled over by all to maximise comparability. Sensor data which corresponded with the aligned coordinates were then extracted in the form of the accelerations and noise levels for the metric computations.

The response of neonatal infants to vibration and noise stimuli has not been thoroughly investigated and therefore a range of metrics were required to provide an indication of the possible comfort and to provide a reference dataset once a response is fully understood. Metrics included the journey duration, average vehicle speed, and different measures of vibration and noise. The journey duration and vehicle speed were included both as a sanity check and to assess their influence on the other metric values. Each metric was first calculated for each journey along an edge before overall edge percentiles were calculated. Values for routes were then aggregated from the constituent edges.

Data were gathered during journeys with both a patient present as well as an empty incubator. Values of noise and vibration were found to be similar for each case and therefore all data was treated equally and used in the assessment of the network.

Greater vehicle speed resulted in an increase in z-axis vibration and noise level, but a reduction in y-axis vibration. The z-axis increase was likely due to the rate at which the ambulance wheels interacted with imperfections in the road surface, while the noise increases were due to the amount of traffic and air resistance in high speed situations. It was theorised that the y-axis vibrations were susceptible to the tight corners and junctions found in urban areas, which therefore were not present at greater speeds.

Duration only had a clear effect on the computed levels of vibration exposure and vi-

bration dose value. This was expected as both measures included functions of time in their calculation. All other metrics were independent of the time spent driving.

Although trends were identified between the type of road used and the resultant metric values, it was unclear how much of this was due to the road and how much to the other factors such as vehicle speed. Unfortunately, vehicle speed itself is entwined in the road classification and therefore statistical methods such as ANCOVA were inappropriate.

The 34 edges of the NCH to LRI route network consisted of roads of various sizes and types. This therefore led to a vast range of durations and vehicle speeds which were compared to the other metric values to see if any causality existed.

The surface of the concrete A-road produced worse results in all vibratory metrics than the motorway. Vibration was at higher levels despite having a slower vehicle speed, with the largest increase at the frequency of wheel hop to further suggest the road surface was the cause. The magnitudes of vibration were such that the action values for both exposure and VDV would be reached in two-thirds and half the time, respectively, than the motorway assuming constant travel along the roads. The A-road also registered a greater proportion of vibration inside every classification of comfort above "not uncomfortable" as well as a greater rate of shocks.

Noise levels were greatest along the motorway, due to the higher vehicle speeds, albeit less than 1 dB greater than those of the concrete A-road.

Thirty-one possible route combinations, constructed from the 34 edges, were assessed in each metric. Thirteen of these road combinations were used by actual journeys travelling from NCH to LRI, which enabled the assessment of the route choices by the transport group.

Durations varied by 13.4 minutes between routes, with the shortest identified route matching that provided by the Google Maps system. Routes with the lowest durations used all 3 of the motorway edges, while the largest durations were found in those which travelled through the centre of Nottingham. Vehicle speeds was inversely related to duration, with the majority of the fastest journeys also having the lowest durations.

Unfiltered vibration was fairly independent of both route duration and vehicle speeds. It also varied minimally, with the optimal route reducing magnitudes by only $0.05\,\mathrm{m{\cdot}s^{-2}}$. The frequency of sudden noise increases was also fairly similar for most routes, but *E11* skewed 8 routes to far greater amounts due to the inconsistent driving patterns (and thus speeds) in highly urban areas.

Duration had a large impact on the resultant route metrics. Vibration exposure and

VDV again increased with duration, along with the time spent at different comfort levels increased with duration although this was to be expected as it is directly linked to duration in calculation. Greater durations resulted in lower noise levels, in part because of the lower speeds along these routes, however all routes only varied by 3 dB(A).

The most frequently driven route, *W*, consistently provided metric values around the average for all routes. It was never the optimum route, but was also never the worst. Less than 10% of journeys took the fastest — and lowest vibration — route, however this route was undergoing roadworks for a substantial part of the collection period.

Although the analysis of edges suggested that vehicle speed increased the overall vibration and noise levels, the assessment of full routes suggested that duration was a key factor to consider. This variation in recommendations was due to the combinations of edges effectively cancelling out any differences identified in individual comparisons. Neither the average vehicle speed nor duration of a route would always indicate the levels of vibration or noise and therefore full analysis is still required.

The variation in routes, along with the ambulance drivers road choices, showed that there is clear scope to using routing to reduce the levels of vibration and noise which would be experienced during neonatal transfers. Recommendations can be made using adult-derived metrics, however the next stage to this work requires thorough investigation into the factors which affect neonatal infants. The results of this investigation should lead to a creation of an index specific to improving outcomes which could then be applied to the data to obtain the true most comfortable route.

CHAPTER 9

Analysis of Trolley Comfort Before and After Ambulance Journeys

9.1 Introduction

While the levels of vibration and noise during ambulance journeys have been investigated, using a variety of metrics, the driven route is only one component of neonatal transfers. The loading and unloading of the trolley onto the ambulance, and the resultant physical excitations, has also been identified as an area of concern with peak values greater than during the journey itself [40, 44]. It is therefore possible that the events that occur during these phases of the transportation process have an influence on the outcomes of the transferred neonatal infants.

A vital part of neonatal transfers are the trolleys that provide the necessary environment required to care for the infants during the journey, of which a core component is a stabilised temperature. The transport incubator is therefore kept at the required temperature at all times to facilitate immediate use in transfers when needed. Maintaining a stable temperature requires a consistent supply of electricity to enable constant monitoring and adjusting, and therefore the trolleys are connected either directly to the mains (on standby in the hospital) or the ambulance (during transfer), or powered by a built-in battery (all other times). This need for continual preparedness, along with being in close vicinity to patients for transferring from ward to transport incubator and vice versa, means the trolley must be moved on and off the ambulance.

Loading and unloading of the ambulance raises logistical problems, in part due to the substantial mass of the trolley. The largest single mass object on the trolley is the incubator at 49 kg (TI500, Dräger, Lübeck, Germany), with up to a further 13 kg for the gas cylinders which are loaded into the base. Combined with all other equipment for life-

support and monitoring, along with the trolley itself, the total mass reaches roughly 100 kg. Although this mass is not significant on the smooth, flat floors of a hospital, it would require increased effort to manoeuvre on potentially uneven pavement and in transitioning to the inside of the ambulance.

Some ambulances possess a powered tail-lift to raise and lower the ambulance horizontally between the ground and the internal height, while others have a section of flooring which folds out to form a ramp. The ambulances operated by CenTre (CenTre Neonatal Transport) are ramp-based and, although there is a winch to assist the loading/unloading process, the staff have reported in personal communication that they tend to simply "take a run-up" to push the trolley onboard. The extra force exerted during the loading of the ambulance using this method, along with the presumed natural acceleration during unloading, may be a root cause of the previously observed peaks.

A further interaction that may distress the neonatal infants is that of the ground surface and the trolley wheels while being rolled in and out of the hospitals. This was flagged by the staff at Nottingham City Hospital (NCH) as they had visually observed infants being shaken when pushed over a section of entrance matting. As could be derived from the name, entrance matting is designed for the entrances of buildings with the aim of removing dirt and moisture from shoes. The matting achieves this using coarse fabric in an often ribbed construction, with the ribs running parallel to the entranceway (perpendicular to the walking direction). It is these ribs, or rather the spacing between the ribs, that would cause vibrations when interacting with the trolley wheels and therefore result in transmission to the infants to be transferred.

Both the method of loading/unloading and the potentially disruptive flooring materials led to the expansion of data collection from the ambulance transfer alone to including the travel within the hospitals. As outlined in Section 5.4.2.3, the transport staff were instructed to begin the recording on the trolley while inside the starting hospital and not to end the recording until inside the destination hospital. The collected data, recorded during the 1,487 successful recordings, will be analysed and compared through the course of this chapter. First, however, the method of identifying and extracting the relevant data will be discussed.

9.2 Identifying data of interest

Before any analysis could be performed, the data that pertained to each hospital needed to be identified within each journey and inserted into the database, with the relevant

meta-data, for future extraction. Database use was required, as outlined in Section 6.2, to enable easy extraction of the data of interest without requiring the processing of multiple Comma-Separated Value (CSV) files each time.

Initially, the journey data were inserted into 3 measurements to contain unadjusted vibration and noise (*raw*), location data (*gps*), and vibration data weighted in accordance with the International Organisation for Standardisation (ISO) standard for the (adult) human response to vibration, ISO 2631 [49] (*iso*). These were supplemented by the equivalent measurements to contain hospital data: (*hosp*, *gps_hosp* & *iso_hosp*). Timestamps for each recording were defined as "at the hospital" in relation to the registered vehicle speeds and the corresponding radial accuracy of the location data (Section 6.3). All timestamps before the first speed above a threshold of 3 $m \cdot s^{-2}$ (7 mph) and a radial accuracy of less than 20 m were deemed as "before the journey", while all timestamps after the final set of location data that met this criteria were determined to be "after the journey". This was found to be more than adequate for route assessment, but problems were apparent when hospital analysis was first attempted.

The main problem with the use of the speed threshold was the interference from the hospital and surrounding buildings. Naturally, when the smartphones were inside the hospitals there was little to no access to the satellites to obtain a clear fix, leading the majority of acquired location data to be subject to substantial drift. The inclusion of the radial accuracy was intended to reduce the effect of the drift as it should signify the confidence in the provided data. Exploring the data, however, there were multiple instances where the thresholds were met earlier (in the case of the departure hospital) than expected, resulting in a loss of data for the analysis here. It was therefore decided that a new method of identifying the hospital data was required.

9.2.1 Re-separating hospital data

An alternative approach to the separation of data was implemented that used solely the distance between recorded coordinates and manually-determined threshold coordinates for each hospital to assign timestamps to "before" or "after" a journey. Before this could be implemented, 8-digit plus codes were utilised to group the hospital location data (in a similar manner to that described in Section 7.3.2) to identify the hospitals present within the data. This information was used to create an initial dictionary of recordings and the hospitals at which the journeys started and ended. Manually examining the recorded coordinates, Google Maps and Google Street View, threshold coordinates at which the recording could be considered "at the hospital" were determined. These are provided in Table 9.1 for each of the 42 different hospitals shown

in Figure 9.1 (note: Birmingham Women's Hospital (BWH) appeared to have separate locations for arrivals and departures, therefore two thresholds were set).

Table 9.1: Determined threshold coordinates for each hospital.

Hospital	Latitude (°)	Longitude (°)
Addenbrooke's Hospital (AH)	52.1732	0.1396
Birmingham Children's Hospital (BCH)	52.4846	-1.8934
Birmingham Heartlands Hospital (BHH)	52.4795	-1.8288
Boston Pilgrim Hospital (BPH)	52.9903	-0.0094
Burton Queen's Hospital (BQH)	52.8181	-1.6560
Bradford Royal Infirmary (BRI)	53.8068	-1.7987
BWH — inbound	52.4534	-1.9439
BWH — outbound	52.4534	-1.9420
Chesterfield Royal Hospital (CRH)	53.2358	-1.3997
Frimley Park Hospital (FPH)	51.3190	-0.7407
George Eliot Hospital (GEH)	52.5127	-1.4769
Glenfield Hospital (GH)	52.6546	-1.1787
Great Ormond Street Hospital (GOS)	51.5219	-0.1201
Hinchingbrooke Hospital (HH)	52.3332	-0.2042
Hull Royal Infirmary (HRI)	53.7441	-0.3559
John Radcliffe Hospital (JR)	51.7627	-1.2173
Kettering General Hospital (KGH)	52.4012	-0.7425
Kingston Hospital (KH)	51.4143	-0.2814
King's Mill Hospital (KMH)	53.1350	-1.2335
Luton & Dunstable University Hospital (LAD)	51.8949	-0.4735
Liverpool Alder Hey Children's Hospital (LAH)	53.4193	-2.8976
Lincoln County Hospital (LCH)	53.2346	-0.5218
Leicester General Hospital (LGH)	52.6308	-1.0778
Leicester Royal Infirmary (LRI)	52.6282	-1.1355
Milton Keynes University Hospital (MKH)	52.0259	-0.7356
NCH	52.9906	-1.1613
Northampton General Hospital (NGH)	52.2357	-0.8874
Norfolk & Norwich University Hospital (NNH)	52.6164	1.2208
Peterborough City Hospital (PCH)	52.5839	-0.2804
Queen's Medical Centre (QMC)	52.9441	-1.1863
Royal Derby Hospital (RDH)	52.9106	-1.5111
Rotherham General Hospital (RGH)	53.4125	-1.3427

Continued on next page

Table 9.1 — *Continued from previous page*

Hospital	Latitude (°)	Longitude (°)
Russells Hall Hospital (RHH)	52.5022	-2.1155
Royal Stoke University Hospital (RSH)	53.0034	-2.2149
Sheffield Children's Hospital (SCH)	53.3810	-1.4913
Stoke Mandeville Hospital (SMH)	51.7974	-0.8001
University College London (UCL)	51.5248	-0.1364
University Hospital Coventry (UHC)	52.4230	-1.4364
Warrington Hospital (WH)	53.3927	-2.6084
Walsall Manor Hospital (WMH)	52.5839	-1.9974
New Cross Hospital (WNC)	52.5980	-2.0952
Worcestershire Royal Hospital (WRH)	52.1909	-2.1814
Warwick Hospital (WWH)	52.2907	-1.5855

Recordings were subsequently processed and inserted into a hospital-specific database with the "before" and "after" timestamps determined by the last and first, respectively, set of coordinates to be within 20 m of the thresholds for the hospitals earlier identified as bookending the journey. Data were again inserted into three distinct measurements with meta-data added as tag values — as specified in Section 6.2.2 (ambulance & trolley code, patient presence, reason for patient transfer, presence of emergency driving, date and whether data was before or after the journey) along with the hospital Identifier (ID) — to enable specific extraction and stratification of the analyses.

Retrieving the required data for analysing hospitals from the database was far simpler than for the edges in Section 8.3 due to the lack of a distance component. Location data inside the hospitals were either highly inaccurate or unavailable, which meant the data could not be aligned and could only be treated as a single entity. Extraction therefore consisted simply of requesting all acceleration and noise data that corresponded to the specific combination of hospital ID and whether before or after the journey was required.

9.2.2 Hospital data overview

Performing a quick overview analysis of the hospital data showed that only 1,440 (97%) of the 1,487 recordings that were available for routing analysis were found to include hospital data after the journey. This was due to recordings (or the final coordinates) ending before reaching the hospital threshold and therefore these could not be included in the hospital analysis as it would not be clear where the journey ended. A greater

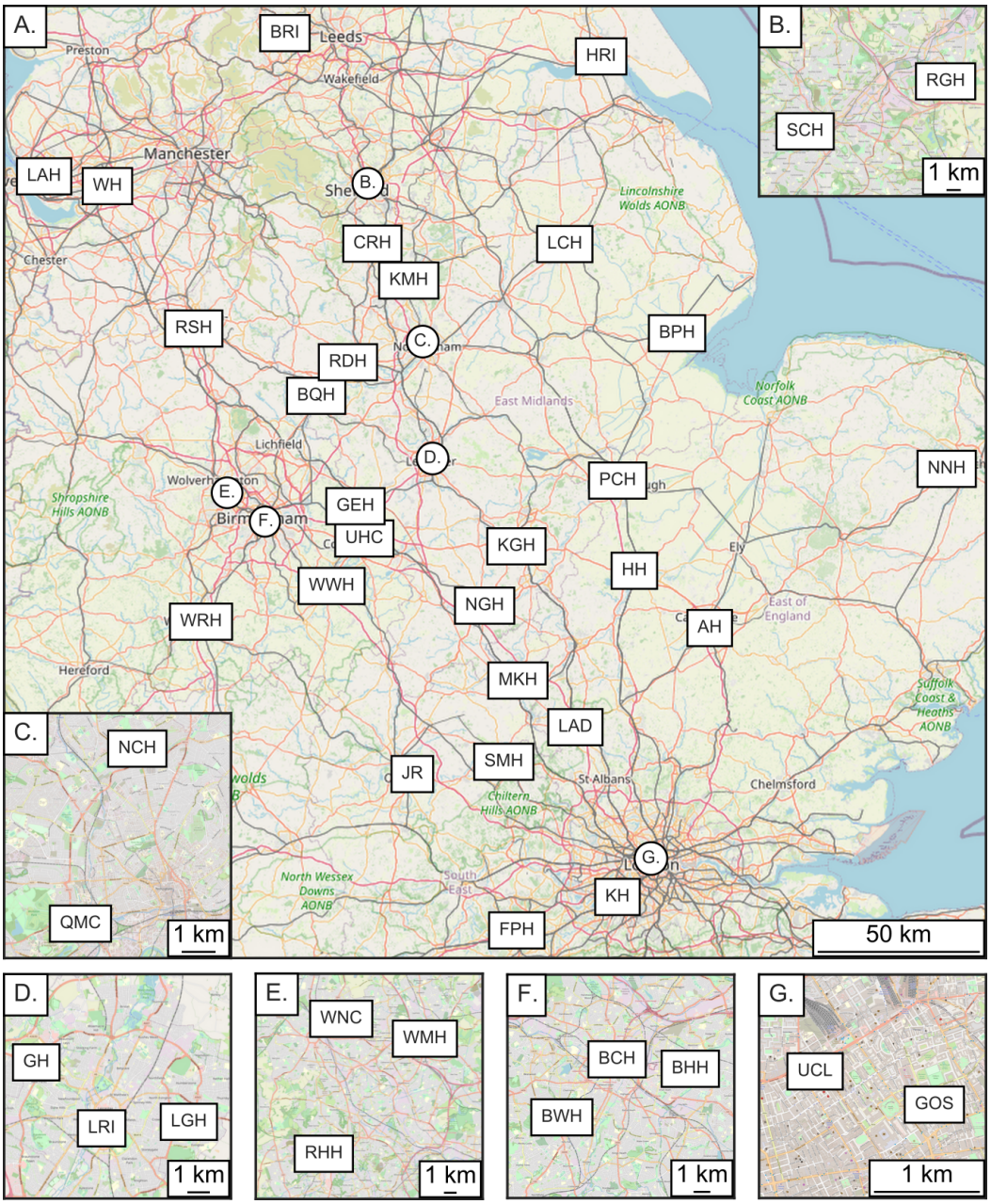

Figure 9.1: Location of all hospitals involved in the data collection.

Table 9.2: Amount of data (amount with patients) recorded at each hospital.

Hospital	Departing	Arriving	Hospital	Departing	Arriving
AH	3 (2)	2 (1)	LGH	56 (23)	85 (59)
BCH	9 (5)	13 (13)	LRI	303 (96)	258 (81)
BHH		3 (2)	MKH	1 (1)	1 (0)
BPH	38 (18)	38 (18)	NCH	272 (58)	251 (73)
BQH	24 (12)	31 (18)	NGH	50 (22)	67 (35)
BRI	1 (1)		NNH		2 (0)
BWH	1 (0)	1 (1)	PCH	2 (2)	1 (0)
CRH	3 (2)	3 (1)	QMC	93 (60)	154 (82)
FPH	1 (1)	1 (0)	RDH	45 (32)	64 (20)
GEH	35 (19)	44 (21)	RGH	1 (1)	2 (0)
GH	26 (12)	25 (23)	RHH		1 (1)
GOS	1 (0)	1 (1)	RSH	1 (1)	3 (1)
HH	1 (0)		SCH	4 (2)	12 (7)
HRI	4 (3)	6 (3)	SMH		2 (1)
JR	10 (8)	15 (5)	UCL	1 (1)	1 (0)
KGH	39 (24)	45 (18)	UHC	104 (68)	127 (52)
KH		1 (1)	WH	1 (1)	
KMH	36 (21)	40 (19)	WMH	1 (0)	2 (2)
LAD	5 (1)	7 (1)	WNC	1 (0)	2 (1)
LAH		1 (1)	WRH	1 (1)	
LCH	70 (41)	79 (42)	WWH	51 (33)	49 (20)

number (13%) of recordings required excluding from analysis of hospitals before the journey started due to late location fixes, resulting in a reduced dataset of 1,295.

There were large variations in the amount of data recorded at each hospital, both as start and end points. The greatest amount of data was recorded at the two hospitals which CenTre use as bases, LRI and NCH, accounting for a combined 44% (before departure) & 35% (after arrival) of the total recordings (Table 9.2). The median number of journeys at a specific hospital were rather low (4.5 departing; 6.5 arriving), however these were skewed by multiple hospitals (13 departing; 9 arriving) only having a single dataset available. Despite the single-occasion hospitals, there remained at least 15 hospitals which registered at least 10 recordings for both before and after the journeys, which should enable consistency of procedures to be explored.

There was similar variation, albeit at reduced magnitudes, in the amount of hospi-

tal data recorded while patients were present, as shown in Table 9.2. Interestingly, there were similar distributions for hospitals before journeys (median: 2.5, Interquartile Range (IQR): 1.0–22.3) and after journeys (median: 2.5, IQR: 1.0–20.0), although the values for each hospital did not often match. The base hospitals of LRI and NCH were again amongst those with the greatest amount of data, along with QMC, UHC and LGH — possibly used for extra ward capacity as they are reasonably close to the bases.

9.3 Assessment of comfort metrics

A range of metrics were defined in Section 8.4 for the assessment of noise and vibration levels during the recorded ambulance journeys. These included analysis of both average and peak values, along with the frequency of certain events occurring. Although the same variables were recorded before and after the journey, not all metrics were suitable for the assessment of hospital loading & unloading procedures.

Firstly, the interactions which were thought to cause some degree of discomfort — rolling over entrance matting and loading onto/unloading from the ambulance — were only expected to be brief events and thus would be evident as shocks. In between these events, little disturbance was expected due to the trolley being either pushed slowly over flat, smooth hospital flooring or stationary while staff prepared other aspects to the transfer. This suggested that the use of averaged metrics would be heavily skewed by the moments of inactivity and thus present a more positive representation than reality.

Secondly, the durations recorded before departure varied greatly both between hospitals and between the recordings at the same hospital. Unlike with the analysis of roads, this was not expected to be due to the trolleys being moved at different speeds. Indeed, upon analysis of the unfiltered waveforms (a subset of which are shown in Figure 9.2) it was clear that events caused sharp spikes, with the differences in duration often consisting of periods of little to no vibration. It is unknown what the causes of the duration variations were, although the presence of a patient was discounted due to the durations with patients (subplots A, C, D & F) varying between themselves.

The problem with extended periods of minimal interaction was further evident within the data recorded after arriving at the destination hospital (Figure 9.3). Along with varying in time between events, as with the data before departure, the recording smartphones appeared to be abandoned until the auto-stop feature (Section 3.5.3) kicked in and ended the recording due to lack of movement. Computing average values would therefore result in greatly reduced values. Analysis of data was restricted to a maxi-

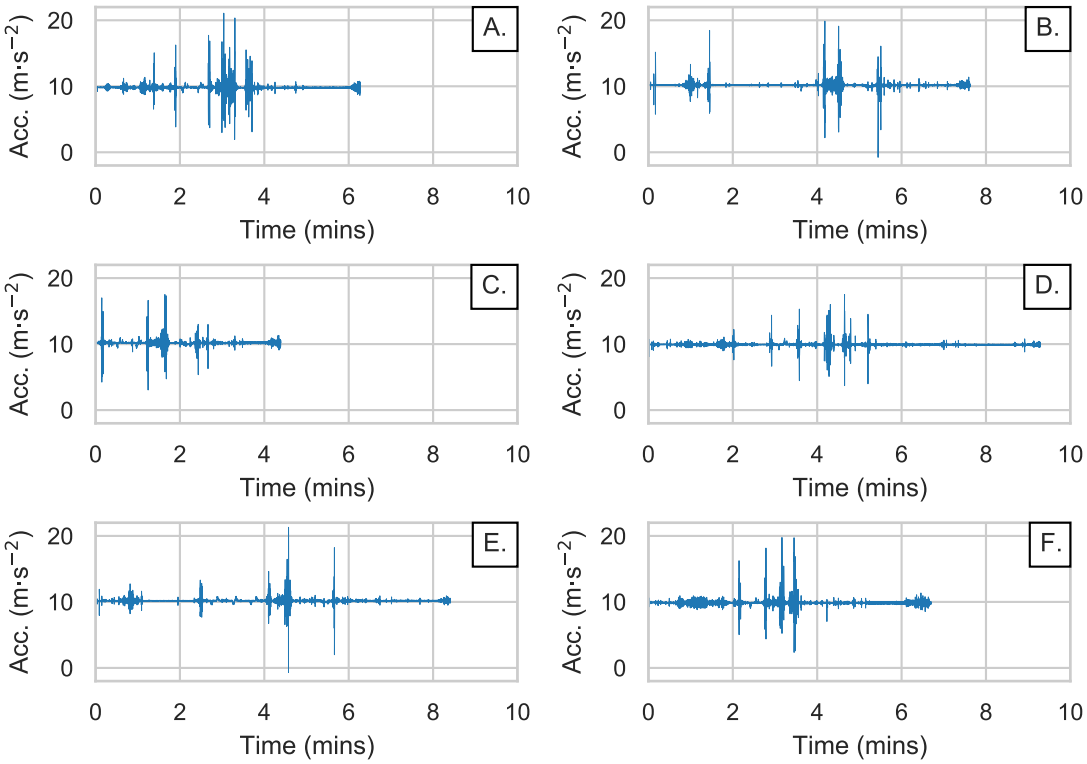

Figure 9.2: Waveforms of as-sampled z-axis accelerations for the first 4 datasets recorded at LRI before departure.

mum of 10 minutes, which was found to remove the excess duration without impacting the number of events. Ultimately, however, due to both the durations of inactivity and the characteristic of the events of interest, metric analysis was constrained mainly to peak measures.

9.3.1 Comfort metrics

The metrics used for assessing the levels of noise and vibration before and after the journeys will be outlined in the following sections. The reasons and brief overviews will be recapped from Section 8.4 for the peak values and shock events used, however the formulae will not be redefined and references will instead be provided to the previous chapter.

9.3.1.1 Unweighted vibration

The first metrics to be assessed were those concerning the raw, unfiltered (apart from removing the 0 Hz components) accelerations. No standards exist that provide guidelines for the use of raw vibration in determining levels of comfort, although a previous

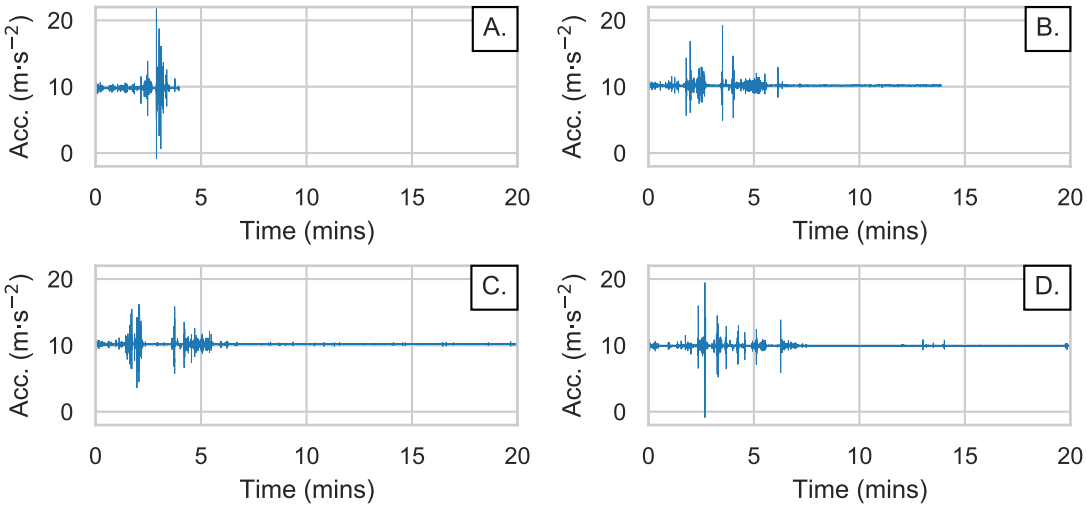

Figure 9.3: Waveforms of as-sampled z-axis accelerations for the first 6 datasets recorded at LRI after arrival.

study by Blaxter et al. [44] did report significant 20 & 100 ms shocks inside the transport incubator during neonatal transfers. In their study, the authors focused on events with magnitudes of at least 2 g with the greatest number of shocks found at the start and end of transfers — thought to be due to the loading/unloading procedures. It was therefore decided to implement the equivalent measure here, both to corroborate the findings and see how hospitals may compare. As outlined in Section 8.4.3.2, additional thresholds were set to provide enhanced clarity, resulting in the final comparison magnitudes being ½, 1, 2 & 3 g. Only time periods of 20 ms were investigated due to the infrequency of high-magnitude events during the journeys.

9.3.1.2 Frequency weighted vibration

Weightings to highlight the acceleration frequencies that have the greatest effect on adults are supplied, along with measures of their impact, by ISO 2631 [49]. Unfortunately, the measures provided mainly focus on the exposure values where duration is a component and as the duration was shown to be inconsistent and untrustworthy, combined with the reduced magnitudes for extended periods, these were not suitable for hospital analysis. Instead, a variety of peak metrics, both derived from previous studies and created for this work, were implemented.

Classifications of vibration according to perceived comfort levels are also supplied within ISO 2631. These ranged from "not uncomfortable" to "extremely uncomfortable" (see Figure 8.3) and were provided to summarise extended periods of vibration. Bouchut et al. [43] applied these in their study, where the number of 1 s average Whole-

Body Vibrations (WBVs) that exceeded 2 $m \cdot s^{-2}$ — the lower bound of "extremely uncomfortable" — were counted. This was furthered here by also tallying the number of 1 s averages that exceeded all of the remaining levels of perceived comfort, to ensure there was data to compare if magnitudes never breached the worst classification.

Finally, the maximum 1 s peak values were reported for each hospital, mainly to provide a comparison with the research performed by Karlsson et al. [40].

9.3.1.3 Noise

Although vibration levels were expected only during significant interactions with the trolley, noise levels tend to be present at all times. The average noise levels were therefore assessed to see how the ambient environments compared.

As with comfort-weighted vibration, Bouchut et al. [43] reported the number of occasions 1 s averaged noise levels were at levels of at least 85 A-weighted dB (dB(A)) during the ambulance journeys in their study. This count was also implemented here, with additional thresholds of 75 & 80 dB(A) included to account for the generally lower levels recorded.

One group of metrics that could give an indication of potential infant disruption are the number of marked increases in noise across consecutive seconds. These formed part of a study by Kuhn et al. [68] where both 5–10 & 10–15 dB increases over the average for the previous second were found to have a negative physiological effect in the Neonatal Intensive Care Unit (NICU). The number of these increases were counted here, along with a further metric that tallied the number of increases of >15 dB.

9.3.2 Assumptions of the data

Before progressing to the results of the analyses, it is important to discuss the assumptions that were required in order to perform assessments. These assumptions were essentially due to inherent limitations with the hospital aspect of data collection using the app.

The first assumption that was required was that the smartphones were either attached to the trolley when recordings were started & ended, or at least were not placed/retrieved with excess force. By making this assumption, all high impact data recorded by the devices could be treated as being caused by interactions with the trolley. It can be seen in the earlier waveforms (Figure 9.2B & C) that some recordings registered significant events shortly after starting, however they resulted in a similar number of

events. It is unknown whether the early peaks were caused by interactions with the smartphones directly, and were therefore left untouched as incorrect removal could also affect final results. Collecting a large quantity of hospital recordings, however, enabled averaging to be performed that would effectively remove the impact of outliers and therefore increase the confidence in the data reflecting solely the departure/arrival procedures.

A further assumption was that all data recorded before the ambulances departed, and after arrival, reflected areas of the hospital that would be traversed when a patient was present. In other words, there were no events that were recorded before the point at which a neonatal infant would be transferred from NICU incubator to the transport incubator on the trolley, and no data after the infant would be removed upon arrival. These periods of time would obviously not be of interest as they would not affect infants, and therefore their inclusion may affect the final results. Unfortunately, due to being inside the hospitals and thus having poor quality location data, it was not possible to determine whether the trolleys were in areas of interest and therefore no data were excluded.

9.4 Influence of patient presence

Large amounts of data provide the greatest confidence in results due to the decreased effect of anomalous recordings when averaging. Therefore, all recorded data collected would ideally be comparable with one another. In a previous study, Blaxter et al. [44] declared a difference was found between the levels of vibration when a patient was transferred compared to journeys with a manikin in its place. Despite this not being reflected in the journey data recorded here (see Section 8.6.1) nor being shown by Blaxter et al. regarding peak values, it was worth checking for the hospital data to remove all doubt.

Considering the staff conducting transfers were all provided by CenTre, rather than by the hospitals at which the journeys were departing or arriving, trends identified within data at one hospital can reasonably be assumed to be similar to those at other hospitals. For example, if it was found that staff tended to induce greater levels of vibration, due to being more forceful, with an empty transport incubator than when a patient was inside at one hospital, it is unlikely that attitudes and results would differ at a different hospital. Despite this hypothesis, all hospital that recorded at least 2 datasets for each condition (with patient and without) were assessed.

Similarities for each metric were assessed using two-tailed t-tests as the directions of

potential differences were unknown, with Welch's t-test performed where the results of Levene's test indicated unequal variances between the two populations. These tests were conducted twice: once for data before departing and again for the data after arrival.

9.4.1 Similarities before departure

The majority of metrics, at each hospital before the journey, were found to have no significant differences whether or not a patient was present. Additionally, although the recorded durations varied there was not a clear relationship that indicated patient presence had an influence. This suggested that the majority of analysis could be performed using the combined datasets, regardless of whether they were recorded with an empty incubator.

The differences that were identified tended to occur at either BPH or GEH, with both significantly different in terms of the number of raw shocks (BPH ½ g: $t(36) = -2.7$, $p < .05$; GEH ½ g: $t(33) = -4.7$, $p < .001$) and ISO-weighted shocks (BPH 1 s "uncomfortable": $t(36) = -2.3$, $p < .05$; GEH 1 s "uncomfortable": $t(33) = -2.3$, $p < .05$), while BPH also differed in the number of 1 s events over 75 dB(A) ($t(36) = -2.4$, $p < .05$) and the number of 10–15 dB increases varied at GEH ($t(33) = -3.6$, $p < .001$). Therefore, only the datasets where an infant was transferred were used for further analysis of these two hospitals due to the numerous differences.

Only a single hospital, KGH, registered significantly different average noise levels when a baby was present compared to empty transfers. This meant that only the patient data could be used for this hospital, however it also suggested that the average noise levels did reflect ambient environments that would not have been expected to differ.

Hospitals that did not register enough data for statistical analysis, either due to too few (or no) patient transfers or too few transfers in general, were assumed to be uninfluenced by patient presence. This was to ensure they could be compared to other hospitals, although it was impossible to verify if the results reflected average expected levels.

9.4.2 Similarities after arrival

Unlike the data from before the journeys, multiple combinations of metrics and hospitals were found to vary with patient presence after arrival. Differences were greatest

in the number of high magnitude shocks, with 9 hospitals (50% of those that recorded enough data) significantly different in either raw shocks over ½ g and/or "extremely uncomfortable" 1 s ISO-weighted shocks. The reasons for these differences are unknown, however the hospitals that were deemed significant all registered a greater number of shocks where the incubator was empty suggesting that less care was taken during manoeuvres.

All metric results that were found to be significantly different were discarded from future analyses of the relevant hospitals, however all others were combined to form a single set of results. Again, hospitals that did not record enough data for statistical analysis were also assumed to have reasonable similarity to ensure data were available for comparison.

9.5 Hospital comfort analysis

Analysis of the metrics recorded during the hospital portions of neonatal transfers were performed in three stages. First, the data recorded before journeys were compared between hospitals to assess whether or not there were differences in the procedures or surface quality that may affect infants. This was then followed by analysis of the equivalent data after arrival at the destination hospital. The final stage of analysis consisted of comparing the levels before journeys with those after to see whether similar artefacts were present (due to entrance matting, for example).

The data from both stages were plotted using the same y-axis limits to enable visual similarities to be identifiable in the metric results.

9.5.1 Analysis of levels before departure

Thirty-six hospitals recorded at least one set of data before the journey segment of a neonatal transfer. The levels of noise and vibration were compared between these hospitals using the metrics described earlier, with the results set out below.

9.5.1.1 Unweighted vibration

The first metrics assessed were the number of unfiltered shocks that were sustained for 20 ms. Shocks of magnitude 2 g, as used by Blaxter et al. [44], were very rare in occurrence, with only 8 (22%) of hospitals registering at least 1 instance. A reduced threshold of 1 g, however, was found to be exceeded in 31 (86%) of hospitals (Figure 9.4)

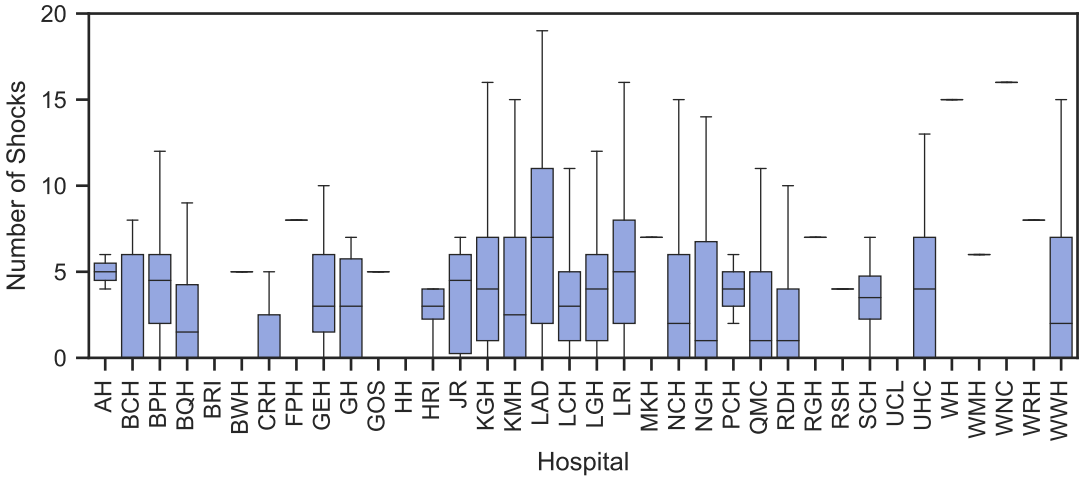

Figure 9.4: Number of 20 ms shocks ≥1 g recorded at each hospital before departure.

with 4 shocks expected on average. Median results varied from 0 to 16 occurrences, although these may have been abnormal events that cannot be accounted for due to only one set of data recorded all hospitals that recorded extremes apart from CRH.

A similar spread of data is evident when examining the number of shocks with a magnitude of at least ½ g. As expected, halving the threshold resulted in a marked increase in the number of events, with an over fourfold increase to an average of 18 (IQR: 13–25). The hospitals with the fewest datasets resulted in either a large number of shocks or no shocks at all, but it is unknown how accurately these single recordings reflect the average expected number.

9.5.1.2 Comfort-weighted vibration

The number of 1 s shocks that breached the lower limits of "a little uncomfortable" and "fairly uncomfortable" followed highly similar trends, albeit with 40% fewer occurrences for the latter (shown in Figure 9.5). On average, a hospital recorded 24 instances of shocks that were at least "fairly uncomfortable", with the worst 25% recording an average of 31 such events.

Increasing the threshold of interest to the "extremely uncomfortable" classification results in some hospitals becoming more prominent (Figure 9.6) suggesting that while the total shocks may have been relatively low, they were dominated by those at larger magnitudes. Two of the most prominent hospitals were LAD and KMH, with the spread of BPH and GH also of concern. While the average number of events was only 2 per hospital, it may be that these events cause severe distress and thus even this amount could cause harm.

The maximum WBV value recorded at each hospital was around 3.5 $m \cdot s^{-2}$ (IQR: 3.0–

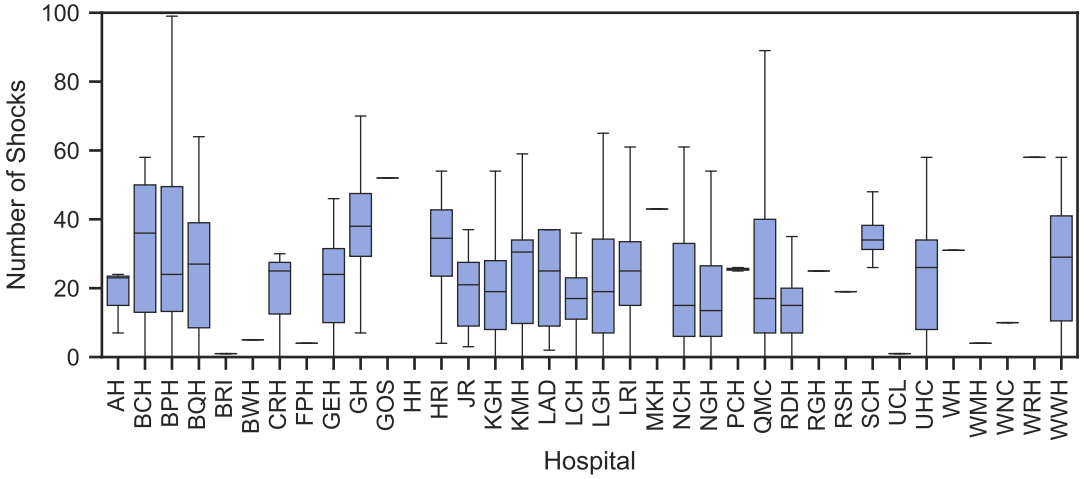

Figure 9.5: Number of 1 s WBV shocks $\geq$0.5 m·s^{-2} (the lower limit of "fairly uncomfortable") recorded at each hospital before departure.

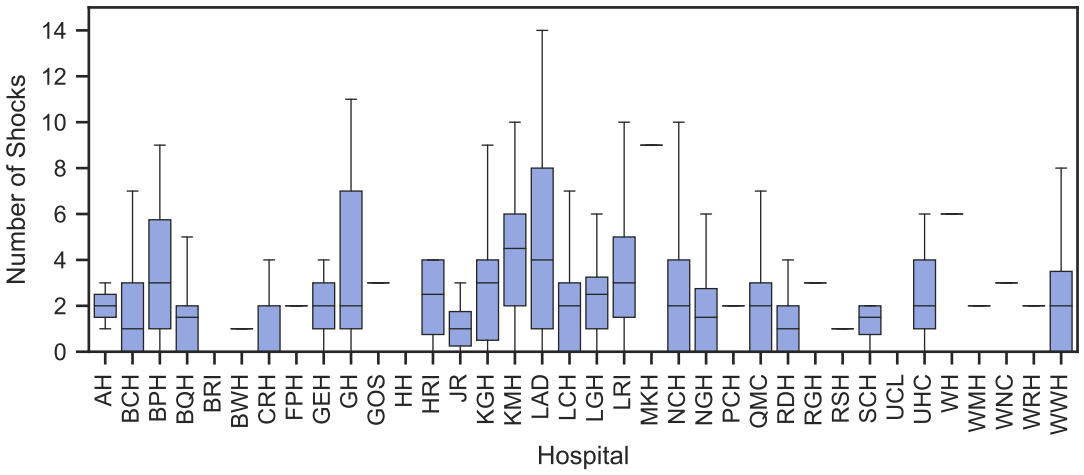

Figure 9.6: Number of 1 s WBV shocks $\geq$2.0 m·s^{-2} (the lower limit of "extremely uncomfortable") recorded at each hospital before departure.

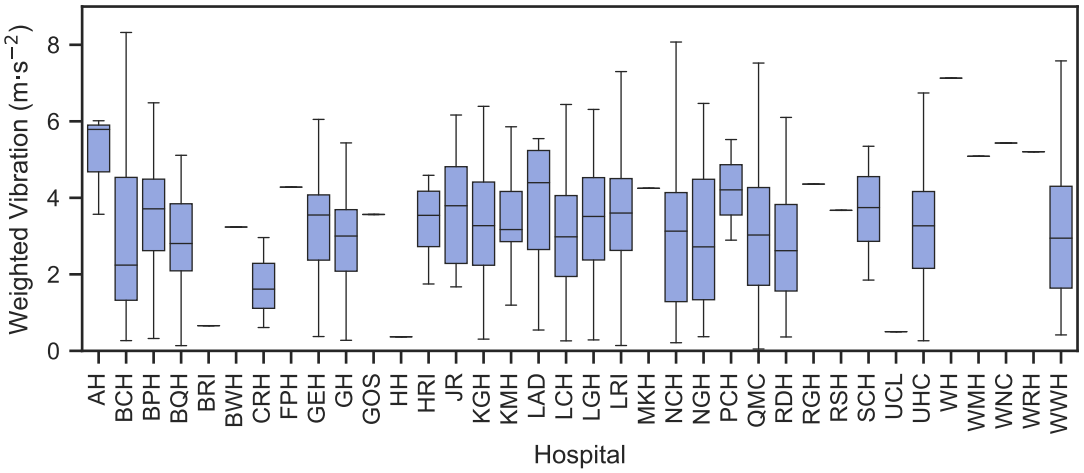

Figure 9.7: Peak 1 s WBV shocks recorded at each hospital before departure.

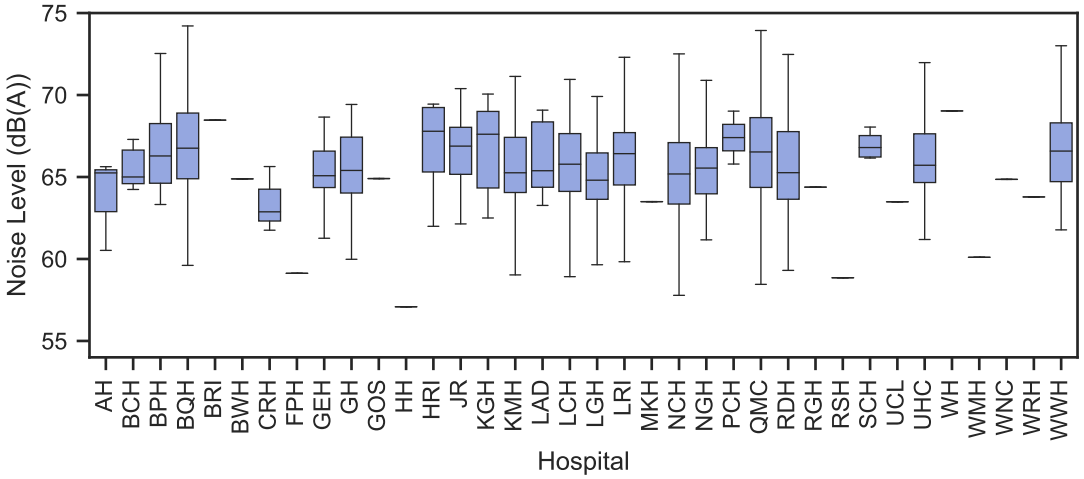

Figure 9.8: Average noise levels recorded at each hospital before departure.

4.2 m·s^{-2}), and would be classified as "extremely uncomfortable" (Figure 9.7). A large spread was visible within the data, with the 25th percentile of hospital data registering an average of 1.8 m·s^{-2} less than the 75th percentile (2.5 vs. 4.3 m·s^{-2}). One of the worst hospitals, AH, resulted in a peak value 2.3 m·s^{-2} greater than the average that suggests there is clear room for improvement. As most hospitals registered high magnitudes, a change in practice may be required.

9.5.1.3 Noise

Average noise levels tended to fluctuate by 3 dB, however were generally very similar with an average of 65 dB(A) (Figure 9.8). Apart from 3 hospitals that only recorded one dataset, hospitals exceeded the 60 dB(A) limit recommended for neonatal transport for the majority of transfers. It can clearly be seen that the NICU recommendation of 45 dB(A) is also significantly breached, suggesting large reductions are required.

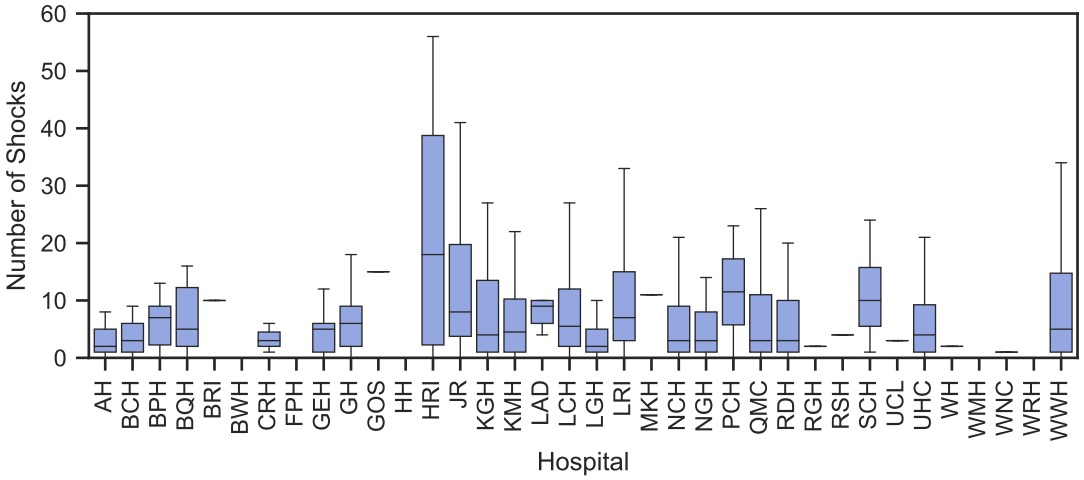

Figure 9.9: Number of 1 s noise levels ≥75 dB(A) recorded at each hospital before departure.

Figure 9.9 shows there were few occasions that hospitals registered 1 s noise levels of at least 75 dB(A), with even fewer found to exceed 80 dB(A) and only 2 datasets that breached the limit of 85 dB(A) used by Bouchut et al. [43]. This was not entirely unexpected, as these levels were only evident in the routing analysis at high vehicle speeds. HRI registered the most events of at least 75 dB(A) at 4.5× the hospital average of 4 and double the 75th percentile. The infrequency of the noise events could suggest that they were caused by factors which could be improved in the future, such as the loading procedure or surface quality, and not the general sound environment at the hospitals.

There was further evidence for manoeuvring events causing bursts of noise with all hospitals registering at least 3 instances where the noise suddenly increased by 5–10 dB (Figure 9.10). Averaging 35 events (IQR: 26–48) for the 50th percentile of hospital data, it is also clear that work is required to stabilise the environment that the infant undergoing transfer is exposed to before departure.

Maximum instantaneous noise levels recorded at each hospital were found to vary only slightly and averaged at 87.2 dB(A), with all datasets registering over 80 dB(A) (Figure 9.11). The peak magnitudes found at each hospital, however brief, are equivalent to being surrounded by heavy traffic or operating a lawnmower and are clearly not suitable for a comfortable environment.

9.5.2 Analysis of levels after arrival

After investigating the levels before the ambulance departed for the journey stage of the transfers, analysis turned to the 38 hospitals that recorded data upon arrival and

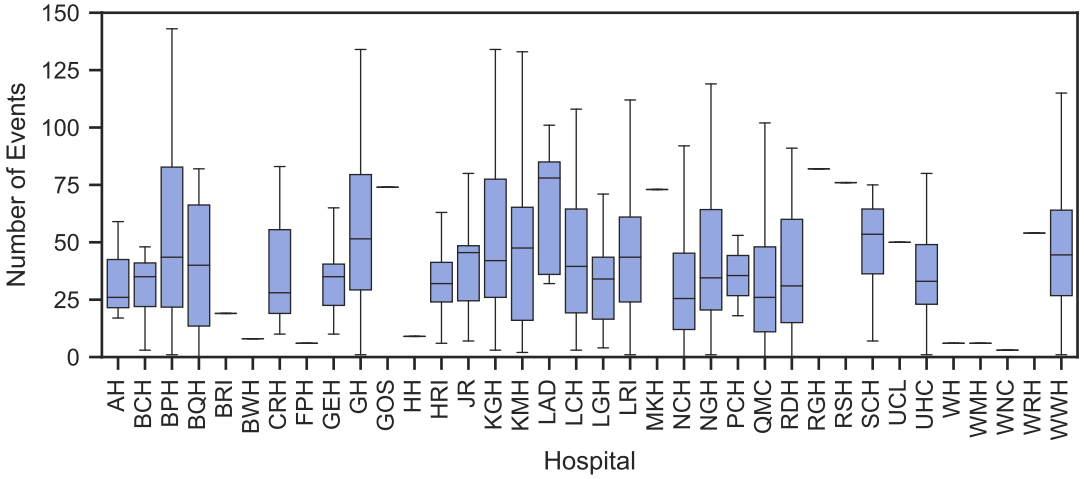

Figure 9.10: Number of 5–10 dB increases in noise level, over subsequent seconds, recorded at each hospital before departure.

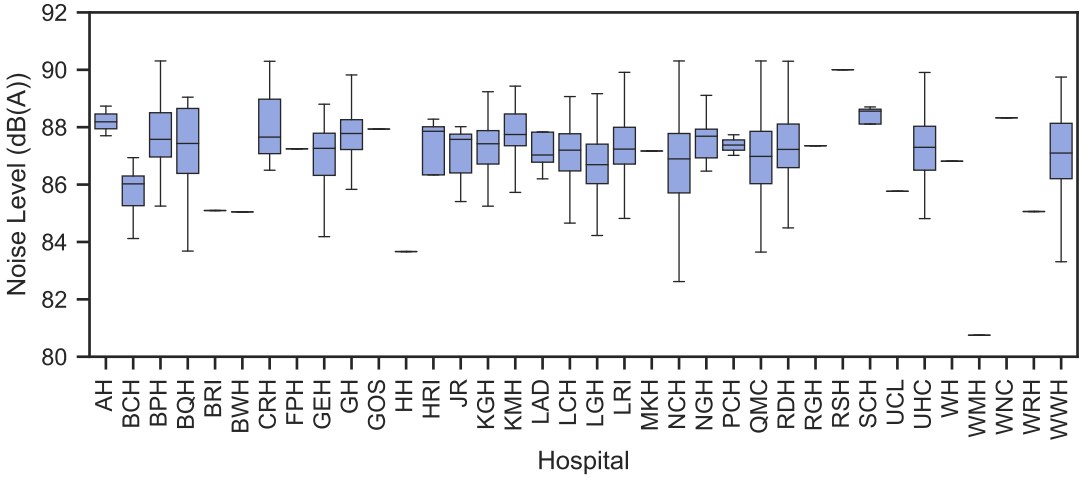

Figure 9.11: Peak instantaneous noise levels recorded at each hospital before departure.

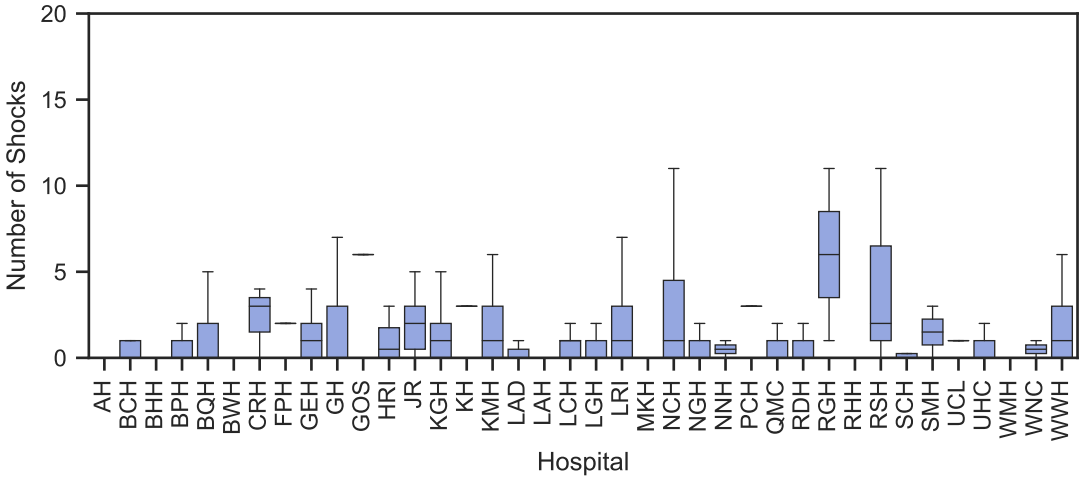

Figure 9.12: Number of 20 ms shocks ≥1 g recorded at each hospital after arrival.

how they compared.

9.5.2.1 Unweighted vibration

In contrast to the data from before departures, recordings after arrivals rarely registered any shocks of at least 1 g (Figure 9.12) with only the worst 25% of hospital datasets averaging at least one event. Eighteen (47%) of the hospitals analysed had no instances of 1 g shocks for at least half of recordings, while the worst hospitals only experienced 6 shocks. From the first metric alone, it appeared that the circumstances involved with loading the ambulance would cause greater discomfort than the unloading procedures.

Comparing data that resulted from the ½ g threshold resulted in an increase in the number of events to 12 (IQR: 7–19) per hospital. This was a third less than the number expected during loading and furthered the case for differences in comfort between both stages.

9.5.2.2 Comfort-weighted vibration

On average, a recording at a hospital after arriving from a transfer resulted in 32 shocks of 1 s WBV above the "fairly uncomfortable" lower bound (Figure 9.13). Again, this differed from the data from before departure, however in this case the average number of shocks upon arrival was greater than the 75th percentile before leaving. Data here also appeared to be more consistent, with reduced IQRs for the majority of hospitals.

In Figure 9.14, it can be seen that 14 (37%) of hospitals registered no WBV shocks, on average, above 2 m·s^{-2} — similar to the lack of 1 g shown above. This led to the overall average expected value of 1 shock per hospital (lower than the 2 per hosptal

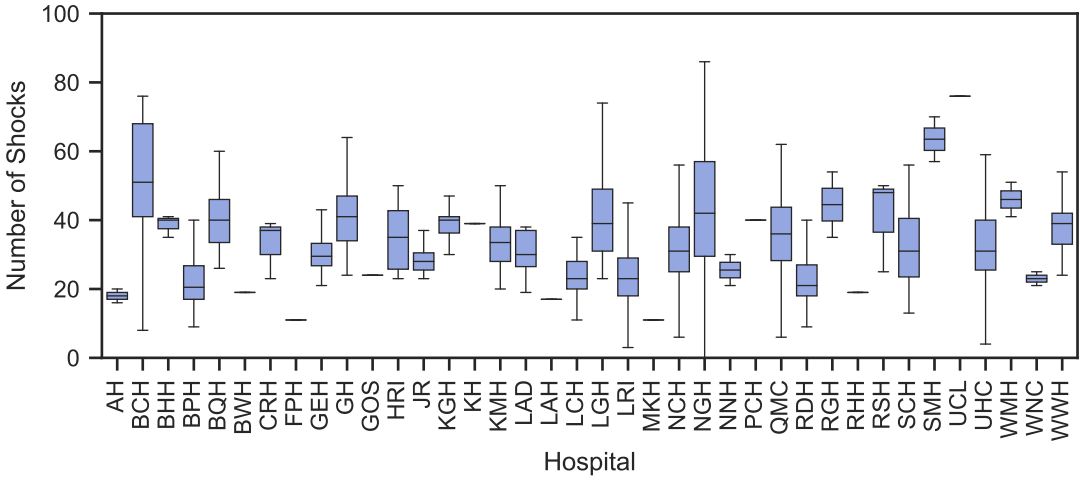

Figure 9.13: Number of 1 s WBV shocks $\geq$0.5 m·s^{-2} (the lower limit of "fairly uncomfortable") recorded at each hospital after arrival.

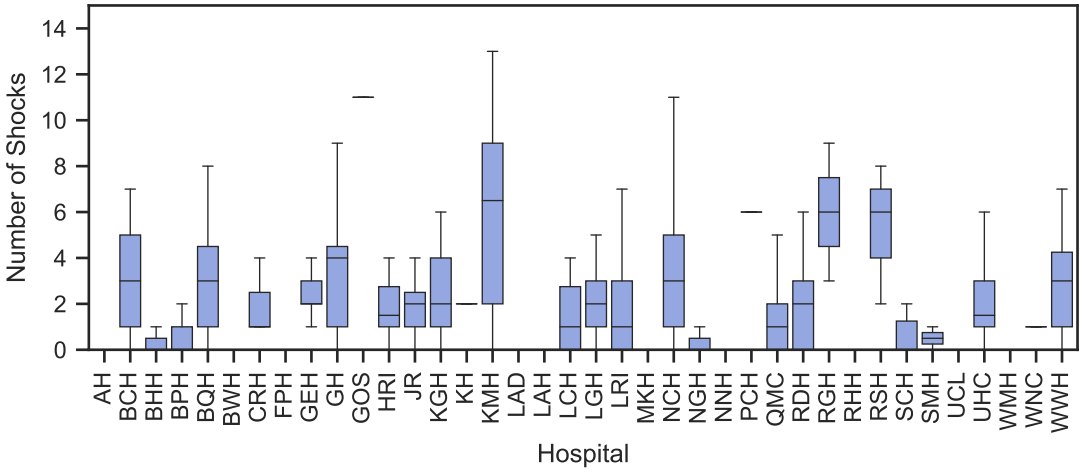

Figure 9.14: Number of 1 s WBV shocks $\geq$2.0 m·s^{-2} (the lower limit of "extremely uncomfortable") recorded at each hospital after arrival.

before departure), despite hospitals such as KMH registering an average of 6.5 over the course of 40 datasets.

Figure 9.15 shows the distribution of 1 s maximum WBV recorded at the different hospitals after the ambulance journey. There was very little variation in each of the hospitals, with the difference in the IQRs being 0.8 m·s^{-2} (less than half the 1.8 m·s^{-2} encountered before departures). Average values for the hospitals remained classed as "extremely uncomfortable" at 2.4 m·s^{-2}, although this was less than the 25th percentile of data before journeys, suggesting the loading procedures require the most improvement.

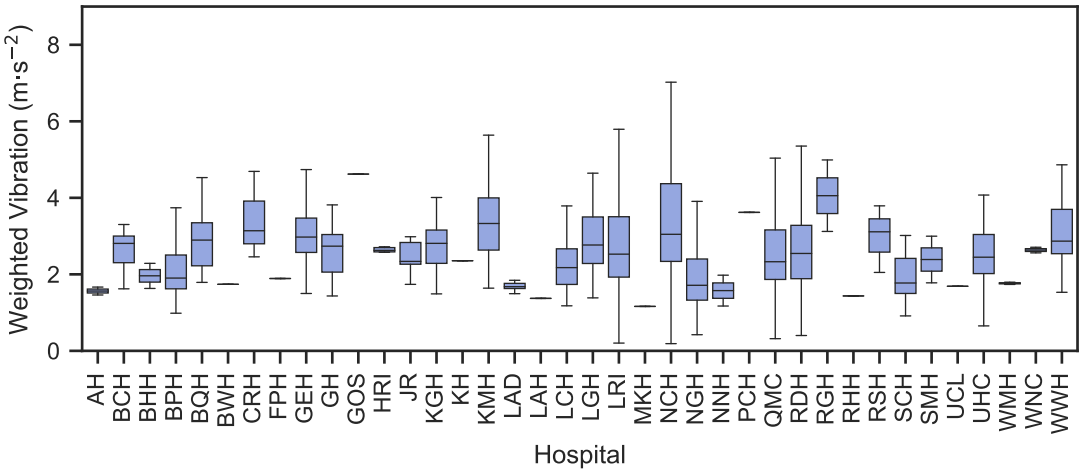

Figure 9.15: Peak 1 s WBV shocks recorded at each hospital after arrival.

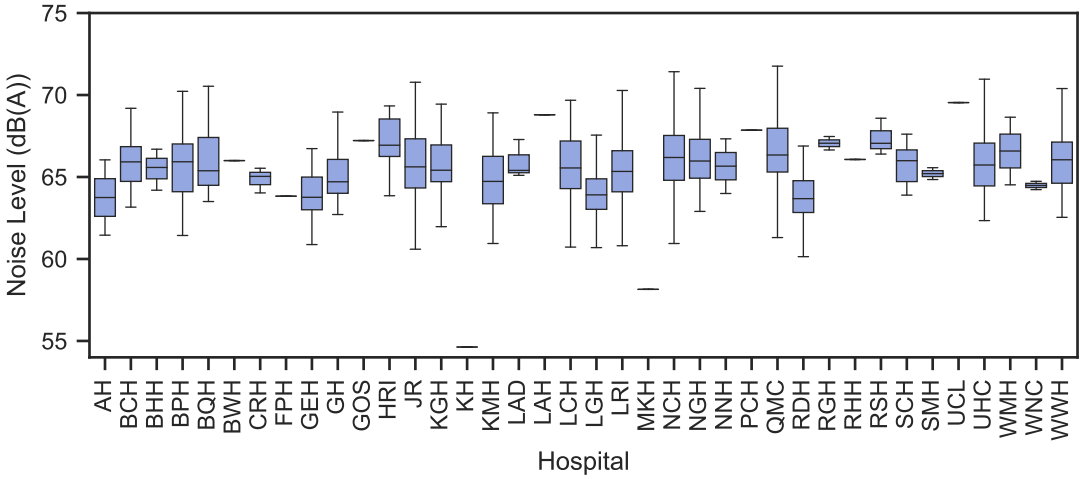

Figure 9.16: Average noise levels recorded at each hospital after arrival.

9.5.2.3 Noise

Figure 9.16 shows the similarity between the average noise levels recorded after arrival at the different hospitals, where the median was around 66 dB(A). This was larger than the levels before departure, however the differences were minimal and could reasonably be assumed negligible. Again, all but two hospitals (that possessed only one dataset) breached the transfer limit of 60 dB(A) with all hospitals clearly exceeding the 45 dB(A) NICU limit.

Noise levels of at least 75 dB(A) (as shown in Figure 9.17) were found to be infrequent at the majority of hospitals with an average of 6 events. This average, along with the IQRs between the hospitals, tended to be 2 events more than the equivalents before departure, however unless these occurrences are shown to have a drastic impact on the infants there appears to be minimal differences between the stages.

Noise levels at each hospital fluctuated frequently after journeys (Figure 9.18), with

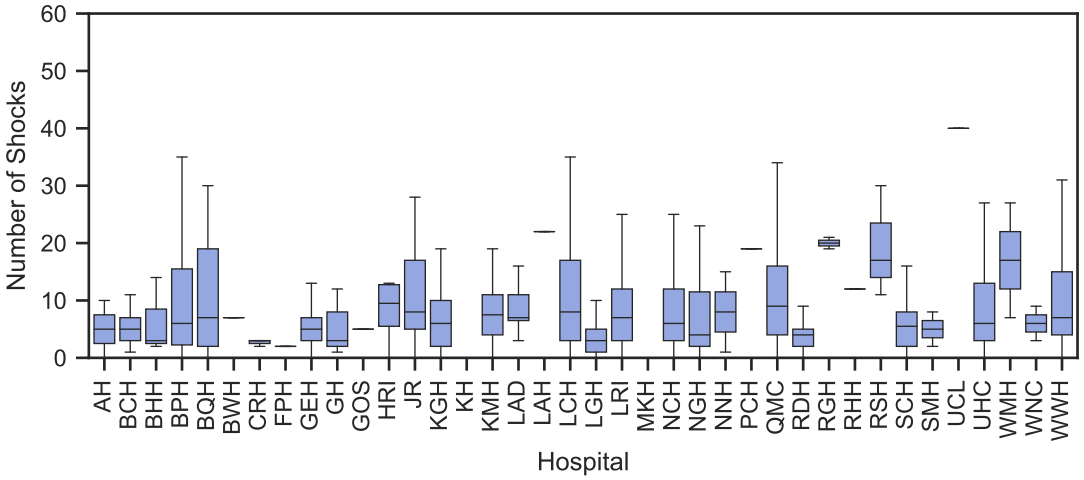

Figure 9.17: Number of 1 s noise levels ≥75 dB(A) recorded at each hospital after arrival.

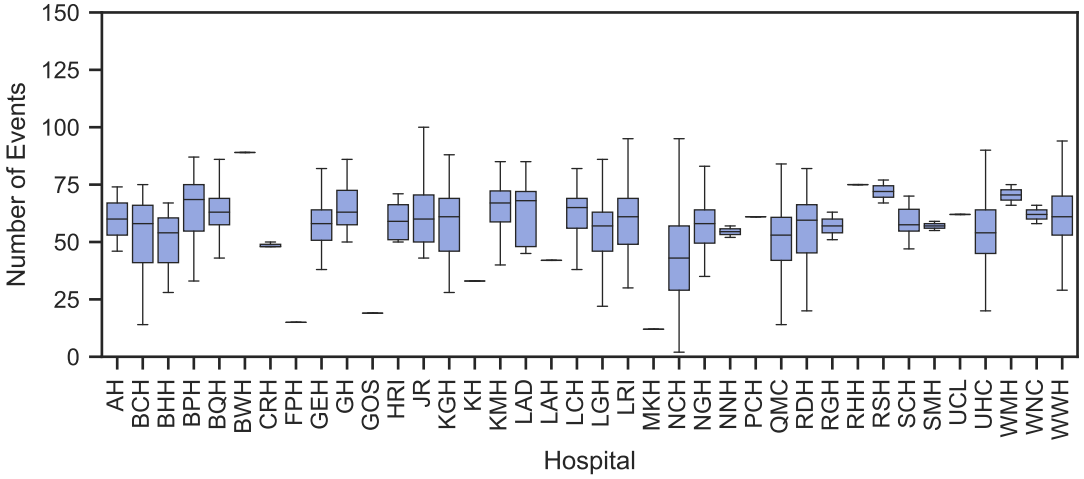

Figure 9.18: Number of 5–10 dB increases in noise level, over subsequent seconds, recorded at each hospital after arrival.

an average of 59 increases of 5-10 dB between seconds. While each hospital registered multiple changes, there were high levels of consistency (IQR: 51–64) suggesting that the values were strongly influenced by the general procedures and environments rather than random events. Compared to before departure, data after journeys could expect 68% more noise fluctuations implying that the latter stage is less stable.

Little variation was seen in the maximum instantaneous noise levels at each hospitals, shown in Figure 9.19. Values were marginally (0.4 dB) less for the post-journey stage than before, however this difference was insignificant and all levels ranged around the same levels.

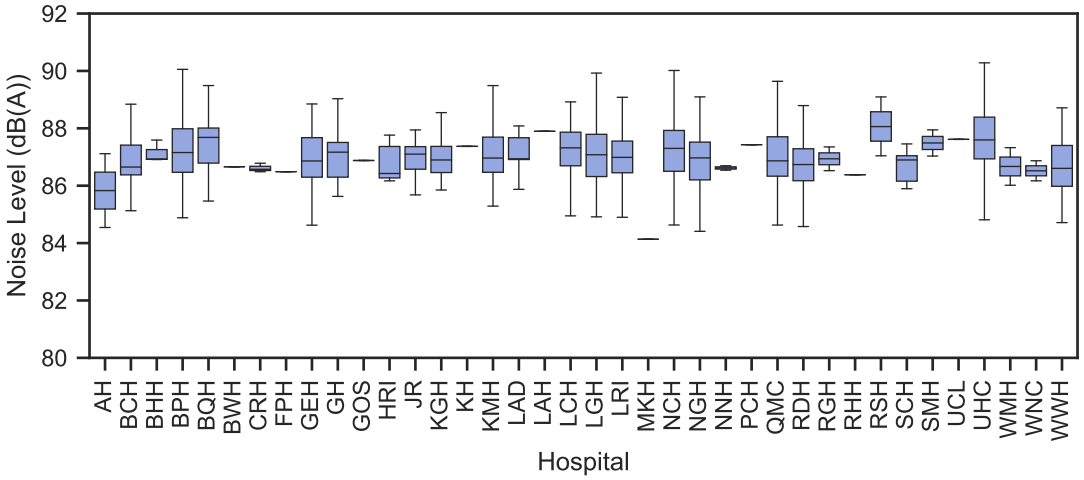

Figure 9.19: Peak instantaneous noise levels recorded at each hospital after arrival.

9.5.3 Loading versus unloading the ambulance

Several differences between the data recorded before and after the ambulance section of transfers became apparent during the examination of metric results at the various hospitals. Analysis was therefore required to see how much of the variation between both stages was due to the procedures involved in the stage itself, and how much could be deemed as random.

Twenty-two (52%) of the hospitals in the data collection had at least 2 recordings before and 2 recordings after the journey segment, and could therefore be included in statistical analysis of any differences between both stages. Data from each hospital was grouped by the stage of transfer and compared using t-tests (either standard or Welch's, dependent on the homoscedasticity) to show the similarity of the averages. These statistics were combined with examination of the absolute differences to determine true significance.

9.5.3.1 Unweighted vibration

In the above sections, it was found that 1 g shocks were more prominent before departure, with just under half of hospitals reporting no occurrences. Thirteen hospitals (59% of the 22 applicable to this section) were found to be statistically different, with all more likely to register a greater number before departure. Although the quantities were low (2–4), these events were rare and it is possible that these are sufficient to cause distress to the infants.

Similarly, 8 hospitals (36%) registered significantly more ½ g shocks (4–22) during the period before departure than after arrival. One hospital, NCH, was also shown to result

in significant differences, however in this case the period after arrival registered almost 18 (43%) additional shocks than the time before departing.

9.5.3.2 Comfort-weighted vibration

Differences were again abundant in the number of WBV shocks that exceeded the lower limit of the "fairly uncomfortable" classification, with 12 hospitals (55%) registering significantly more (8–25, 24–59%) shocks during the unloading phase than during loading. This was the inverse of what was seen in the unfiltered shocks at greater magnitudes, suggesting that the general composition of events were varied between the stages. It should be noted that one hospital, BPH, was found to result in 13 shocks fewer during the unloading stage than the loading.

Far fewer hospitals significantly differed in the number of shocks which would be perceived as "extremely uncomfortable". Those that did followed similar trends to the ½ g shocks, with 5 hospitals (23%) resulting in a decrease (1–4) from loading to unloading, whereas NCH increased by an average of 3 occurrences.

This trend is again seen in the 1 s peak WBV values, where 7 hospitals (32%) were at significantly higher magnitudes before departure by 0.6–3.6 $m \cdot s^{-2}$, although two of the worst increases were at hospitals with few datasets and therefore the expected values cannot be confidently determined. NCH was again significant and an outlier to the general pattern, resulting in peak values that were an average of 0.7 $m \cdot s^{-2}$ (18%) greater during the post-journey period.

9.5.3.3 Noise

Average noise levels tended to be similar for both stages of journeys that were conducted at hospitals. Although 6 hospitals (27%) were determined to be statistically significant in their differences (5 hospitals louder during loading, NCH louder during unloading) these differences were only around 1 dB in magnitude and therefore would likely have a minimal impact upon the patient.

All hospitals apart from NCH were found to be similar in the number of sudden noises of at least 75 dB(A). Although the number of events after arrival at NCH were determined to be significantly greater than before departure ($t(517) = -2.8$, $p < .01$), this only resulted in an increase of 3 events to an expected total of 6.

Fifty-five percent (12) of the hospitals that recorded sufficient data both before and after the journey were found to result in significantly more 5–10 dB increases over con-

secutive seconds during the unloading period. This resulted in an increase of 19–40% (9–24) over the number of occurrences during loading, suggesting that the final stage of transfers was less stable. The trend was also evident in the number of 10–15 and >15 dB increases, with 10 and 3 hospitals, respectively, resulting in significant increases.

Similar to the average noise levels, although 7 hospitals were statistically different between the two stages these amounted to less than 2 dB and were therefore negligible.

9.6 Conclusion

Data obtained before and after the ambulance segment of neonatal transfer were recorded by CenTre to see whether the events during these stages may contribute to levels of discomfort and, consequently, the outcomes of transported infants. All data were collected by smartphones attached to the transport trolleys, using the developed app that was proven to accurately report noise and vibration.

The phases of neonatal transfer that occurred at the hospitals (i.e. the periods that were before and after the ambulance transfer) were initially identified using a combination of vehicle speed and radial location accuracy. These were used due to the locations of hospitals that would be involved in the neonatal transfers being unknown at the start of collection, and it was thought the transfer could be determined as having departed from a hospital once the speed had sufficiently increased. Upon further investigation, considerable chunks of data were either incorrectly removed or not removed from analysis — both of which would affect resultant analysis. Therefore, the coordinates were manually derived for each site with data being classed as at the hospital accordingly.

Due to setting more precise thresholds for determining hospital data, the 1,487 recordings that contributed to the routing analysis were reduced to 1,295 datasets of the time before departure and 1,440 datasets afterwards. In total, this resulted in data from 42 separate hospitals that were involved in at least one stage of neonatal transfers.

A range of metrics were used to provide an indication of possible discomfort and variation between hospitals as a detailed investigation is still required into the response of neonatal infants to vibration and noise stimuli. Events were expected to be sporadic during the periods of interest due to the nature of the physical interactions with the trolley and the variable periods of inactivity observed within the data. Therefore, metrics for assessment consisted of counting the number of occurrences of specific shocks, along with the average noise level to indicate the ambient environment.

Collection of data was undertaken both when a patient was inside the transport incu-

bator and where the incubator was empty. Analysis focused on the comfort of infants and thus both groups of data were compared to see if the empty datasets could be included. A large proportion of the combined hospitals and metrics were found not to have significant differences between both groups and were therefore treated as one. The instances where metric results were significantly different, those specific measures at the relevant hospitals were restricted in analysis to solely the patient transfers. Hospitals with insufficient data for comparison were treated as similar to ensure they were included in comparisons.

The 36 hospitals for which data before departure was available varied greatly in the number of recordings. Despite this, metric results tended to be quite similar with the number of events that breached the various thresholds quite sparse.

These findings were repeated after arriving at the 38 hospitals where data were collected. Results from metrics were again similar across the different hospitals, with no site particularly standing out.

Comparing data from the 22 hospitals that recorded at least 2 datasets during both loading and unloading phases resulted in the identification of a few trends. High magnitude vibratory shocks, be they as-sampled or frequency-weighted, were found to be significantly more frequent during loading than unloading suggesting a difference in the interactions with the trolley. The loading phase was also found to result in the greatest peak values of vibration. These findings were the case with all but one of the hospitals that were statistically different between both phases, with NCH found to have worse levels during the unloading phase for each.

Low impact vibrations, however, were found to be more abundant during the unloading phase compared to the loading. This suggested that the interactions in both phases had stark contrasts in their compositions.

Similarly, the number of sudden increases in noise were greater during the unloading phases. The reasons for this are unclear, as the average ambient noise levels were determined to be consistent for both stages.

While it was possible to assess metric values at the different hospitals, some assumptions were required with the data. The main assumption was that all data before or after (depending on the phase in question) the determined threshold coordinates corresponded to interactions with the transport trolley between the hospital and the ambulance. The app was originally designed with the intention of logging journeys, with the inclusion of satellite data to identify location and movement. This data was not reliably available inside the hospital buildings and therefore there was no spatial

awareness regarding what was happening. Therefore, no specific recommendations for future changes can be provided.

Additionally, although metric results could be compared between hospitals and phases, there was no feasible method of knowing what may have caused particular events both due to the lack of location data and the variation in timings involved.

Using the limited knowledge of neonatal response, combined with approximations of adult standards, it can reasonably be declared that work to reduce levels of noise and vibration during loading and unloading is required. Further work is required to determine exactly what physical stressors affect the infants being transferred to provide a clearer goal, however this data could be used as a reference. Similarly, more work is required in the recording of the noise and vibration of these phases in a more controlled manner. Although the results for each hospital averaged out, there was considerable variability and too many unknowns in the data that led to assumptions. By performing a more controlled collection where each interaction can be correlated with results the comfort during the hospital phases of transfers can then be optimised.

CHAPTER 10

Discussion and Conclusions

10.1 Introduction

This chapter summarises the overall thesis and presents comparisons of the recorded levels of vibration and noise with those from previous studies, along with the potential implications of the resultant values.

10.2 Thesis summary

The concept of neonatal transport was introduced in Chapter 2 along with the associated increase in Intraventricular Haemorrhage (IVH) that is the underlying concern of most previous research. Factors, such as noise and vibration, that are thought to influence this increase in IVH were outlined, along with the levels at which they have been observed, although conclusive links have yet to be identified. Both theoretical and experimental methods to reduce these factors were detailed and reviewed, with the idea of routing ambulances along smoother, quieter roads selected as having the greatest potential. This was due to the speed at which it could be implemented and the relatively small improvements that could be provided by other means. Methods of acquiring vibration and noise levels along roads were then assessed as this data were not available and would need collecting. There were no systems that existed that were suitable for the purpose of data collection, however it appeared possible that smartphones could be used.

Chapter 3 outlined the steps taken in the development of an Android app to gather the data required to inform routing choices. A simple interface with minimal interaction was implemented that both promoted continued use and reduced the risk of user error. The app recorded accelerations in all axes, maximum noise levels, and location data to

identify the roads of input. All variables were sampled at the fastest possible rates and logged to Comma-Separated Value (CSV) files that were automatically uploaded to a remote server for retrieval and analysis. Initial testing showed that data were repetitive for multiple journeys along the same stretch of road, strongly suggesting the road was the dominant input and not the vehicle.

Chapter 4 looked at the accuracy of the values registered by the app and the effects of different factors on results. The Inertial Measurement Unit (IMU) inside the smartphone used for testing was found to produce accelerations that were accurate up to 40 Hz, although there was significant jitter within the timestamps. Noise levels correlated well with the A-weighted dB (dB(A)) values from a precision sound meter, requiring a constant dB adjustment to match. Both the presence of additional audio (such as via the radio) and the use of an alternative smartphone were shown to affect the levels recorded by the app. This suggested that data collection via crowdsourcing was not a valid option and a more controlled method was instead required.

The use of the ambulances, operated by CenTre (CenTre Neonatal Transport), for data collection was outlined in Chapter 5. Four new smartphones were purchased for use during neonatal transfers after identifying a model that was capable of sampling at 200 Hz (and thus meeting the requirements for vibration analysis set out in International Organisation for Standardisation (ISO) 2631 [49]) while being reasonably priced to facilitate future expansion around the UK. Acceleration and noise were again similar to reference sensors, however with far reduced IMU sampling jitter and no adjustment required in the noise levels. Battery life was sufficient to record continuously for over 30 hours, reducing the risk of recordings being interrupted mid-journey. An asymmetric pattern of magnets were implemented for repeatable positioning of each device on the transport trolleys with minimal disruption to vibration transmission. Finally, modifications were made to the app to enable input of information describing each transfer, along with the ability to record the loading and unloading of the trolley onto the ambulance.

Chapter 6 described the steps taken to prepare the data, gathered over the course of 12 months inside the ambulances, for analysis. Data were extracted from the resultant 1,890 CSV files and inserted into a database for efficient storage and to enable quick extraction where conditions of interest were met. Upon investigation, some problems were identified within the data as recorded, due to the planned procedure not being followed or mistakes being made. Data were subsequently cleaned to result in a final dataset consisting of 1,487 distinct journeys that totalled over 81,000 km, or 1,318 hours, of driving.

The division of travelled roads into a network of unique segments was detailed in Chapter 7, with a view to using the maximum amount of data available to assess each road. Initially, a programmatic approach was attempted to enable the network to adapt to new roads/junctions, both during the data collection period with CenTre and for continuous analysis in future expanded collection. An algorithm was created using plus codes that grouped locations based on a global grid, rather than through repeated, computationally-heavy distance calculations. Despite modifying area sizes, road networks were found to be too complex due to varying intricacies (rural vs. urban, for example), combined with the non-linear and bidirectional tendencies. Instead, roads were split using manually identified coordinates of decision points. This resulted in the routes used during the 39 journeys from Nottingham City Hospital (NCH) to Leicester Royal Infirmary (LRI) being divided up into 34 edges that could form 31 different combinations. Including all data that traversed these roads produced further splits to increase the total to 114 unique edges.

Chapter 8 investigated the levels of vibration and noise recorded along roads of the NCH to LRI network, identified in the previous chapter, to see how they varied, what may influence values, and whether routing would provide a more comfortable experience. A range of metrics were defined, covering both averages and peak events, to give an indication of the potential expected discomfort in lieu of a known neonatal response. Data from all 588 journeys that traversed at least one edge were included in the analysis, with the differences between empty journeys and those transferring infants found to be largely insignificant. Increases in speed were found to lead to increases in both vertical vibration and noise level, and a reduction in side-to-side accelerations. Although there appeared to be trends relating metric results to road classifications, it is unknown whether this was a reflection of the speed differences and as such could not be assessed due to the lack of overlapping data. Results for edges varied, although the concrete-based section that was used by half of all possible routes was found to produce worse vibrations than the motorway that would otherwise be taken, despite slower speeds. Metric values were also found to vary between routes, however it appeared that the majority of the variability of the edges cancelled out. Ultimately, there was clear scope to change levels during transfers, with the most frequently driven route never minimising values.

Chapter 9 assessed the equivalent metric values recorded at hospitals both before and after the ambulance journey to see if the loading and unloading procedures may adversely affect infants undergoing transfer. Data of interest were identified using the proximity to manually-determined coordinates at each hospital, resulting in a reduced

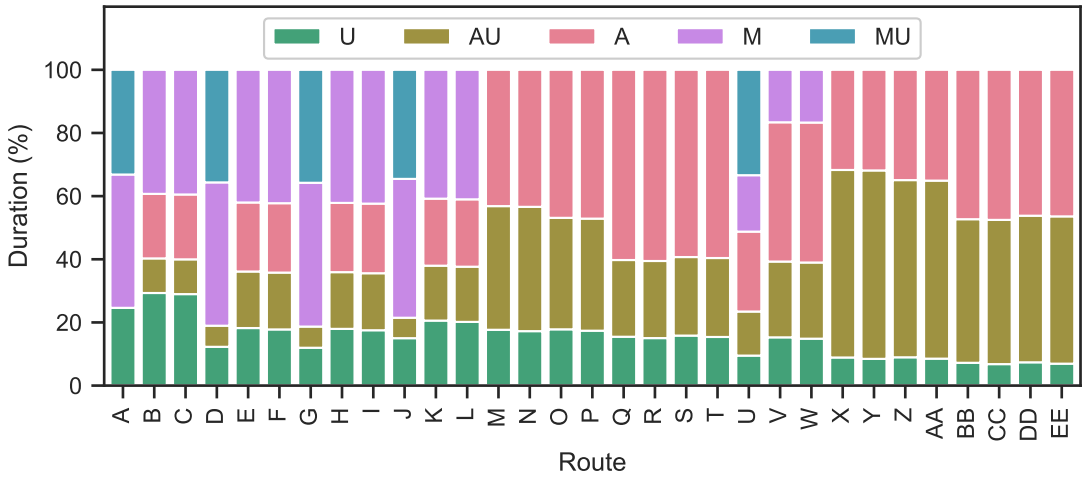

Figure 10.1: Road classification composition of each route in the NCH to LRI network, as a proportion of total duration.

dataset of 1,295 recordings before journeys and 1,440 after. These comprised of data from 42 hospitals, comprising of 36 before and 38 after. Durations of interaction with the transport trolley varied greatly between recordings, effectively restricting metrics to peak events. There was little variation between the results from different hospitals, however there were significant differences between the two phases and all data indicated there was room for improvement during transfers.

10.3 Further discussion

The similarity between the metric results for the routes was unexpected after the edges were shown to produce a range of values. It was thought that the similarities may have been due to equal proportions of the different road classifications within each route, however Figure 10.1 shows little conformity. Instead, it appears that it may simply be a coincidence that the differences in edge metrics were cancelled out for all routes of the chosen network, and more stark contrasts may be apparent when assessing roads between other hospitals. Additionally, while differences may have been minimal between the routes tested, further knowledge on the neonatal response may prove these differences to be significant.

None of the roads of the NCH to LRI network resulted in unfiltered shocks of the magnitudes observed by Blaxter et al. [44], while occurrences at hospitals were very rare. It was unclear, however, whether the data presented by Blaxter et al. were recorded by accelerometers attached to the incubator frame or the infant/manikin inside, although the latter may explain the disparity with the data here as the authors noted a gain in cer-

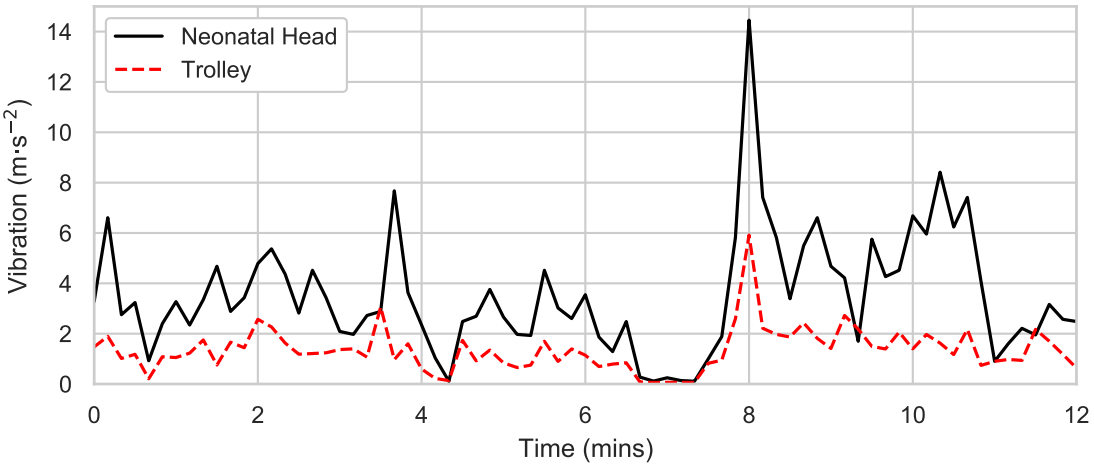

Figure 10.2: Comparison between 10 s r.m.s. vibrations recorded on the head of an infant inside a transport incubator and the vibrations on the trolley, recorded concurrently during a single transfer.

tain frequencies from the incubator chassis. This increase in magnitude could also be seen within the data here when compared with the data from a concurrent study that recorded accelerations on the heads of neonatal infants (Figure 10.2). The 10 s r.m.s. (Root-Mean-Square) values recorded on the head by the 3-axis accelerometer (3093B, Dytran, Chatsworth, CA, USA) connected to a human vibration meter (SV 106, Svantek, Warsaw, Poland) tended to be 47–70% greater than those recorded on the trolley, although this comparison was limited due to the meter being set to output values every 10 s and thus discarding all frequency data. Despite the smartphones resulting in lower levels to those experienced by the neonatal infants, recording on the trolley benefited from not requiring ethical approval due to having no contact with the infants and therefore provided an easier route towards collecting data that remained directly correlated.

The average Whole-Body Vibrations (WBVs) along the roads of the NCH to LRI network, calculated in accordance with ISO 2631 [49], were also found to vary from the previously found values. Deriving neonatal WBV from presented figures, Blaxter et al. [44] registered values between 0.37 & 0.44 m·s^{-2}, that provide some overlap with the route averages here of 0.43–0.49 m·s^{-2}. Both Bouchut et al. [43] and Karlsson et al. [40] reported average levels far lower at 0.35 m·s^{-2} and 0.25–0.30 m·s^{-2}, respectively, however these are likely underestimates due to calculations using median or mean methods rather than the r.m.s. that should be implemented.

Although the average WBV values were greater than those in previous work, the maximum 1 s values recorded by the smartphones during ambulance journeys were at significantly reduced levels. Bouchut et al. [43] and Karlsson et al. [40] both observed

values that breached 2 $m \cdot s^{-2}$, with Karlsson et al. reaching magnitudes of 2.9 $m \cdot s^{-2}$ that were around twice as large as those recorded here (1.4–1.6 $m \cdot s^{-2}$). Peaks during loading and unloading events, on the other hand, were quite similar between the collected data and the results of Karlsson et al., being in the region of 2–4 $m \cdot s^{-2}$.

Average noise levels recorded by the app were found to closely match those of previous studies. Two values were given by Buckland et al. [73] for roads in urban areas (65 dB(A)) and those in more rural settings (74 dB(A)), both of which matched the averages of the edges classed as urban or A-road/motorway respectively. Other studies [40, 42, 43, 47] all reported only single values between 65 & 73 dB(A), however it is unknown what roads or speeds were used to obtain these readings and thus they cannot be definitively determined to be different to the overall NCH to LRI averages of 73–75 dB(A) due to the significant influence of these factors.

Peak noise levels were at reduced values compared to most previous studies. Maximum magnitudes registered by Bouchut et al. [43], Karlsson et al. [40] and Prehn et al. [47] were all in the region of 85–100 dB(A), whereas no sustained levels above 85 dB(A) were observed by the smartphones. Although there was an upper limit of just over 90 dB(A) that could be reported by the smartphones, this did not explain the lack of events over 85 dB(A). It is possible that some noises, or specific frequencies, may be magnified by the incubator enclosure as all of the previous work sampled levels inside, however the differences are unknown. Unlike the other work, Bailey et al. [42] reported maximum noise levels of only 71 dB(A) that were far smaller even than those found in the smartphone data, however not enough data were provided to determine reasoning.

Although the levels of vibration and noise during ambulance are able to be altered through route choice, the resultant effects on infants remain uncertain. Very few studies have investigated levels of these inputs along with the physiological responses, while those that did reported conclusions using extremely small populations. Some of these studies have suggested that increased average values may influence neonatal comfort [40, 66] and others have determined that fluctuations are problematic [41, 68], but it is highly likely that a combination of all events are required to approximate infant distress. Additionally, there are two thresholds for noise that are provided for the Neonatal Intensive Care Unit (NICU) [62] and for transport [72], however these appear to have no medical backing specific to neonates.

There are multiple potential applications to routing by comfort, alongside the neonatal transfers that were the focus of this thesis. In fact, any journey that is not time-critical could benefit from an increase in comfort (or, rather, a reduction in discomfort). For

example, it may be beneficial to transport a patient with a neck injury along roads with reduced vibration to avoid any further damage being caused. A further potential use of a most comfortable route option is in the field of self-driving cars. These vehicles are becoming increasingly autonomous, leading to a potential future where the occupants have no interaction with the vehicle and instead are able to work or read during the journey. If this were possible, occupants would not necessarily require the quickest route due to the ability to utilise the journey better, and therefore increased comfort would be more enticing.

10.4 Future work

All future research into neonatal comfort require some level of standardisation to enable comparability between results. This was identified as a problem in Chapter 2 and is currently hindering the progress of knowledge in this field. Initial suggestions have been made and are presented in Appendix A.

A thorough understanding of the factors present during transport that affect neonatal infants is required to be able to confidently determine, and therefore provide to the transport teams, optimum routes. One possible method would be to compare the medical records of the infants transferred during the data collection in this thesis with the vast amount of noise and acceleration values. Performing these comparisons, with ethical approval, would hopefully identify some trends in the data of events that need to be minimised, etcetera.

As mentioned at the end of Chapter 9, there is a need for a more detailed analysis of the events during loading and unloading procedures. Data here provided an overview of the levels of vibration and noise during these phases, however further information is required to be able to give recommendations as to what surfaces or interactions may need improvement.

Although recommendations could be provided based on the levels of physical factors recorded on the trolley, vibration has been shown to be at greater magnitudes once it reaches the patient inside the incubator. Therefore, a transfer function to apply to the trolley values should be developed to accurately determine what the infants are experiencing without any obstructions to normal care practices.

Data collection should also be expanded to the whole of the UK to enable all transport teams to be provided with optimal routes for all transfers. There has already been expressed interest from members of the UK Neonatal Transport Group (NTG) (aside

from CenTre) with staff signalling concern of the potential noise and vibration to which the infants are exposed.

APPENDIX A

Standardising Research within Neonatal Transport

A.1 Introduction

The area of neonatal transport research regarding physical comfort is, by its nature, an interdisciplinary topic. First, there is the medical aspect that concerns the physiological responses of the infants. Second, there is the engineering focus of vibration, along with methods of reduction. Third, there is the intermediate ground of noise, that is a defined field in medicine as well as engineering. Both medicine and engineering disciplines require years of training and education before reaching a level where research is performed, however the differences between them often appear overlooked. Ideally, teams would consist of members with both medical and engineering backgrounds when researching an aspect of neonatal transport where either noise or vibration are involved, which should enable appropriate collection, analysis and interpretation of both physiological and physical data. Unfortunately, as discussed in Section 2.4, several studies have been performed by teams without the required knowledge.

There are no studies involving physiological data that were performed solely by engineers. This is because several studies conducted by engineers were investigating the mechanical aspects of vibration, and therefore physiological data were not pertinent or even available (due to dummy runs), while it is fully understood that a medical professional is required for any interactions with infants and for any interpretation of the subsequent results.

Research performed on physical factors by teams of medical professionals alone, however, appear to be of greater concern. Often, these studies provide limited information to the readers regarding the equipment used for data collection, the sample rates and

details about processing. This is not too great an issue with noise measurements as the outputs are highly standardised, although the frequent use of the term "mean" [40, 42, 73] suggests averages are underestimated (see Section 8.6.5.1). Vibration, on the other hand, has been logged and reported in a variety of ways, with sampling rates often unknown [41, 42] or requiring further interpretation [45, 47], insufficient or incorrect processing [41, 42] and the use of non-standard [41] or incoherent [45] units, all of which contribute to a reduction in comparability.

The lack of consistency in the results within this narrow field of research is hindering the progress of knowledge and, subsequently, the advancement of care. Recommendations for future research into the state (and subsequent effects) of noise and vibration will therefore be outlined over the subsequent sections. These sections will cover areas such as the pertinent information to report about the devices used, how data should be processed and what additional information would be useful to describe the conditions in which data were collected. A summary of all points is also provided in the final section.

A.2 Recording device details

The first piece of information that researchers should provide is the model of the device used for measuring the levels of noise and/or vibration, accordingly. Providing this information enables future readers to look up the relevant data sheets that detail the capabilities of the device, reducing the amount of that needs to be presented.

Next, the location of the microphone or accelerometer need to be outlined. Vibration differs depending on whether the accelerometer is within the incubator or attached to the trolley, as shown in previous studies and in the earlier discussion. Similarly, it is likely that noise levels differ inside the incubator to outside, especially if soundproofing materials are used as covers. It is therefore essential to provide the location information for readers to fully visualise the experiment and to provide possible insight as to reasons for or against data similarity between studies.

Specific to accelerometers are the need to report both the orientation and the method of attachment. The use of different axes orientations between studies would increase the likelihood of comparison errors, and therefore there is a need to standardise. Many studies already appear to use the same axes, however they will be repeated here to ensure consistency in future work. Each axis of the accelerometer should be aligned with those of the ambulance, where 'x' relates to the direction of travel, 'y' defines side-to-side motions and 'z' refers to the vertical components. It is important to use

the ambulance axes instead of the incubator to account for potential differences in / testing of interior layouts. Similar to the orientation, the method of attachment can influence results by altering the transmission of accelerations to the device (see Section 5.4.1). The method used should thus be reported and, where not recommended by the manufacturer and facilities are available, tested.

The rate at which the device samples the data must be reported to enable readers to understand what frequencies may impact results. Studies that have too low a sample rate may unknowingly ignore frequencies that have high magnitudes and therefore would result in lower overall values. Reporting this figure, along with the location mentioned earlier, would aid in the formation of hypotheses during comparisons between work.

There are multiple purpose-built devices that offer the capture of either noise or vibration that meet relevant standards, often for testing levels within the workplace. While these devices are capable of highly accurate readings, they do not tend to provide the raw, as-sampled data for analysis and instead perform various computations. To ensure full transparency in results, researchers that use such devices must report the use of any weighting settings (and specify which), the form of aggregation used, and the output period for the aggregated values. Weighting and aggregation can be defined as processing, which will be covered in the following section, while the time periods help describe the extent of the aggregation. The aggregations are typically performed to reduce the amount of data to analyse, however come with a trade-off against resolution where the larger the period, the less visible variability.

A.3 Processing of data

In the absence of neonatal-specific metrics that accurately predict the effects of noise and vibration on the patients, three filters are recommended (as a minimum) to enable comparison with both the known adult response and with research that has already been performed. The first recommendation is the use of A-weighted dB (dB(A)) in the analysis of noise resembles the perceived loudness of different frequencies to the adult ear and is already widely used. The remaining two filters both target vibration. Health impact and perceived comfort for adults is subject to the ISO 2631 standard, therefore comparisons should be made after applying the W_d filter to the horizontal (x & y) axes and W_k to the vertical. Additionally, while the response of infants remains unknown, the unweighted vibration may provide useful information and even result in new weightings being formulated. Apart from in the investigation of specific frequencies, it is strongly recommended to use a 0.5 Hz high-pass filter before analysis to

remove constant components such as gravity that would otherwise strongly skew results. Note that, while the use of A-weighting for noise and both ISO weighting & high-pass filtering for vibration are recommended for all studies to perform, this should not restrict the additional use of alternative frequency weightings that may show causality.

Along with filtering data appropriately, it is important to apply the relevant averaging methods. The choice of method is dependent on the data to be averaged, with methods for processing vibration being identical to those of noise. Where all data are from a single journey the averages should be computed using the r.m.s. method. Only where data from multiple journeys along the same section of road are involved should averaging be performed via mean or median (the latter useful to reduce the effect of outliers). It is not wholly appropriate to perform averaging across a range of roads due to the differences in input, but the mean/median could be used as long as the individual results are also reported. These methods are applicable to the individually sampled data, the device aggregates, and the data averaged along a journey, as long as the time intervals remain constant, however noise levels must be converted to Sound Pressure Level (SPL) [61] for all averaging to prevent underestimating results (see Section 8.6.5.1).

Information that describes each variable should be provided to give a snapshot of the recorded environment. This can be in the form of average values (using the appropriate method as outlined above) along with maximum recorded magnitudes. These two quantities indicate both the base levels and an indication of sudden shocks. As before, these are the minimum recommended values to be reported and, depending on the findings within data or future knowledge of infant response, they can be supplemented as the authors see fit.

The final step to the processing of data is to ensure that the correct units are used, both in calculations and in the reporting of results, to maximise comparability with minimal complication. This simply requires the use of the International System of Units (SI units) that, as the name suggests, are used globally as the preferred system. In terms of the quantities here: noise levels require reporting in dB (A-weighted or otherwise); vibrations in $m \cdot s^{-2}$ (not in g, as this involves unnecessary computations with no benefit); and frequencies should be reported in Hz.

A.4 General recording environment

Where the use of different devices may affect results, the same may be true regarding the core components of the transfer such as the ambulance, trolley and incubator configurations. As an example, research conducted in North American ambulances would

likely result in different levels of vibration compared to UK ambulances due to vehicle construction and suspension characteristics that are designed for the specific road networks. Similarly, the form of trolley and placement of components was strikingly different in the work by Green et al. [46] compared to here (Figure 2.1) and could reasonably be assumed to differ in vibration transmission; thus the models for all relevant components should be listed.

Significant differences were shown in the resultant levels of both noise and vibration depending on both the road classification and speeds at which journeys were traversed (Sections 8.6.2 & 8.6.3). It is therefore illogical to try and check for similarity between sets of data from research where the conditions were different. It is also partly for this reason that data should not be averaged across different roads from different journeys. Future work should provide details that pertain to the types of road used, and speeds along these roads, in order to further determine reasons of differences/similarity.

A.5 Summary

All of the above recommendations for future work were generated with the aim of providing a foundation for comparability after observing multiple problems within the current literature. The items detailed are recommended as a minimum that should be included by authors, as there is scope for additional information where new effects on infants are investigated. A collated summary of all recommendations is provided in Table A.1 for easy reference.

Table A.1: Summary checklist of the recommendations for future studies.

Details required regarding the recording device

a. Device model / part number
b. Location on trolley / within ambulance
c. Orientation of accelerometer (where relevant)
d. Method of attachment of accelerometer (where relevant)
e. Sample rate in Hz
f. Weighting settings (where set up on the device)
g. Aggregation (where present in device output)
h. Output rate / period (where aggregated)

Steps to standardise processing of data

i. Apply filters appropriately
   - dB(A) for noise
   - 0.5 Hz high-pass for unweighted vibration
   - W_d / W_k for compliance with ISO 2631 [49]
j. Conduct averaging using the correct methods
   - Within a single recording: r.m.s. (Root-Mean-Square)
   - Between recordings: mean or median
   - Convert noise levels to SPL before processing
k. Calculate both average and peak values
l. Present data in the appropriate units
   - Noise: dB(A)
   - Vibration: $m \cdot s^{-2}$
   - Frequency: Hz

General information to provide about recording conditions

m. Models of ambulance / trolley / incubator used
n. Type(s) of roads traversed
o. Vehicle speeds

Bibliography

[1] *Concise Medical Dictionary*. Oxford University Press, 2020.

[2] Office for National Statistics. Births in England and Wales: summary tables 2019. Technical report, 2020.

[3] NHS. Neonatal Critical Care Specifications, 2013.

[4] National Institute for Health and Care Excellence. Preterm labour and birth final scope. Technical report, 2013.

[5] Manuck TA, Rice MM, Bailit JL et al. Preterm neonatal morbidity and mortality by gestational age: A contemporary cohort. *American Journal of Obstetrics and Gynecology* 2016; 215(1): 103.e1–103.e14.

[6] Johnson S, Fawke J, Hennessy E et al. Neurodevelopmental disability through 11 years of age in children born before 26 weeks of gestation. *Pediatrics* 2009; 124(2).

[7] Pierrat V, Marchand-Martin L, Arnaud C et al. Neurodevelopmental outcome at 2 years for preterm children born at 22 to 34 weeks' gestation in France in 2011: EPIPAGE-2 cohort study. *BMJ* 2017; 358: 3448.

[8] Twilhaar ES, Wade RM, De Kieviet JF et al. Cognitive outcomes of children born extremely or very preterm since the 1990s and associated risk factors: A meta-analysis and meta-regression. *JAMA Pediatrics* 2018; 172(4): 361–367.

[9] Larroque B, Ancel PY, Marret S et al. Neurodevelopmental disabilities and special care of 5-year-old children born before 33 weeks of gestation (the EPIPAGE study): a longitudinal cohort study. *The Lancet* 2008; 371(9615): 813–820.

[10] Yeo KT, Thomas R, Chow SS et al. Improving incidence trends of severe intraventricular haemorrhages in preterm infants <32 weeks gestation: A cohort study. *Archives of Disease in Childhood - Fetal and Neonatal Edition* 2020; 105(2): F145–F150.

[11] Chen F, Bajwa NM, Rimensberger PC et al. Thirteen-year mortality and morbidity in preterm infants in Switzerland. *Archives of Disease in Childhood - Fetal and Neonatal Edition* 2016; 101(5): F377–F383.

[12] Watson SI, Arulampalam W, Petrou S et al. The effects of designation and volume of neonatal care on mortality and morbidity outcomes of very preterm infants in England: retrospective population-based cohort study. *BMJ* 2014; 4(7): e004856.

[13] Marlow N, Bennett C, Draper ES et al. Perinatal outcomes for extremely preterm babies in relation to place of birth in England: the EPICure 2 study. *Archives of Disease in Childhood - Fetal and Neonatal Edition* 2014; 99(3): F181–F188.

[14] Lasswell SM, Barfield WD, Rochat RW et al. Perinatal Regionalization for Very Low-Birth-Weight and Very Preterm Infants. *JAMA* 2010; 304(9): 992.

[15] Phibbs CS, Baker LC, Caughey AB et al. Level and Volume of Neonatal Intensive Care and Mortality in Very-Low-Birth-Weight Infants. *New England Journal of Medicine* 2007; 356(21): 2165–2175.

[16] Department of Health. Report of Department of Health Working Group on Neonatal Intensive Care Services. Technical report, 2003.

[17] Gale C, Santhakumaran S, Nagarajan S et al. Impact of managed clinical networks on NHS specialist neonatal services in England: population based study. *BMJ* 2012; 344.

[18] Fenton AC, Leslie A and Skeoch CH. Optimising neonatal transfer. *Archives of Disease in Childhood - Fetal and Neonatal Edition* 2004; 89(3): F215—-F219.

[19] Working together to optimise care for babies and their families during transportation. URL `https://ukntg.net/` (Accessed: 2020/12/16).

[20] Cornette L. Transporting the sick neonate. *Current Paediatrics* 2004; 14(1): 20–25.

[21] Jackson A and Devon C. UK Neonatal Transport Data 2019. In *UK Neonatal Transport Group Meeting*. Southampton.

[22] Helenius K, Longford N, Lehtonen L et al. Association of early postnatal transfer and birth outside a tertiary hospital with mortality and severe brain injury in extremely preterm infants: observational cohort study with propensity score matching. *BMJ* 2019; 367: l5678.

[23] Shipley L, Gyorkos T, Dorling J et al. Risk of Severe Intraventricular Hemorrhage in the First Week of Life in Preterm Infants Transported Before 72 Hours of Age. *Pediatric Critical Care Medicine* 2019; 20(7): 638–644.

[24] Mohamed MA and Aly H. Transport of premature infants is associated with increased risk for intraventricular haemorrhage. *Archives of Disease in Childhood - Fetal and Neonatal Edition* 2010; 95(6): F403—-F407.

[25] Watson A, Saville B, Lu Z et al. It is not the ride: inter-hospital transport is not an independent risk factor for intraventricular hemorrhage among very low birth weight infants. *Journal of Perinatology* 2013; 33(5): 366–370.

[26] Stevens B, Johnston C, Petryshen P et al. Premature infant pain profile: development and initial validation. *The Clinical journal of pain* 1996; 12(1): 13–22.

[27] Harrison C and McKechnie L. How comfortable is neonatal transport? *Acta Paediatrica* 2012; 101(2): 143–147.

[28] Stevens BJ, Gibbins S, Yamada J et al. The Premature Infant Pain Profile-Revised (PIPP-R): Initial Validation and Feasibility. *The Clinical Journal of Pain* 2014; 30(3): 238–243.

[29] Zwissig M, Rio L, Roth-Kleiner M et al. Measurement of stress in stable neonates during ambulance transportation: A feasibility study. *Australian Critical Care* 2019; 32(1): 28–33.

[30] van Dijk M, Peters JW, van Deventer P et al. The COMFORT Behavior Scale. *AJN The American Journal of Nursing* 2005; 105(1): 33–36.

[31] Lee SK, Zupancic JA, Pendray M et al. Transport risk index of physiologic stability: A practical system for assessing infant transport care. *Journal of Pediatrics* 2001; 139(2): 220–226.

[32] Grass B, Ye XY, Kelly E et al. Association between Transport Risk Index of Physiologic Stability in Extremely Premature Infants and Mortality or Neurodevelopmental Impairment at 18 to 24 Months. *Journal of Pediatrics* 2020; 224: 51–56.e5.

[33] Snedec N, Simoncic M, Klemenc M et al. Heart rate variability of transported critically ill neonates. *European Journal of Pediatrics* 2013; 172(12): 1565–1571.

[34] Grosek S, Mlakar G, Vidmar I et al. Heart rate and leukocytes after air and ground transportation in artificially ventilated neonates: a prospective observational study. *Intensive Care Medicine* 2009; 35(1): 161–165.

[35] Vesoulis ZA and Mathur AM. Cerebral autoregulation, brain injury, and the transitioning premature infant. *Frontiers in Pediatrics* 2017; 5.

[36] British Standards Institution. BS EN 13976-1:2018 Rescue systems - Transportation of incubators - Interface Requirements. Technical report, British Standards Institution, 2018.

[37] Top 10 most comfortable cars on sale | Auto Express. URL `https://www.autoexpress.co.uk/best-cars/101118/top-10-most-comfortable-cars-on-sale` (Accessed: 2021/05/31).

[38] Ford Motor Company. *Ford Fiesta Owner's Manual*. Detroit, MI, USA, 2014.

[39] Škoda Auto. *Škoda Kodiaq Owner's Manual*. Mladá Boleslav, Czech Rupublic, 2016.

[40] Karlsson BM, Lindkvist M, Lindkvist M et al. Sound and vibration: Effects on infants' heart rate and heart rate variability during neonatal transport. *Acta Paediatrica* 2012; 101(2): 148–154.

[41] Hall V. *A study to investigate whether speed and road conditions have an effect on the physiological stability of sick and preterm babies undergoing inter-hospital transfer by ambulance*. Dprof, University of Salford, 2017.

[42] Bailey V, Szyld E, Cagle K et al. Modern Neonatal Transport: Sound and Vibration Levels and Their Impact on Physiological Stability. *American Journal of Perinatology* 2019; 36(4): 352–359.

[43] Bouchut JC, Van Lancker E, Chritin V et al. Physical Stressors during Neonatal Transport: Helicopter Compared with Ground Ambulance. *Air Medical Journal* 2011; 30(3): 134–139.

[44] Blaxter L, Yeo M, McNally D et al. Neonatal head and torso vibration exposure during inter-hospital transfer. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine* 2017; 231(2): 99–113.

[45] Shah S, Rothberger A, Caprio M et al. Quantification of impulse experienced by neonates during inter-and intra-hospital transport measured by biophysical accelerometery. *Journal of perinatal medicine* 2008; 36(1): 87–92.

[46] Green J, Langlois R, Chan A et al. Investigating Vibration Levels in a Neonatal Transport System. *Canadian Medical and Biological Engineering Society Proceedings* 2019; 42: 1–4.

[47] Prehn J, McEwen I, Jeffries L et al. Decreasing sound and vibration during ground transport of infants with very low birth weight. *J Perinatol* 2015; 35(2): 110–114.

[48] Gajendragadkar G, Boyd JA, Potter DW et al. Mechanical vibration in neonatal transport: a randomized study of different mattresses. *Journal of Perinatology* 2000; 20(5): 307.

[49] British Standards Institution. BS ISO 2631-1:1997 Mechanical vibration and shock. Evaluation of human exposure to whole-body vibration. Technical report, 1997.

[50] European Parliament and the Council of the European Union. Directive 2002/44/EC on the minimum health and safety requirements regarding the exposure of workers to the risks arising from physical agents (vibration). *Official Journal of the European Communities* 2002; (OJ L177): 13.

[51] Health and Safety Executive (HSE). *Whole-body vibration. The Control of Vibration at Work Regulations 2005*. HSE Books, 2005.

[52] Shah S, Hudak J, Gad A et al. Simulated transport alters surfactant homeostasis and causes dose-dependent changes in respiratory function in neonatal Sprague-Dawley rats. *Journal of Perinatal Medicine* 2010; 38(5): 535–543.

[53] Clifford M, Brooks R, Howe A et al. *An introduction to Mechanical Engineering: Part 1*. Hodder Education, 2009.

[54] Sayers M, Gillespie T and Paterson W. Guidelines for conducting and calibrating road roughness measurements. Technical Report 46, Washington, D.C., 1986.

[55] Shenai JP, Johnson GE and Varney RV. Mechanical Vibration in Neonatal Transport. *Pediatrics* 1981; 68(1): 55–57.

[56] British Standards Institution. BS ISO 2041: 2018 Mechanical vibration, shock and condition monitoring — Vocabulary. Technical report, 2018.

[57] Bandak FA, Ling G, Bandak A et al. *Chapter 6 — Injury biomechanics, neuropathology, and simplified physics of explosive blast and impact mild traumatic brain injury*, volume 127. 1 ed. Elsevier B.V., 2015.

[58] Mendez-Figueroa H, Dahlke JD, Vrees RA et al. Trauma in pregnancy: An updated systematic review. *American Journal of Obstetrics and Gynecology* 2013; 209(1): 1–10.

[59] British Standards Institution. BS ISO 226:2003 Acoustics — Normal equal-loudness-level contours. Technical report, 2003.

[60] British Standards Institution. BS EN 61672-1:2013 Electroacoustics - Sound level meters - Specifications. Technical report, 2013.

[61] British Standards Institution. BS EN ISO 9612:2009 Acoustics — Determination of occupational noise exposure — Engineering method. Technical report, 2009.

[62] American Academy of Pediatrics. Noise: A hazard for the fetus and newborn. Committee on Environmental Health. *Pediatrics* 1997; 100(4): 724–727.

[63] Information on Levels of Environmental Noise Requisite to Protect Public Health and Welfare With an Adequate Margin of Safety. Technical report, Environmental Protection Agency, Office of Noise Abatement and Control, Washington, D.C., USA, 1974.

[64] Knutson AJ. *Acceptable Noise Levels for Neonates in the Neonatal Intensive Care Unit*. PhD Thesis, Washington University School of Medicine, 2013.

[65] Altuncu E, Akman I, Kulekci S et al. Noise levels in neonatal intensive care unit and use of sound absorbing panel in the isolette. *International Journal of Pediatric Otorhinolaryngology* 2009; 73(7): 951–953.

[66] Kramarić K, Šapina M, Milas V et al. The effect of ambient noise in the NICU on cerebral oxygenation in preterm neonates on high flow oxygen therapy. *Signa Vitae* 2017; 13: 52–56.

[67] Williams AL, Sanderson M, Lai D et al. Intensive Care Noise and Mean Arterial Blood Pressure in Extremely Low-Birth-Weight Neonates. *American journal of perinatology* 2009; 26(05): 323–329.

[68] Kuhn P, Zores C, Pebayle T et al. Infants born very preterm react to variations of the acoustic environment in their incubator from a minimum signal-to-noise ratio threshold of 5 to 10 dBA. *Pediatric Research* 2012; 71(4-1): 386–392.

[69] Trapanotto M, Benini F, Farina M et al. Behavioural and physiological reactivity to noise in the newborn. *Journal of Paediatrics and Child Health* 2004; 40(5-6): 275–281.

[70] Kuhn P, Zores C, Langlet C et al. Moderate acoustic changes can disrupt the sleep of very preterm infants in their incubators. *Acta Paediatrica* 2013; 102: 949–954.

[71] Wachman EM and Lahav A. The effects of noise on preterm infants in the NICU. *Archives of disease in childhood - Fetal and neonatal edition* 2011; 96(4): F305–F309.

[72] British Standards Institution. BS EN 13976-2:2018 Rescue systems - Transportation of incubators - System Requirements. Technical report, British Standards Institution, 2018.

[73] Buckland L, Austin N, Jackson A et al. Excessive exposure of sick neonates to sound during transport. *Archives of Disease in Childhood - Fetal and Neonatal Edition* 2003; 88(6): F513—-F516.

[74] Hohlagschwandtner M, Husslein P, Klebermass K et al. Perinatal mortality and morbidity. *Archives of Gynecology and Obstetrics* 2001; 265(3): 113–118.

[75] Chen S, Wang D, Zuo A et al. Vehicle Ride Comfort Analysis and Optimization Using Design of Experiment. In *2010 Second International Conference on Intelligent Human-Machine Systems and Cybernetics*, volume 1. pp. 14–18.

[76] TheJournal.ie. How does wheel size affect my car and driving¿. URL `https://www.thejournal.ie/how-car-wheel-size-affects-driving-4047484-Jun2018/` (Accessed: 2021/06/03).

[77] Bailey-Van Kuren M and Shukla A. System design for isolation of a neonatal transport unit using passive and semi-active control strategies. *Journal of Sound and Vibration* 2005; 286(1-2): 382–394.

[78] Bailey-Van Kuren M and Shukla A. Isolation device for shock reduction in a neonatal transport apparatus, 2007.

[79] Arnold CA, Rogers DD, Kerry Guarino et al. Transport Incubator System, 2014.

[80] Berglund B, Lindvall T and Schwela D. Guidelines for Community Noise. Technical report, World Health Organization, Geneva, Switzerland, 1999.

[81] Natus Newborn Care. MiniMuffs Datasheet, 2017.

[82] Duran R, Çiftdemir NA, Özbek ÜV et al. The effects of noise reduction by earmuffs on the physiologic and behavioral responses in very low birth weight preterm infants. *International Journal of Pediatric Otorhinolaryngology* 2012; 76: 1490–1493.

[83] Khalesi N, Khosravi N, Ranjbar A et al. The effectiveness of earmuffs on the physiologic and behavioral stability in preterm infants. *International Journal of Pediatric Otorhinolaryngology* 2017; 98: 43–47.

[84] Abou Turk C, Williams AL and Lasky RE. A randomized clinical trial evaluating silicone earplugs for very low birth weight newborns in intensive care. *Journal of Perinatology* 2009; 29(5): 358–363.

[85] Ludington-Hoe SM and Abouelfettoh A. Light Reduction Capabilities of Home-made and Commercial Incubator Covers in NICU. *ISRN Nursing* 2013; 2013(Table 1): 1–9.

[86] Hellström-Westas L, Inghammar M, Isaksson K et al. Short-term effects of incubator covers on quiet sleep in stable premature infants. *Acta Paediatrica* 2001; 90(9): 1004–1008.

[87] Macnab AJ, Schweers D, Kendall MD et al. Improved transport incubator temperature control with insulating thermal cover. *Air Medical Journal* 1995; 14(2): 65–68.

[88] Kellam B, Waller J and Bhatia J. Results of an Acoustic Incubator Cover among Preterm Infants: Incubators and Modes of Respiratory Therapy. *Clinical Medicine Review* 2018; : 1–12.

[89] OpenStreetMap. URL `https://www.openstreetmap.org/about` (Accessed: 2020/12/22).

[90] CEPA and TRL. Measuring pavement condition. Technical report, Office of Rail and Road, 2018.

[91] Nottinghamshire County Council. Highway Inspection & Risk Manual. Technical Report July, Nottinghamshire County Council, 2018.

[92] TRL Ltd. SCANNER surveys for local roads. Technical report, UK Roads Board, 2011.

[93] Gallagher KA, Wright A, King PC et al. Smaller, quicker, cheaper automated carriageway surveys. Technical report, TRL, 2009.

[94] Eriksson J, Girod L, Hull B et al. The pothole patrol: using a mobile sensor network for road surface monitoring. In *Proceedings of the 6th international conference on Mobile systems, applications, and services*. ACM, pp. 29–39.

[95] González A, O'Brien E, Li YY et al. The use of vehicle acceleration measurements to estimate road roughness. *Vehicle System Dynamics* 2008; 46(6): 483–499.

[96] Li Z, Kolmanovsky IV, Atkins EM et al. Road Disturbance Estimation and Cloud-Aided Comfort-Based Route Planning. *IEEE Transactions on Cybernetics* 2016; PP(99): 1–13.

[97] British Standards Institution. BS ISO 8608:2016 - Mechanical vibration — Road surface profiles — Reporting of measured data. Technical report, 2016.

[98] Boyle M. Mobile internet statistics - 2020 | Finder UK, 2020. URL `https://www.finder.com/uk/mobile-internet-statistics` (Accessed: 2021/01/20).

[99] Position sensors | Android Developers. URL `https://developer.android.com/guide/topics/sensors/sensors_position` (Accessed: 2021/06/01).

[100] SensorManager | Android Developers. URL `https://developer.android.com/reference/android/hardware/SensorManager.html` (Accessed: 2019/05/20).

[101] Carrera F, Guerin S and Thorp JB. By the People, for the People: the Crowdsourcing of "STREETBUMP", An Automatic Pothole Mapping App. In Ellul C, Zlatanova S, Rumor M et al. (eds.) *29th Urban Data Management Symposium*. London: International Society for Photogrammetry and Remote Sensing, pp. 19–23.

[102] Delpriori S, Freschi V, Lattanzi E et al. Efficient algorithms for accuracy improvement in mobile crowdsensing vehicular applications. In *UBICOMM 2015: The Ninth International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies*. Nice, France: IARIA, pp. 145–150.

[103] Sensors Overview | Android Developers. URL `https://developer.android.com/guide/topics/sensors/sensors_overview.html` (Accessed: 2019/07/24).

[104] Feng M, Fukuda Y, Mizuta M et al. Citizen sensors for SHM: Use of accelerometer data from smartphones. *Sensors (Switzerland)* 2015; 15(2): 2980–2998.

[105] Feldbusch A, Sadegh-Azar H and Agne P. Vibration analysis using mobile devices (smartphones or tablets). In *Procedia Engineering*, volume 199. Elsevier Ltd, pp. 2790–2795.

[106] Astarita V, Vaiana R, Iuele T et al. Automated Sensing System for Monitoring of Road Surface Quality by Mobile Devices. *Procedia - Social and Behavioral Sciences* 2014; 111: 242–251.

[107] Mukherjee A and Majhi S. Characterisation of road bumps using smartphones. *European Transport Research Review* 2016; 8(2): 13.

[108] Mohamed A, Fouad MMM, Elhariri E et al. RoadMonitor: An Intelligent Road Surface Condition Monitoring System. In *Intelligent Systems'2014. Advances in Intelligent Systems and Computing*, volume 323. Springer International Publishing, 2015. pp. 377–387.

[109] Sebestyen G, Muresan D and Hangan A. Road quality evaluation with mobile devices. In *Proceedings of the 2015 16th International Carpathian Control Conference (ICCC)*. IEEE, pp. 458–464.

[110] Badurowicz M and Montusiewicz J. Identifying Road Artefacts with Mobile Devices. In Dregvaite G and Damasevicius R (eds.) *Information and Software Technologies: 21st International Conference, ICIST 2015*. Druskininkai, Lithuania: Springer International Publishing, pp. 503–514.

[111] Forslöf L and Jones H. Roadroid: Continuous Road Condition Monitoring with Smart Phones. *Journal of Civil Engineering and Architecture* 2015; 9: 485–496.

[112] Seraj F, Meratnia N and Havinga PJ. RoVi: Continuous transport infrastructure monitoring framework for preventive maintenance. In *2017 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, pp. 217–226.

[113] Alessandroni G, Carini A, Lattanzi E et al. A Study on the Influence of Speed on Road Roughness Sensing: The SmartRoadSense Case. *Sensors* 2017; 17(2): 305.

[114] Kumar R, Mukherjee A and Singh VP. Community Sensor Network for Monitoring Road Roughness Using Smartphones. *Journal of Computing in Civil Engineering* 2017; 31(3): 04016059.

[115] Cantisani G and Loprencipe G. Road roughness and whole body vibration: Evaluation tools and comfort limits. *Journal of Transportation Engineering* 2010; 136(9): 818–826.

[116] Kardous CA and Shaw PB. Evaluation of smartphone sound measurement applications. *The Journal of the Acoustical Society of America* 2014; 135(4): EL186–EL192.

[117] Murphy E and King EA. Smartphone-based noise mapping: Integrating sound level meter app data into the strategic noise mapping process. *Science of the Total Environment* 2016; 562: 852–859.

[118] Murphy E and King EA. Testing the accuracy of smartphones and sound level meter applications for measuring environmental noise. *Applied Acoustics* 2016; 106: 16–22.

[119] Ventura R, Mallet V, Issarny V et al. Evaluation and calibration of mobile phones for noise monitoring application. *The Journal of the Acoustical Society of America* 2017; 142(5): 3084–3093.

[120] Serpanos YC, Renne B, Schoepflin JR et al. The accuracy of smartphone sound level meter applications with and without calibration. *American Journal of Speech-Language Pathology* 2018; 27(4): 1319–1328.

[121] Celestina M, Hrovat J and Kardous CA. Smartphone-based sound level measurement apps: Evaluation of compliance with international sound level meter standards. *Applied Acoustics* 2018; 139: 119–128.

[122] Kumar R, Mukherjee A and Singh VP. Traffic noise mapping of Indian roads through smartphone user community participation. *Environmental Monitoring and Assessment* 2017; 189(1): 1–14.

[123] British Standards Institution. BS EN 1789:2007+A2:2014 Medical vehicles and their equipment. Road ambulances, 2007.

[124] Shannon CE. Communication in the Presence of Noise. *Proceedings of the IRE* 1949; 37(1): 10–21.

[125] What is personal data? | ICO. URL `https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/key-definitions/what-is-personal-data` (Accessed: 2021/06/03).

[126] Dignited. Why are most Android phones cheaper than iPhones. URL `https://www.dignited.com/36659/why-are-most-android-phones-cheaper-than-iphones/` (Accessed: 2021/06/03).

[127] Alkan H and Celebi H. The Implementation of Positioning System with Trilateration of Haversine Distance. In *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC*.

[128] Activity | Android Developers. URL `https://developer.android.com/reference/android/app/Activity` (Accessed: 2020/10/30).

[129] Services overview | Android Developers. URL https://developer.android.com/guide/components/services (Accessed: 2020/10/26).

[130] IntentService | Android Developers. URL https://developer.android.com/reference/android/app/IntentService (Accessed: 2020/10/27).

[131] PowerManager.WakeLock | Android Developers. URL https://developer.android.com/reference/android/os/PowerManager.WakeLock?hl=en (Accessed: 2020/10/27).

[132] Welch PD. The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics* 1967; 15(2): 70–73.

[133] MediaRecorder | Android Developers. URL https://developer.android.com/reference/android/media/MediaRecorder (Accessed: 2020/10/26).

[134] soundrecorder/VUMeter.java | GitHub. URL https://github.com/aosp-mirror/platform_packages_apps_soundrecorder/blob/master/src/com/android/soundrecorder/VUMeter.java (Accessed: 2020/10/26).

[135] BlockingQueue | Android Developers. URL https://developer.android.com/reference/java/util/concurrent/BlockingQueue (Accessed: 2020/10/30).

[136] GNU Gzip. URL https://www.gnu.org/software/gzip/manual/gzip.html (Accessed: 2021/06/01).

[137] OkHttp. URL https://square.github.io/okhttp/ (Accessed: 2020/10/26).

[138] JobScheduler | Android Developers. URL https://developer.android.com/reference/android/app/job/JobScheduler (Accessed: 2020/10/30).

[139] British Standards Institution. BS EN ISO 8041:2017: Human response to vibration — Measuring instrumentation, 2017.

[140] Instrument Control Toolbox Documentation - MathWorks United Kingdom. URL https://uk.mathworks.com/help/instrument/ (Accessed: 2020/11/05).

[141] Data Acquisition Toolbox Documentation - MathWorks United Kingdom. URL https://uk.mathworks.com/help/daq/ (Accessed: 2020/11/05).

[142] ANSI. American National Standard: Acoustical Terminology. ANSI/ASA S1.1-2013, 2013.

[143] British Standards Institution. BS EN 61043:1994 - Electroacoustics — Instruments for the measurement of sound intensity — Measurement with pairs of pressure sensing microphones, 1993.

[144] CenTre Neonatal Transport. Home. URL http://www.centreneonataltransport.nhs.uk/ (Accessed: 2020/10/22).

[145] Fraden J. *Handbook of modern sensors: physics, designs, and applications*. 5th ed. ed. Cham: Cham : Springer, 2016.

[146] Magnet Expert. Technical advice. URL https://www.magnetexpert.com/technical-advice-i685 (Accessed: 2020/09/09).

[147] First4Magnets. Questions about magnets answered. URL https://www.first4magnets.com/tech-centre-i61/frequently-asked-questions-i69 (Accessed: 2020/09/09).

[148] The Highway Code. Technical report, Department for Transport, 2021.

[149] Save key-value data | Android Developers. URL https://developer.android.com/training/data-storage/shared-preferences (Accessed: 2020/09/11).

[150] InfluxDB Open Source Time Series Database | InfluxDB | InfluxData. URL https://docs.influxdata.com/influxdb/v1.7/ (Accessed: 2020/09/28).

[151] InfluxDB glossary | InfluxDB v1.7 Documentation. URL https://docs.influxdata.com/influxdb/v1.7/concepts/glossary/ (Accessed: 2020/09/28).

[152] InfluxDB schema design and data layout | InfluxDB v1.7 Documentation. URL https://docs.influxdata.com/influxdb/v1.7/concepts/schema_and_data_layout/ (Accessed: 2020/09/28).

[153] The Open Group. General Concepts, 2018. URL https://pubs.opengroup.org/onlinepubs/9699919799/xrat/V4_xbd_chap04.html#tag_21_04_16 (Accessed: 2021/02/03).

[154] Recents Screen | Android Developers. URL https://developer.android.com/guide/components/activities/recents (Accessed: 2020/09/16).

[155] Romano E, Kaufmann M, Clark JR et al. Quantification of vibration forces experienced by the newborn during ambulance transport. In *2012 Critical Care Transport Medicine Conference*, volume 31. Nashville, TN, pp. 167–173.

[156] CenTre Neonatal Transport. Our Hospitals. URL http://www.centreneonataltransport.nhs.uk/about-us/our-hospitals/ (Accessed: 2020/10/06).

[157] Gustavo Niemeyer. Geohash - geohash.org. URL http://geohash.org/ (Accessed: 2020/10/22).

[158] About | what3words. URL https://what3words.com/about-us/ (Accessed: 2020/10/22).

[159] Doug Rinckes and Philipp Bunge. Open Location Code: An Open Source Standard for Addresses, Independent of Building Numbers And Street Names. URL https://github.com/google/open-location-code/blob/master/docs/olc_definition.adoc (Accessed: 2019/06/20).

[160] Gallier J. *Discrete Mathematics*. New York, NY: Springer, 2011.

[161] Built-in Types — Python 3.9.1 documentation, 2021. URL https://docs.python.org/3/library/stdtypes.html#set (Accessed: 2021/02/04).

[162] singleton, n.2 : Oxford English Dictionary. URL https://www.oed.com/view/Entry/180156?rskey=iTzKfT&result=2#eid (Accessed: 2021/06/02).

[163] NetworkX — NetworkX documentation. URL https://networkx.org/.

[164] homo-, comb. form : Oxford English Dictionary. URL https://www.oed.com/view/Entry/87990?redirectedFrom=homoscedastic#eid1500813 (Accessed: 2021/06/03).

[165] Miller GA and Chapman JP. Misunderstanding analysis of covariance. *Journal of Abnormal Psychology* 2001; 110(1): 40–48.