HUMAN RESPONSE TO ENVIRONMENTAL NOISE AND VIBRATION FROM FREIGHT AND PASSENGER RAILWAY TRAFFIC

Calum SHARP

Acoustics Research Centre

School of Computing, Science and Engineering

College of Science and Technology

University of Salford, Salford, UK

Submitted in Partial Fulfilment of the Requirements of the

Degree of Doctor of Philosophy, October 2014

Contents

List of figures

List of tables

Nomenclature

%A	Percentage annoyed (%)
%HA	Percentage highly annoyed (%)
%LA	Percentage little annoyed (%)
$\bar{a}$	Mean acceleration (m s^{-2})
$\bar{p}$	Mean pressure (Pa)
β	Regression coefficient
β_0	Regression intercept coefficient
β_{tj}	Threshold coefficient of the jth category in an ordinal regression model
$\boldsymbol{\beta}$	Vector of β regression coefficients
χ^2	Chi-square distribution
δ_{ij}	Pairwise dissimilarity measured between stimuli i and j
$\mathbf{E_b}$	Covariance matrix of regression parameters
$\mathbf{E}$	Multidimensional Euclidean configuration
$\mathbf{V}$	Vector of vibration exposures
$\mathbf{X}$	Vector of predictor variables in a regression model
$\mathbf{Y}$	Vector of response data

μ	Mean
Φ	Cumulative normal distribution function
ϕ	Standard normal probability density function
ψ	Psychological stimulus magnitude
ρ_{ij}	Correlation coefficient between i and j
σ	Standard deviation
σ^2	Diagonals of the covariance matrix
σ_1	Stress-1 for a multidimensional scaling configuration
σ_r	Raw stress for a multidimensional scaling configuration
σ_v	Standard deviation of the maximum 1 second average weighted velocity for all passbys in a measurement period (mm s^{-1})
σ_{SE}	Standard error of the grouped regression model
τ_j	Cut-point of the jth category in a grouped regression model
$\tilde{A}_{ij}$	Paired annoyance score for stimuli i and j estimated from averaged single figure annoyance scores
$\triangle L$	Perceived temporal masking depth
ε	Random error component
φ	Physical stimulus magnitude
A	Self reported annoyance utilised in the development of exposure-response relationships
a	Vibration acceleration (m s^{-2})
A^*	Latent annoyance variable utilised in the development of exposure-response relationships

A_p	Predicted single figure annoyance
A_T	Thurstone's Case V single figure annoyance
A_{av}	Average single figure annoyance
$a_{b,24hr}$	W_b weighted acceleration time history composed of all passbys in a 24 hour period (m s^{-2})
A_{ij}	Measured paired annoyance score for stimuli i and j
$a_{k,24hr}$	W_k weighted acceleration time history composed of all passbys in a 24 hour period (m s^{-2})
$a_{w,95}$	Statistical 95 percentile weighted acceleration (mm s^{-2})
C	Annoyance category cut-point
C_r	Crest factor
C_{95}	Upper and lower 95% confidence intervals
C_{ij}	Count matrix
c_{mf}	Magnitude Fourier coefficient of the nth spectral bin
d_{ij}^n	Pairwise distance between stimuli i and j in the private space of subject n in a multidimensional scaling configuration
D_m	Position of stimulus on the mth dimension of a multidimensional scaling configuration
d_{ij}	Pairwise distance between stimuli i and j in a multidimensional scaling configuration
e_{ij}	Squared error of representation between distances in a multidimensional space for stimuli i and j
f	Transformation function
$f(n)$	Central frequency of the nth spectral bin (Hz)

F_1	Harmonic mean of precision and recall
f_{max}	Dominant frequency of the power spectral density of a noise or vibration signal (Hz)
f_{mod}	Modulation frequency (Hz)
$g(z)$	Weighting function that increases with the critical band rate, z, used in the calculation of sharpness
i	Represents various indices throughout the thesis
j	Represents various indices throughout the thesis
k	Represents various indices and constants throughout the thesis
K_b	Frequency weighting curve for vibration velocity as recommended by DIN 4150-2:1999
K_t	Kurtosis
$KB^2_{FTM,j}$	Average of the maximum 0.125 s running exponential rms velocity for each 30 second period of an event (mm s^{-1})
KB_{Fmax}	0.125 s running exponential rms K_b weighted velocity value (mm s^{-1})
L	Likelihood function
$L(z)$	Maximal level of the tonal components as a function of the critical-band rate z (sone)
L_{10}	Sound pressure level exceeded for 10% of the evaluation period (dB)
L_{A10}	A-weighted sound pressure level exceeded for 10% of the evaluation period (dBA)
L_{Aeq}	A-weighted equivalent sound pressure level, (dBA)
L_{Amax}	Peak A-weighted sound pressure level of a noise signal (dBA)
L_{dn}	Day-night average sound pressure level (dB)

L_{eq}	Equivalent sound pressure level (dB)
L_{max}	Peak sound pressure level of a noise signal (dB)
L_{pA}	Time-varying A-weighted sound pressure level (dBA)
LL	Log-likelihood function
LL_{con}	Log likelihood of a constrained regression model
LL_{full}	Log likelihood of a full regression model
LL_{int}	Log likelihood of an intercept only regression model
LL_{uncon}	Log likelihood of an unconstrained regression model
m	Represents various indices and constants throughout the thesis
MCC	Matthews Correlation Coefficient
n	Represents various indices and constants throughout the thesis
N'	Specific loudness (sone)
P	Probability
$p(t)$	Pressure time history of a noise signal (Pa)
$p_A(t)$	A-weighted Pressure time history of a noise signal (Pa)
P_{95}	Probability confidence limits for exposure-response relationships
p_{ij}	Probability of stimulus i falling in the jth category in an ordinal regression model
p_{max}	Peak pressure (Pa)
R	Roughness (asper)
R^2_{pseudo}	McFadden's pseudo R^2
$R_{S/L}$	Ratio utilised in the STA/LTA algorithm

rms	Root mean square of a noise pressure signal (Pa) or vibration acceleration signal (m s^{-2})
rms_A	A-weighted root mean square pressure (Pa)
rms_k	W_k weighted root mean square acceleration (m s^{-2})
$rms_{16\text{Hz}}$	rms acceleration energy contained within the 16 Hz octave band of a vibration signal (m s^{-2})
$rms_{32\text{Hz}}$	rms acceleration energy contained within the 32 Hz octave band of a vibration signal (m s^{-2})
$rms_{5\text{Hz}}$	Normalised proportional rms energy contained within the 5 Hz 1/3rd octave band
$rms_{k,24hr}$	W_k weighted 24 hour root mean square acceleration (m s^{-2})
S	Sharpness (acum)
s	Scale parameter
SEL	Sound exposure level (dB)
SEL_A	A-weighted sound exposure level (dBA)
T	Total time of a noise or vibration signal (s)
t	Continuous time of a noise or vibration signal (s)
T_e	Normalised event signal duration defined by the duration of the signal that exceeds the top 1/3rd of its dynamic range
T_L	Width of the long term window in the STA/LTA triggering algorithm (s)
T_S	Width of the short term window in the STA/LTA triggering algorithm (s)
T_{10dB}	Duration defined by the 10 dB downpoints of a signal (s)

T_{3dB}	Duration defined by the 3 dB downpoints of a signal (s)
$T_{4/5,v}$	Duration for which a vibration signal exceeds the top 4/5ths of its dynamic range (s)
$T_{e,j}$	Exposure period of the jth event (s)
$TETC$	Total energy of the tonal components within critical bands from 12 to 24 barks (sone)
V	Vibration exposure magnitude
$V_{dir,max}$	Maximum W_k weighted fast exponentially filtered rms velocity (mm s^{-1})
V_{eq}	Equivalent continuous vibration level (dB)
v_{rms}	Root mean square vibration velocity (mm s^{-1})
$v_{w,95}$	Statistical 95 percentile weighted velocity (mm s^{-1})
VAP	Variance of time-varying A-weighted pressure normalised by root mean square A-weighted pressure
var	Finite sample variance operator
VEL	Vibration exposure level (dB)
W	Wald test statistic
w_m^n	Weighting attributed by subject n on the mth dimension in a multidimensional scaling configuration
W_b	Frequency weighting curve for vertical vibration acceleration as recommended by BS 6472-1:2008
W_d	Frequency weighting curve for horizontal vibration acceleration as recommended by BS 6472-1:2008
W_k	Frequency weighting curve for vertical vibration acceleration as recommended by BS ISO 2631-1:1997

W_m	Frequency weighting curve for vibration acceleration in any direction as recommended by ISO 2631-2:2003a
x_A	Subscript A represents an A-weighted quantity of parameter x
x_b	Subscript b represents a W_b weighted quantity of parameter x
x_k	Subscript k represents a W_k weighted quantity of parameter x
x_n	Numerical value of feature n in a logistic regression model
x_{jm}	Coordinate of the jth stimulus on the mth dimension in the group space of a multidimensional scaling configuration
y	Response variable utilised in grouped regression model
y_{jm}^n	Coordinate of the jth stimulus on the mth dimension in the private space of subject n in a multidimensional scaling configuration
Y_m	Observed class of the mth observation in a logistic regression model
Z	Multiple linear model utilised in logistic regression
z	Critical-band rate (bark)
CER	Community of European railways
CTL	Community tolerance level
ECG	Electrocardiogram
FN	False negative
FP	False positive
ICBEN	International commission on the biological effects of noise
INDSCAL	Individual difference scaling algorithm
IUPT	International union of public transport
MDS	Multidimensional scaling

MTVV	Maximum transient vibration value
PROXSCAL	Computerised multidimensional scaling program
REM	Rapid eye movement
RSS	Residual sum of squares
SE	Standard error
STA/LTA	Short time average/long time average event identification triggering algorithm
TN	True negative
TP	True positive
TVANE	Train vibration and noise effects
UIC	Union internationale des chemins de fer (international union of railways)
UNIFE	Union des industries ferroviaires Européennes (union of European railway industries)
VDV	Vibration dose value (m $s^{-1.75}$)
VDV_b	W_b weighted vibration dose value (m $s^{-1.75}$)
$VDV_{k,24hr}$	W_b weighted 24 hour vibration dose value (m $s^{-1.75}$)

Acknowledgements

There is a large list of those who I need to thank for the help, support, advice and friendship given to me during my time at Salford. Firstly, I must thank my supervisor, Dr. David Waddington, whose support throughout my MSc and PhD has been immeasurable. It is thanks to him that I pursued the PhD in the first place. Special thanks must also go to the original V-Team: James, Eulalia and Gennaro. Without their hard work on the DEFRA project, there would be no thesis to write! I would also like to thank Dr. Bill Davies and Dr. Mikael Ögren for taking the time to examine my thesis. Their excellent advice has only made the thesis stronger.

And of course I must thank my friends at Salford. There are far too many to mention, but I would be remiss if I didn't list a few names. Thanks to Andy, Toby, Charlie, Rodolfo, Linda, Jenna, Phil, Tim, Paul, Nanu, Josh and James. A special mention must go to G10 where most of the work was carried out and to The Crescent, where most of the play was carried out. Thanks also to my family, who have always supported me, even when I decided to completely change my career path and study acoustics.

Finally, most importantly, thanks to Meg for everything, always.

Declaration

This research makes extensive use of data collected during the field study of Waddington et al. (2014). Though all the analysis of the data present in this thesis is original work by the author, the author did not take part in performing the vibration measurements themselves, or the collection of the questionnaire data. Aspects of this research have previously been published in the following papers:

Sharp, C., Woodcock, J., Sica, G., Peris, E., Waddington, D. C., and Moorhouse, A. T. (2012). Developing exposure-response relationships for annoyance caused by vibration from freight and passenger railway traffic. In *Proceedings of Euronoise*, Czech Republic.

Sharp, C., Woodcock, J., Sica, G., Peris, E., Moorhouse, A. T., and Waddington, D. C. (2012). Developing exposure-response relationships for annoyance caused by vibration from freight and passenger railway traffic. In *Proceedings of the 47th United Kingdom Conference on Human Responses to Vibration*, Southampton, UK.

Sharp, C., Woodcock, J., Peris, E., Sica, G., Moorhouse, A. T., and Waddington, D. C. (2013). Analysis of railway vibration signals using supervised machine learning for the development of exposure-response relationships. *Proceedings of Meetings on Acoustics*, 19:040107.

Sharp, C., Woodcock, J., Sica, G., Peris, E., Moorhouse, A. T., and Waddington, D. C. (2014). Exposure-response relationships for annoyance due to freight and passenger railway vibration exposure in residential environments. *The Journal of the Acoustical Society of America*, 135(1):205-212.

Abstract

There is currently a push to increase the proportion of freight traffic that is transported by rail, which is argued to be a safer, more sustainable and more climate friendly means of freight transportation when compared with road or air transportation. This will result in increased noise and vibration from freight railway traffic, the potential impacts of which are not well known. The aim of this research, therefore, is to further the understanding of the human response to freight railway noise and vibration.

Data for this research comes from a field study comprising interviews with respondents and measurements of their vibration exposure. A logistic regression model was created and optimised, and is able to accurately classify 96% of these measured railway vibration signals as freight or passenger signals based on two signal properties that quantify the duration and low frequency content of each signal. Exposure-response relationships are then determined using ordinal probit modelling with fixed thresholds and cumulative ordinal logit models. The results indicate that people are able to distinguish between freight and passenger railway vibration, and that the annoyance response due to freight railway vibration is significantly higher than that due to passenger railway vibration, even for equal levels of exposure.

To further investigate this disparity in response, a laboratory study was performed in which subjects were exposed to combined noise and vibration from freight and passenger railway traffic. Through the technique of multidimensional scaling, the subjective responses to these stimuli were analysed to investigate the specific attributes of the stimuli that may lead to the difference in human response. The results of this study suggest that the perception of combined railway noise and vibration takes into account not only the exposure magnitude of the noise and vibration stimuli, but also signal properties such as duration, spectral distribution and envelope modulation. These parameters, and in particular the duration parameter, appear to account for the difference in the human response to freight and passenger railway vibration.

Chapter 1

Introduction

1.1 General introduction

Railway transport is generally argued to be a safer, more sustainable and more climate friendly mode of freight transportation when compared with road and air transport (Wiebe et al., 2011). This, in combination with a need to decrease road congestion by addressing the imbalance between transportation modes, has influenced European policy to direct movement of freight transport from the roads and onto the rails (Wiebe et al., 2011). Specifically, the International Union of Railways (UIC), the Community of European Railways (CER), the International Union of Public Transport (IUPT) and the Union of European Railway Industries (UNIFE) have all agreed, within the White Paper for European Transport, to attempt an increase in the market share of goods traffic on rail from 8% in 2001 to 15% in 2020 (Commission of the European Communities, 2001). Indeed, an increase in freight railway transportation has already been observed in Germany, amongst other places, where the amount of freight transported by rail has increased by 25% between 2002 and 2010 (Elmenhorst et al., 2012). This increase in freight railway transport will lead to an increase in the resulting noise and vibration and the potential effects that this may have on residents living in the vicinity of railway lines needs to be understood. Research has shown that increasing levels of noise and vibration can lead to negative human response in the form of annoyance and sleep disturbance (Klæboe et al., 2003b; Miedema and Vos, 1998). In addition, it has been shown that annoyance due to vibration and annoyance reactions due to noise are higher during evenings and night-time, when freight railway traffic tends to be more prevalent (Öhrström, 1997; Peris et al., 2012). Due to the planned increase in freight traffic, and the potential adverse effect that the resulting increase in noise and vibration may have on local residents, the human response to railway noise and vibration, and specifically that created by freight railway traffic, should be better understood. The human response to environmental noise is well researched, but the response to environmental vibration, and particularly the response to freight railway vibration, has been less documented. With greater knowledge of the human response to freight railway vibration, measures to ensure acceptable combined levels of noise and vibration can be developed, minimising the degree of annoyance and sleep disturbance experienced by residents living in the vicinity of freight railway lines.

The primary aim of this research therefore, is to further the understanding of the human response to noise and vibration from freight railway traffic. In order to achieve this, an existing database of railway vibration measurements, and resulting human responses, collected during the field study of Waddington et al. (2014) are analysed. Firstly, a logistic regression machine learning algorithm is developed in order to classify unknown railway vibration events within this measurement database as freight or passenger vibration events. With these vibration events classified, it is then possible to determine exposure data for these two sources of environmental vibration. These exposures can then be paired with the annoyance responses collected during the field study in order to investigate the human response to vibration from freight and passenger railway traffic.

The analysis of the exposure-response relationships for annoyance due to exposure to freight and passenger railway traffic reveals that the human response to these two sources is significantly different, with the annoyance response to freight vibration being higher than that due to passenger vibration, even for equal levels of vibration exposure magnitude. This suggests that the human response to these two sources of vibration takes into account more than just the vibration exposure magnitude, quantified using existing vibration exposure descriptors. The remainder of the research, therefore, focuses on an investigation into the perception of freight and passenger railway noise and vibration. This is achieved through the means of a subjective test with subjects exposed to combined railway noise and vibration. The perceptual results of the subjective tests are analysed using multidimensional scaling, in the hopes of determining other aspects of the vibration signals, and accompanying noise, that may lead to the increased annoyance response due to freight railway vibration.

1.2 Scope of thesis

Following a general introduction in Chapter 1, Chapter 2 contains a summary of the current knowledge and literature concerning the human response to railway noise and vibration. This chapter highlights the relative dearth of research that has focussed on the human response to railway vibration, and freight railway vibration in particular, providing the motivation for the current research.

Chapter 3 contains a description of the methods used to classify unknown railway vibration signals within the measurement database from the field study of Waddington et al. (2014). A logistic regression model is created and the methods used to optimise and test the classification accuracy of the model are described.

Chapter 4 details how exposure-response relationships are derived for annoyance due to exposure to freight and passenger railway vibration, utilising the signals classified using the logistic regression model as described in Chapter 3. The difference in the annoyance response is investigated via several methods, including creating grouped models with a source type dummy variable, and models derived for freight and passenger exposures and responses separately. The regression models used to develop these relationships are described in detail and the resulting exposure-response relationships are presented and discussed.

In order to further investigate the differences in human response to freight and passenger railway traffic, a subjective test on the human response to railway noise and vibration is performed. The development and methodology of this subjective test is detailed in Chapter 5. Annoyance responses due to the noise and vibration stimuli, a mixture of freight and passenger noise and vibration events, are computed from paired comparison annoyance data collected during the subjective test.

In Chapter 6, the results of the subjective test are analysed using the technique of multidimensional scaling. This allows the perception of railway noise and vibration to be analysed as a multidimensional phenomenon, helping to further the understanding of the human response to railway noise and vibration, and potentially revealing aspects of the noise and vibration signals that lead to an increase in the annoyance response due to freight railway vibration.

Finally, overall conclusions are drawn in Chapter 7, and potential avenues of further work are suggested.

1.3 Novel aspects of work

The following outcomes, which are presented throughout the thesis, are considered by the author to be novel contributions to the field of human response to noise and vibration:

- A logistic regression classification model has been developed and optimised, and is capable of classifying unknown railway vibration signals as freight or passenger railway signals with an accuracy of 96%.
- Grouped regression analyses with dummy variables for source type have shown that there is a difference in the human response to freight and passenger railway vibration, and that it is valid to derive separate exposure-response relationships for these two sources of environmental vibration.
- Exposure-response relationships for annoyance due to exposure to freight and passenger railway vibration in residential environments have been derived, demonstrating that freight railway vibration is significantly more annoying than passenger railway vibration, even for equal levels of vibration exposure.
- The perception of railway noise and vibration has been shown to be a multidimensional phenomenon which can be quantified using a small number of objective parameters of the noise and vibration signals.
- The results of the multidimensional analysis have been used to derive a new model for the prediction of annoyance due to railway noise and vibration, as a function of objective parameters of the noise and vibration signals.

Chapter 2

Literature review

2.1 Introduction

This section contains a summary of the current knowledge and literature concerning the human response to environmental noise and vibration, as well as more specific research into the human response to noise and vibration from freight and passenger railway traffic. Firstly, laboratory studies on the perception of whole-body vibration will be presented, with information on the physiological mechanisms of vibration perception, as well as the perceptual effects of vibration magnitude, frequency and duration. Following this is a summary of the mechanisms of groundborne vibration as well as recommendations on the measurement and assessment of groundborne vibration in residential environments. A summary of relevant field studies on the community response to environmental noise and vibration is then presented, as well as a summary of laboratory studies on the effects of combined railway noise and vibration. Finally, a discussion of a small number of studies investigating the differences in the human response to freight and passenger railway traffic is presented.

2.2 The perception of whole-body vibration

Although much work has been performed on the human response to noise in the ambient environment, the relative amount of research on the human response to environmental vibration has fallen somewhat behind. This is perhaps due to the absence of perceivable vibration in the everyday lives of most people. However, there are many avenues through which one can be exposed to whole-body vibration, ranging from experiencing vibration in a moving vehicle to vibration experienced from operating industrial equipment to vibration exposure in the home from external or internal sources. Vibration exposure can be extremely complicated and may exist in combination with airborne noise from the vibration source or other sources, vibration induced low frequency noise and rattling noise from loose objects in the immediate surroundings. The potential effects of vibration exposure depend on many factors including the characteristics of the motion, the characteristics of the exposed person, the activities of the exposed person and other aspects of the environment. These factors can come together to influence how the vibration is perceived and whether it may cause annoyance, discomfort, activity disturbance, impaired health or motion sickness (Griffin, 1996). This research focuses primarily on vibration that is experience as whole-body vibration, defined by Griffin (1996) as vibration which occurs “when the body is supported on a surface which is vibrating”.

In order to further the understanding of the human response to vibration, and to predict the effects that these vibrations may have on humans, it is necessary to form an understanding of the measurable objective characteristics of vibrations and how these characteristics are perceived by humans. This section, therefore, provides a summary of psychophysical studies on the human response to vibration.

2.2.1 Physiological mechanisms of vibration perception

Physiologically, vibration perception is primarily perceived through three mechanisms: cutaneous, kinaesthetic and visceral (Mansfield, 2005). Cutaneous perception is determined by mechanoreceptors in the skin which respond to vibratory excitations as well as sensations of touch and pressure. These mechanoreceptors respond to vibration by producing a pulse train of action potentials, the density of which are linearly related to the amplitude of the excitation. The four major mechanoreceptors, and the frequency range to which they are sensitive, are presented in Table 2.1. These mechanoreceptors are often collectively referred to as “low-threshold”, since even weak mechanical stimulation of the skin causes them to produce action potentials (Purves et al., 2001).

Kinaesthetic perception of whole-body vibration is perceived via forces and movements within the body; information of the position and forces in joints, muscles and tendons are picked up by proprioceptors and sent to the brain. Visceral perception occurs via receptors in the abdomen (Mansfield, 2005). Vibration can also

Mechanoreceptor	Frequency Range (Hz)
Merkel Disk Receptors	5 - 15
Meissner's Corpuscles	20 - 50
Pacinian Corpuscles	60 - 400
Ruffian Endings	100 - 500

Table 2.1: Summary of the four main mechanoreceptors in the skin and the vibratory frequency range to which they are sensitive (Kandel et al., 2000)

be perceived via the auditory system above 20 Hz, via airborne pathways and bone conduction.

Several studies have demonstrated that the whole-body response to vibration is non-linear. Studies by Fairley and Griffin (1990), Matsumoto and Griffin (2002) and Nawayseh and Griffin (2003) all show a lowering of the resonant frequency of the human body in the seated position with increasing magnitude of vertical vibratory excitation. Non-linearities have also been demonstrated for the horizontal vibratory excitations (Nawayseh and Griffin, 2005) and for the human body in the standing position (Matsumoto and Griffin, 1998).

2.2.2 Perceptual effects of vibration magnitude

An increase in perceived intensity and discomfort is usually associated with an increase of vibration magnitude. Many studies have attempted to quantify this increase in perceived intensity or discomfort using Steven's power law (Stevens, 1975), which suggests that the psychological magnitude, ψ of a stimulus is related to its physical magnitude, φ by the following expression:

$$\psi = k\varphi^n \tag{2.1}$$

where k is a constant and the growth in sensation is determined by the exponent n. A study by Miwa (1968a) was one of the earliest to try and estimate the growth constant n. Using the corrected ratio method, subjects were presented with pairs of vibration stimuli and the magnitude of the second stimulus was adjusted until the subject perceived it to have half the magnitude of the first reference stimulus. This procedure was repeated for sinusoidal vibration at three frequencies (5 Hz, 20 Hz and 60 Hz) and at six magnitudes of reference stimuli in the vertical and horizontal directions. Their results suggested the growth constant n did not differ significantly with frequency, however they reported a reduction in the constant with an increase in vibration magnitude. For magnitudes below 1 m s^{-2} they reported a growth constant of 0.60 and for magnitudes above 1 m s^{-2} they reported a growth constant of 0.46.

In the years following this work, several studies attempted to investigate the relationship between stimulus magnitude and perceived intensity and discomfort. Leatherwood and Dempsey (1976) provide a summary of these studies, the results of which show much variability and contradiction. It is suggested that these inconsistent results may be due to poor experimental design, unrealistic laboratory experiments, use of inadequate rating scales or adjectives, small subject samples and lack of information regarding the nature of the relationship between subjective ratings and vibration stimuli. Two exceptions are noted, however, in the work of Shoenberger and Harris (1971) and Jones and Saunders (1974) who presented data to support the hypothesis that magnitude estimates of subjective intensity obey a power law with respect to the physical magnitude of the vibration stimulus. These two studies suggested growth constant exponents ranging between 0.86 and 1.04 depending on the frequency of the vibration stimulus. The result that the growth constant was found to be close to unity led Leatherwood and Dempsey (1976) to question whether the power law is really the best way to represent the data, or whether other relationships such as linear, logarithmic or exponential might be more appropriate. A growth constant of unity corresponds to a linear function (i.e. perceived intensity or discomfort doubles if the vibration magnitude doubles).

To investigate the functional form of the psychophysical relationship, Leatherwood and Dempsey (1976) performed their own study with subjective tests of perceived intensity and discomfort. The method applied was that of magnitude estimation, whereby a subject is exposed to a reference vibration stimulus with a pre-assigned numerical value. The subject is then exposed to a comparison stimulus and asked to evaluate this stimulus relative to the reference stimulus, in terms of intensity in one test and discomfort in another, by assigning it a numerical value. Vertical sinusoidal exposures were assessed at frequencies ranging between 2 and 28 Hz and at reference magnitudes ranging between 0.49 and 4.41 m s^{-2}. Four potential relationships to describe the resulting relationship were investigated: power, exponential, logarithmic and linear. No significant differences were found between the correlation coefficients for these relationships, leading Leatherwood and Dempsey (1976) to suggest that a linear psychophysical relationship be adopted (Equation 2.2) since there appeared to be no scientific basis for adopting the more complicated power law relationship (Equation 2.1).

$$\psi = k + n\varphi \tag{2.2}$$

Subsequent studies have also shown growth constants close to unity, for certain vibratory stimuli, further suggesting that the psychophysical relationship can be described by a linear model (Hiramatsu and Griffin, 1984; Howarth and Griffin, 1988b). Several studies, however, have demonstrated that the growth constant is not equal for all frequencies and vibratory excitation directions. Shoenberger and Harris (1971) found the growth constant to be greater than unity (1.04) for the vibration stimuli of 5 Hz, but found it to be less than unity (0.86 to 0.98) for vibration stimuli of 3.5, 7, 9, 11, 15 and 20 Hz. Howarth and Griffin (1988b) found the growth constant to also change for the direction of excitation, with the growth constant being higher for excitation in the vertical direction than in the horizontal direction for frequencies of 8 Hz and lower, whilst the opposite is true for frequencies of 11.3 Hz and higher. In more recent studies it has been shown that the rate of growth of discomfort is greater at lower frequencies than it is at higher frequencies (Morioka and Griffin, 2006, 2010; Wyllie and Griffin, 2007)

2.2.3 Perceptual effects of vibration frequency

Since the early days of research into the human response to vibration, perception of vibration has been shown to be frequency dependent (Miwa, 1967a). Investigations into the effects of vibration frequency have typically sought to quantify equal levels of comfort, or the perception threshold of vibration, as a function of frequency. Griffin (1996) describes the effects of vibration frequency as follows: at low frequencies (below 1 or 2 Hz depending on the vibration direction and body orientation) the forces acting on the body are approximately proportional to the input acceleration and this movement is transmitted throughout the entire body. At slightly higher frequencies body resonances begin to amplify this motion and overall discomfort is influenced by sensations in different parts of the body. At even higher frequencies, the body provides increasing attenuation of vibration, eventually reducing the perceived location of discomfort to that in close proximity to the vibration input position (Whitham and Griffin, 1978). These general observations imply that at low frequencies, with the body acting as a virtually rigid system, discomfort tends to be proportional to vibration acceleration (Griffin, 1996).

A wide variety of curves showing the effects of vibration frequency on (dis)comfort can be found in the literature. A summary of these equivalent comfort contours is presented by Griffin (1996) and reproduced here in Figure 2.1. Though there is some degree of variability in the curves, most contours fall, showing an increase to sensitivity with acceleration, with an increase of frequency from about 2 or 3 Hz to about 5 or 6 Hz, where the greatest sensitivity is observed in most studies. At higher frequencies the contours generally rise, though the rate of rise and the frequency at which the rise begins varies between studies.

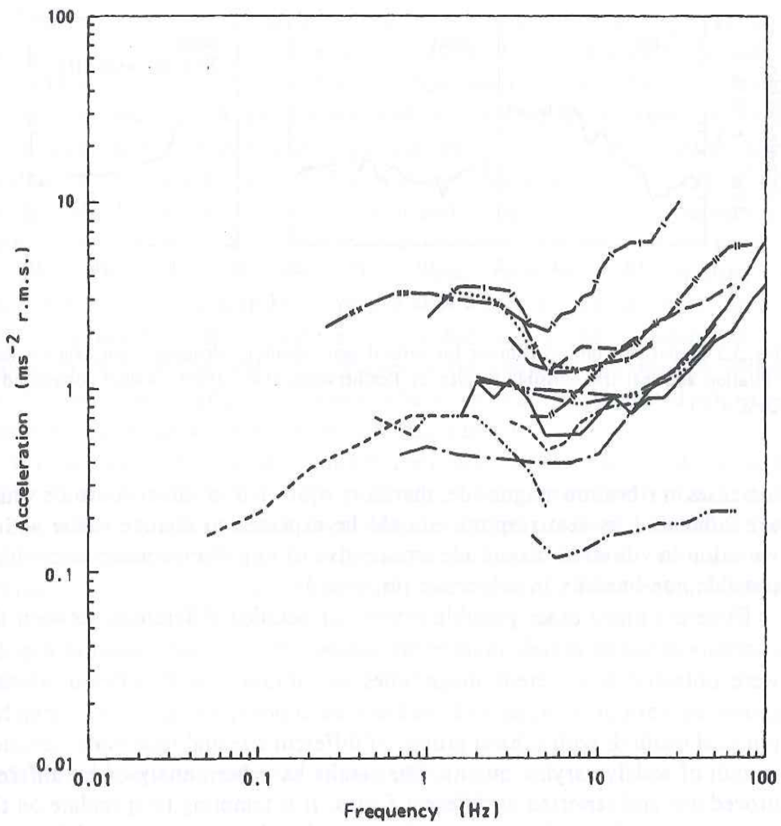

FIGURE 2.1: Equivalent comfort contours for vertical vibration of seated persons (Source: Griffin, 1996). $-\cdot-$ (Miwa, 1967a), $-\times-$ (Shoenberger and Harris, 1971), - - - (Yonekawa and Miwa, 1972), $-|-$ (Dupuis et al., 1972a,b,c), $-\|-$ (Jones and Saunders, 1972), $-\times\times-$ (Shoenberger, 1975), $--\cdot--$ (Griffin, 1976), $-$ (Griffin et al., 1982), $-\cdots-$ (Parsons et al., 1982), $---$ (Oborne and Boarer, 1982), $\cdots$ (Donati et al., 1983), - - ·· - - (Corbridge and Griffin, 1986), $-\cdot\cdot-$ (Howarth and Griffin, 1988b)

The perception threshold of vibration as a function of frequency has also been investigated in several studies. A summary of these perception thresholds is presented by Griffin (1996) and reproduced here in Figure 2.2. Again, though there is some variability in the perception threshold curves across the studies, likely due to variations in posture, characteristics and imperfections of the vibration stimuli and inter-subject variability, some general trends can be observed. For perception thresholds, similarly to equal comfort contours, the greatest sensitivity for vertical vibratory excitation is generally observed in the 5 to 6 Hz region. For horizontal vibration, the greatest sensitivity is generally observed in a lower frequency region of 1 to 2 Hz. More recent studies have suggested that the threshold of perception in the vertical direction is relatively flat with vibration acceleration above 10 Hz (Bellmann, 2002; Morioka and Griffin, 2006).

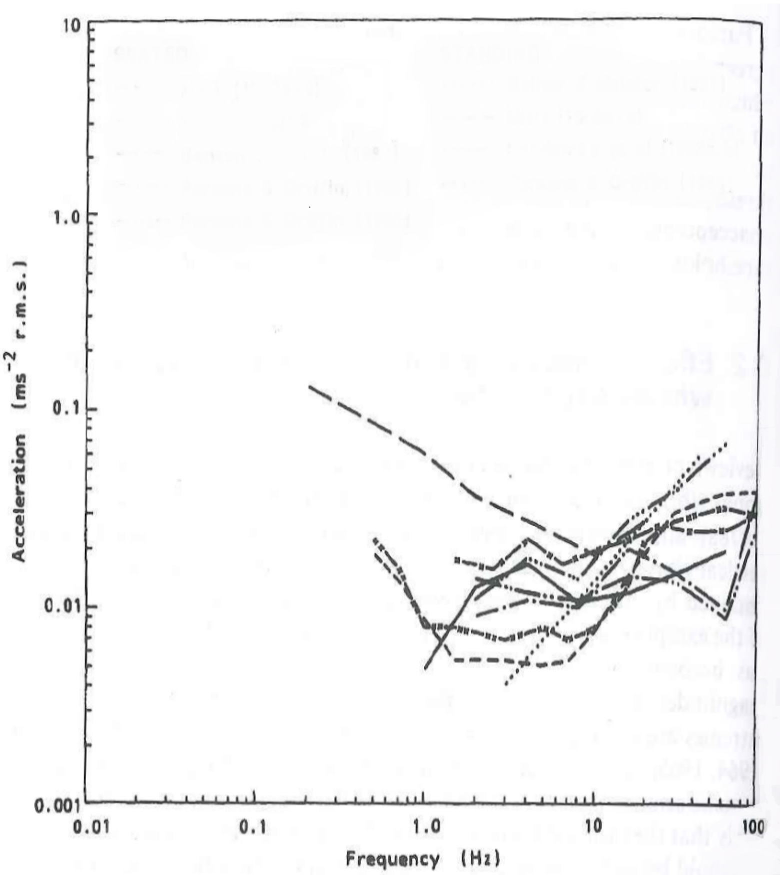

FIGURE 2.2: Perception thresholds for vertical axis whole-body vibration of seated and standing persons (Source: Griffin, 1996). SEATED: $- \times \times -$ (Miwa, 1967a,b,c), $- \times - \times$ (McKay, 1971), $--$ (Benson and Dilnot, 1981), $- \cdot - \cdot -$ (Parsons and Griffin, 1988), $- \cdots -$ (Parsons and Griffin, 1988). STANDING: $\cdots$ (Reiher and Meister, 1931), $- - -$ (Miwa, 1967a,b,c), $-$ (Landström et al., 1983a,b), $- \cdot \cdot -$ (Parsons and Griffin, 1988)

2.2.4 Perceptual effects of vibration duration

It is perhaps intuitive that a vibration stimulus of longer duration will elicit a higher level of discomfort. However, relatively few studies have attempted to quantify this perceptual effect of vibration duration. In some of the earliest experimental studies, subjects were exposed to a few seconds of a vibration stimulus and asked to predict the time after which they would expect the vibration to become uncomfortable (Magid et al., 1960; Miwa et al., 1973; Simic, 1974). In other early studies where limits are suggested for vibration environments, lower limits are suggested for longer durations (Eldick Thieme, 1961; Magid et al., 1960; Notess, 1963).

In an experiment performed by Miwa (1968b), subjects were asked to judge the relative discomfort produced by short periods of sinusoidal vibration and pulsed sinusoidal vibrations with various pulse shapes and durations up to 6 s. He reported that discomfort increases with duration up to a certain limit and that for increasing duration beyond approximately 2 s for vibration in the range 2 to 60 Hz, and beyond approximately 0.8 s for vibration in the range 60 to 200 Hz, there may be no further increase in sensation, suggesting there may be a "time constant" for the evaluation of response.

However, studies conducted by Griffin and Whitham (1980a,b) failed to find substantial evidence of a time constant. They reported that subjects' judgements indicated increasing discomfort with duration from 0.03 to 32 s. They also found some evidence that the relation between discomfort and duration may be frequency dependent. Relationships obtained between duration and discomfort for 4, 8, 16 and 32 Hz vibrations up to 4 s are shown in Figure 2.3. For all frequencies studied, the magnitude of a fixed duration vibration stimulus which produced equivalent discomfort to the various reference stimuli was found to change at a lower rate than would be expected from a squared relationship between acceleration and time (i.e. a^2t = constant) which is implied by root mean square measures of vibration magnitude averaging. Instead, Griffin and Whitham (1980a,b) proposed the alternative of root mean quad averaging (i.e. a^4t = constant). This averaging method approximately reflects the reduced slopes found during their experiments and captures more accurately the greater effect of high magnitude peaks occurring in irregular vibration stimuli and impulses. Results close to the fourth power relationship were also found by Howarth and Griffin (1988a) in a study investigating the change in annoyance due to variations in the rate at which simulated railway-induced building vibrations occurred. This fourth power relationship was later used as the basis for the development of the vibration dose value metric (see Section 2.3.3).

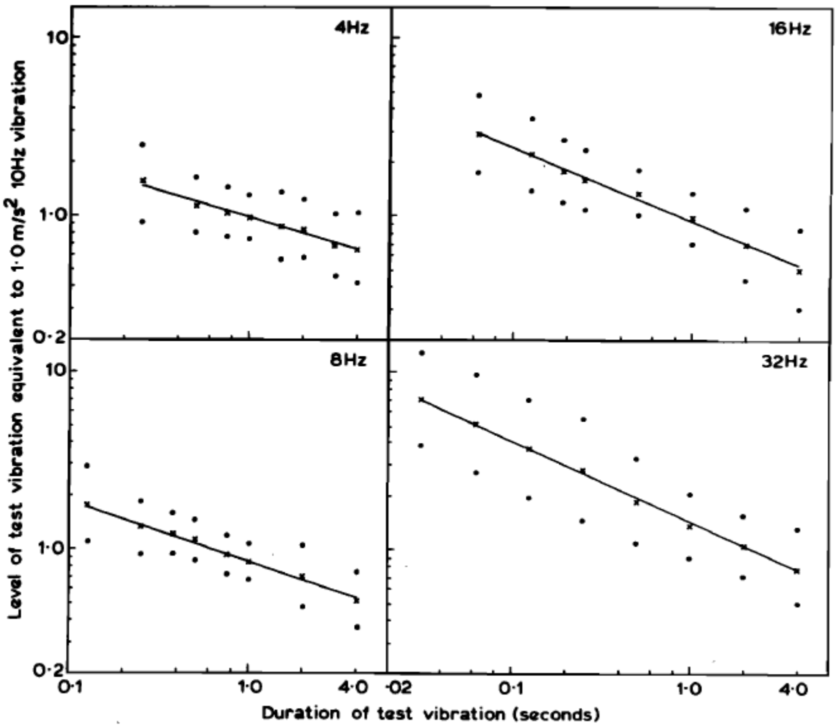

FIGURE 2.3: Effect of vibration duration on vibration discomfort for 4, 8, 16 and 32 Hz (Source: Griffin and Whitham, 1980a). × median values for 20 subjects, • 10th and 90th percentiles.

In a recent study, Huang and Griffin (2014) investigated the effects of stimulus duration on the relative discomfort of noise and vibration. In their study, the subjective equivalence of noise and vibration was investigated with all 49 combinations of 7 levels of noise and 7 levels of whole-body vertical vibration magnitudes for five different durations (2, 4, 8, 16 and 32 s). They found the rate of increasing discomfort with increasing duration to be similar for noise and vibration. As stimuli duration increased from 2 to 32 seconds, they found the influence of vibration on the judgement of noise discomfort to decrease, whereas the influence of noise on the judgement of vibration discomfort was unchanging with duration.

2.2.5 Multidimensional perception of vibration

In a subjective study, Woodcock et al. (2014a) investigated the perception and annoyance caused by railway induced vibration as a multidimensional phenomena, using the technique of multidimensional scaling. In their laboratory test, twenty one subjects were exposed to 14 railway vibration stimuli, measured during the field study performed by Waddington et al. (2014). For more details on the field study, see Section 2.5.1. Through multidimensional scaling analysis, Woodcock et al. (2014a) demonstrated that the perception of railway induced vibration is dependent on up to four perceptual dimensions. These dimensions were found to relate to root mean square acceleration in the 16 Hz and 32 Hz 1/3rd octave bands (rms_{16Hz} and rms_{32Hz} respectively), the duration of the train passby defined by the 10 dB downpoints of the signal (T_{10dB}) and the modulation frequency of the vibration envelope (f_{mod}). They developed a relationship for predicted single figure annoyance (A_p) as a function of these perceptual dimensions using the following relationship:

$$A_p = -0.40 + 4.57rms_{16Hz} + 3.18rms_{32Hz} + 0.02T_{10dB} + 0.02f_{mod} \tag{2.3}$$

This relationship was able to successfully predict single figure annoyance as determined from paired comparison ratings of annoyance in the subjective test. The multidimensional perceptual relationship takes into account properties of the vibration signal, such as spectral content and duration, that have been shown to have an influence on the perception of whole-body vibration, as described in the previous sections. It should be noted, however, that a stepwise regression on the above model resulted in a reduced model containing only the rms_{16Hz} and rms_{32Hz} terms, leading Woodcock et al. (2014a) to suggest that further work is needed to find objective correlates which better describe the third and fourth perceptual dimensions.

2.3 Groundborne vibration: mechanisms, measurement and assessment

Environmental whole-body vibration from railways is experienced as groundborne vibration which is transmitted from the source to the receiver (the body). This section aims to provide an overview of the mechanisms of vibration transmission, as well as the current knowledge and guidance related to the measurement and assessment of groundborne vibration.

2.3.1 Mechanisms of groundborne vibration transmission

Vibration travels through the ground with three types of wave motion:

1. Longitudinal (often referred to as compressive waves or P waves)
2. Transverse (often referred to as shear waves or S waves)
3. Rayleigh

The motion of the longitudinal wave is in the direction of travel of the wave, whereas the motion of the transverse wave is normal to the propagation direction. Rayleigh waves are more complex, with a component in the direction of propagation and a component normal to the surface, resulting in elliptical particle acceleration in the vertical plane through the direction of propagation. Rayleigh and transverse waves propagate at lower velocities, for example around 200 m s^{-1}, than compressive waves which can travel at around 1000 m s^{-1} (Griffin, 1996; Thompson, 2009). The wave magnitudes decrease with increasing propagation distance, though the Rayleigh waves have the least reduction with distance and tend to be the dominant motion at further distances from the vibration source. Since the work in this thesis is primarily concerned with propagation distances of up to 100 m, i.e. the distances from railway lines to houses located close to the railway, Rayleigh waves are likely to dominate in the vibration transmission.

The transmission of groundborne vibration is extremely complex, and generally outside the scope of this thesis, varying significantly according to numerous variables, such as the nature and location of the source, ground type, the horizontal and vertical ground continuity, the water table and the presence of any frozen material in the ground (Griffin, 1996). This makes the modelling of groundborne vibration difficult, resulting in inherently large uncertainties. Considering these uncertainties, measurement of groundborne vibration is generally preferred to modelling, especially if the propagation path also includes transmission through buildings, which provides a further means of uncertainty.

2.3.2 Measurement of groundborne vibration

When considering human response, the objective of vibration measurements is to quantify vibration exposure as close as possible to the point of entry to the human body. However, this is often not practical and guidance suggests that measurements should be made at a location that would expect to give rise to the highest levels of vibration to which the persons would be exposed (BS 6472-1:2008). There is general agreement among various guidance documents and standards that this can be achieved by measuring vibration at the mid-span of the floor of the room which is the most exposed to the source of vibration (Association of Noise Consultants, 2012). The standard BS EN ISO 8041:2005 provides guidance on instrumentation requirements for the measurement of building vibration with respect to human response, suggesting many recommended design features, including the minimum requirements that the instrumentation shall record the following:

1. Time-averaged weighted vibration acceleration value of the measurement duration

2. Band-limited time-averaged vibration acceleration value over the measurement duration

3. Measurement duration

Two standards which are referred to extensively in this thesis, BS 6472-1:2008 and BS ISO 2631-1:1997, require the measurement of vibration acceleration time histories in three orthogonal directions in the frequency range 0.5 to 80 Hz. The orthogonal directions are defined differently between the two standards, with BS 6472-1:2008 defining the directions using a geo-centric coordinate system (i.e. the principal axes are earth centred) and BS ISO 2631-1:1997 defining the directions using a basi-centric coordinate system (i.e. the principle axes are defined with respect to the position of the human body).

When measurements are made on the floor, the nature of the floor covering can create uncertainties in the measurement, meaning that the mounting of the measurement instrumentation is an important consideration. Ideally, accelerometers should be coupled to the vibrating medium such that they faithfully record the motion relative to the focus of the investigation (Association of Noise Consultants, 2012). This can be achieved when floor covering such as carpet is either removed or penetrated by spikes on the base of the transducer mount. However, this is often not practical. Alternative solutions up to at least 50 Hz have been demonstrated either by mounting accelerometers on thin plates approximately 150 mm square or by supporting accelerometers on a three-legged mount with a mass of approximately 0.5 kg (Griffin, 1996).

2.3.3 Vibration exposure metrics

Several national and international standards provide guidance on metrics which can be used to quantify the human exposure to vibration. Guidance typically takes the form of single figure exposure metrics which are functions of frequency dependent weightings. The standard BS 6472-1:2008 recommends two frequency weighting curves, W_b for the vibration acceleration in the vertical direction and W_d for the horizontal direction. This weighting curve, as well as others presented in this section, intend to reflect the sensitivity of the human body to vibration as a function of frequency (see Section 2.2.3). The W_b weighting curve shows maximum sensitivity to vertical vibration in the frequency range 4 to 12.5 Hz and the W_d weighting curve shows maximum sensitivity to horizontal vibration in the range 1 to 2 Hz.

Other frequency weighting curves include the W_k weighting curve for vertical vibration acceleration as recommended by BS ISO 2631-1:1997, which also recommends the use of the W_d weighting curve for horizontal vibration. The W_b and the W_k weighting curves, which are utilised extensively throughout the thesis, are presented in Figure 2.4. The two weighting curves differ slightly, though the difference is less than the inter-subject variability recorded in the laboratory studies from which the weighting curves were developed (see Section 2.2.3). A further weighting curve, W_m, which applies for vibration acceleration in any direction, is recommended in ISO 2631-1:2003. The W_m weighting curve is derived from a combination of the W_k and W_d curves. The German national standard, DIN 4150-2 (1999), which is used as the basis for guidance in many parts of continental Europe, recommends the use of the K_b weighting curve for vibration velocity, as opposed to acceleration. The K_b weighting curve is similar to the W_b weighting curve when transformed and applied to acceleration (Woodcock, 2013).

The vibration metric that is perhaps most commonly used throughout the United Kingdom to quantify human exposure to vibration is the vibration dose value (VDV) as recommended by BS 6472-1:2008. VDV is defined using either the W_b weighting curve for vertical vibration or the W_d weighting curve for horizontal vibration and can be derived over a 16 hour daytime period (07:00 to 23:00) and an 8 hour night-time period (23:00 to 07:00). VDV is a fourth power integration of vibration acceleration and is defined as follows:

$$\mathrm{VDV} = \sqrt[4]{\int_0^T a(t)^4 dt} \tag{2.4}$$

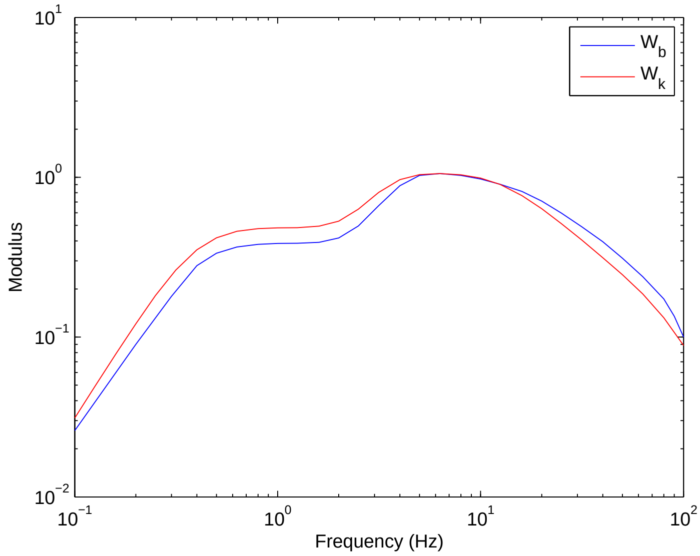

FIGURE 2.4: W_b and W_k frequency weighting curves for vertical vibration acceleration

where $a(t)$ is vibration acceleration varying with time t and total duration T. Due to the fourth power integration, VDV has unconventional units of m s$^{-1.75}$. The fourth power relationship, which accounts more accurately for the greater perceptual effect of high magnitude peaks in a vibration signal, is based in part upon the laboratory studies of Griffin and Whitham (1980a,b) and Howarth and Griffin (1988a) as described in Section 2.2.4.

Another commonly utilised vibration metric, particularly in the United States, is the frequency weighted root mean square (rms) acceleration as recommended by BS ISO 2631-1:1997. The recommended weightings are W_k for vertical acceleration and W_d for horizontal acceleration. The rms acceleration is analogous to rms pressure that is widely used in the measurement of noise signals and is defined as follows:

$$rms = \sqrt{\frac{1}{T}\int_0^T a(t)^2 dt} \tag{2.5}$$

The recommendation of BS ISO 2631-1:1997 is that the rms acceleration should only be used to quantify signals with a relatively low crest factor. The crest factor of an acceleration signal is the ratio of the peak acceleration to the rms acceleration. A higher crest factor indicates the signal may be dominated by peaks. For signals with a crest factor greater than 9, BS ISO 2631-1:1997 recommends the use of VDV or the maximum transient vibration value (MTVV), defined as the maximum value of the slow weighted running rms over the evaluation period.

The Norwegian standard NS 817 (2005) recommends the use of a statistical 95 percentile weighted velocity ($v_{w,95}$) derived from 1 second averages of the velocity signals. This metric is calculated as follows:

$$v_{w,95} = \bar{v}_{w,max} + 1.8\sigma_v \tag{2.6}$$

where $\bar{v}_{w,max}$ is the mean value of the maximum weighted velocity for all train passbys and and σ_v is the standard deviation of the maximum 1 second average weighted velocity for all passbys.

The German national standard DIN 4150-2 (1999), used as the basis for guidance in much of continental Europe, suggests an evaluation procedure based on two vibration exposure descriptors. The first step of the procedure is to evaluate the metric KB_{Fmax} which is a 0.125 s running exponential rms K_b weighted velocity value of the evaluation period. This metric is then compared to context sensitive thresholds, taking into account time of day, the vibration source and location of the building, and if the thresholds are exceeded, a second metric, KB_{FTr} is calculated:

$$KB_{FTr} = \sqrt{\frac{1}{T_r}\sum_j T_{e,j} KB^2_{FTM,j}} \tag{2.7}$$

where T_r is the evaluation period (16 hours for daytime, 8 hours for night-time), $T_{e,j}$ is the exposure period of the jth event and $KB^2_{FTM,j}$ is the average of the maximum 0.125 s running exponential rms velocity for each 30 second period of an event.

There are several other national standard which provide guidance for the evaluation of vibration exposure. A summary of these standards and their key features is provided by the Railway-Induced Vibration Abatement Solutions Collaborative Project (RIVAS, 2011). Many of the guidelines, including the Netherlands standard SBR Richtlijn - Deel B (2006), the United States FTA guidelines (2006), the Swedish standard SS 460 48 61 (1992), the Spanish standard Real Decreto 1367/2007, the Italian standard UNI 9614:1990, the Japanese Vibration Regulation Law and the Austrian standard ÖENORM S 9012:2010 recommend the use of some variation of the maximum running average *rms* velocity or acceleration.

In a guidance document produced for the CargoVibes project, Woodcock et al. (2014b) state that there is insufficient evidence to recommend any one vibration exposure metric over another. They suggest that guidance should be provided, where possible, in terms of three metrics that can be easily related to other metrics found in national and international standards. These three metrics are $V_{dir,max}$ which is the maximum W_k weighted fast exponentially filtered *rms* velocity over the entire assessment period, and the W_k weighted *rms* acceleration and VDV which are defined above. They also recommend retaining raw acceleration time histories of measurements for future analysis.

2.3.4 Guidance for assessment of vibration exposure

As well as the guidance on suitable metrics for measurement of vibration exposure, some of the standards previously mentioned provide guidance as to the probable annoyance caused by a given vibration exposure. BS 6472-1:2008 provides the "probability of adverse comment" for different ranges of VDV, which are reproduced here in Table 2.2. The standard also states that the levels in Table 2.2 should be multiplied by factors of 2 and 4 for offices and workshops respectively. Note that the levels in the table are halved for the 8 hour night-time period, suggesting a doubling in sensitivity to vibration during the night. However, there is no indication as to how these values were derived and no definition provided for "adverse comment".

Location And Time	Low Probability of Adverse Comment	Adverse Comment Possible	Adverse Comment Probable
Residential Buildings 16 Hour Day	0.2 to 0.4	0.4 to 0.8	0.8 to 1.6
Residential Buildings 8 Hour Night	0.1 to 0.4	0.2 to 0.4	0.4 to 0.8

TABLE 2.2: Vibration dose value ranges (m $s^{-1.75}$) which may result in various probabilities of adverse comment within residential buildings (Source: BS 6472-1:2008)

BS ISO 2631-1:1997 does not provide any specific *rms* ranges which are likely to incite annoyance, but does state that "occupants of residential buildings are likely to complain if the vibration magnitudes are only slightly above the perception threshold".

The Norwegian standard 8176 (2005) defines four classes of comfort for dwellings with respect to vibration exposures expressed using the metric $v_{w,95}$ (Equation 2.6) and a similar metric $a_{w,95}$ which is an equivalent metric using vibration acceleration rather than velocity. The guidance levels for these dwelling classes are based on the results of a socio-vibrational survey which is described in further detail in Section 2.5.1. The guidance levels are reproduced here in Table 2.3. A Class A dwelling is one in which it is expected that no occupants will notice vibration, a Class B dwelling is one in which it is expected that occupants will be disturbed to some extent by vibration, a Class C dwelling is one in which it is expected that 15% of occupants will be disturbed by vibration and a Class D dwelling is one in which it is expected that at least 25% of occupants will be disturbed by vibration.

Vibration Metric	Class A	Class B	Class C	Class D
$v_{w,95}$ (mm s^{-1})	0.10	0.15	0.30	0.60
$a_{w,95}$ (mm s^{-2})	3.6	5.4	11.0	21.0

TABLE 2.3: Guidance classification of dwellings with the upper limits for the statistical maximum value for weighted velocity, $v_{w,95}$, or acceleration, $a_{w,95}$ (Source: Turunen-Rise et al., 2003)

2.4 Community response to environmental noise

The effects of environmental noise on the community has been extensively researched in a number of studies. Some of the key studies of the human response to environmental noise will be discussed in this section, with a particular focus on exposure-response relationships developed for annoyance due to transportation noise.

2.4.1 The concept of quantifiable annoyance

When quantifying the negative effects of an environmental stimulus such as noise or vibration, it is necessary to define a measure by which the negative effects are quantified. However, since the human response is inherently subjective, much care must be taken when choosing and defining the measure that will be used to quantify the response. Response data in field studies into the community response to noise and vibration is generally recorded in terms of "annoyance".

Annoyance is the most commonly reported problem as a result of exposure to transportation noise and it is most often the factor used to determine a community response to noise exposure (Clark and Stansfeld, 2007). The World Health Organisation refers to noise annoyance as a "global phenomenon" and considers it to be an important health effect of noise (World Health Organization, 2000). Guski et al. (1999) states that annoyance is a diverse concept that is associated with disturbance, aggravation, dissatisfaction, concern, bother, displeasure, harassment, irritation, nuisance, vexation, exasperation, discomfort, uneasiness, distress and hate.

As well as annoyance, studies have demonstrated a moderate effect of transportation noise on sleep disturbance, hypertension, cardiovascular disease, catecholamine secretion and an impairment of cognitive performance in schoolchildren (Basner et al., 2014; Clark and Stansfeld, 2007). Although these effects may be relatively modest, they are becoming increasingly more important as more people are being exposed to higher levels of environmental noise. There are several variables that can influence an individual's response to a particular noise source. Source-specific variables such as source type, exposure level and time of day are certainly important, but so are many receiver-specific factors such as the extent of interference experienced, ability to cope, expectation, fear associated with the source, visibility of the source, sensitivity, anger and beliefs about whether the noise could be reduced by responsible parties (Job, 1988; World Health Organization, 2000).

Measurements of noise exposure have been standardised (e.g. ISO/TS 15666:2003 (2003b)) and typically take the form of socio-acoustic questionnaires. Fields et al. (2001) recommend standardised general questions that take the form of asking a respondent how bothered, annoyed or disturbed they are by a particular noise source. The degree to which the respondents are annoyed is typically recorded on a five point semantic scale ("not at all", "slightly", "moderately", "very" or "extremely") and an 11 point numeric scale (0 to 10). An important breakthrough in exposure-response relationships occurred when Schultz (1978) performed a meta-analysis on existing socio-economic noise annoyance surveys and suggested the use of a "highly annoyed" threshold for a community's subjective response. Percentage highly annoyed (%HA) is a percentile-based metric which describes the proportion of respondents who express levels of annoyance within the upper 28% of the annoyance scale. Investigating only the highly annoyed responses ensures that the respondents have indicated a specific and conscious identification and subsequent reaction to a noise source and reduces the impact of non-acoustic variables, resulting in a high correlation between exposure and response.

2.4.2 Exposure-response relationships for the human response to environmental noise

Schultz's (1978) work was an attempt to collate and summarise the world literature on community response to transportation noise at the time. The exposure-response relationship developed by Schultz was an informal fit to 161 data points taken from five aircraft and six rail and road noise surveys. The original Schultz curve used percentage highly annoyed (%HA) as the dependent variable to represent the response, and the day-night average sound pressure level (L_{dn}) as the descriptor used to quantify noise exposure (see Figure 2.5).

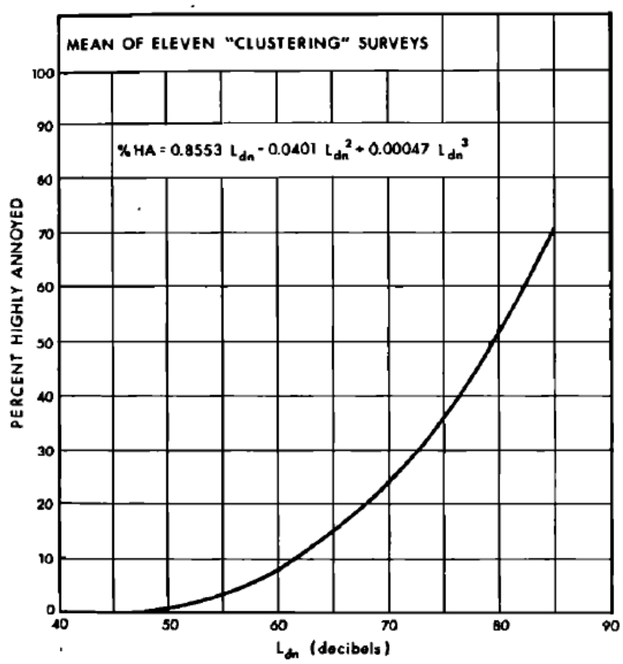

FIGURE 2.5: Exposure-response relationship for percentage highly annoyed persons, derived from 11 noise surveys (Source: Schultz, 1978)

Following the publication of Schultz's 1978 paper, a public discussion began between Schultz and Kryter (Kryter, 1982, 1983; Schultz, 1982). Kryter (1982) cast doubt over the adequacy of using a single curve to describe the exposure-response for road, railway and air traffic noise sources together. He argued that using separate curves for ground (road and railway) transportation and air transportation gives a better representation of the data originally presented by Schultz (1978) and that, by fitting a single curve to all three noise sources, Schultz's exposure-response relationship significantly underestimates the annoyance associated with aircraft noise and significantly overestimates the annoyance associated with ground transportation noise. Fidell et al. (1991) continued Schultz's work by developing a new least squares quadratic fit to the original 161 data points considered by Schultz, combined with an additional 292 data points from air, road and railway traffic noise studies, published after Schultz's original analysis ($\%\mathrm{HA} = 0.036L_{dn}^2 - 3.26L_{dn} + 79.92$). Although the new model was derived using almost triple the amount of data points used, Schultz's original 1978 curve still provided a reasonable fit to the data. However, Miedema and Vos (1998) argue that the additional data of Fidell et al. (1991) appears to support Kryter's argument that, at the same exposure level, aircraft noise is more annoying than ground transportation noise. In recent papers, Fidell, writing with other authors, has produced separate exposure-response relationships for aircraft noise and road and railway noise (Fidell et al., 2011; Schomer et al., 2012).

Miedema and Vos (1998) derived exposure-response relationships for transportation noise using all 21 datasets used by Schultz (1978) and Fidell et al. (1991), for which acceptable L_{dn} and %HA measures could be derived, augmented with an additional 34 datasets. They produced separate exposure-response relationships for aircraft, road and railway traffic, stating that there is a "systematic and substantial difference" between these three noise sources and concluding that different relationships should be used for different transportation sources. They found that the rate of increase of %HA as a function of noise exposure was higher for aircraft noise than for road transport noise, and higher for road transport noise than for railway transport noise (see Figure 2.6). For a given noise exposure, %HA is highest for aircraft noise, followed by road traffic noise and then railway traffic noise. They found a difference between sources for all studies combined and for studies in which respondents evaluated two separate sources, giving confidence that the different exposure-response relationships for each source are a result of differences in the source type, rather than differences in study methodology.

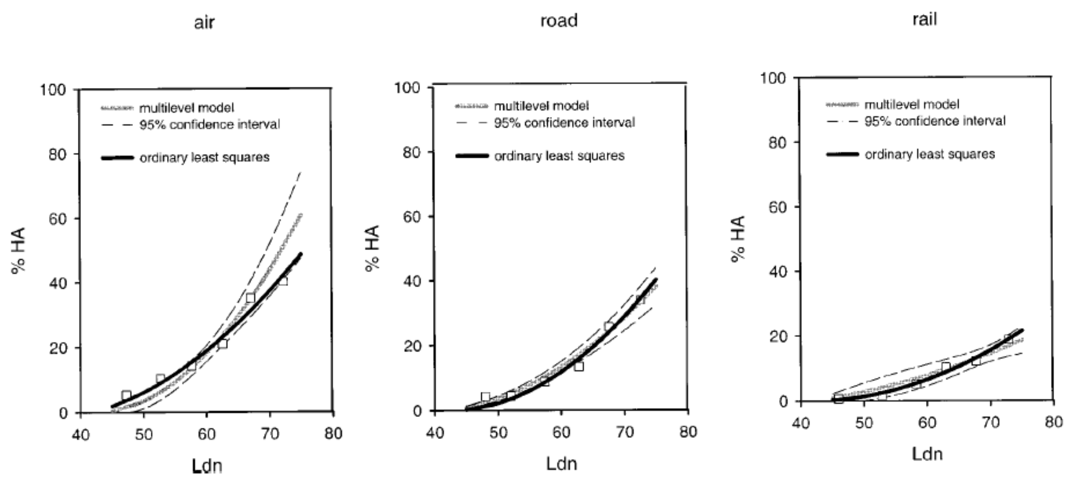

FIGURE 2.6: Percentage highly annoyed persons (%HA) as a function of L_{dn} for air, road and rail traffic (Source: Miedema and Vos, 1998)

Miedema and Oudshoorn (2001) used the same datasets as Miedema and Vos (1998), but derived a new improved model using grouped regression. Using this model, they were able to model the entire annoyance distribution, providing better estimates of the confidence intervals and allowing different annoyance measures to be calculated. In particular, they introduced specific cut-offs for "percentage little annoyed" (cut-off at 28 on a 0 to 100 scale) and "percentage annoyed" (cut-off at 50 on a 0 to 100 scale), as well as the "percentage highly annoyed" (cut-off at 72 on a 0 to 100 scale) as previously defined by Schultz (1978). These exposure-response relationships are reproduced here in Figure 2.7.

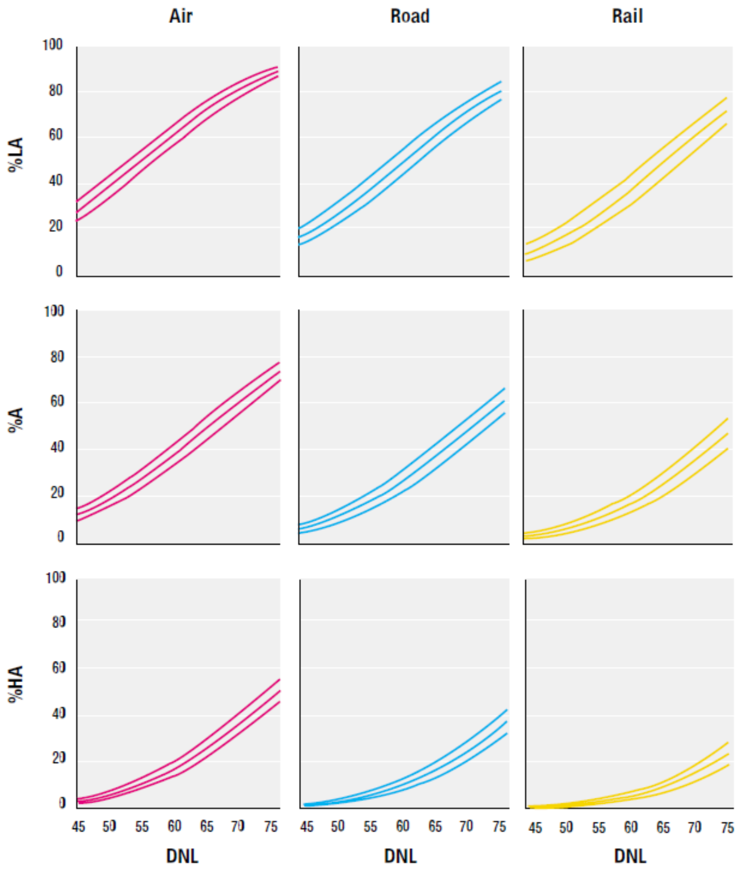

FIGURE 2.7: Percentage little annoyed persons (%LA), percentage annoyed persons (%A) and percentage highly annoyed persons (%HA) as a function of L_{dn} for air, road and rail traffic (Source: Miedema and Oudshoorn, 2001)

2.4.3 The community tolerance level

Fidell et al. (2011) developed a different approach to the traditional curve-fitting methods described above, based on the findings of 43 studies of annoyance due to aircraft noise. They found that the rate of change of annoyance with L_{dn} closely resembled the rate of change of loudness with sound level. Their models agreed with findings of curve fitting exercises, such as those derived by Miedema and Vos (1998), despite the differences in analytical methods and the disparate data sets. As with previous curve fitting methods, they discovered that, although annoyance prevalence rates within communities increase consistently in proportion with duration-adjusted loudness, annoyance rates across communities still vary greatly. To account for this variance, they introduced a model in terms of a single parameter expressed in L_{dn} units, the "community tolerance level" (CTL), which accounts for the aggregate influence of non L_{dn} related factors (i.e. non acoustic factors) on annoyance prevalence rates in different communities. The CTL is defined by the midpoint of the effective loudness function, i.e. the L_{dn} exposure which corresponds to 50% highly annoyed persons. Figure 2.8 shows an example of this process applied to six aircraft noise surveys, showing a range of community tolerance levels of approximately 30 dB.

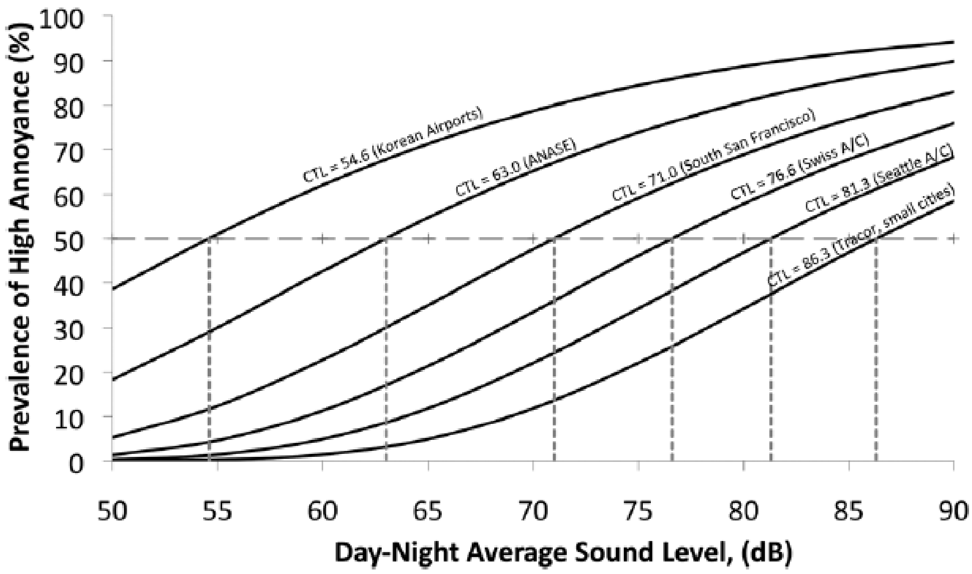

Figure 2.8: Community tolerance levels computed from the findings of six aircraft noise surveys (Source: Fidell et al., 2011)

Fidell et al. (2011) conclude that CTL values appear to be little influenced by airport size, but may be influenced by airport type (i.e. regional/major airports). In addition, they appear to be unrelated to climate variables, but economic factors, such as median housing values and household incomes may be influential. Schomer et al. (2012) applied the same model to predict the prevalence of noise-induced annoyance of road and railway traffic noise. They found that the model applies well for road traffic noise, with the loudness function and a CTL of 78.3 dB closely resembling the results presented by Miedema and Vos (1998). They also found that the prevalence of annoyance due to railway noise is more accurately predicted when sites are separated into those with and without high levels of vibration and/or rattle, suggesting that the annoyance of railway noise is strongly related to the presence or absence of vibration and/or vibration induced rattle. According to their results, noise from conventional trains without appreciable vibration are approximately 9 to 10 dB less annoying than road traffic noise, and noise from conventional trains accompanied by appreciable vibration are approximately 2 to 3 dB more annoying than road traffic noise.

2.5 Community response to environmental vibration

Compared to the community response to environmental noise, the community response to environmental vibration has been less studied. However, there have been several key field studies which have investigated the human response to environmental vibration. These studies, with a particular focus on studies looking at railway vibration, are summarised in this section.

2.5.1 Field studies investigating the human response to environmental vibration

A significant portion of literature focusing on the human response to environmental vibration concerns research on annoyance due to railway vibration. Fields (1979) summarised some early work on the human response to vibration in residential environments, based on the results of the 1975 British railway noise study. The results of this study indicated that railway vibration is an important source of annoyance. Although it was deemed infeasible to measure physical vibration data within the study, an effort was made to extract as much information as possible from the survey respondents' reports of experiencing vibration. To this end, four different measures of vibration were collected: perception of vibration, extent bothered by it, judgement as to whether it is a "problem", and belief about whether any damage is caused by vibration. In the absence of vibration measurement data, these vibration measures were represented as a function of distance from the railway. From this relationship it was determined that all four vibration measures decrease with distance from the railway, presumably as vibration exposure also decreases. They also found some influence of other factors on vibration reactions, such as the speed of trains and the visibility of the railway. In terms of the type of trains which especially cause vibration, freight trains were mentioned more often by respondents than passenger trains.

In another early field study on this topic, Woodruff and Griffin (1987) collected responses to railway vibration from 459 residents in Scotland, and measured vibration exposures in 52 of the dwellings in which respondents claimed to perceive vibration. Analysis of their questionnaire data revealed that 34.8% of respondents living within 100 m of a railway line noticed railway-induced building vibration. After correlating several different measures of vibration exposure with reported annoyance, it was found that the most appropriate vibration exposure descriptor for predicting annoyance in this study was the number of train passbys occurring within a 24-hour period, where annoyance was found to increase with the number of train passbys. This increase in annoyance, represented by a decrease of the percentage of persons "not at all" annoyed, is reproduced here in Figure 2.9.

The Transport and Road Research Laboratory conducted a field study in which approximately 30 residents at each of 50 sites in the United Kingdom were interviewed about nuisance related to road traffic induced vibrations and airborne noise (Watts, 1984, 1987, 1990). As well as the subjective response data collected by the interviews, measurements of airborne noise were conducted to quantify the noise exposure for each of the respondents. Although questions were asked about vibration, no vibration measurements were conducted. A relationships was derived between these noise measurements, expressed as the noise level exceeded for 10% of the evaluation period, L_{10}, and the percentage of respondents that were bothered by traffic-induced noise and vibration. This relationships is reproduced

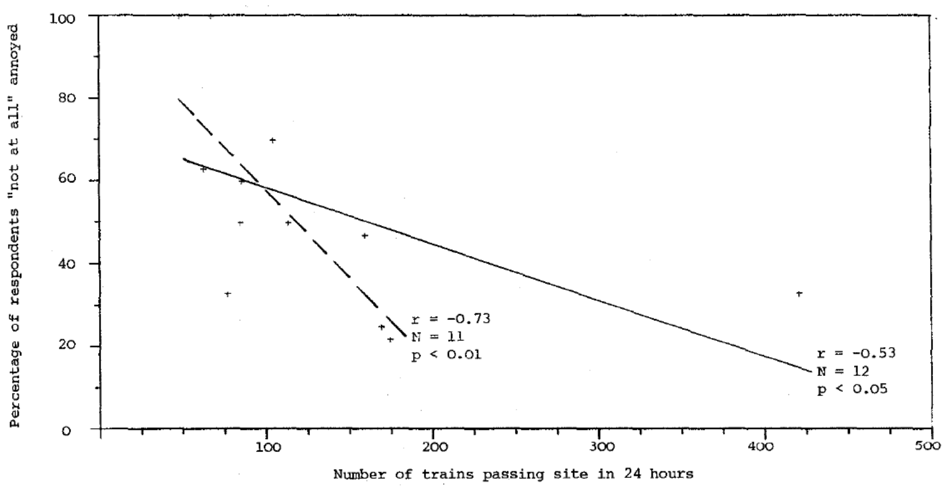

FIGURE 2.9: Relationship between number of trains passing a site in 24 hours and the percentage of respondents who are "not at all" annoyed by vibration (Source: Woodruff and Griffin, 1987)

here in Figure 2.10 and suggests that noise exposure correlates reasonably well with bother caused by traffic-induced vibration.

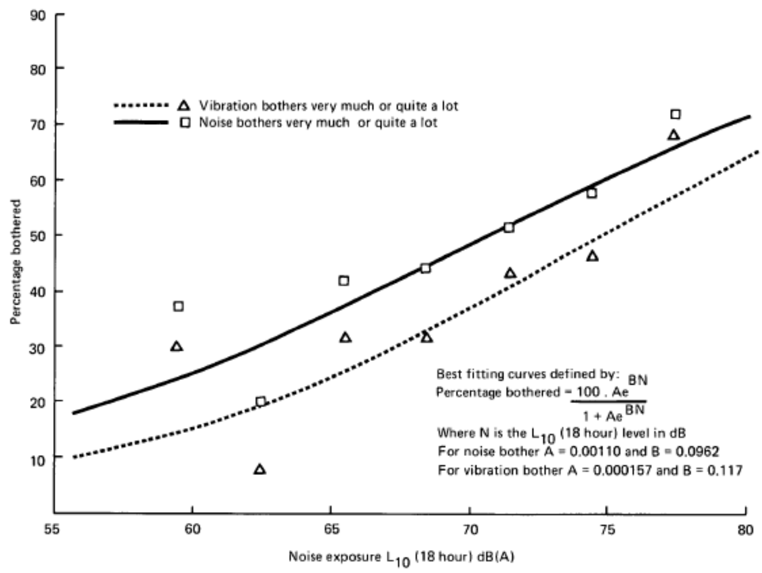

FIGURE 2.10: Percentage of respondents bothered by noise and vibration caused by traffic (Source: Watts, 1990)

In field studies conducted by Öhrström and Skånberg (1996), the effects of exposure to noise and vibration from railway traffic were investigated. The study took place over fifteen sites located near railway lines in Sweden, covering areas with different number of trains in a 24 hour period and areas with strong vibration levels (exceeding 2 mm s^{-1}), low vibration levels (below 2 mm s^{-1}) and no measurable vibration. Effects on annoyance, sleep disturbance, psychosocial well-being and activity disturbance were evaluated via a postal questionnaire with 2833 participants. Noise exposure was determined according to the Nordic calculation model for railway noise, with some control measurements made at different distances from the railway line. Minimal vibration measurements were conducted to identify areas with strong vibration. The results of the study indicated that railway noise is perceived as more annoying in areas in which there is simultaneous exposure to vibration from railway traffic. This is illustrated in Figure 2.11, which shows the percentage of "rather" and "very" annoyed respondents in two areas which have approximately equal number of trains in 24 hours, but different vibration levels. In this figure, noise exposure is quantified by the maximum A-weighted sound pressure level, L_{Amax}. In areas with vibration, the annoyance levels due to railway noise are consistently higher (by up to 35%) than in areas without vibration. In further analysis of the field data, Öhrström (1997) states that in areas with simultaneous exposure to noise and strong vibration, to ensure acceptable environmental quality, action against vibration, or a longer distance between houses and the railway line is needed.

In a German field study, summarised by Knall (1996), the effect of vibration levels, frequency of train passbys and noise levels on humans were investigated. The subjective effects were determined using a standardised questionnaire, with noise and vibration exposures determined by measurements. A total of 1056 questionnaires were completed from 565 households. Railway noise was considered to be more annoying in the studied households, with 76% of respondents stating that railway noise is "more annoying" or "much more annoying" than railway vibration. However, though respondents were less annoyed by vibration, there was a greater feeling that respondents could not avoid or protect themselves from vibration, though they could with noise. They also found that the proportion of events

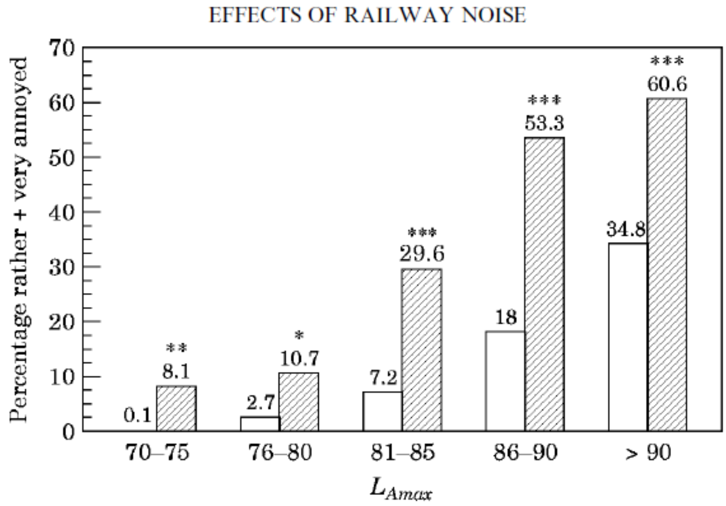

FIGURE 2.11: The relationship between annoyance and noise exposure from railway traffic in areas with vibration (shaded bars) and without vibration (white bars) (Source: Öhrström and Skånberg, 1996)

exceeding the vibration perception threshold appeared to be more important than the absolute number of train passbys.

Turunen-Rise et al. (2003) and Klæboe et al. (2003a,b) derived exposure-response relationships using the 1998 Norwegian Socio-Vibrational Survey datasets. In this survey, telephone interviews were conducted with 1503 individuals spread amongst 14 survey sites in order to determine people's reactions to vibrations in dwellings from a variety of sources including road and railway traffic. Vibration exposures quantified by statistical maximum weighted vibration velocity (see Equation 2.6) were estimated within 1427 of the dwellings using a semi-empirical vibration prediction model (Madshus et al., 1996). Logistic and ordinal logit regression models were then used to develop exposure-response relationships for annoyance as a function of exposure to railway and road traffic induced vibration, reproduced here in Figure 2.12. They observed the general trend that the proportion of people expressing certain levels of annoyance increases as the vibration exposure increases. They also found no significant difference in people's responses to vibration from different sources, concluding that, with appropriate frequency weightings, an exposure-response relationship can be derived without considering whether the vibration source is railway or road traffic. However, they made no attempt to distinguish between different sources of railway vibration. They conclude with the recommendation that further work is needed to determine whether or not the vibration exposure metric should take into account the number and duration of vibration events, as recommended by Griffin (1996).

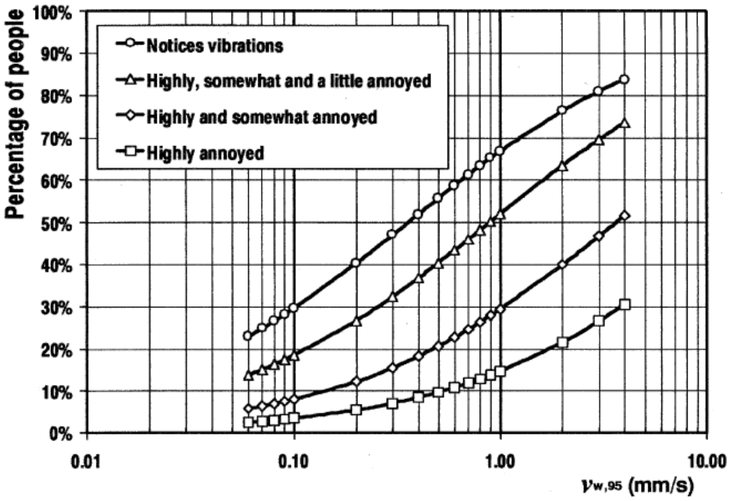

FIGURE 2.12: Estimated cumulative percentages of people reporting different degrees of annoyance as a function of vibration exposure (Source: Klæboe et al., 2003b)

Another field study was performed by the American Transit Cooperative Research Program (Zapfe et al., 2009). The study took place in five North American cities with 1306 interviews conducted via telephone about annoyance from vibration induced by railway systems, along with measurements of external vibration. With these measurements and responses, and using a logistic regression model, the authors were able to derive exposure-response relationships predicting the percentage of annoyed people as a function of vibration exposure. Many different descriptors were considered to quantify the measured vibration exposure, all of which were highly correlated with each other, leading Zapfe et al. (2009) to conclude that any of the metrics could be used as a good predictor of annoyance. The exposure-response relationship reproduced here in Figure 2.13 uses the passby maximum vibration *rms* vibration velocity level as a measure of vibration exposure, showing the proportion of people expressing annoyance to increase monotonically with vibration exposure.

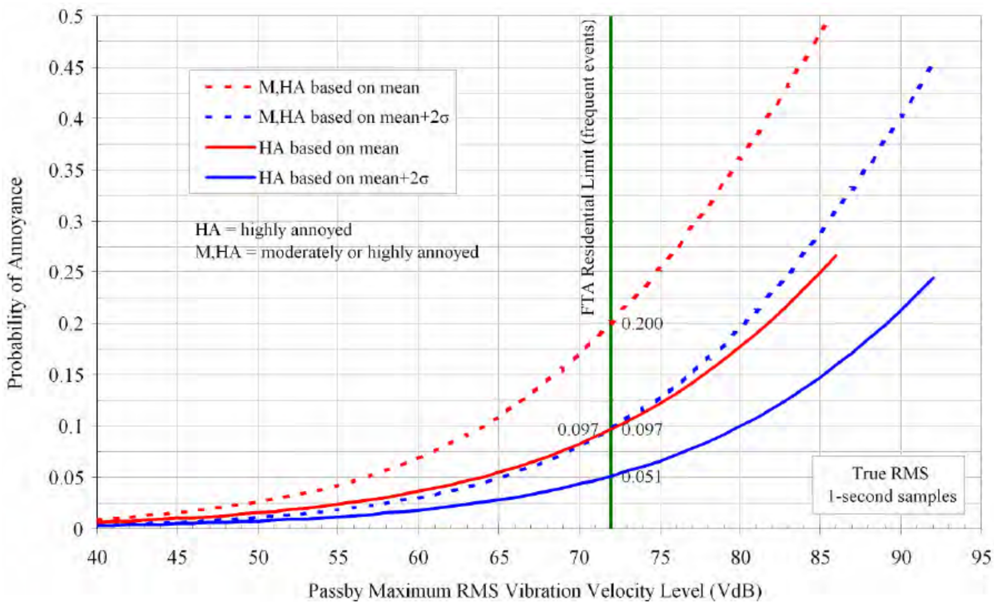

Figure 2.13: Probability of annoyance based on passby maximum *rms* vibration velocity level (Source: Zapfe et al., 2009)

The Swedish research project Train Vibration and Noise Effects (TVANE), summarised by Gidlöf-Gunnarsson et al. (2012), studied the effects of railway noise and vibration in residential environments, focusing on the effects of number of trains, the presence or absence of groundborne vibration and building situational factors such as orientation. A field study was performed in which questionnaires were conducted with 1695 respondents living within 451 m of a railway line. The respondents were spread across areas which were classified as having no vibration (521 respondents), some vibration (459 respondents) and a high frequency of train passbys (715 respondents). The questionnaires collected annoyance responses due to noise and vibration from the railway, and estimates of noise and vibration exposures were predicted for each respondent using combined measurement and prediction methods. For the same magnitudes of noise exposure, it was found that a higher proportion of respondents expressed high annoyance in areas classified as having vibration than those who were in areas without vibration. They derived an exposure-response relationship for the respondents living in areas with vibration, reproduced here in Figure 2.14, with the vibration exposure quantified as vibration velocity.

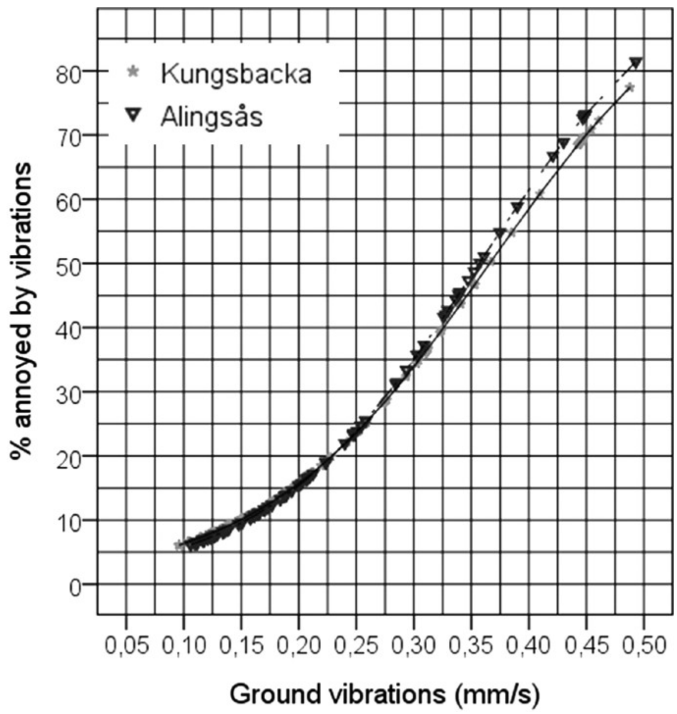

FIGURE 2.14: Estimated exposure-response relationship for annoyance due to railway-induced vibration quantified by vibration velocity at two sites (Source: Gidlöf-Gunnarsson et al., 2012)

Another recent large scale field study focusing on the human response to railway and construction vibration was carried out in the United Kingdom and is summarised by Waddington et al. (2014). In this study, annoyance data was collected using questionnaires conducted face-to-face with residents in their own homes for those exposed to railway induced vibration (931 respondents) and for those exposed to vibration from the construction of a light rail system (350 respondents). Sources of internal vibration (e.g. neighbour activity, door slams and domestic machinery) were also considered, though exposure-response relationships could not be developed for these sources, with the authors suggesting that annoyance from internal sources is better considered on a case-by-case basis than as a community response. For railway vibration exposure, measurements of vibration were conducted at internal and external positions, allowing the estimates of 24 hour vibration exposures to be derived for 1073 of the case studies. Sixty different vibration descriptors were considered to quantify the vibration exposure and, though none were found to be a better predictor of annoyance than any other, it was found that the use of relevant frequency weightings improved the correlation between vibration exposure and annoyance response. Exposure-response relationships for annoyance due to exposure to railway vibration and exposure to construction vibration were successfully derived, with the vibration exposure quantified using either the 24 hour W_b weighted VDV or the 24 hour W_m weighted rms acceleration. An exposure-response relationship for different degrees of annoyance due to exposure to railway vibration, quantified by W_b weighted VDV, is reproduced here in Figure 2.15. The data collected during this field study is used extensively in the research presented throughout this thesis.

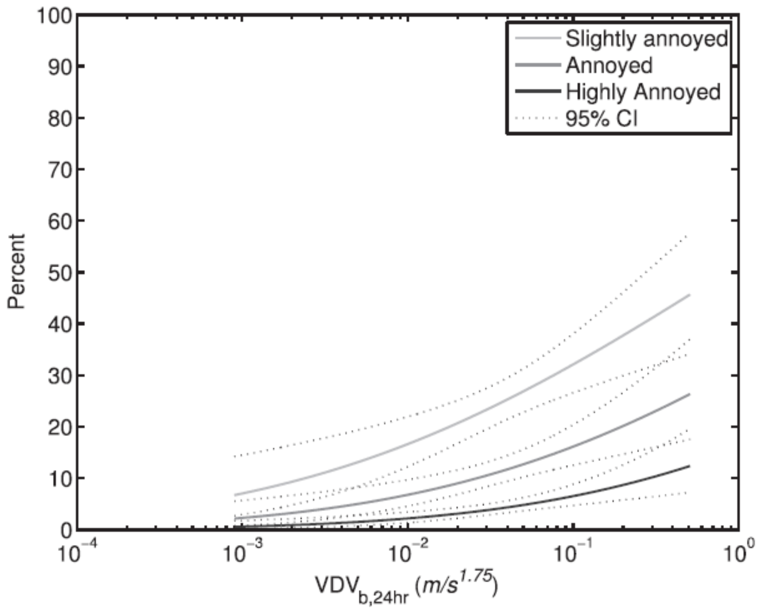

FIGURE 2.15: Exposure-response relationship showing the proportion of people reporting different degrees of annoyance for a given vibration exposure from railway. Dashed lines indicate the 95% confidence intervals (Source: Waddington et al., 2014)

In a follow up study using the data from the field study of Waddington et al. (2014), Peris et al. (2012) investigated the effects of time of day on the annoyance response due to railway vibration. They found that the annoyance response to

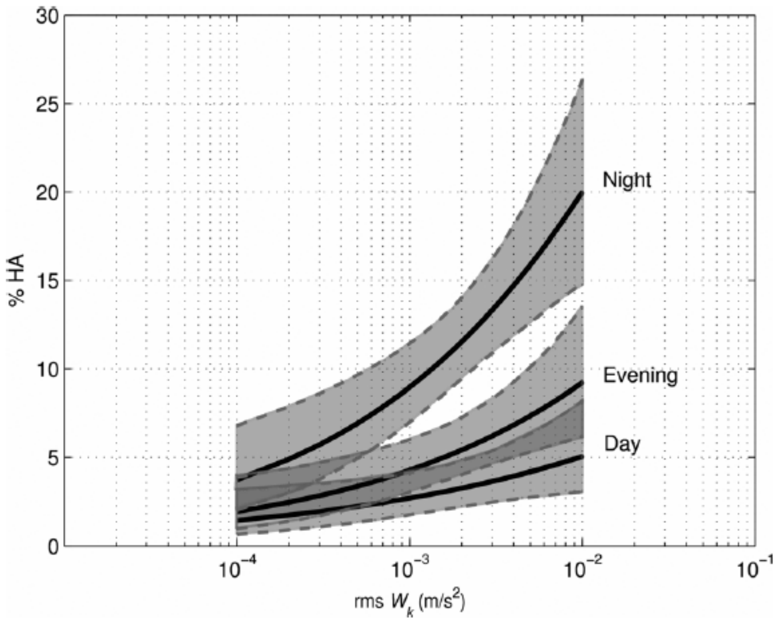

FIGURE 2.16: Exposure-response relationships for day (07:00 to 19:00), evening (19:00 to 23:00) and night (23:00 to 07:00), showing the percentage of people reporting high annoyance for a given vibration exposure. Curves are shown with their 95% confidence intervals (Source: Peris et al., 2012)

vibration exposure is higher during the night-time (23:00 to 07:00) than during the evening time (19:00 to 23:00), and higher during the evening time than during the daytime (07:00 to 19:00). The exposure-response relationships showing this difference in response for different times of the day is reproduced here in Figure 2.16.

In another follow up study using the same data, Peris et al. (2014) investigated the effects of situational, attitudinal and demographic factors on annoyance due to railway vibration. They found that annoyance scores were strongly influenced by concern of property damage and expectations about future levels of vibration, i.e. respondents were more likely to report a higher level of annoyance if they also felt concerned that their property was being damaged due to the vibration levels, or if they thought that the vibration levels were going to get worse in the future. They also found the type of residential area to have an effect, with the same vibration exposure level leading to more than twice as many respondents being highly annoyed by railway vibration in rural areas, such as small towns or villages, than those in urban areas. Respondent age was also shown to have an effect on annoyance, with respondents in the middle age range of around 45 years old showing the highest levels of annoyance for a given vibration exposure. Other factors with a significant but small influence on the annoyance response were the respondent's visibility of the railway and time spent at home, with respondents reporting higher levels of annoyance when the railway is visible from their residence or when the respondent spends less than 10 hours at home during a weekday.

Several field studies, mostly in Japan, have investigated the effects on the community of noise and vibration from high speed railway traffic but these sources are quite different from conventional railways and are generally outside the scope of this thesis. For more information on these studies see, for example, Yano et al. (2005), Yokoshima and Tamura (2005) and Yokoshima et al. (2011). Some key results of these studies showed that high speed train noise was more annoying than conventional railway noise (Yano et al., 2005) and that the presence of vibration from high speed trains led to greater noise annoyance (Yokoshima et al., 2011; Yokoshima and Tamura, 2005).

2.5.2 Sleep disturbance due to vibration exposure

A small number of studies have also investigated the effects of vibration on sleep disturbance. A laboratory study by Arnberg et al. (1990) looked at the effects on sleep of whole-body vibration and noise caused by heavy road traffic. During the study, nine participants slept in a specially designed room and were subjected to 140 vibration exposures with a dominant frequency of approximately 12 Hz and a duration of 2 s in both the vertical (peak levels of 0.24 m s^{-2}) and horizontal (peak levels of 0.17 m s^{-2}) directions simultaneously. Five of these participants were exposed to different vibration levels and the other four were exposed to a combination of noise (peak levels 50 dB(A)) and vibration. Their results suggested that when traffic noise is accompanied by vibration, sleep is more disturbed than when the noise is presented in the absence of vibration. They also found that the duration of rapid eye movement (REM) sleep, the subjectively rated sleep quality and morning performance was negatively affected by higher vibration levels.

More recent studies with a higher number of participants were performed as part of the Swedish Train Vibration and Noise Effects (TVANE) project (Ögren et al., 2009; Öhrström et al., 2009). These studies also involved subjects sleeping in specially designed rooms whilst being exposed to railway noise and vibration throughout the night. The results of these studies suggested that self-reported sleep disturbance was higher for increased vibration amplitude, irrespective of the noise level. A decrease in subjective sleep quality was observed when vibration amplitudes were increased from 0.4 to 1.4 mm s^{-2}. Limitations of this study, however, were identified by Smith et al. (2013) as the use of only two vibration amplitudes, with no intermediate levels, and the reliance on subjectively reported data.

Partly due to these limitations, further laboratory tests were performed by Smith et al. (2013) focusing specifically on variations arising from different vibration amplitudes, and incorporating physiological measurements of sleep disturbance as well as self-reported data. Again, subjects slept in a specially designed room and were exposed to a range of vibration exposures and fixed noise exposures from freight railway traffic. Cardiac accelerations of the subjects were assessed throughout the night using a combination of polysomnography and electrocardiogram (ECG) recordings and sleep was assessed subjectively using questionnaires. The results of the study indicated that nocturnal vibration has a negative impact on sleep and that the impact increases with vibration amplitude. With increasing vibration amplitude the authors found a decrease in latency and an increase in amplitude of heart rate as well as a reduction in sleep quality and increase in sleep disturbance.

2.6 Laboratory studies on the effects of combined railway noise and vibration

Vibration often exists alongside airborne noise and this is certainly the case for the vibration that is experienced in residential environments due to railway traffic. Many of the field studies described in the previous sections have shown that perceivable levels of vibration in combination with noise from railways can have an effect on the overall annoyance response (Fields, 1979; Gidlöf-Gunnarsson et al., 2012; Schomer et al., 2012; Waddington et al., 2014). Indeed, Schomer et al. (2012) suggest the need to develop separate predictions for annoyance due to railway noise for railway sources that produce perceivable vibrations and for those that do not. They demonstrate that, even though railway noise is generally believed to be less annoying than road traffic noise (Miedema and Vos, 1998; Moehler, 1988; Moehler et al., 2000), when perceivable vibration is present, railway noise can actually cause more annoyance than road traffic noise.

The first laboratory studies on the subjective response to noise and vibration from railway traffic were performed by Howarth and Griffin (1990). In these studies, subjects were presented with stimuli composed of different combinations of six magnitudes of railway noise and vibration. Twenty four subjects took part in the three part magnitude estimation study. In the first session, the assessment of vibration in the presence of noise was investigated. In the second session, the assessment of noise in the presence of vibration was investigated. In the third session, the combined effects of noise and vibration was investigated. The results of these tests indicated that, within the range of stimuli magnitudes investigated, vibration does not significantly influence the judgement of noise, but the judgement of vibration may be affected by the presence of noise, depending on the magnitudes of the stimuli. They discovered that a reasonable approximation of the total annoyance caused by combined noise and vibration stimuli can be determined from a summation of the effects of the individual stimuli, using the following relations:

$$\psi = 15.9 + 260\varphi_v^{1.04} + 0.167\varphi_n^{0.039} \quad (2.8)$$

$$\varphi_v = \mathrm{VDV}_b \quad (2.9)$$

$$10\log_{10}\varphi_n = SEL_A \quad (2.10)$$

where ψ is the annoyance response, VDV_b is the W_b weighted vibration dose value and SEL_A is the A-weighted sound exposure level. In a follow up study, Howarth and Griffin (1991) investigated the effects of duration, magnitude and frequency of vibration stimuli in combination with noise. This was achieved by exposing subjects to combinations of noise and vibration stimuli with different magnitudes, duration and vibration frequency. Their results suggested that an annoyance relation involving a summation of the individual magnitudes of the noise and vibration stimuli provides a more accurate means of prediction of overall annoyance due to combined noise and vibration than relations based on either the noise or vibration stimuli alone. They derived a new relationship for predicting overall annoyance, which is similar to Equation 2.8 and is shown below:

$$\psi = 22.7 + 243\varphi_v^{1.18} + 0.265\varphi_n^{0.036} \quad (2.11)$$

where φ_v and φ_n are defined by Equations 2.9 and 2.10 respectively. Although the relationship shown in Equation 2.11 was derived for stimuli with varying durations, the maximum duration was 29 s, and Howarth and Griffin (1991) state that the relationship may not be appropriate for predicting annoyance for combined noise and vibration stimuli with durations longer than this.

In another laboratory study, Paulsen and Kastka (1995) investigated the effects of combined noise and vibration, from a passing tram and a hammermill, on rated intensity and annoyance. Four levels of noise and vibration levels were presented to subjects in every possible combination. The results indicated that the presence of vibration influences the evaluation of noise annoyance and has a greater influence on the evaluation of total annoyance. They developed a predictive relationship for total annoyance caused by combined tram noise and vibration exposure, which takes the same form as those developed Howarth and Griffin (1990, 1991), and is shown below:

$$\psi = -0.15 + 1.58\log_{10}(v_{rms}) + 0.11L_{Aeq} \tag{2.12}$$

where v_{rms} is the rms vibration velocity and L_{Aeq} is the A-weighted continuous sound pressure level.

More recently, Jik Lee and Griffin (2013) performed a laboratory study looking at the combined effects of noise and vibration produced by high speed trains on annoyance. In this test, subjects were exposed to six levels of noise and six levels of vibration exposure, for both windows open and windows closed scenarios. The experiment was divided into four sessions:

1. Evaluation of noise annoyance in the absence of vibration
2. Evaluation of total annoyance from simultaneous noise and vibration
3. Evaluation of noise annoyance in the presence of vibration
4. Evaluation of vibration annoyance in the presence of noise

The results indicated that vibration did not influence ratings of noise annoyance, but that total annoyance for combined noise and vibration was significantly greater than that due to noise alone. The authors developed predictive models of the total annoyance resulting from combined noise and vibration based on two classical models: the dominance model (Rice and Izumi, 1984) and the independent effect model (Taylor, 1982). In the dominance model, the total annoyance is a function of the maximum of the single source (noise or vibration) annoyance, whereas the independence model is a function of of the annoyance of both sources, assuming that the separate sources make independent contributions to the total annoyance.

Jik Lee and Griffin (2013) found that both models provided useful predictions of the total annoyance caused by simultaneous noise and vibration from high speed trains.

2.7 Differences in response to freight and passenger railway traffic

Although the results of several field and laboratory studies indicate that respondents rate freight railway noise as more annoying than passenger railway noise (Andersen et al., 1983; Fields, 1979; Fields and Walker, 1982), a relatively small number of studies have addressed this difference in human response directly. De Jong and Miedema (1996) performed an investigation into whether freight traffic noise is more annoying than passenger traffic noise as a direct result of concern over a new freight only railway route, the "Betuweroute", that was planned to carry freight traffic between Rotterdam and Germany. They concluded that, although residents are more likely to report annoyance with freight traffic generally, after the effects of differing noise levels are removed, no consistent differences were found between routes with small or high proportions of freight traffic trains. Other relevant issues, such as the effects of vibration and late night freight traffic on sleep disturbance, were not considered.

Saremi et al. (2008) performed a study looking at how different types of train influence the effects of railway noise on sleep fragmentation, citing the general rule that noise disturbances vary according to the physical characteristics of the noise events. It is therefore possible that different train types, which produce different noise and vibration signals, would result in similar differences. Using different train type signals, they discovered that the type of train had a significant effect on awakenings (arousals lasting longer than 10 s), with freight trains causing a much higher awakening rate when compared to passenger or automotive trains. Despite this significance difference, no difference on induced micro-arousals (arousals lasting between 3 and 10 s) was found between the train types. The type of train also had an effect on the arousal onset latency, with arousal reactions occurring much later after the passby of freight trains when compared to the passby of passenger or automotive trains.

Saremi et al. (2008) theorise that the greater degree of nocturnal disturbance caused by freight trains could be due to the duration of the noise signal, as it has been shown that stimuli of longer duration elicit more physiological responses with greater amplitude in waking subjects (Blumenthal, 1988; Blumenthal et al., 2005; Putnam and Roth, 1990). The increased duration of freight train noise signals could therefore cause the increased awakening rate observed by Saremi et al. (2008). It has been theorised that humans respond to stimuli of increased duration using temporal summation, responding to a greater extent to stimuli of longer duration which contain more energy (Blumenthal and Berg, 1986; Blumenthal and Goode, 1991). This temporal summation could cause humans to respond more to freight train passbys of longer duration, leading to sleep disturbance during the night and increased annoyance during the day. A further physical property that Saremi et al. (2008) observed to be different for freight and passenger train signals was that of rise time. Their results in terms of rise time were in agreement with Blumenthal (1988) who discovered that auditory stimuli with longer rise times, i.e. freight trains, elicit responses with greater onset latency. It is likely that these physical characteristics combine together to influence sleep disturbance and annoyance.

In a German field study, Pennig et al. (2012) investigated annoyance and self-reported sleep disturbance due to night-time railway noise, with specific attention made to differences in response to freight and passenger railway sources. They found that annoyance was primarily determined by freight trains, with annoyance ratings increasing significantly with the total number of trains and freight trains per night, but non-significantly with increasing numbers of passenger trains. The total number of trains and freight trains were also found to significantly affect the frequency of self-reported awakenings. In providing some possible explanations for the difference in annoyance response, they cite the typically longer durations of freight train passbys, higher maximum sounds levels, their increased occurrence during night-time hours and the potential for the presence of groundborne vibrations and accompanying low frequency noise. As part of the same study, Elmenhorst et al. (2012) investigated the difference in response between nocturnal railway noise and air traffic noise using polysomnography. It was found that nocturnal freight railway noise accounted for more awakenings than passenger railway noise and aircraft noise.

2.8 Summary

To summarize, a great deal of field and laboratory studies have focused on the human response to railway noise, yet the human response to railway vibration has been somewhat less examined. In particular, the differences in human response to freight and passenger railway noise has been studied relatively little and studies on the difference in human response to freight and passenger railway vibration are almost non-existent. In light of the fact that freight railway traffic is increasing, it is important that the human response to freight railway noise and vibration be better understood. The aim of this research, therefore, is to develop separate exposure-response relationships for annoyance caused by exposure to vibration from freight and passenger railway vibration, allowing the difference in response to these two sources of railway vibration to be better understood.

Following this is an investigation into the human response to combined noise and vibration from railway traffic, in the hopes of revealing characteristics of the freight and passenger stimuli which lead to a difference in the human response to these two sources of environmental noise and vibration. The results of this analysis are used to develop models that are able to predict annoyance due to railway noise and vibration, regardless of the whether the source is freight or passenger trains.

Chapter 3

Classification of unknown railway vibration signals

3.1 Introduction

The literature review in Chapter 2 has highlighted that relatively little work has been performed on the human response to railway vibration, and in particular the differences in the human response to freight and passenger railway vibration. With the proposed increases in freight traffic outlined in the European Union White Paper (2001), it is important to understand the potential impacts that the resulting increase in vibration may have on residents living in the vicinity of railway lines. The aim of this research, therefore, is to develop exposure-response relationships for annoyance caused by exposure to freight and passenger railway vibration. In order to develop these exposure-response relationships, it is necessary to first determine freight and passenger vibration exposures from a database of measurements recorded as part of a field study on the human response to vibration, performed by Waddington et al. (2014), a brief summary of which is provided in Section 2.5.1.

This chapter begins by providing the motivation for developing separate exposure-response relationships for freight and passenger railway vibration, then goes on to detail the methods that will be used to achieve this. The main focus of this chapter is to describe the creation, optimisation and utilisation of a logistic regression classification model that is used to classify unknown railway vibration signals within the database of measurements. The features of the model are optimised using a combination of correlation testing, univariate and multivariate likelihood ratio testing and accuracy testing. The final optimised model is a function of only two parameters quantifying the vibration signal's duration and low frequency content and is able to correctly classify, on average, 96% of unknown signals which are introduced to the model independently of the training, fitting and optimisation of the model. This is a promising result and suggests that the model can be successfully used to classify the unknown signals in the measurement database, in order to determine separate exposure-response relationships for annoyance caused by passenger and freight railway vibration.

3.2 Motivation for developing separate exposure-response relationships for freight and passenger railway vibration

Though little work has focussed on the difference in the human response to freight and passenger railway traffic directly, several laboratory and field studies have indicated that respondents rate freight railway noise and vibration as more annoying than passenger railway noise and vibration (see Section 2.7). A preliminary analysis can be performed on the responses collected during the field study by Waddington et al. (2014) to see if this is also true for this field study.

In the field study, respondents were asked, via face-to-face questionnaires, how bothered, annoyed or disturbed they were by freight and passenger railway vibration (the questionnaire and method of response collection is described in further detail in Section 4.3). The percentage of respondents reporting annoyance categories of being slightly annoyed and above due to freight and passenger railway vibration is shown in Figure 3.1. Clearly, a higher number of respondents reported that they found freight railway vibration to be "moderately", "very" or "extremely" annoying than those reporting the same annoyance levels for passenger railway vibration. The percentage of respondents who reported freight railway vibration to be very or extremely annoying is approximately 14%, compared to only 4% of respondents who report the same levels of annoyance due to passenger railway vibration.

The difference in the human response to freight and passenger railway vibration can be tested by performing a Kruskal-Wallis test, a non-parametric one-way analysis of variance which tests the the null hypothesis that independent samples from the two response groups come from distributions with equal medians. The results of the test indicate that the null hypothesis can be rejected and that the annoyance response due to freight railway vibration is significantly higher than that due to passenger railway vibration ($\chi^2 = 16.3$, $p < 0.001$). The mean rank of annoyance scores for freight and passenger railway vibration, as determined by

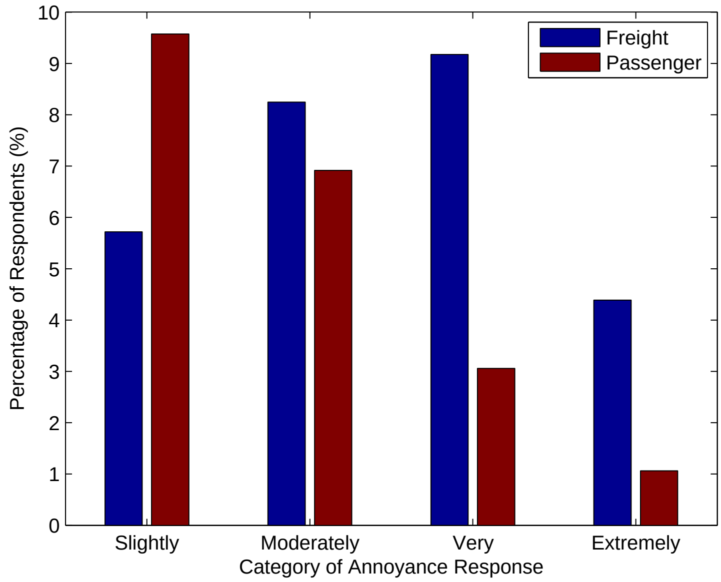

FIGURE 3.1: Percentage of respondents reporting different levels of annoyance for freight and passenger railway vibration

the Kruskal-Wallis test, are presented in Figure 3.2. The different mean ranks, and their non-overlapping standard errors, indicate a significant difference in annoyance responses for these to sources of railway vibration.

The response data collected during the field study, and the results of the Kruskal-Wallis statistical tests give confidence that different relationships may exist for the human response to freight and passenger railway vibration, providing motivation to investigate these relationships separately. Though separate responses data were collected for freight and passenger vibration, measurements were made for all railway traffic in a 24 hour period, meaning that separate exposure data for these two sources cannot be directly determined. In order to determine separate exposure data, and subsequent exposure-response relationships, the railway vibration signals captured during the measurements need to be accurately classified as either freight or passenger signals. The topic of this chapter is the development of a supervised machine learning algorithm that is capable of achieving this classification of unknown railway vibration signals.

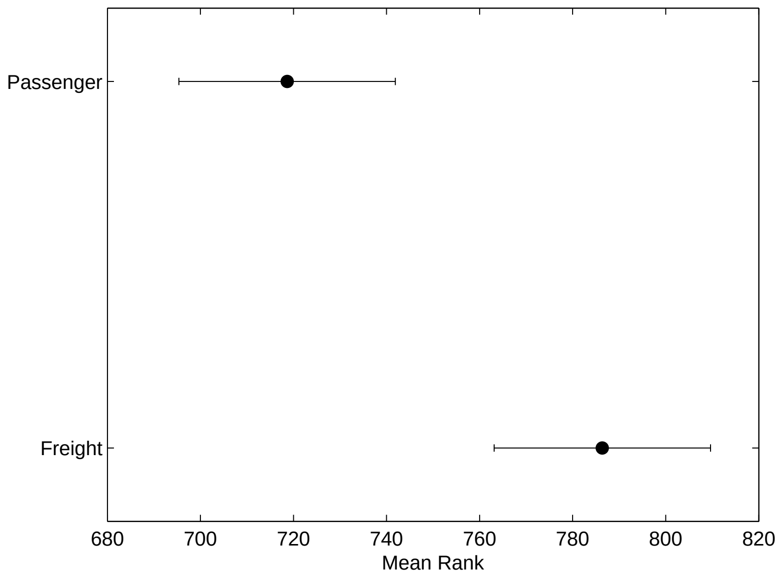

FIGURE 3.2: Comparison of mean rank annoyance for exposure to freight and passenger railway vibration

3.3 Source classification using supervised machine learning

The method used to classify unknown railway vibration signals as either freight or passenger signals in this research is that of supervised machine learning. Supervised machine learning is a powerful tool whereby known data is used to develop an algorithm capable of classifying unknown data based on several properties, or features, of the data. In this case, the known data is a set of freight and passenger vibration signals, the unknown data is vibration signals in the database of field study measurements and the features are objective parameters of the vibration signals. In essence, the model "learns" typical values of features for known examples of the classes to be identified, then makes a comparison between these values and values of the same features for unclassified data, before making a decision of in which class the unclassified data best belongs.

Within the field of acoustics, machine learning, both supervised and unsupervised, has been used to accomplish many varied tasks. For example, Brown (1999) performed a k-means clustering algorithm for the classification of musical oboe and saxophone sounds. The resulting algorithm, which uses cepstral coefficients of the sounds as the model features, was found to be better than humans at identifying oboe sounds, and as good as humans at identifying saxophone sounds. For more information on the k-means clustering algorithm, and many other machine learning methods, see Alpaydin (2010) or any other introductory machine learning text.

An example of the varied potential applications of machine learning can be seen in a series of papers involving the same author as the paper on classification of musical sounds, in which machine learning is used to identify killer whales, based on the cepstral coefficients of their calls (Brown and Smaragdis, 2009; Brown et al., 2010). The method of classification used in these papers were hidden Markov models and Gaussian mixture models. In a more recent paper, Shamir et al. (2014) also used machine learning, in combination with crowd sourcing from an online citizen science project, to classify whale calls.

3.4 Classification using logistic regression

There are many methods of machine learning for classification of unknown data, but due to the dichotomous nature of the classification required for this research, the method of logistic regression shall be employed. Logistic regression models have been used for many years and gained increased popularity after Truett et al. (1967) successfully used the model to provide a multivariate analysis of the Framingham heart study data. Since this work, logistic regression has become a standard method for regression analysis of dichotomous data in many fields (Hosmer and Lemeshow, 1989). This section presents the details of the logistic regression model, as well as methods for testing the goodness of fit, significance and accuracy of the model.

3.4.1 The logistic regression model

The logistic regression model makes its predictions based on the logistic function (Equation 3.1, Figure 3.3). As Z tends towards positive infinity, the probability $P(\mathbf{X})$ tends towards 1 and as Z tends towards negative infinity, $P(\mathbf{X})$ tends towards 0. The logistic function, therefore, is a useful model as it can be used to predict a probability between 0 and 1 that an object belongs to a certain class from two possible choices. In this case the logistic function predicts the probability that a vibration event signal belongs in the freight railway class, rather than the passenger railway class. The logistic function is a function of the model Z, which in turn is a function of several properties or "features" of an object. In Equation 3.2, x_n refers to the numerical value of the independent feature n, and β_n refers to initially unknown regression coefficients corresponding to the same feature. Since x_0 is the intercept term, β_0 is always equal to 1. $\boldsymbol{\beta}$ and $\mathbf{X}$ are vectors containing all values of β and x.

$$P(\mathbf{X}) = \frac{1}{1 + e^{-Z}} \tag{3.1}$$

$$Z = \beta_0 x_0 + \beta_1 x_1 + +\beta_2 x_2 + ... = \boldsymbol{\beta}\mathbf{X} \tag{3.2}$$

3.4.2 Fitting the logistic regression model

With the logistic model defined, it is necessary to fit the predictions of the model to known or "observed" data, by estimating the coefficient $\boldsymbol{\beta}$ which result in the best fit of the model to the observed data. In linear regression, the most commonly used method of regression coefficient estimation is that of least squares. This method determines the values of $\boldsymbol{\beta}$ that minimise the sum of the squared deviations of the observed values from the values predicted by the model. However, when this method is applied to models with dichotomous outcomes, it results in coefficients with undesirable statistical properties (Hosmer and Lemeshow, 1989).

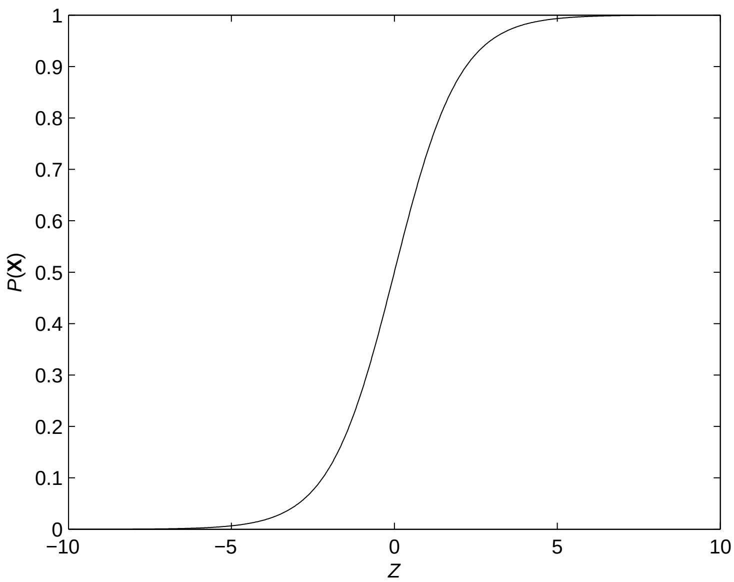

FIGURE 3.3: The logistic function

The preferred method for fitting the logistic regression model is that of maximum likelihood which determines values for the regression coefficients that maximise the probability of obtaining the observed data set. This method is based on maximising the likelihood function, which expresses the probability of the observed data as a function of the estimated coefficients $\boldsymbol{\beta}$. The likelihood function for logistic regression is shown in Equation 3.3, where m is the number of observations and Y_m is the observed class (0 or 1) of the mth observation (Alpaydin, 2010).

$$L(\boldsymbol{\beta}) = \prod_m P(\mathbf{X}_m)^{Y_m}[1 - P(\mathbf{X}_m)]^{1-Y_m} \tag{3.3}$$

Mathematically, it is preferable to convert the likelihood function into an error function to be minimised, where $LL = -\ln L$. This function is known as the log likelihood or cross-entropy and is shown in Equation 3.4.

$$LL(\boldsymbol{\beta}) = -\sum_m Y_m \ln\left[P(\mathbf{X}_m)\right] + [1 - Y_m]\ln[1 - P(\mathbf{X}_m)] \tag{3.4}$$

Due to the non-linearity of the logistic function, it is not possible to solve Equation 3.4 directly. Instead, the method of gradient descent is used to minimise the log likelihood, which is mathematically equivalent to maximising the likelihood. Gradient descent is an iterative optimisation method which utilises both the value of the log likelihood function, and its derivative, or gradient (Equation 3.5), to iteratively step in the direction of the gradient by changing the values of $\boldsymbol{\beta}$ until the derivative is equal to zero and a minimum has been reached. This method can be problematic for models with local minima, but since logistic regression has only one minimum, gradient descent is a suitable optimisation method (Alpaydin, 2010).

$$\Delta LL = \sum_m [Y_m - P(\mathbf{X}_m)]\mathbf{X}_m \tag{3.5}$$

3.4.3 Goodness of fit and significance of regression coefficients

For ordinary least squares regression, the goodness of fit is generally assessed using the R^2 value, which is commonly interpreted as a description of the proportion of variance in the variable explained by the model. However, since logistic regression is determined using maximum likelihood, rather than minimisation of variance, the R^2 value is not an appropriate descriptor for the goodness of fit of the model. Instead, many descriptors similar to the R^2 value have been developed in an attempt to assess the goodness of fit of a regression model determined using maximum likelihood (Long, 1997). In this work, McFadden's R^2_{pseudo} value (Equation 3.6) will be used to assess the goodness of fit of the regression model. This R^2_{pseudo} value considers the log likelihood of the full model with all its coefficients, $\boldsymbol{\beta}$, (LL_{full}) and the log likelihood of the model with only the intercept term, β_0, considered (LL_{int}). If all the coefficients $\boldsymbol{\beta}$ remain zero after the fitting of the model, then $LL_{full} = LL_{int}$ and $R^2_{pseudo} = 0$ (i.e. the model is very poorly fit). As the fit of the model improves, the difference between LL_{full} and LL_{int} increases and R^2_{pseudo} approaches 1 (though it can never exactly equal 1).

$$R^2_{pseudo} = 1 - \frac{LL_{full}}{LL_{int}} \tag{3.6}$$

As well as testing the goodness of fit of the model, the significance of the coefficients, $\boldsymbol{\beta}$, can be examined. The results of these tests can be used to determine which features contribute significantly to the model, using a likelihood ratio test. For this test, an unconstrained model with log likelihood LL_{uncon} is defined containing the intercept term and a single feature x and corresponding coefficient β. A constraint is then imposed on this model in which $\beta = 0$ and the model is reduced to the intercept only model with log likelihood LL_{con} (this is equivalent to LL_{int}). The null hypothesis is then that the constraint is valid, and the feature does not significantly contribute to the model. The likelihood ratio test statistic for this test is defined as follows (Hosmer and Lemeshow, 1989; Long, 1997):

$$\chi^2 = 2\left(\frac{LL_{uncon}}{LL_{con}}\right) \tag{3.7}$$

If the null hypothesis is true, then χ^2 is asymptotically distributed as chi-square with degrees of freedom equal to the number of independent constraints. The significance of this hypothesis can then be tested and rejected if $p < 0.01$. If the null hypothesis is rejected, then it can be assumed that the feature being tested does indeed significantly contribute to the model. The likelihood ratio test can also be used to determine the significance of adding a new feature to an existing multivariate model by defining the constrained model as the existing model and the unconstrained model defined as the existing model with the addition of the new feature. Finally, the likelihood ratio test can also be used to determine the combined significance of the coefficients of a full model, by defining the unconstrained model as the full model with all features $\mathbf{X}$ and coefficients $\boldsymbol{\beta}$ included (this is equivalent to LL_{full}) and defining the constrained model as the intercept only model (this is equivalent to LL_{int}).

Once the model has been optimised, the standard error and significance of the coefficients can be determined. The standard error is calculated from the diagonals of the covariance matrix, $\sigma^2(x_n)$, where n represents the nth parameter, as follows:

$$\mathrm{SE}(x_n) = \sqrt{\sigma^2(x_n)} \tag{3.8}$$

The univariate Wald test statistic can then be used to test the hypothesis that each of the coefficients in the optimised model are non-zero and significant. The Wald test statistic of the nth coefficient is defined as:

$$W_n = \frac{x_n}{\mathrm{SE}(x_n)} \tag{3.9}$$

The Wald test statistic can then be tested against a standard normal distribution for an indication of the significance of each coefficient.

3.4.4 Classification accuracy of the model

Since the logistic regression model is used to classify unknown signals, it is also necessary to test the ability of the model to accurately achieve this purpose. Theoretically, a model could be perfectly fit to a set of training data, but if it is unable to generalise to classify new unknown signals that it has not seen before, then it will perform very poorly at classifying unknown signals correctly, a situation referred to as "over-fitting". It is necessary therefore, to also test the accuracy of classification that the fitted model achieves. In order to test this, the set of known vibration signals is split into three data subsets: training, cross validation and test. First, the model is fit to the labelled training data set by maximum likelihood. Next, the model with the maximum likelihood estimated coefficients is applied to the cross validation data set and the probability that each event signal belongs to the freight class is predicted. All signals with a predicted probability greater than 0.5 are sorted into the freight class. Since the cross validation data set is also labelled, the accuracy of classification can be calculated. Investigations can then be performed to find the optimal conditions of the model, such as the selection of features that make up the model, that results in the greatest prediction accuracy. Finally, the optimised model is applied to the test set, which until that point has not been introduced to the model in any way, making this data set entirely independent of the fitting or optimising of the model. The accuracy is then reported based on the success of classification of the test data set.

There are several ways to determine the accuracy of the model, either on the cross validation or test set. The simplest is to calculate the classification accuracy, defined as the proportion of total signals tested that are correctly classified. However, this can be misleading for cases of skewed classes, where there are many more examples of one class than the other in the test set. This is the case for this work, where approximately 80% of the signals belong to the passenger class. The model could therefore achieve a classification accuracy of 80% simply by predicting that every signal belongs to the passenger class, regardless of its features. It is more appropriate therefore, especially for skewed classes, to determine accuracy based on true positives, true negatives, false positives and false negatives. These terms are defined by the confusion matrix as shown in Table 3.1.

		Predicted Class	
		Freight	Passenger
Actual Class	Freight	True Positive (TP)	False Negative (FN)
	Passenger	False Positive (FP)	True Negative (TN)

Table 3.1: Confusion matrix

Determining the true and false positives and negatives allows the calculation of precision and recall. Precision is defined as the proportion of predicted freight signals that truly belong to the freight class (Equation 3.10) and recall is defined as the proportion of true freight signals that are correctly classified as such (Equation 3.11). Accuracies of binary classification algorithms are often reported using the F_1 score (Equation 3.12), which is the harmonic mean of precision and recall, and ranges from 0 to 1, with 1 representing perfect classification. The F_1 score can change depending on which class is defined as the "positive" class and it does not take into account true negatives. Another descriptor, the Matthews Correlation Coefficient (MCC) (Equation 3.13), does take into account true negatives and is independent of which class is defined as the positive class. Its values range from -1 to 1 with 1 representing perfect classification. Both the F_1 score and the Matthews Correlation Coefficient will be considered when assessing the accuracy of the logistic regression model.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3.10}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3.11}$$

$$F_1 = 2\left(\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}\right) \tag{3.12}$$

$$MCC = \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}} \tag{3.13}$$

3.5 Building the logistic regression model

In the previous section, the theory behind the logistic regression machine learning model is outlined, along with methods used to fit the model and to determine the significance of its regression coefficients and the classification accuracy of the model. In this section, these methods will be used to build and optimise a logistic regression model for the classification of unknown railway vibration signals in the database of field measurements performed by Waddington et al. (2014).

3.5.1 Data extraction and labelling

Machine learning using logistic regression relies on labelled input data in order to "train" the model so that it can subsequently make predictions about the class of unknown and unlabelled data. Therefore, it is necessary to extract and label known vibration signals from the field study measurement database. This was made possible since those who performed the field study noted the details of passbys that occurred during certain periods of vibration measurements. In many cases, the operators noted the type of train and the time of the passby, allowing the resulting vibration signals to be identified within the acceleration time histories, and then extracted and labelled as either freight or passenger signals. As there were several inconsistencies in the handwritten logs from which the signal labels were determined, only logs that contained examples of freight trains were used, as this indicates that the operator was able and willing to distinguish between freight and passenger trains. For example, logs which had all passbys marked simply "train" were ignored as it is unclear whether all the passbys that occurred during this measurement periods were passenger trains, or if the operator did not, or was unable to, determine whether the passbys were freight or passenger trains. In total, 194 passenger and 44 freight railway vibration signals were extracted and labelled. These vibration signals were taken from 27 separate external control position measurements spread over 7 sites along the West Coast Main Line in the North West and Midlands regions of England. Figure 3.4 shows examples of known freight and passenger railway vibration signals identified using the above methods.

3.5.2 Initial feature selection

One of the most important aspects of classification model design is the selection of features that will make up the model. Finding features that can be used to effectively differentiate between two classes not only results in a highly accurate classification model, but also provides information about the differences in the two

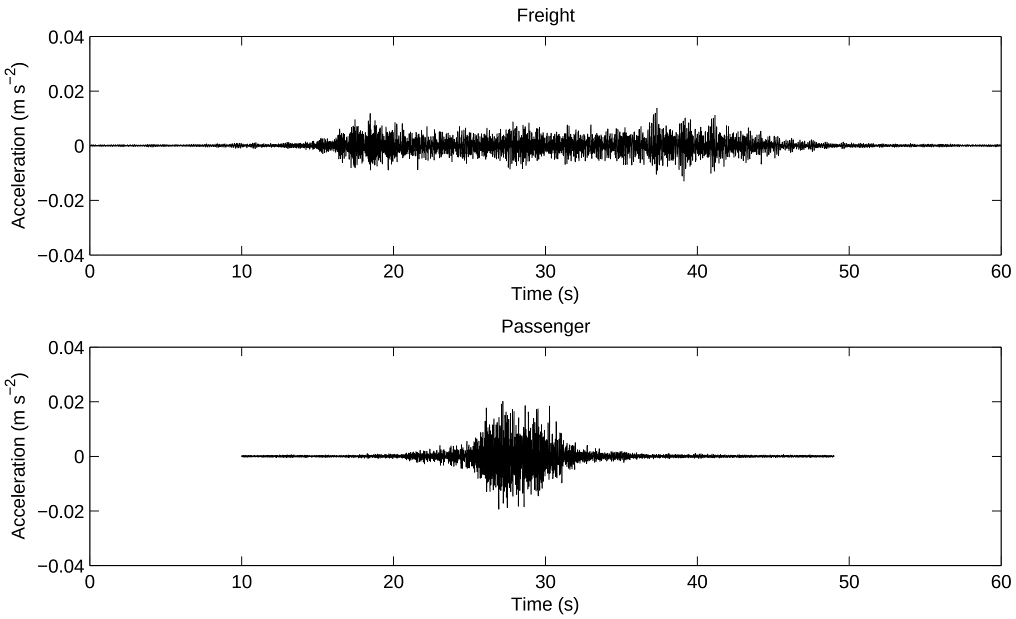

FIGURE 3.4: Examples of known signals of freight and passenger railway vibration

classes. When optimising the features used in the logistic regression model for this work, over 130 features were initially introduced to the model. Through a combination of univariate and multivariate significance testing, testing of correlation between features and accuracy testing, the number of features was reduced to only 2. To avoid the model being over-fitted to a certain set of signals, each test was performed over 1000 randomised splits of training, cross-validation and test sets and decisions were made based on the mean value of these repeated tests. The process of selecting these features is described in this section.

As was demonstrated in Section 3.2, the annoyance responses collected in the field study by Waddington et al. (2014) indicate that the annoyance response to freight railway vibration is different than that to passenger railway vibration. A sensible starting point for feature selection therefore is to consider vibration magnitude descriptors that have previously been shown by Waddington et al. (2014) to correlate well with annoyance . One such descriptor is the vibration dose value, recommended as a quantifier of vibration exposure by the British Standard BS 6472-1:2008. For more information on this metric, and the relevant British

Standard, see Section 2.3.3. Vibration dose value is calculated as follows:

$$\mathrm{VDV} = \sqrt[4]{\int_0^T a(t)^4 dt} \tag{3.14}$$

where $a(t)$ is the vibration acceleration time history with total duration T. Another vibration magnitude descriptor that was shown to correlate with annoyance is the rms acceleration as recommended by the standard BS ISO 2631-1:1997. Again, for more information on this metric, and the standard which recommends it, see Section 2.3.3. The rms acceleration is calculated as follows:

$$rms = \sqrt{\frac{1}{T}\int_0^T a(t)^2 dt} \tag{3.15}$$

Other vibration magnitude descriptors include the equivalent continuous vibration level, which is analogous to the equivalent continuous sound pressure level, L_{eq} and is calculated as follows:

$$V_{eq} = 20\log_{10}\left(\frac{rms}{1 \times 10^{-6}}\right) \tag{3.16}$$

The vibration exposure level is analogous to the sound exposure level, SEL, and is the vibration level of duration 1 second that would have the same energy content as the whole event. The vibration exposure level is calculated using the following equation:

$$VEL = V_{eq} + 10\log_{10}(T_s) \tag{3.17}$$

The above descriptors should sufficiently quantify any differences between freight and passenger vibration signals that may exist in exposure magnitude. As well as differences in exposure magnitude between these sources, there may be differences in the spectral content which could lead to differences in the annoyance response, since the perception of whole-body vibration has been shown to be influenced by vibration frequency (see Section 2.2.3). Therefore, it would be sensible to include some descriptors which are measures of the frequency content of the vibration signals in the initial feature selection. One such feature is the spectral centroid, which is a weighted mean of the frequency content in the signal, with the magnitudes of the Fourier transform coefficients as weights, defined as follows:

$$f_{sc} = \frac{\sum f(n) c_{mf}(n)}{\sum c_{mf}(n)} \tag{3.18}$$

where $f(n)$ is the central frequency of the nth spectral bin and $c_{mf}(n)$ is the magnitude Fourier coefficient of the nth spectral bin. As well as the spectral centroid, spectral energy is quantified by determining the proportional rms acceleration contained within all 1/3rd octave bands between 0.5 and 80 Hz. This proportional rms acceleration is defined as the rms acceleration within a particular 1/3rd octave band divided by the rms acceleration in the entire signal. Different spectral distributions may also be captured by the W_b and W_k frequency weightings that can be applied to calculations of VDV and rms respectively.

Freight railway passbys are typically longer in duration than passenger railway passbys, which may account for differences in the annoyance response, since the perception of whole-body vibration has also been shown to be influenced by duration (see Section 2.2.4). Descriptors of the duration of the signal will therefore also be included in the initial feature selection. These descriptors include the duration defined by the 3 dB downpoints of the signal (T_{3dB}), the duration defined by the 10 dB downpoints of the signal (T_{10dB}) and the “event signal duration” (T_e) defined here as the duration of the signal that exceeds the top 1/3rd of the signal’s dynamic range. All downpoints are defined as the “first” and “last” points of the signal which exceed the specified threshold, in order to avoid capturing only strong peaks in the signal. The event signal duration was defined to capture the duration of signals regardless of their dynamic range, since the T_{3dB} and T_{10dB} descriptors can under-estimate or over-estimate the duration of signals with high or low dynamic ranges respectively. This is important to consider, since the dynamic range of vibration signals will be affected by the propagation distance from the railway to the residence. Other temporal descriptors include the rise time, defined as the duration between the first 10 dB and 3 dB downpoints, and the fall time, defined as the duration between the last 3 dB and 10 dB downpoints.

Some statistical parameters that are included in the initial feature selection include the crest factor, which is a function of the peak vibration acceleration, and is a measure of the "peakiness" of the vibration signal:

$$C_r = \frac{a_{max}}{rms} \tag{3.19}$$

where a_{max} is the peak vibration acceleration. Another descriptor which quantifies the peakiness of the vibration signal is the kurtosis, K_t, which is defined as follows:

$$K_t = \frac{1}{T\sigma^4}\sum_{t=0}^{T}[a(t) - \bar{a}]^4 \tag{3.20}$$

where σ is the standard deviation of the vibration acceleration (equivalent to the rms acceleration when the mean of the acceleration signal, $\bar{a}$, is zero). Another statistical descriptor considered by Waddington et al. (2014) as a potential metric to quantify exposure is the skewness, S_k, which is a measure of the temporal distribution of the acceleration signal envelope and is calculated as follows:

$$S_k = \frac{1}{T\sigma^3}\sum_{t=0}^{T}|a(t) - \bar{a}|^3 \tag{3.21}$$

All the above descriptors cover the exposure magnitude, spectral and temporal characteristics of the vibration signal and the use of these descriptors in the building of the logistic regression model should allow any potential differences in freight and passenger vibration signals to be determined, and subsequently utilised in a classification algorithm for unknown vibration signals. Since features in a regression model may have vastly different magnitudes, it is often recommended that the features be normalised in order to make computation easier, and the regression coefficients more interpretable. Due to potential differences in ground conditions and source to receiver distances between measurement positions, each signal feature was normalised against the mean value of the same feature of all event signals recorded at the same control position, with the same ground conditions and source to receiver distances.

3.5.3 Optimising the feature selection

All these features combined, calculated over the measured vibration signals in three orthogonal directions (vertical, North/South and East/West), results in over 100 features being calculated. There is likely to be a great deal of redundancy in this data. As an initial investigation of redundancy, correlation coefficients were determined from the covariance matrix between all pairs of features. Unsurprisingly, the highest and most significant correlations ($p < 0.01$) exist between the three orthogonal directions of each feature. For this reason, and since vibration magnitude is greater in the vertical direction, only features calculated in the vertical direction will be considered, reducing the number of features to 76. The significance of each of these features was then tested using a univariate likelihood ratio test (Equation 3.7). Disregarding features that do not significantly contribute to the model ($p < 0.01$) results in a reduction of the number of features to 17. The features that were retained are: event signal duration, 3 dB envelope duration, 10 dB envelope duration, rise time, fall time and proportional rms acceleration (both W_k weighted and un-weighted) in all 1/3rd octave bands between 2.5 and 8 Hz. A likelihood ratio test of this reduced model against the model of 76 features suggests that that the null hypothesis can be accepted and that the reduced model has no significant reduction in goodness of fit ($p \approx 1$).

Although these 17 features have all been found to significantly contribute to the model when compared with the intercept only model, it is possible that there is further redundancy between the features when considered in a model in combination. A multivariate likelihood ratio test can be performed between a constrained and unconstrained model when adding features to the model. To test this, features are added one by one to the intercept only model, in order of ascending

p-value determined by the univariate significance test, calculating the likelihood ratio between the model with and without each new feature. The results of this test suggests that only the addition of the event signal duration, 10 dB envelope duration and proportional rms acceleration in the un-weighted 5 Hz 1/3rd octave band are significant. A likelihood ratio test of this further reduced model against the previous model of 17 features suggests that the null hypothesis can be accepted and that the reduced model has no significant reduction in goodness of fit ($p = 0.91$).

However, although the addition of the event signal duration and 10 dB envelope duration increases the significance of the fit to the training data, the accuracy of the model, in terms of the F_1 score and the MCC, does not change, suggesting that the model with these features included may be better fit to the training data set but is not any better at classifying new unseen data (i.e. the model may be have been over-fitted). In addition, the correlation between the three duration descriptors is significant ($p < 0.01$), suggesting that multiple descriptors of signal duration may be redundant. Of the three duration descriptors, the event signal duration provides the most significant contribution according to the univariate likelihood ratio test. Therefore, the number of features was reduced to a final number of 2: the event signal duration (T_e) and the proportional energy in terms of the rms acceleration of the 5 Hz 1/3 octave band ($rms_{5\text{Hz}}$). A likelihood ratio test of this fully reduced model against the model of 3 features again suggests that the fully reduced model has no significant reduction in goodness of fit ($p = 0.09$). In addition, this model of two features results in the highest, or at least comparative, accuracy when compared with all other tested models, suggesting that the selection of features for the logistic regression model has been optimised. Reducing the model further, i.e. considering only T_e or $rms_{5\text{Hz}}$ on their own, results in a decrease of the model's accuracy and significance. The likelihood ratio p-value and accuracies of each version of the model are summarised in Table 3.2.

Model	Initial	Univariate Significant	Multivariate Significant	Optimised
Number of Features	76	17	3	2
p-value	0.033	< 0.001	< 0.001	< 0.001
R^2_{pseudo}	0.94	0.92	0.86	0.81
Classification Accuracy	0.92	0.93	0.96	0.96
Precision	0.81	0.85	0.91	0.91
Recall	0.77	0.81	0.88	0.88
F_1	0.79	0.83	0.89	0.89
MCC	0.74	0.79	0.87	0.87

Table 3.2: Goodness of fit and measures of accuracy for four tested logistic regression models with different features

3.6 The optimised logistic regression model

The feature optimised logistic regression model is a function of only two features, one of which quantifies the duration of the vibration event, (T_e), with the other quantifying the proportion of the signal's rms acceleration contained within the 5 Hz 1/3rd octave band ($rms_{5\mathrm{Hz}}$). Both these features are normalised against the mean feature value for all signals recorded at the same measurement position using the same instrument, accounting for distance attenuation and other potential differences in ground conditions between measurement positions. Using only these two features, the logistic regression model is able to correctly classify, on average, 96% of all signals tested. The mean precision of the model is 0.91, meaning that 91% of the signals that are classified as freight railway signals truly are freight railway signals and the mean recall of the model is 0.88, meaning that 88% of all the freight railway signals are correctly classified as such. When considering passenger as the positive class, the precision and recall are both equal to 0.98.

With the parameters of the model defined, the logistic regression was performed using all 238 known examples of freight and passenger vibration signals, in order to determine the regression coefficients. The result of this logistic regression is shown in Equation 3.22, where $P(Y = 1)$ is the predicted probability that a vibration signal with normalised event signal duration, T_e, and normalised proportional 5 Hz rms acceleration, $rms_{5\text{Hz}}$, is a freight railway vibration signal, i.e. its class, Y, is equal to 1. Further details of the logistic regression model are presented in Table 3.3, where standard errors are calculated using Equation 3.8 and the parameter p-values are determined from the Wald test statistic (Equation 3.9).

$$P(Y = 1) = \frac{1}{1 + \exp(10.7 - 5.01T_e - 2.25rms_{5\text{Hz}})} \tag{3.22}$$

Parameter	β Estimate	Standard Error	p-value	Overall Model	
Intercept	-10.73	1.76	< 0.001	N	238
T_e	5.01	0.89	< 0.001	p-value	< 0.001
$rms_{5\text{Hz}}$	2.25	0.70	< 0.010	R^2_{pseudo}	0.79

TABLE 3.3: Parameter estimates and other details of the optimised logistic regression model

Figure 3.5 shows all of the 238 known examples of freight and passenger vibration signals plotted in the two-dimensional feature space of T_e and $rms_{5\text{Hz}}$. Also shown is the decision boundary for which $P(Y = 1) = 0.5$ and above which signals are classified as freight vibration signals by the logistic regression model. With this fit of the regression model to the data, 4 passenger railway signals and 4 freight railway signals exist in the wrong prediction regions and would be incorrectly classified if introduced to the model as unlabelled signals. However, the remaining 190 passenger railway signals and 40 freight railway signals exist in the correct region and would be correctly classified if introduced to the model as unlabelled signals. This is commensurate with the reported 96% accuracy of the model when tested on unlabelled data completely independent to that used for the training, fitting and optimizing of the model. Most passenger railway vibration signals are clustered together in a region of low event signal duration and low proportional 5 Hz 1/3rd octave band energy. The freight railway vibration signals show more variation, but tend to have longer event signal durations and greater proportional rms acceleration in the 5 Hz 1/3rd octave band, allowing these signals to be classified with confidence using only these two signal properties.

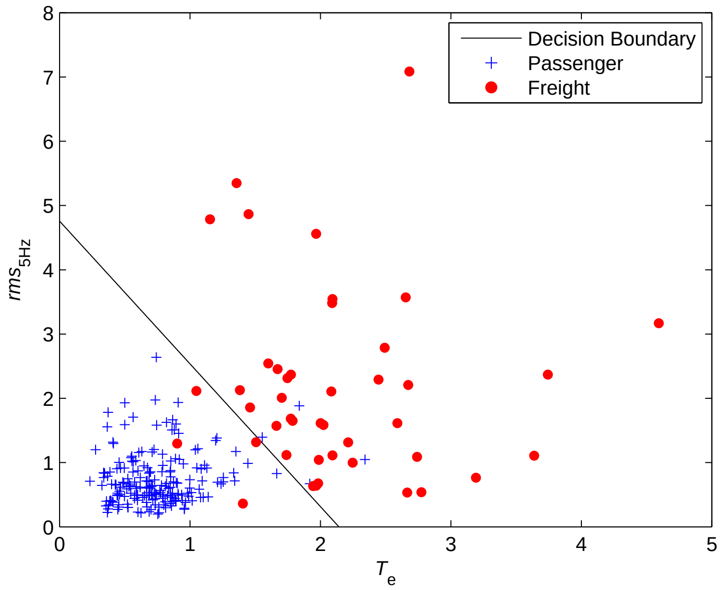

FIGURE 3.5: Decision boundary of logistic regression model as a function of the normalized event signal duration (T_e) and normalized proportional 5 Hz 1/3rd octave band rms acceleration ($rms_{5\text{Hz}}$)

As a final check of the model accuracy, 500 vibration signals were randomly selected from the field study measurements of Waddington et al. (2014) and their class was predicted using the logistic regression model. These vibration signatures were visually inspected by the author who judged their class, based on previous experience, and made the same class judgements as the logistic regression model for 94% of the vibration signals inspected.

The optimised logistic regression model has been shown to be able to correctly classify, on average, 96% of unknown railway vibration signals that are completely independent of the training, fitting and optimising of the model. In addition, both features of the optimised logistic regression model are normalised to other signals recorded at the same position using the same instrument. This allows the model to be more applicable to other sets of data in the measurement database than if absolute properties were used. For example, it avoids the problem of misclassification occurring because all railway traffic moves slowly close to a measurement position, perhaps due to a proximity to a station or tight bend, resulting in longer passbys relative to signals measured at other measurement positions. However, the model would not be applicable to freight only railway lines, as this would significantly skew the normalised values of T_e and $rms_{5\text{Hz}}$. This is not a cause for concern in this work as no measurements were made near freight only railway lines. In the extreme case where there are no freight pass bys in a 24 hour measurement period, each signal property will be of similar magnitude to the mean properties of the passenger passbys (since all passbys will be passenger traffic) and all the signals will cluster around $T_e = rms_{5\text{Hz}} = 1$ and should mostly be correctly classified. The high level of classification accuracy and the normalised nature of the features of the model suggests that the model can be confidently applied to the remaining data from the field study by Waddington et al. (2014) in order to classify the unknown signals and determine exposure-response relationships for annoyance caused by exposure to freight and passenger railway vibration separately. This will be useful in furthering the understanding of the human response to freight railway vibration in light of current proposals to increase the proportion of freight traffic on rail.

3.7 Validation of model with new measurement data

During this research, new measurements of railway vibration were performed as part of a laboratory study on the perception of combined railway noise and vibration (see Chapters 5 and 6 for more detail on these measurements and the laboratory study). As these were observed measurements, they resulted in a set of known freight and passenger vibration signals which can be used to validate the logistic regression model derived in this chapter. During these new measurements, 53 passenger and 9 freight vibration signals were recorded. Applying the logistic regression model (Equation 3.22) to these measurements resulted in 98% of the signals being correctly classified, with a precision of 90% and a recall of 100%. The two dimensional feature space of the new measurements, along with the decision boundary of the logistic regression model, is shown in Figure 3.6.

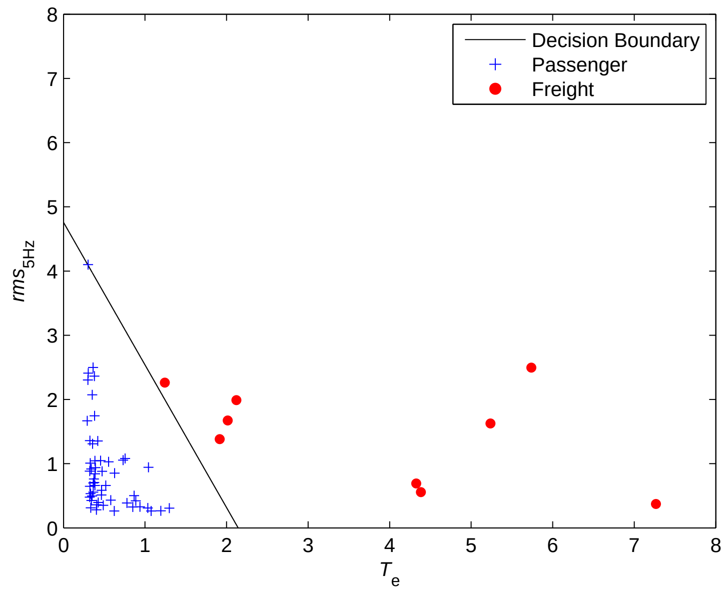

FIGURE 3.6: New measurement data plotted in the two dimensional feature space of the logistic regression model

Once again, Figure 3.6 shows the passenger vibration signals mostly clustered in a region of low event signal duration and low proportional 5 Hz *rms* acceleration. Again the freight vibration signals exhibit more variation, but tend to have higher than average durations and 5 Hz *rms* energy, allowing all of these signals to be correctly classified in this case. The only signal that is incorrectly classified is a passenger train signal that lies almost exactly on the decision boundary ($P(Y) = 0.507$), meaning that it is only just incorrectly classified. This is due to its high normalised proportional 5 Hz *rms* acceleration. Its $rms_{5\text{Hz}}$ value of 4.10 is higher than that of all the new measurements, including the freight signals, and higher than that of all the passenger signals used to build the logistic regression model (see Figure 3.5), suggesting that this signal may in fact be an outlier.

Regardless of the incorrectly classified signal, the predictions of the logistic regression model on these new measurements is still extremely high, with a classification accuracy of 98% and a precision and recall of 90% and 100% respectively. This is a very encouraging result as it gives further confidence that the logistic regression model can be successfully applied to unknown railway vibration signals.

3.8 Summary

In this chapter, it has been demonstrated that the annoyance response to freight and passenger railway vibration, collected during the field study of Waddington et al. (2014), is significantly different. In order to investigate these differences, it is necessary to determine separate exposures for freight and passenger railway vibration to pair with the separate responses collected during the field study. This can be achieved through the use of logistic regression model for classification of unknown railway vibration signals in the measurement database.

The logistic regression model has been developed using a number of known examples of freight and passenger railway vibration signals measured during the field study. Using a combination of univariate and multivariate significance testing, testing of correlation between features and analysis of classification accuracy, the model has been optimised to contain only two features. The two features quantify the duration and low frequency energy content of each railway vibration signal and the optimised model is able to classify, on average, 96% of known railway vibration signals from the field study measurements. A validation of the model on new measurements showed an even higher classification accuracy of 98%.

The high classification accuracy of the logistic regression model gives confidence that it can be applied to unknown vibration signals in the database of field study measurements, in order to classify each measured railway vibration signal as either freight or passenger. This allows the determination of separate source exposures, which in turn allows the development of exposure-response relationships for annoyance caused by exposure to freight and passenger railway vibration in residential environments. The development of these separate source exposure-response relationships is presented in Chapter 4.

Chapter 4

Exposure-response relationships for annoyance due to exposure to freight and passenger railway vibration

4.1 Introduction

In Chapter 3, it is demonstrated that in the field study of Waddington et al. (2014), the annoyance response due to freight and passenger railway exposure was significantly different. In order to further investigate this difference, exposure-response relationships are developed for the annoyance due to freight and passenger railway vibration in residential environments. This is made possible by applying the logistic regression classification model, the development of which is detailed in Chapter 3, to unknown railway vibration signals in the field study measurement database. With the unknown railway signals classified as freight or passenger signals using this method, it is possible to determine separate freight and passenger vibration exposures for each case study in the database. These exposures can then be paired to the responses collected in the field study in order to develop exposure-response relationships. The development of these exposure-response relationships is described in this chapter.

4.2 Determining freight and passenger railway vibration exposure

The data used to develop the separate exposure-response relationships comes from vibration exposure measurements and questionnaire responses collected during the field study of Waddington et al. (2014). In this section, the measurement protocol applied during this field study is summarised, and the methods used to determine separate exposures for freight and passenger railway vibration is presented.

4.2.1 Field study measurement protocol

In the field study performed by Waddington et al. (2014), vibration exposure was determined by measurement. The vibration measurement protocol for the field study involved long term vibration monitoring at external control positions, for a period of at least 24 hours. Where possible, the instruments for the control position measurements were placed were they were unlikely to be interfered with, such as in a shed or a garage. Whilst these long term measurements were taking place, short term, time synchronised, internal measurements were conducted within the properties of residents who had agreed to take part in the study. A diagram illustrating this measurement setup is reproduced here in Figure 4.1.

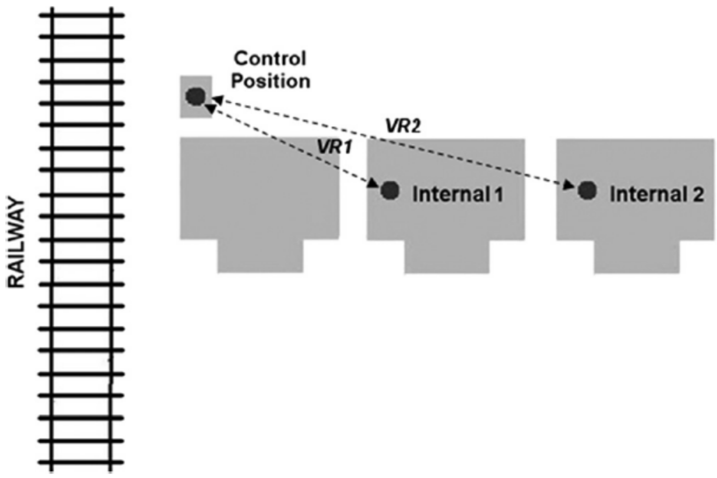

FIGURE 4.1: Schematic of measurement approach, showing control position and internal measurements (Source: Waddington et al., 2014)

The short term measurements were typically around 30 minutes in duration, or for a period covering approximately 5 to 10 train passes. For the internal measurements, the measurement was taken as close as possible to the centre of the room in which the resident stated that they could feel the strongest magnitude of vibration from the railway, as recommended by the Association of Noise Consultants (2012). The transmissibility between the paired control position and internal measurements were used to estimate 24-hour vibration time histories within dwellings. In cases where a measurement of internal vibration was either not conducted or unavailable due to data corruption, the internal vibration exposure was assumed to be equivalent to that which had been successfully determined for a similar property type in the same measurement area and at a similar distance from the railway. Using this process, it was possible to estimate 24 hour internal vibration exposures for 752 properties, 497 of which were based on the transmissibility method with the remaining 255 exposures based on estimations of internal vibration in a similar property type. Sica et al. (2014) estimate that the uncertainty associated with this method is ±2.2 dB when internal and external measurements were performed and ±11.4 dB when only external measurements were performed. The equipment used for the vibration measurements were Guralp CMG-5TD strong motion accelerometers with a sampling rate of 200 Hz and a 100 Hz low pass filter, capable of measuring vibration in three orthogonal directions (vertical, North/South and East/West). For more detailed information on the measurement protocol, and the methods used to estimate the vibration exposures, see Sica et al. (2014).

4.2.2 Railway vibration event identification

Railway vibration events (i.e. train passes) were identified by Waddington et al. (2014) from the 24 hour control position time histories using a process based on a short time average/long time average (STA/LTA) algorithm. The STA/LTA algorithm is an event identification process commonly used in seismology and is based on the ratio, $R_{S/L}$, between short term averaging and long term averaging of a vibration acceleration time history:

$$R_{S/L} = \frac{\frac{1}{T_S}\sum_{t=0}^{T_S} a(t)^2}{\frac{1}{T_L}\sum_{t=0}^{T_L} a(t)^2} \tag{4.1}$$

where T_S and T_L are the widths of the short term and long term averaging windows respectively and $a(t)$ is the acceleration time history. Waddington et al. (2014) found that window lengths of $T_S = 1$ s and $T_L = 15$ s, and a triggering threshold of $R_{S/L} = 0.80$ were effective parameter values for correctly identifying train passes. They also employed the use of the crest factor (Equation 3.19) to analyse triggered signal identifications, rejecting events with a crest factor greater than 10, as these tended to be clusters of short term vibration transients (such as several seconds of footfalls). The STA/LTA algorithm, combined with the crest factor analysis and an integrity check, in which triggered events were manually inspected and rejected if the data was deemed to be contaminated, resulted in a highly accurate automated event identification system. For more information on this event identification process, see Sica et al. (2011).

Applying this identification system across all the 24 hour acceleration time histories for the 752 case studies allowed the identification of 99 997 railway passby events. It is this database of unknown railway vibration events that must be classified as either freight or passenger events in order to determine separate exposures and hence separate exposure-response relationships for these two sources.

4.2.3 Classification of freight and passenger vibration signals

99 997 railway passbys were identified by Waddington et al. (2014) using the methods outlined above. In order to classify each of these unknown railway events as either freight or passenger railway passbys, the logistic regression classification algorithm, the development of which is outlined in Chapter 3, was applied to each event. The logistic regression classification model has a reported classification accuracy of 96% and is defined as follows:

$$P(Y=1) = \frac{1}{1+\exp(10.7 - 5.01T_e - 2.25rms_{5\mathrm{Hz}})} \tag{4.2}$$

where T_e is the normalised event signal duration and $rms_{5\mathrm{Hz}}$ is the normalised proportional rms acceleration contained within the 5 Hz 1/3rd octave band. These two features were calculated for each of the unknown railway event signals, and then substituted into Equation 4.2 in order to determine a predicted probability that the event is a freight railway passby (as opposed to a passenger railway passby). All signals with a predicted probability greater than 0.5 were subsequently classified as freight railway passbys, with the remaining signals classified as passenger railway passbys. This resulted in 14 874 signals being classified as freight with the remaining 85 103 signals classified as passenger. Figure 4.2 shows an example of two signals that have been classified as freight and passenger vibration signals. Visually, these examples show a great deal of similarity to examples of known freight and passenger vibration signals, as shown in Figure 3.4, once again providing confidence that the logistic regression classification model is able to correctly identify freight and passenger vibration signals.

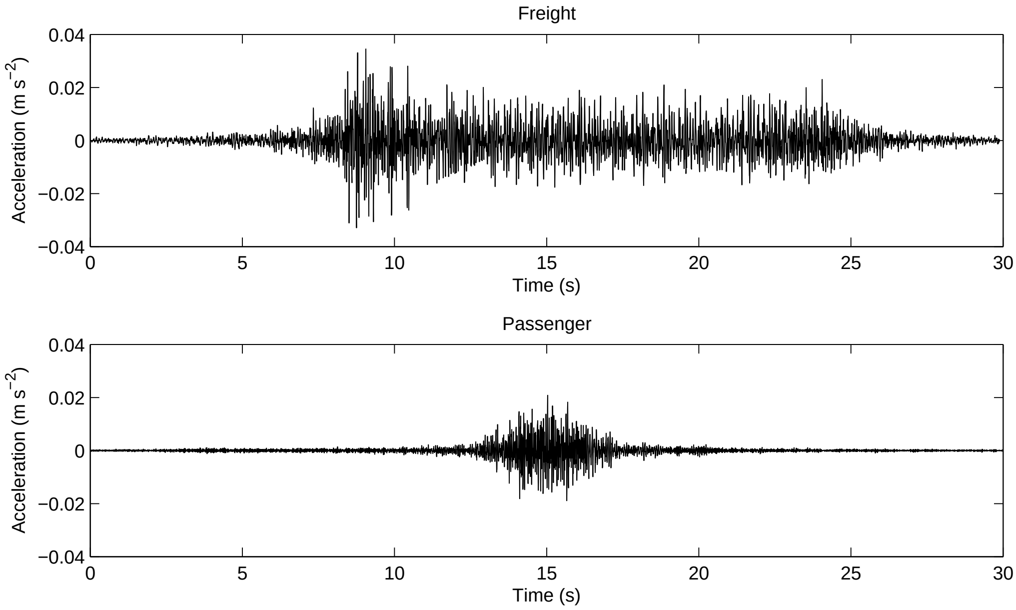

FIGURE 4.2: Examples of unknown signals that have been classified as freight and passenger using the logistic regression classification model

4.2.4 Determining separate freight and passenger railway vibration exposures

With all the railway vibration signals in the measurement database classified as either freight or passenger vibration signals, it is possible to determine 24 hour vibration exposures for these two sources, for each of the 752 case studies. In terms of quantifying the 24-hour vibration exposures, several different vibration descriptors were investigated during the field study by Waddington et al. (2014). Descriptors investigated included the root mean square acceleration, the vibration dose value, the equivalent continuous vibration level, the vibration exposure level, the standard deviation, the skewness, the kurtosis and the peak particle acceleration, all of which are summarised in Section 3.5.2. In investigating these parameters, Waddington et al. (2014) found similar magnitudes of correlation to exist between all descriptors and the self reported annoyance (the only exceptions being the skewness and the kurtosis). They concluded that, for their dataset of exposures and responses, the single figure descriptors that they investigated were equally effective predictors of annoyance, a conclusion that is consistent with the findings of Zapfe et al. (2009) in their field study. However, Waddington et al. (2014) did find a marginal improvement in the magnitude and significance of correlation when applying appropriate frequency weightings to the root mean square acceleration and vibration dose value descriptors. These findings led them to report their exposures using both the weighted root mean square acceleration and weighted vibration dose value. Similar correlations with different descriptors were found in this research for separate freight and passenger vibration exposures and responses.

Though vibration exposure in the field study was measured in three orthogonal directions, the vibration magnitude dominates in the vertical direction. British Standard BS 6472-1:2008 suggests that when the magnitude of vibration is dominant in one axis, only the direction with the highest magnitude need be considered when quantifying the exposure. For this reason, and those outlined above, 24 hour vibration exposures will be quantified in this research using the vertical vibration, W_k weighted for root mean square acceleration (as recommended by BS ISO 2631-1:1997) and W_b weighted for vibration dose value (as recommended by BS 6472-1:2008). The 24 hour W_k weighted root mean square (rms) acceleration is defined as follows:

$$rms_{k,24hr} = \sqrt{\frac{1}{T}\int_0^T a_{k,24hr}(t)^2 dt} \tag{4.3}$$

where $a_{k,24hr}(t)$ is a W_k weighted acceleration time history of duration T, composed of all passbys in a 24 hour period. The 24 hour W_b weighted vibration dose value (VDV) is defined as follows:

$$\mathrm{VDV}_{b,24hr} = \sqrt[4]{\int_0^T a_{b,24hr}(t)^4 dt} \qquad (4.4)$$

where $a_{b,24hr}(t)$ is a W_b weighted acceleration time history of duration T, composed of all passbys in a 24 hour period. For more information on these descriptors, and the standards which recommend their use, see Section 2.3.3. Using Equations 4.3 and 4.4, and the signals classified as freight and passenger by the logistic regression classification model (Equation 4.2) it is possible to determine separate 24 hour vibration exposures for freight and passenger railway vibration, in terms of weighted *rms* and VDV, for each case study.

4.3 Determining the annoyance response to freight and passenger railway vibration

In Section 4.2, the methods used to determine separate exposure data for 24 hour exposure to freight and passenger railway vibration is presented. In order to determine exposure-response relationships for these two sources of railway vibration, the annoyance response due to the two sources must also be determined. In this section, the methods used to determine the annoyance response to freight and passenger railway vibration is presented.

4.3.1 Field study social survey questionnaire

In the field study of Waddington et al. (2014), the human response to various environmental factors were collected via means of a social survey questionnaire, conducted face-to-face with residents in their own homes, usually whilst the short term measurements of internal vibration were being conducted. To avoid biasing response to questions on noise and vibration, the social survey was presented as a neighbourhood satisfaction survey, collecting data not only on noise and vibration but also on general satisfaction with the home and neighbourhood. The survey was also designed to measure a variety of factors which may affect the response to noise and vibration, such as self-reported sensitivity to noise and vibration as well as situational, attitudinal and demographic factors. For a brief summary of the effects of these non-exposure factors, see Section 2.5.1.

The social survey questionnaire was developed by a team of social scientists and is based on a pilot questionnaire developed for the project (TRL et al., 2007), the Nordtest method for the development of socio-vibration surveys (Nordtest Tekniikantie, 2001), best practice guidelines for the measurement of annoyance due to noise provided by Team 6 of the International Commission on the Biological Effects of Noise (ICBEN) (Fields et al., 2001) and guidance from the International Standard ISO/TS 15666:2003. A summary of considerations in the development of the social survey is provided by Whittle et al. (2015).

For field studies on the community response to environmental noise and vibration, the response is usually measured in terms of annoyance, with this type of response seen as an overall concept for the negative evaluations of environmental conditions (Guski et al., 1999). Therefore, the primary aim of the survey was to measure self-reported annoyance due to various environmental conditions, including railway noise and vibration. It is this measurement of annoyance response due to railway vibration that is relevant to this research. For more information on the use of annoyance as a quantifiable measurement of response, see Section 2.4.1.

4.3.2 Determining response to freight and passenger railway vibration

For this research, the response measure of interest is the annoyance caused by vibration from freight and passenger railway traffic. As part of the social survey, respondents were asked the following question:

"Thinking about the last 12 months or so, when indoors at home, how bothered, annoyed or disturbed have you been by feeling vibration or shaking or hearing or seeing things rattle, vibrate or shake caused by [source]?"

The above question was asked four times, with the source term being replaced by "passing passenger trains", "passing freight trains", "railway maintenance" or "other railway activity". Following the guidance of ICBEN (Fields et al., 2001) and ISO/TS 15666:2003, the answer to the above question was recorded on a five point semantic scale ("not at all", "slightly", "moderately", "very" or "extremely") and an 11 point numeric scale (0 to 10). Respondents who reported that they had not felt any vibration that they thought was caused by the railway had their annoyance responses recorded as the lowest category on the semantic and numeric scales.

For the exposure-response relationships developed in this chapter, the response of the above question, where the source is either "passing passenger trains" or "passing freight trains" and the response is recorded using the semantic scale, was collected for each case of the 752 case studies in order to quantify the annoyance response to freight and passenger railway vibration.

4.4 Exposure-response relationships derived from an ordinal probit grouped regression model

Sections 4.2 and 4.3 show how separate exposures and separate responses can be determined for freight and passenger railway vibration. Using these methods, separate exposures and responses for freight and passenger vibration were determined for all 752 case studies in the field study of Waddington et al. (2014). Exposure-response relationships can then be determined for these separate sources using a statistical grouped regression model presented by Groothuis-Oudshoorn and Miedema (2006) and adapted by Waddington et al. (2014). This section presents the details of the grouped regression model, as well as the methods used to test the goodness of fit and significance of the model. Exposure-response relationships are then determined for both sources combined, with the use of a dummy source variable, and for separate sources.

4.4.1 The ordinal probit grouped regression model

The ordinal probit grouped regression model takes the form of a curve indicating the percentage of respondents expressing annoyance above a certain threshold, in a similar manner to many of the exposure-response relationships presented in the literature for exposure to noise and vibration and summarised in Sections 2.4 and 2.5 respectively. The model is derived from vibration exposure data and self reported annoyance recorded on a scale ranging from 0 to 100. Annoyance response scales with a certain number of categories, either semantic or numeric, are re-scaled to this range using the following relation:

$$\tau_j = 100j/m \tag{4.5}$$

where τ_j are the category cut-points, j is the rank number of each category (the lowest category is assigned a value of 0) and m is the total number of categories. The annoyance data, A can then be centred on the midpoints of these categories. For the data utilised in this work, self reported annoyance, A was recorded on an ordinal semantic scale with $J = 5$ categories. It can be assumed that a latent variable A^*, a linear combination of vibration exposure ($\mathbf{V}$) and a random error component ε, underlies the categorical annoyance variable as follows:

$$A^* = \boldsymbol{\beta}\mathbf{V} + \varepsilon \tag{4.6}$$

where $\boldsymbol{\beta}$ is a vector of regression parameters. The latent variable A^* is then linked to the observed variable A, with boundaries of 0 and 100, using the following relationship:

$$A = \begin{cases} 0 & \text{if } A^* < 0 \\ A^* & \text{if } A^* \in [0, 100] \\ 100 & \text{if } A^* > 100 \end{cases} \tag{4.7}$$

Community responses of annoyance are often represented as the proportion of a population that is likely to respond above a certain level of annoyance, C. Three commonly used values of C are $C = 28$ (percent slightly annoyed), $C = 50$ (percent annoyed) and $C = 72$ (percent highly annoyed) (Miedema and Oudshoorn, 2001). The probability of an individual exposed to a certain magnitude of vibration exposure (V) responding with an annoyance level above a cut-off C can be expressed using the following relations:

$$
\begin{aligned}
P(V) &= \text{Prob}(A^* > C) \\
&= \text{Prob}(\boldsymbol{\beta}\mathbf{V} + \varepsilon \geq C) \\
&= \text{Prob}(\varepsilon \geq C - \boldsymbol{\beta}\mathbf{V})
\end{aligned}
\tag{4.8}
$$

Assuming that the error term ε is normally distributed, Equation 4.8 becomes:

$$
P(V) = \left(1 - \Phi\left[\frac{C - \boldsymbol{\beta}\mathbf{V}}{\sigma_{SE}}\right]\right) \tag{4.9}
$$

where Φ represents the cumulative normal distribution function and σ_{SE} represents the standard error. The distribution of responses at different annoyance levels can be expressed by varying the cut-off, C. The regression coefficients of this model can be estimated using maximum likelihood with the following likelihood equation:

$$
L(\boldsymbol{\beta}) = \prod_{j=1}^{J} \prod_{y_i=j} [\Phi(\tau_j - \mathbf{V_i}\boldsymbol{\beta}) - \Phi(\tau_{j-1} - \mathbf{V_i}\boldsymbol{\beta})] \tag{4.10}
$$

where $\mathbf{V_i}$ is a vector of exposures that result in response y_i. The upper and lower 95% confidence intervals as a function of exposure are determined as:

$$
C_{95}(V) = \boldsymbol{\beta}\mathbf{V} \pm Z\sqrt{\mathbf{V}\mathbf{E_b}\mathbf{V^T}} \tag{4.11}
$$

where $\mathbf{E_b}$ is the covariance matrix of the $\boldsymbol{\beta}$ parameters, $\mathbf{V}^{\mathrm{T}}$ is the transpose of the exposure vector $\mathbf{V}$ combined with an intercept term and $Z = 1.96$ for a standard normal distribution. The confidence limits for the exposure-response relationship can then be represented as:

$$P_{95}(V) = \left(1 - \Phi\left[\frac{C - C_{95}}{\sigma}\right]\right) \tag{4.12}$$

4.4.2 Goodness of fit and significance of the grouped regression model

For an indication of how well the exposure-response model fits the data, the goodness of fit and significance of the model can be tested, in a similar manner to that of testing the goodness of fit and significance of the logistic regression classification model (see Section 3.4.3). Once again, the R^2 value, commonly used to assess the goodness of fit for ordinary least squares regression, is not appropriate for models that are determined using maximum likelihood rather than minimisation of variance. Instead, many similar indicators have been developed in an attempt to describe the goodness of fit of a regression model estimated with maximum likelihood. Similarly to the logistic regression model, Mcfadden's R^2_{pseudo} value (Equation 3.6) will be used to assess the goodness of fit of the grouped regression model. The overall significance of the model can be calculated using the likelihood ratio test statistic (Equation 3.7), the standard error of individual parameters can be calculated from the covariance matrix (Equation 3.8) and the significance of individual parameters can be calculated using the Wald ratio test statistic (Equation 3.9).

4.4.3 Exposure-response relationship with a dummy variable for source type

Before deriving separate exposure-response relationships for exposure to freight and passenger railway vibration, the applicability of using separate exposure-response relationships can be investigated by introducing a dummy variable to the exposure-response model. This method of introducing a dummy variable to the exposure-response relationship has previously been used by Klæboe et al. (2003b) to investigate different vibration sources, differences in study regions and different distances from the vibration source. Waddington et al. (2014) also used the dummy variable method to investigate different vibration sources, concluding that different exposure-response relationships should be derived for vibration due to construction and for vibration due to the railway.

For the dummy variable analysis, separate exposures and responses to freight and passenger railway vibration from all 752 case studies were pooled and a dummy variable for source type was created (passenger = 0, freight = 1). An exposure response relationship was created, using the five point semantic annoyance scale as the dependent variable and vibration exposure and the source type dummy variable as the independent variables. For both exposure quantified as 24 hour weighted *rms* acceleration and VDV, the fitting of this exposure-response relationship resulted in a highly significant parameter estimate for the source dummy variable ($p < 0.001$) and a significant increase in the likelihood ($p < 0.001$) when compared to models fitted without the dummy variable. Figure 4.3 shows plots of the exposure-response relationships with the dummy variable, for both freight railway exposure (source type = 1) and passenger railway exposure (source type = 0), whilst parameter estimates and other details of the pooled exposure-response relationships with the dummy variables are presented in Table 4.1.

The positive parameter value for the source type dummy variable has the effect of shifting the freight annoyance response along the vibration exposure axis, suggesting that freight vibration exposure results in higher annoyance responses when compared with the same magnitude of passenger vibration exposure. This result gives confidence that the difference in the response to these two sources is not just due to differences in the magnitude of vibration exposure, and that it is justifiable to derive separate exposure-response relationships for freight and passenger sources.

Parameter	β Estimate	Standard Error	p-value	Overall Model	
Intercept	41.22	12.15	< 0.001	N	1504
$10\log_{10}(rms_{k,24hr})$	2.47	0.46	< 0.001	p-value	< 0.001
Source Type	15.81	3.57	< 0.001	R^2_{pseudo}	0.02
σ_{SE}	52.24	2.28	< 0.001		
Intercept	11.37	7.03	n.s.	N	1504
$10\log_{10}(\mathrm{VDV}_{b,24hr})$	2.36	0.44	< 0.001	p-value	< 0.001
Source Type	19.24	3.62	< 0.001	R^2_{pseudo}	0.02
σ_{SE}	52.28	2.27	< 0.001		

TABLE 4.1: Parameter estimates and other details of the pooled ordinal probit grouped regression model with a dummy source variable

4.4.4 Exposure-response relationships for separate sources

The positive and significant regression coefficient for the dummy source variable gives confidence that separate exposure response relationships can be determined for exposure to freight and passenger railway traffic. Therefore, separate exposure-response relationships are derived for these two sources of railway vibration. These separate exposure response relationships, for freight and passenger railway vibration, and for vibration quantified as 24 hour *rms* acceleration and VDV, are shown in Figure 4.4. Parameter estimates and other model details are presented in Table 4.2 for models derived for freight and passenger vibration, and for exposures quantified as *rms* acceleration and VDV.

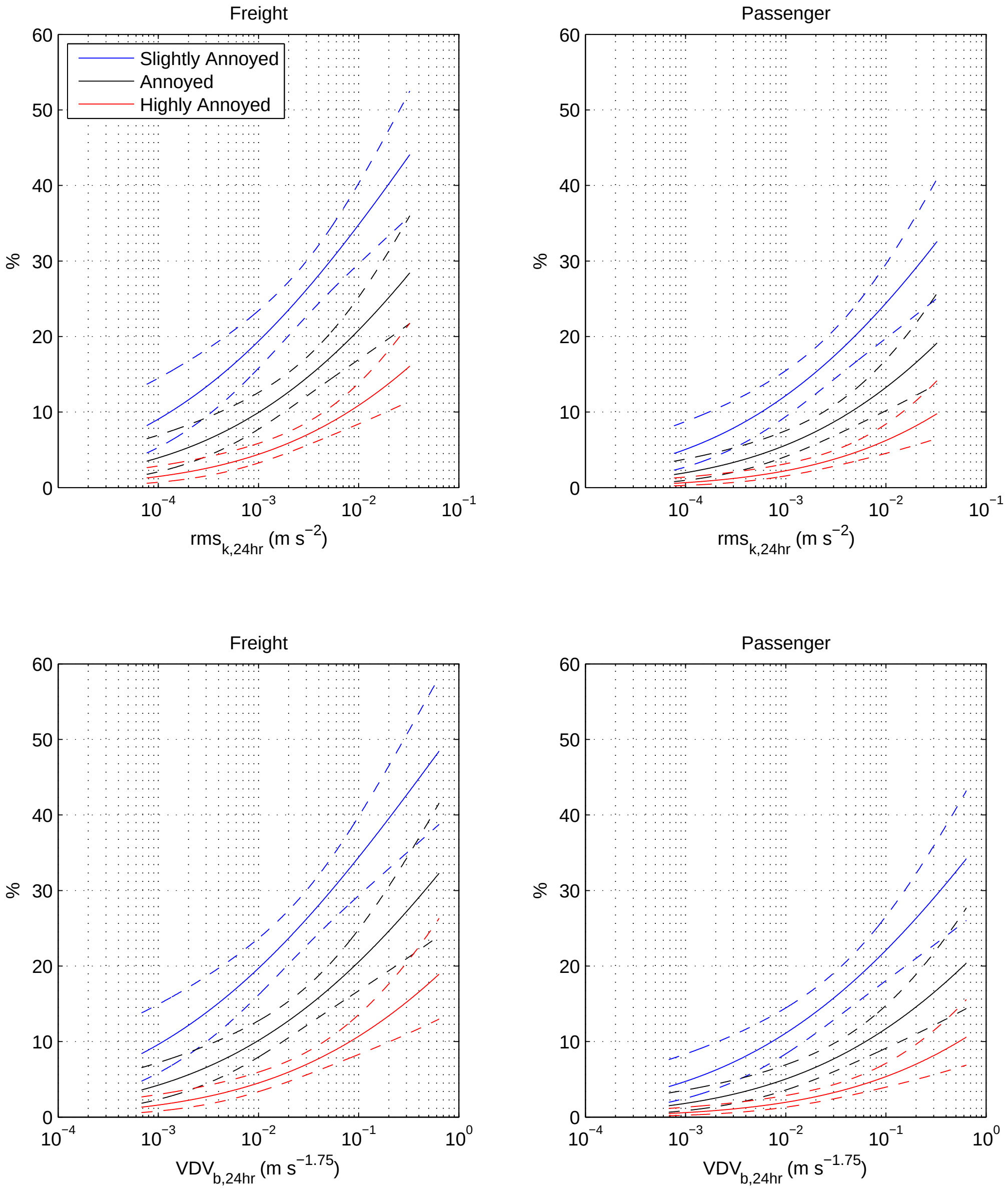

FIGURE 4.3: Exposure-response relationship derived using the ordinal probit grouped regression model, with a dummy source type variable for freight exposure (source type = 1) and passenger exposure (source type = 0). Broken lines indicate the 95% confidence intervals.

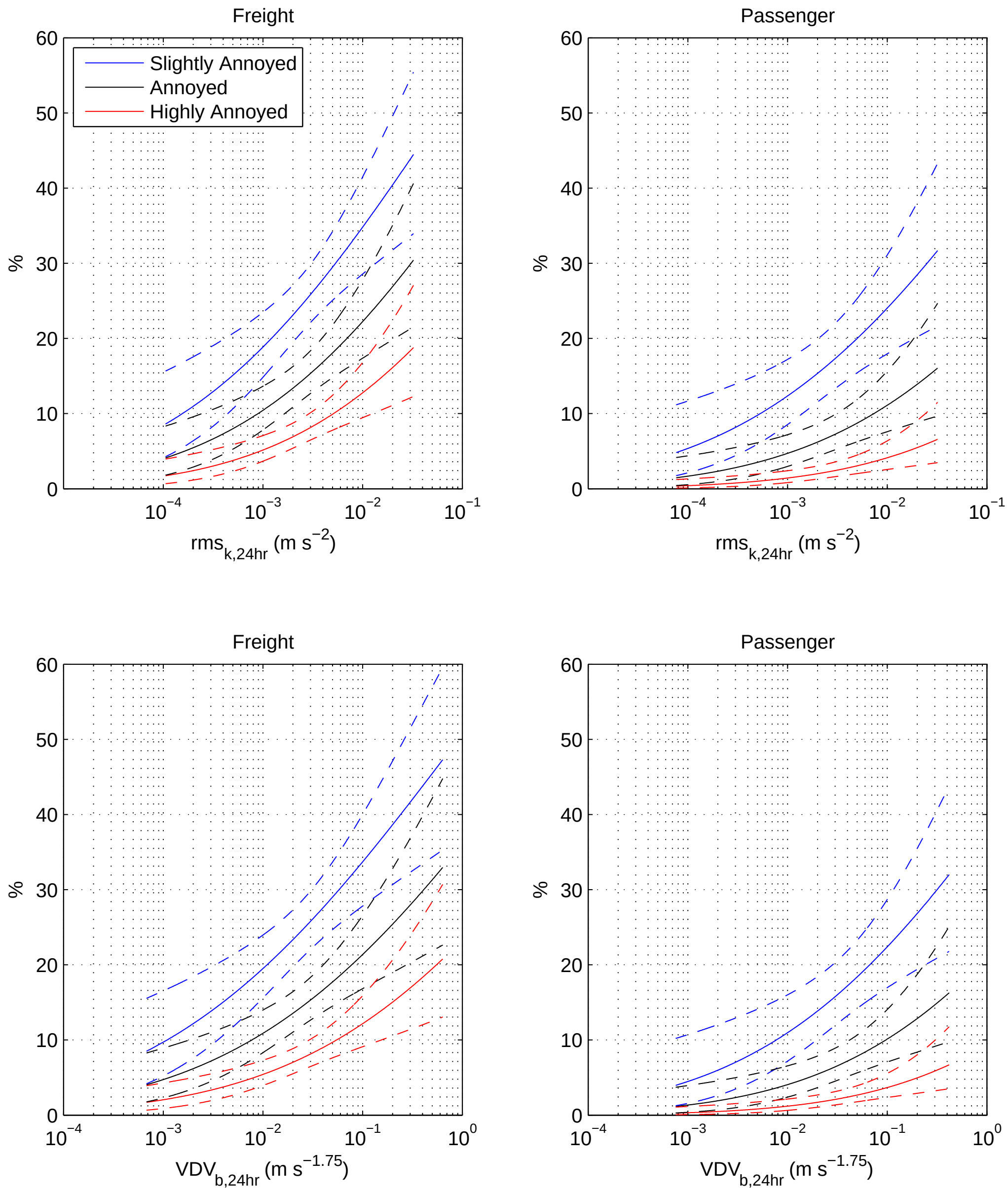

FIGURE 4.4: Exposure-response relationships derived using separate ordinal probit grouped regression models for annoyance due to exposure to freight and passenger railway traffic. Broken lines indicate the 95% confidence intervals.

	Parameter	β Estimate	Standard Error	p-value	Overall Model	
Freight	Intercept	63.06	18.29	< 0.001	N	752
	$10\log_{10}(rms_{k,24hr})$	2.90	0.70	< 0.001	p-value	< 0.001
	σ_{SE}	58.90	3.19	< 0.001	R^2_{pseudo}	0.01
Freight	Intercept	29.07	10.47	< 0.010	N	752
	$10\log_{10}(\mathrm{VDV}_{b,24hr})$	2.59	0.63	< 0.001	p-value	< 0.001
	σ_{SE}	58.97	3.19	< 0.001	R^2_{pseudo}	0.01
Passenger	Intercept	36.52	14.37	< 0.050	N	752
	$10\log_{10}(rms_{k,24hr})$	1.93	0.56	< 0.001	p-value	< 0.010
	σ_{SE}	42.66	3.09	< 0.001	R^2_{pseudo}	0.01
Passenger	Intercept	15.62	8.68	$n.s.$	N	752
	$10\log_{10}(\mathrm{VDV}_{b,24hr})$	2.00	0.57	< 0.001	p-value	< 0.010
	σ_{SE}	42.68	3.07	< 0.001	R^2_{pseudo}	0.01

Table 4.2: Parameter estimates and other details of the ordinal probit grouped regression models for annoyance due to exposure to freight and passenger railway vibration

For comparison, the exposure-response relationships for percentage highly annoyed due to both freight and passenger railway vibration exposure are presented together in Figure 4.5. The exposure-response relationships show that freight railway vibration results in a greater annoyance response, even for equal levels of vibration exposure. For example, given a 24 hour rms_k vibration exposure of 0.01 m s^{-2}, approximately 4% of the studied population is likely to be highly annoyed if the source is passenger railway vibration, whereas approximately 13% of the studied population is likely to be highly annoyed if the source is freight railway vibration. In terms of equal annoyance response, 0.0100 m s^{-2} of passenger railway vibration exposure is equivalent to only 0.0006 m s^{-2} of freight railway vibration.

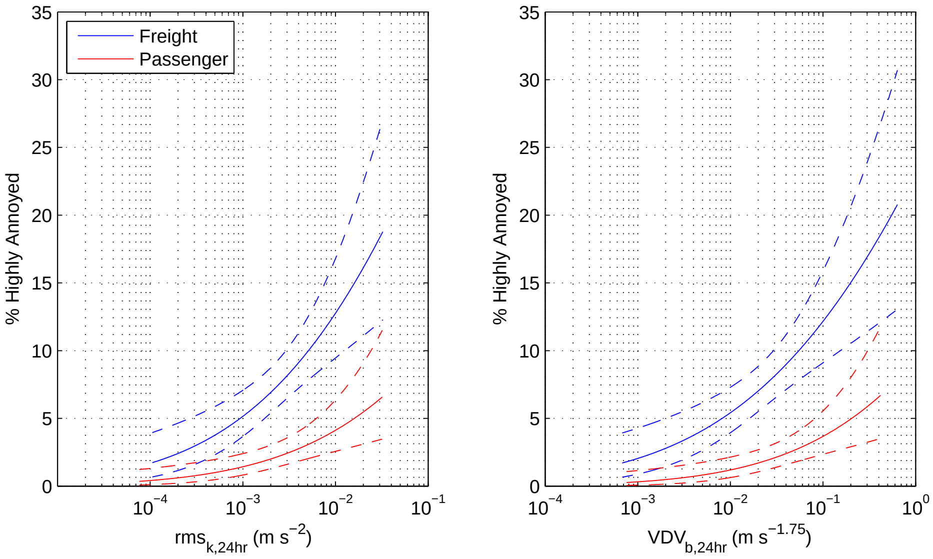

FIGURE 4.5: Exposure-response relationships derived separately for freight and passenger exposure using the ordinal probit grouped regression model, showing the percentage of high annoyance caused by exposure to freight and passenger railway vibration. Broken lines indicate the 95% confidence intervals.

Fidell et al. (2011) provide a method of quantifying differences in community response to different noise sources using a community tolerance level (CTL). For more information on the CTL, and its use in quantifying community responses to different sources, see Section 2.4.3. Fidell et al. (2011) define the CTL as the noise level at which 50% of a community population is likely to be highly annoyed. Applying this method to vibration and to the population studied results in a CTL level of 78 dB for passenger railway vibration and 63 dB for freight railway vibration (re 10^{-6} m s^{-2}). In other words, the population studied appears to be 15 dB more tolerant to passenger railway vibration than they are to freight railway vibration.

4.5 Exposure-response relationships derived with a cumulative ordinal logit model

The exposure-response relationships derived in the previous sections have shown that the human response to freight and passenger railway vibration is significantly different. This has been demonstrated with the use of a dummy source variable in a grouped regression model, showing a significant shift along the exposure axis to account for the difference in the response to these two sources of environmental vibration (see Section 4.4.3). The difference in response is further demonstrated by developing separate exposure response relationships for exposure to freight and passenger railway vibration (see Section 4.4.4). An additional method to investigate the difference in response is to develop exposure-response relationships using a cumulative ordinal logit model with a source dummy variable. The parameter estimate of the source dummy variable can then be interpreted as an odds ratio, showing the likelihood of a higher degree of annoyance being experienced due to freight railway vibration when compared to passenger railway vibration. The development of this model, and an analysis of the odds ratio, is presented in this section.

4.5.1 The cumulative ordinal logit model

The ordinal logit regression model is based on the inverse of the logistic function used to classify unknown railway vibration signals in this work (see Chapter 3). The ordinal logit model is particularly suited to the analysis of ordinal variables (Long, 1997), such as the ordinal response variables that were collected during the field study of Waddington et al. (2014) and analysed in this work. For an ordinal response variable y_i that can fall into $j = 1, ..., J$ categories, y_i follows a multinomial distribution where p_{ij} represents the probability that the ith observation falls into response category j. The cumulative probability P_{ij} that the ith observation falls into category j or lower can then be defined:

$$P_{ij} = P(y_i \leq j) = p_{i1} + p_{i2} + ... + p_{ij} \tag{4.13}$$

The link function used in ordinal logit regression is the logit function:

$$\text{logit}(P) = \ln\left(\frac{P}{1-P}\right) \tag{4.14}$$

Combining Equations 4.13 and 4.14 gives the cumulative logit:

$$\text{logit}(P_{ij}) = \ln\left(\frac{P(y_i \leq j)}{1-P(y_i \leq j)}\right) \tag{4.15}$$

The cumulative logit model can then take the form of a regression model as follows:

$$\text{logit}(P_{ij}) = \beta_j - \boldsymbol{\beta}\mathbf{X} \tag{4.16}$$

where β_j is the threshold coefficient for the jth category, $\mathbf{X}$ is a vector of independent predictor variables and $\boldsymbol{\beta}$ is a vector of corresponding regression coefficients to be estimated by maximum likelihood. The goodness of fit and significance of the model, and the significance of the regression parameters can be determined using the same methods as for the ordinal probit grouped regression model (see Section 4.4.2).

4.5.2 Exposure-response relationships with the cumulative ordinal logit model

The ordinal logit model was used to develop an exposure response relationship with a source dummy variable, similarly to the method that was employed with the grouped regression model with a source dummy variable in Section 4.4.3. Exposures and responses for both freight and passenger railway vibration were pooled together and a dummy variable for source type was created (passenger =

0, freight = 1). An exposure-response relationship was then determined using the ordinal logit model, with the regression coefficients estimated via maximum likelihood. These exposure response relationships are presented in Figure 4.6. As the model is of a cumulative form, the exposure-response curves represent the probability of a respondent expressing annoyance in the given category or higher. Parameter estimates and other details of the cumulative ordinal logit model are presented in Table 4.3. For reasons of clarity, the 95% confidence intervals are omitted from Figure 4.6 and are instead presented in Table 4.4.

Parameter	β Estimate	Standard Error	p-value	Overall Model	
Threshold				N	1504
Not at All/Slightly	-0.86	0.42	< 0.050	p-value	< 0.001
Slightly/Moderately	-0.38	0.42	*n.s.*	R^2_{pseudo}	0.02
Moderately/Very	0.35	0.42	*n.s.*		
Very/Extremely	1.60	0.44	< 0.001		
Location					
$10\log_{10}(rms_{k,24hr})$	0.09	0.02	< 0.001		
Source Type	0.48	0.12	< 0.001		
Threshold				N	1504
Not at All/Slightly	0.18	0.24	*n.s.*	p-value	< 0.001
Slightly/Moderately	0.66	0.24	< 0.010	R^2_{pseudo}	0.02
Moderately/Very	1.38	0.24	< 0.001		
Very/Extremely	2.64	0.27	< 0.001		
Location					
$10\log_{10}(\mathrm{VDV}_{b,24hr})$	0.08	0.02	< 0.001		
Source Type	0.59	0.12	< 0.001		

TABLE 4.3: Parameter estimates and other details of the cumulative ordinal logit models for annoyance due to exposure to freight and passenger railway vibration

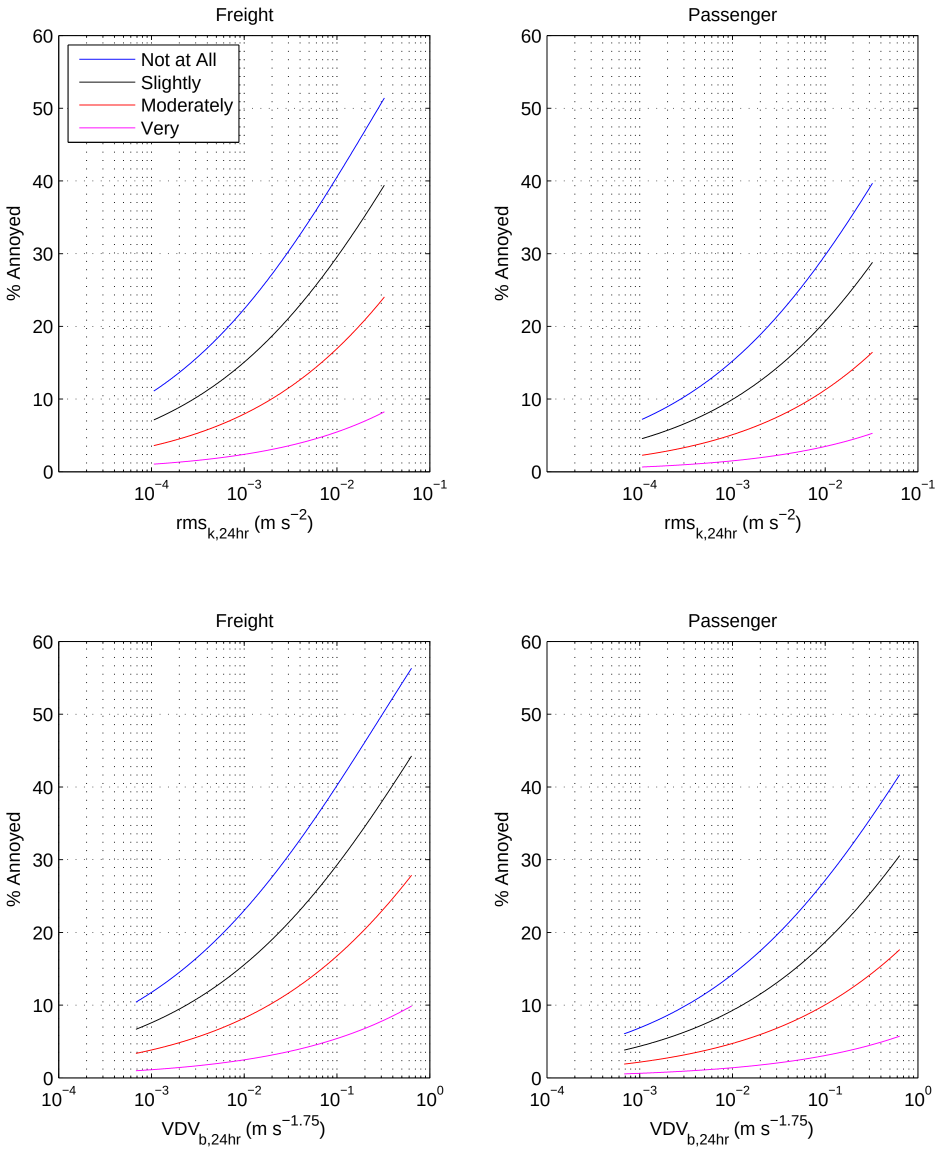

FIGURE 4.6: Exposure-response relationships derived using the cumulative ordinal logit model, with a dummy source type variable for source type for freight exposure (source type $= 1$) and passenger exposure (source type $= 0$)

Parameter	β Estimate	95% Confidence Intervals	
		Lower Bound	Upper Bound
Threshold			
Not at All/Slightly	-0.86	-1.68	-0.05
Slightly/Moderately	-0.38	-1.19	0.44
Moderately/Very	0.35	-0.47	1.16
Very/Extremely	1.60	0.75	2.46
Location			
$10\log_{10}(rms_{k,24hr})$	0.09	0.06	0.12
Source Type	0.48	0.24	0.71
Threshold			
Not at All/Slightly	0.18	-0.29	0.64
Slightly/Moderately	0.66	0.20	1.13
Moderately/Very	1.38	0.91	1.86
Very/Extremely	2.64	2.11	3.18
Location			
$10\log_{10}(\mathrm{VDV}_{b,24hr})$	0.08	0.05	0.11
Source Type	0.59	0.35	0.83

Table 4.4: Parameter estimates and their 95% confidence intervals for the cumulative ordinal logit models for annoyance due to exposure to freight and passenger railway vibration

The dummy source type variable is positive and significant, for both models derived with rms acceleration and VDV exposure, once again giving confidence that the annoyance response to freight and passenger railway vibration is different, and that it is therefore viable to derive separate exposure-response relationships for these two vibration sources. Setting the source type to 1 and 0 allows the determination of exposure-response relationships for exposure to freight and passenger railway vibration exposure respectively, showing that the annoyance response for the population studied is consistently higher for exposure to freight vibration across the range of exposure studied. One of the main benefits in using the ordinal logit model, is that its regression coefficients can be interpreted as odds ratios, which is further detailed in the next section.

4.5.3 Interpreting the source type as an odds ratio

The advantage of using an ordinal logit model is that its regression coefficients can be intuitively interpreted as odds ratios, which is not the case for the ordinal probit grouped regression model derived in Section 4.4. The odds ratios are particularity useful in examining the influence of dummy variables, such as the source type dummy variable used in the ordinal logit model in this section. Examining the odds ratio for the source type variable allows the influence of a change in conditions on an outcome to be determined, i.e. the influence of changing the source from passenger to freight on the exposure-response relationship. For the ordinal logit model, the odds ratio associated with a particular variable can be computed as the exponential of its regression coefficient, as follows (Agresti, 2002):

$$\text{Odds Ratio} = e^{\beta_n} \tag{4.17}$$

where β_n is the regression coefficient of the nth variable. When all other variables are held constant, the odds ratio computed using Equation 4.17 represents the relative odds of the modelled outcome occurring, given the condition represented by the nth variable. Since the ordinal logit model used in this section is cumulative, the odds ratio is also cumulative, representing the odds of respondents reporting annoyance of a higher category (Agresti, 2002). Substituting the source type regression coefficients shown in Table 4.3 into Equation 4.17 allows the determination of the odds ratio associated with the source type variable. The results indicate that, all other variables being held equal, for the same 24 hour *rms* acceleration exposure, respondents are 1.6 times more likely to express annoyance in a higher annoyance category due to freight railway vibration than for passenger vibration. Likewise, for the same 24 hour VDV exposure, respondents are 1.8 times more likely to express annoyance in a higher annoyance category due to freight railway vibration than due to passenger railway vibration. Once again, this shows that the annoyance response due to freight railway vibration is significantly higher than that due to passenger railway vibration, even for equal magnitudes of vibration exposure.

4.6 Differences in the human response to freight and passenger railway vibration

Exposure-response relationships presented in this chapter suggest that the human response to freight railway vibration is significantly different than that due to passenger railway vibration, even for equal levels of vibration exposure. Exposure-response relationships for these two sources of environmental vibration show that the annoyance response increases monotonically with vibration exposure magnitude, but the rate of increase of annoyance is much higher for freight vibration than for passenger vibration. This difference in the responses suggest that people are able to differentiate between these two sources of railway vibration, and that freight railway vibration is significantly more annoying than passenger railway vibration, even for equal levels of vibration magnitude.

Though no previous research has specifically investigated the differences in human response to different sources of railway vibration, previous field studies have reported differences in response to different sources of railway noise. In Fields and Walker's 1982 field study, freight trains were mentioned as being most bothersome approximately three times more often than passenger trains. Similarly, more than half of the interviewees from a field study by Andersen et al. (1983) mentioned freight trains as being particularly disturbing. In their field study, Pennig et al. (2012) found that the annoyance response due to railway noise increased significantly with increasing total number of trains and number of freight trains but not with increasing number of passenger trains. Freight railway noise has also been shown to have a greater effect on sleep disturbance than passenger railway noise

(Saremi et al., 2008) and even aircraft noise in some cases (Elmenhorst et al., 2012). These studies suggest a difference in the human response to different sources of railway noise, so it is not a surprising result that differences also exist between sources of railway vibration. In terms of both noise and vibration, it appears that freight railway traffic is more annoying than passenger railway traffic. Although exposure-response relationships derived in this chapter have demonstrated that this is true even for equal levels of vibration exposure, previous research has shown that the same may not be true for equal levels of noise exposure (De Jong and Miedema, 1996).

Differences in response to freight and passenger railway noise are often attributed to the increased duration of freight passbys and the greater proportion of low frequency noise (Pennig et al., 2012). Similar conclusions can be drawn for differences between freight and passenger railway vibration. Indeed, it has been shown that the logistic regression model can accurately distinguish between freight and passenger railway vibration signals based only on their duration and low frequency energy content. In this work, the mean duration of freight vibration events is 23.8 s ($\sigma = 5.4$) and the mean duration of passenger vibration events is 16.1 s ($\sigma = 6.4$). As well as being longer in duration, freight trains are typically heavier and can more easily elicit groundborne vibrations (De Jong, 1979). BS 6472-1:2008 defines a region of high sensitivity for humans for vertical vibration from 4 to 12.5 Hz. For this study, the mean proportion of a signal's rms acceleration that is contained within this region is 18.4% ($\sigma = 12.4$) for freight signals and 14.4% ($\sigma = 9.5$) for passenger signals.

An additional factor may be found in the tendency of freight railway traffic to be more frequent during evening and night-time hours. In this work, the mean proportion of freight traffic during day time hours (07:00 to 19:00) evening time hours (19:00 to 23:00) and night-time hours (23:00 to 07:00) is 10.1%, 18.2% and 21.5% respectively (SD = 4.7, 12.0, 11.6). Peris et al. (2012) demonstrated that annoyance due to equal levels of railway vibration exposure is greater during night-time than during evening time, and greater during evening time than during day time. The fact that freight traffic is more prevalent during periods in which sensitivity to railway vibration is higher is therefore likely to affect the annoyance response to freight railway vibration.

4.7 Summary

In this chapter, exposure-response relationships have been developed for annoyance due to exposure to freight and passenger railway traffic, in order to investigate whether there are any differences in the human response to these two sources of environmental vibration. Firstly, exposure-response relationships with a dummy source type variable were developed using an ordinal probit grouped regression analysis. The regression resulted in a significant and positive source type variable, giving confidence that it is necessary to develop separate exposure-response relationships for freight and passenger railway vibration.

Separate exposure-response relationships were then developed, showing that the annoyance response to freight railway vibration is significantly higher than that due to passenger railway vibration, even for equal levels of vibration exposure magnitude. In terms of a community tolerance level, the population studied appears to be 15 dB (re 10^{-6} m s^{-2}) more tolerant to passenger railway vibration than to freight railway vibration.

The difference in response was further investigated by developing exposure-response relationships using a cumulative ordinal logit model. This model allowed the investigation of the odds ratio associated with the source type, showing that for equal 24 hour exposures of *rms* acceleration, respondents are 1.6 more likely to express annoyance in a higher annoyance category for freight vibration than they are for passenger vibration. Likewise, for equal 24 hour exposures of VDV, respondents are 1.8 times more likely to express annoyance in a higher annoyance category due to freight vibration than they are due to passenger vibration.

The indication that the human response to freight and passenger railway vibration is different for equal vibration magnitudes, suggests that respondents are perceiving more than just the vibration exposure magnitude when making their annoyance judgements. There may be other factors that contribute to the annoyance response (for example, event duration and simultaneous noise exposure) that are not sufficiently quantified in the exposure metrics studied in this chapter. In an attempt to investigate these contributing factors, a subjective test was designed to further investigate the human response to railway noise and vibration. The subjective test is described in detail in Chapter 5, and the analysis of the human response to noise and vibration as a multidimensional phenomenon is presented in Chapter 6.

Chapter 5

A subjective test on the perception of combined railway noise and vibration

5.1 Introduction

The exposure-response relationships derived in Chapter 4 demonstrate that there is a disparity in the human response to freight and passenger railway vibration. Different relationships, or a source type penalty for freight signals, are required to estimate the human response to freight and passenger railway vibration when the independent variable is a measure of the vibration exposure magnitude alone (i.e. *rms* acceleration or VDV). It is possible that other objective features of the vibration signals from freight and passenger trains contribute to the overall human response to these sources, and identifying these features could help to explain the difference in the response relationship. Identifying these factors, and subsequently including them in a relationship to describe the human response to vibration from freight and passenger railway traffic, could allow the prediction of the human response to railway vibration, without the need to identify the vibration source or apply a source penalty to freight railway vibration exposure.

In addition, environmental vibration very often exists in concert with environmental noise from the same source. When investigating the response relationship to environmental vibration, therefore, it is also important to give consideration to the effects of combined noise and vibration. The perception of noise or sound has widely been accepted as a complex and multidimensional phenomenon. That is, the perception of noise is a function of several perceptual dimensions which relate to objective parameters of the noise. This has also been shown to be the case for environmental railway vibration (Woodcock et al., 2014a). Several laboratory studies have shown that the perception of combined noise and vibration is a complex one, and that annoyance due to noise is affected by the presence of vibration and vice versa (Howarth and Griffin, 1990, 1991; Jik Lee and Griffin, 2013; Paulsen and Kastka, 1995) however, no attempt has yet been made to investigate the perception of combined noise and vibration as a multidimensional phenomena. Considering the perception of combined railway noise and vibration as a multidimensional phenomena may allow the identification of objective features of the noise and vibration signals that can be used to predict the human response with a greater degree of accuracy than those which only take into account the magnitude of noise and vibration exposure.

In order to investigate the perception of combined railway noise and vibration a subjective laboratory test was designed. The design and methodology of the subjective test will be presented in this chapter, followed by an analysis of the perceptual results as a multidimensional phenomena in Chapter 6.

5.2 Measurement methodology

The subjective test involved exposing subjects to combined railway noise and vibration signals in a controlled setting. Since the test required stimuli of noise and vibration in combination, and the database of measurements utilised in previous chapters contains only vibration measurements, it was necessary to perform new field measurements. These field measurements of noise and vibration were performed over a period of two days in June 2014 at a location that had previously been identified as a good location for measurements due to the lack of other noise and vibration sources in the area. The measurement location was on the West Coast Mainline, between the Leyland and Preston stations. Noise was recorded using two AKG C214 microphones, with a sensitivity of 20 mV/Pa and a signal to noise ratio of 81 dB(A), paired in an X-Y stereo configuration. These microphones have a cardioid polar pattern. Vibration was recorded using a Guralp CMG-5TD tri-axial strong motion accelerometer with a sampling rate of 200 Hz and a 100 Hz low pass filter. Both the stereo microphone configuration and the Guralp were positioned 10 m from the near-side rail and 20 m from the far-side rail. Noise and vibration was recorded continuously for several hours, with a note taken of the time and details of each train passby so that the signals could later be identified, extracted and labelled from the recordings.

5.3 Selection of noise and vibration stimuli

Across the two days of recordings, 9 freight train and 53 passenger train passbys were recorded. Due to the nature of the subjective test, the selection of the stimuli had to be quite strict. The subjective test took the form of a paired comparison test, which can result in extremely long tests when the number of stimuli is large, or the stimuli are of long duration. Since the nature of the research requires that some of the stimuli will be freight train passbys, which typically are of longer duration than passenger trains, it was necessary to significantly reduce the number of stimuli.

Firstly, the 9 freight passby noise recordings were inspected for contamination by wind noise or other noise sources since these passbys were of significantly longer duration and hence had a greater chance of being contaminated. Audible inspection of the freight passby noise recordings revealed that 6 of the 9 recordings had some noise contamination and so were discarded. Since the nature of the research is focuses on the difference in response to freight and passenger noise and vibration, it was deemed important to include as many freight passbys as feasibly possible, so the 3 remaining freight passbys were included in the subjective test. The remaining signals in the test are made up of passenger passbys and the selection criteria was to include as many as possible, without the test being prohibitively long. For a complete paired comparison test, the subject is required to make judgements on $\frac{N(N-1)}{2}$ pairs of stimuli, where N is the number of stimuli included in the test. A selection of 10 stimuli, comprised of 7 passenger and 3 freight passbys, would therefore result in 45 pairs in the subjective test. With the duration of the signals known, the duration of a test containing these 45 pairs was estimated to be approximately 25-35 minutes depending on how quickly the subjects were able to make their judgements. Adding an extra passenger signal to the stimuli set would increase the number of pairs to 55 and the estimate of the test duration to approximately 30-40 minutes. To avoid fatigue due to a higher number of paired judgements and an increased test duration, the stimuli set with 7 passenger and 3 freight passbys was chosen.

The 7 passenger signals selected for the subjective test were taken from the set of 53 recorded passbys. In order to obtain a varied stimulus set, signals were selected based on the results of a principle component analysis on various calculated objective properties of the vibration signals. The calculated properties included the *rms* acceleration and the VDV. The spectral centroid was calculated as a measure of the signal's spectral distribution. Temporal factors of the signal were quantified by the kurtosis, the skewness, the crest factor and the duration defined by the 3 dB and 10 dB downpoints of the signal. For further details on these vibration descriptors, see Section 3.5.2. Further aspects of the signal envelope were quantified by the modulation depth and modulation frequency. The modulation depth is defined as the average difference between the maxima and minima of the signal envelope, and the modulation frequency is the inverse of the average period between the maxima of the signal envelope. The scree plot of the principal component analysis on these descriptors, showing the percentage of variance explained by each of the recovered principle components, is shown in Figure 5.1. A four dimensional principal component solution, accounting for approximately 94% of the explained variance in the descriptor space, was chosen for examination.

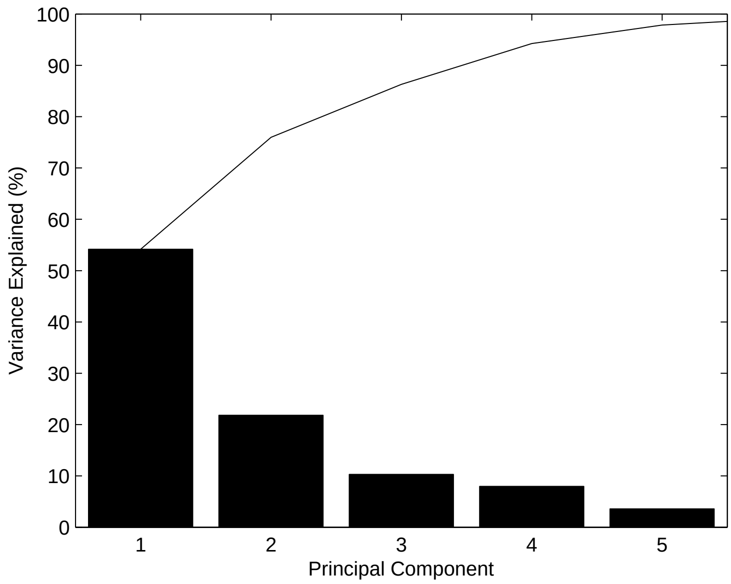

FIGURE 5.1: Scree plot of the percentage of variation explained by each of the recovered principal components

Figures 5.2 to 5.4 show the positions of the 53 passenger vibration signals on the first four principal components, along with the weighting of each of the calculated descriptors on the components, indicated by the projections in the figures. Since the solution is of more than two dimensions, some of the projections are shorter than others, indicating that they are partially projecting into other dimensions not represented in the two-dimensional figures.

Though it can be hard to visually interpret the weightings of the descriptors on each principal component, and the interpretation is something of a subjective nature, some judgements can be made about which principal components could be related to which descriptors. For example, the first principal component could be related to the duration of the signal, the second component could be related to exposure magnitude and peaks in the signal, the third component could be related to envelope modulation and the fourth component could be related to envelope modulation and peaks in the signal. To select a varied passenger signal stimulus set, each principal component was equally divided into two portions and a signal was randomly selected from each half, resulting in a set of 8 passenger signals. One signal was then randomly removed, resulting in the desired selection of 7 passenger train signals. The noise pressure time histories and the vibration acceleration time histories of the 10 selected stimuli are presented in Figures 5.5 and 5.6 respectively. Note that the time axes are extended for the three freight train signals: stimuli 8, 9 and 10.

5.4 Subjective test design

The subjective test involved exposing 30 subjects to pairs of combined noise and vibration stimuli in an acoustically controlled setting and asking them to make paired comparison judgements on the annoyance and similarity of the stimulus pair, as well as judgements of categorical annoyance of each individual stimulus. The design of the subjective test is presented in this section.

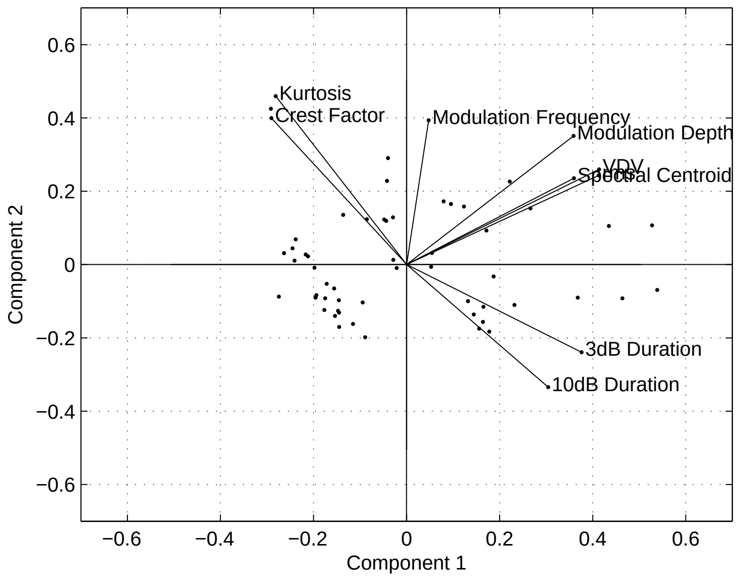

FIGURE 5.2: Position of the 53 passenger vibration signals on the first and second principal components as well as the weighting each descriptor has on the components

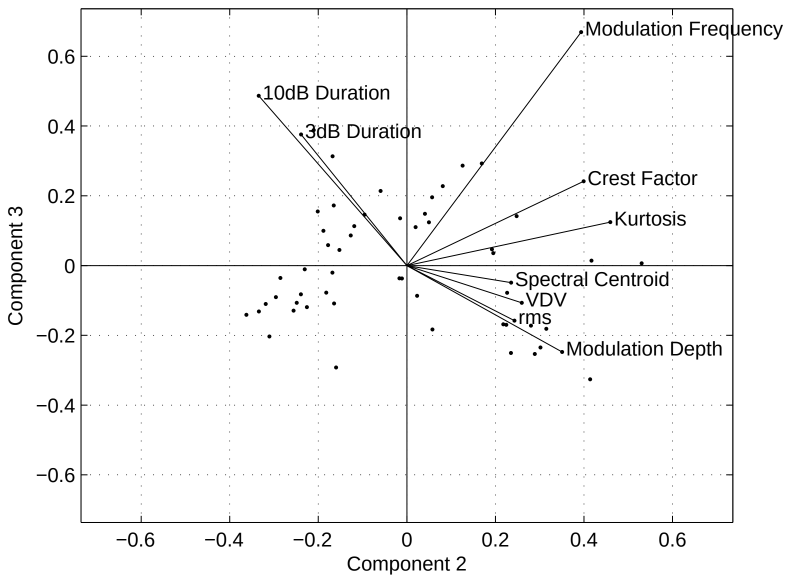

FIGURE 5.3: Position of the 53 passenger vibration signals on the second and third principal components as well as the weighting each descriptor has on the components

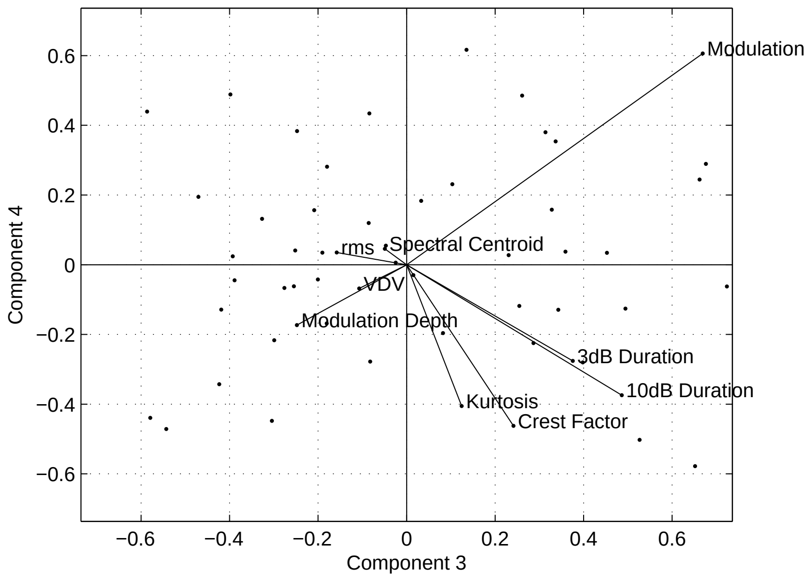

FIGURE 5.4: Position of the 53 passenger vibration signals on the third and fourth principal components as well as the weighting each descriptor has on the components

5.4.1 Reproduction of combined noise and vibration stimuli

The subjective test took place in the University of Salford's Listening Room, a room specifically designed for subjective testing. It meets the stringent requirements for the subjective assessment of small impairments in audio systems laid out in ITU-R BS 1116-2 (2014), as well as the requirements for listening tests on loudspeakers specified by BS 6840-13 IEC 60268-13 (1998). The room dimensions are 6.6 m $\times$ 5.8 m $\times$ 2.8 m and the background noise level is 5.7 dB(A).

The noise and vibration reproduction and the user interface for the subjective test were controlled using the Max visual programming language (Cycling '74, 2013). Stimuli were reproduced as three data channels - stereo audio plus a third channel for the vibration signal. Channel data was passed from a Macbook Pro to an RME ADI-8 DS ADAT/TDIF AD/DA converter via an M-Audio Profire Lightbridge interface. Stereo audio was passed directly from the ADAT converter to a pair of bi-amplified loudspeakers (Genelec 8030A). Vibration data was passed

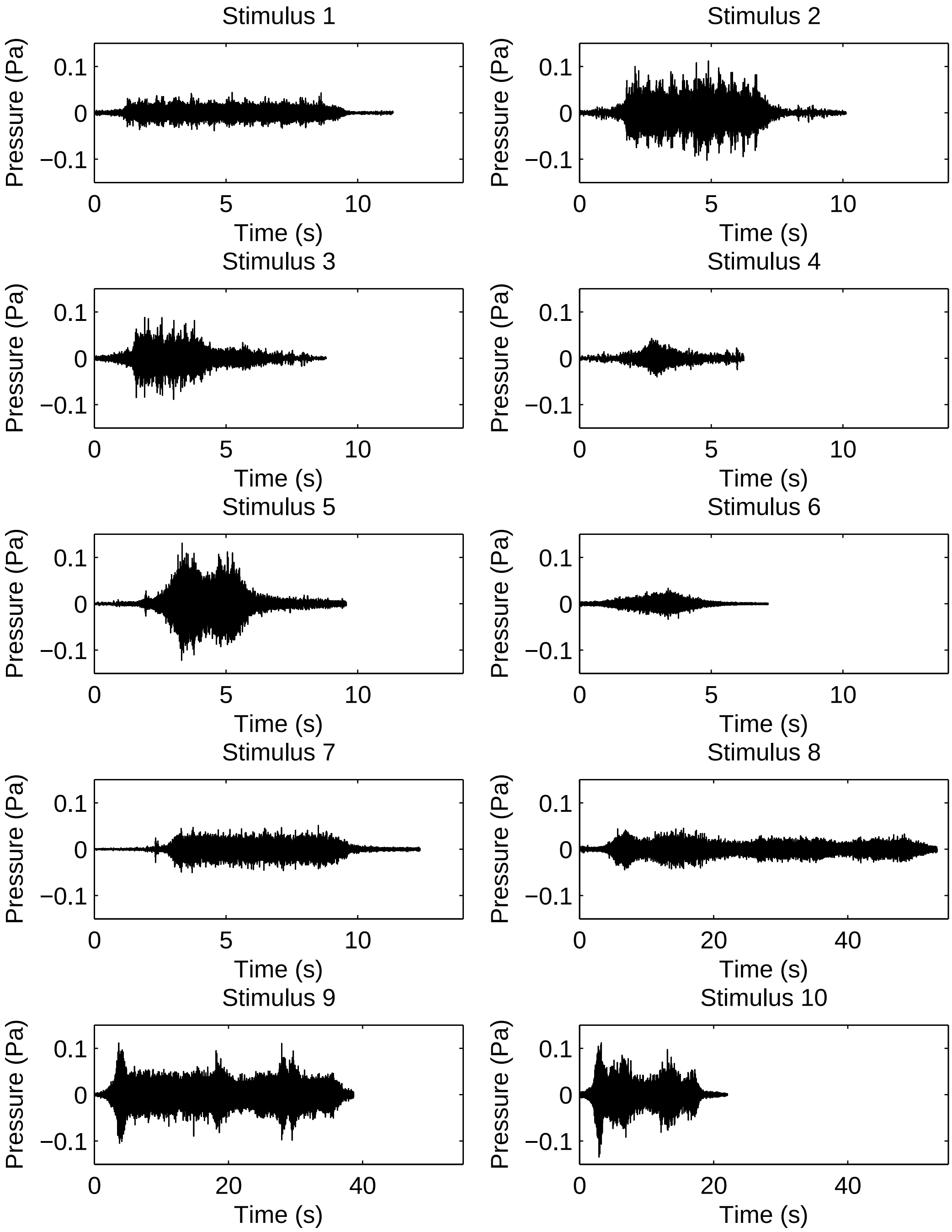

Figure 5.5: Pressure time histories of the 10 noise stimuli selected for the subjective test

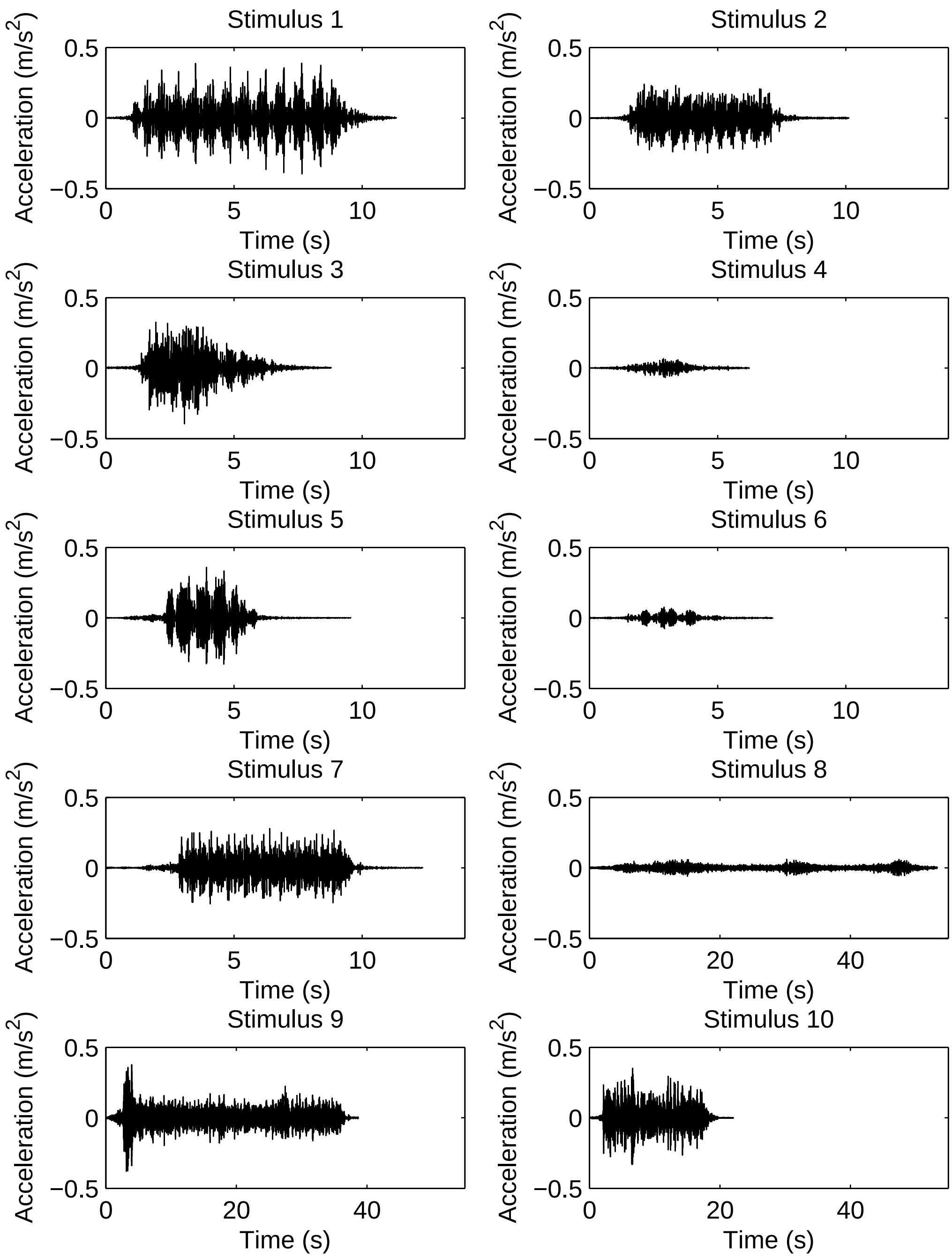

FIGURE 5.6: Acceleration time histories of the 10 vibration stimuli selected for the subjective test

to a tactile transducer (BK-LFE-KIT) rigidly mounted to the underside of a chair via a 1000 W amplifier (BKA1000-N).

The listener and loudspeakers were positioned to comply with ITU-R Recommendation BS 1116-1 (2014) and EBU 3276 (1998) as shown in Figure 5.7. The loudspeakers and listener formed an equilateral triangle with the distance between the two speakers and the distance between the listener and each speaker equal to 2 m. The loudspeakers were placed at a height of 1.2 m and the distance between each loudspeaker and the closest wall was 1.3 m. A photograph of the setup, with a subject sitting in the listener position, is shown in Figure 5.8.

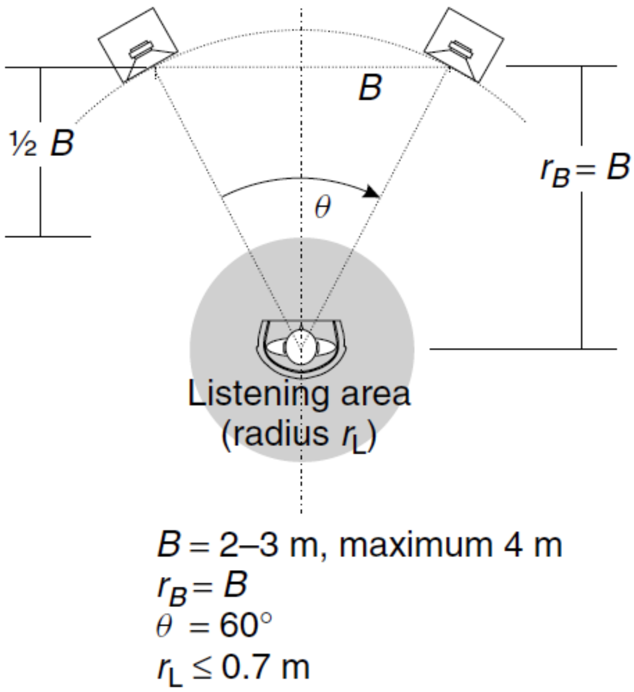

FIGURE 5.7: Recommended stereophonic listener and speaker locations adapted from ITU-R Recommendation BS 1116-1 (2014) by Bech and Zacharov (2006). Shaded region represents the recommended listening area.

The reproduced noise was recorded using a 01 dB Symphonie measurement system combined with an MCE212 microphone with a sensitivity of 50 mV/Pa and a frequency range of 6.3 Hz to 20 kHz. The reproduced vibration was measured

Figure 5.8: Photograph of the subjective test setup, showing a subject sitting in the recommended listener position

using a Svantek SV38 seat pad accelerometer with a sensitivity of 100 mV/(m s^{-2}) and a frequency range of 0.1 to 100 Hz, paired with a Svantek SVAN 957 sound and vibration analyser. Objective descriptors of the 10 noise and vibration stimuli, as reproduced and measured using the subjective test setup, are presented in Table 5.1. The noise and vibration stimuli were reproduced at levels that are comfortable for the subject, realistic for levels experienced at residential environments within 100 m of a railway line, and commensurate with levels used in previous subjective tests of combined railway noise and vibration (Howarth and Griffin, 1990, 1991; Paulsen and Kastka, 1995).

5.4.2 Limitations of the noise and vibration reproduction

There are some limitations of the noise and vibration reproduction that are worth discussing. Though the transducer used in the experiment has a relatively flat response in the region of 10 to 80 Hz, it is highly non-linear and can generate a high degree of harmonic distortion (Abercrombie and Braasch, 2009; Woodcock, 2013). Woodcock (2013), noting the limitations of this particular transducer, built a new test rig, consisting of a chair mounted on an electrodynamic shaker, which

Stimulus	L_{Aeq} (dBA)	SEL_A (dBA)	VDV$_b$ (m s$^{-1.75}$)	rms_k (m s^{-2})	Duration (s)
1	51.6	62.6	0.0843	0.0305	11.3
2	50.7	59.8	0.0522	0.0217	10.1
3	57.6	68.3	0.0514	0.0219	8.8
4	52.8	62.9	0.0694	0.0288	6.2
5	60.0	70.6	0.0611	0.0258	9.6
6	53.2	64.4	0.0511	0.0202	7.2
7	57.1	68.0	0.0664	0.0258	12.4
8	58.7	72.4	0.0954	0.0226	53.3
9	58.2	74.2	0.0685	0.0140	38.7
10	62.5	75.0	0.0723	0.0230	22.1
Mean	56.2	67.8	0.0672	0.0234	18.0
Standard Deviation	3.9	5.3	0.0145	0.0047	15.8

Table 5.1: Objective descriptors of the noise and vibration stimuli as reproduced and measured in the subjective test

was designed to be more capable of faithful reproduction of vibration signals. An initial study was performed to determine if this shaker could be used for the tests in this research, however, the shaker and its 6000W amplifier produced an excessive amount of noise, making it impossible to use this piece of equipment for multi modal tests. It was decided therefore, to use the tactile transducer combined with a seat pad accelerometer. This meant that, though the vibration signals may not be accurately reproduced, the vibration experienced by the subject is being accurately measured, and any analysis of the test results uses the measured data from the seat pad accelerometer. Therefore, any models of the perception of combined noise and vibration derived from these results are representative of the vibration experienced by the subject, even if that vibration is not necessarily highly representative of railway vibration. Similarly, the reproduction of the noise as stereo audio could be perceived as a limitation. Ambisonic reproduction may have result in a more realistic experience for the subject. For future experiments, more robust means of reproducing the noise and vibration should be considered, however, the set up used for these experiments, though limited, should still allow an investigation into the perception of combined noise and vibration, and in particular should allow an investigation into the applicability of using multidimensional scaling to analyse the perception of combined noise and vibration.

5.4.3 Subject training

Thirty subjects participated in the subjective tests described in this chapter. The majority of subjects were members of the Acoustics Research Centre at the University of Salford, and had experience in taking part in various subjective tests. However, though many of the subjects had taken part in subjective tests to do with noise, few had experience with subjective tests involving vibration, and so would perhaps respond more like the general population for this experiment. Prior to the start of the test, each subject was provided with written and verbal instructions of how the test would proceed, and were asked to sign a consent form agreeing to take part in the test. The subjects were encouraged to imagine that they were at home, living in the vicinity of a railway line, and that the noise and vibration stimuli would be something that they would experience on a day to day basis. An opportunity was provided for the subjects to ask any questions that they may have and they were then asked to sit comfortably on the chair on top of the seat pad accelerometer. Once the subjects had made themselves comfortable, they were encouraged to keep their posture consistent as much as possible throughout the test. The 10 noise and vibration stimuli were then played in order to allow the subjects to familiarise themselves to the stimuli playback and to give them a feel for the levels that they would experience throughout the test. During this process, the vibration from the transducer was measured using the seat pad accelerometer in order to check the variability of the vibration transmitted to each subject according to differences in their posture and body mass. Subjects were then allowed to familiarise themselves with the test user interface presented in the following section with two trial pairs of noise and vibration stimuli. Following these processes of familiarisation and practice, the subjects then embarked on the paired comparison test of 45 pairs.

5.4.4 Paired comparison judgements of annoyance and similarity

The test user interface used in the paired comparison test is shown in Figure 5.9. The user interface was presented on an iPad held by the subject and paired to the Macbook Pro using a local wireless network. This allowed the subject to control the test using an intuitive touch screen interface which was a preferred option to requiring the subject to sit holding a laptop, potentially causing discomfort. A photograph of a subject using the touch screen interface is shown in Figure 5.10. The subjects were allowed to play stimulus A or B by touching the respective buttons, and control playback using a play/stop button. Subjects were encouraged to play each stimulus as many times as they liked in order to make their judgements, though they were encouraged to make their judgements based on their initial reactions as far as possible. The ordering of the pairs was randomised for each subject according to a Ross series (Ross, 1934) which ensures the greatest separation of pairs with common stimuli (David, 1988). For each pair, subjects were asked to make two paired comparison judgements:

1. Which of the pair would bother, annoy or disturb you most if you experienced them in your own home?
2. How similar do you perceive the pair to be?

Each subject marked their response to both of these questions via continuous 101 point sliders, which were coded between -0.5 to 0.5 for question 1 and 0 to 100 for question 2. Subjects were encouraged to use the entire scale. Each slider was initially blank to avoid biasing results, and the black marker line only appeared on the slider when the subject touched the scale at any point. The instance of the interface shown in Figure 5.9 represents a situation in which the subject has last played stimulus A (hence the button for stimulus A is highlighted) and has marked their judgements on both scales. Once the subject had made their judgements, they were required to touch the box marked "Confirm your choice" before the button to move to the next pair was activated, in order to prevent this button being accidentally pressed before the subject had finished with their judgements.

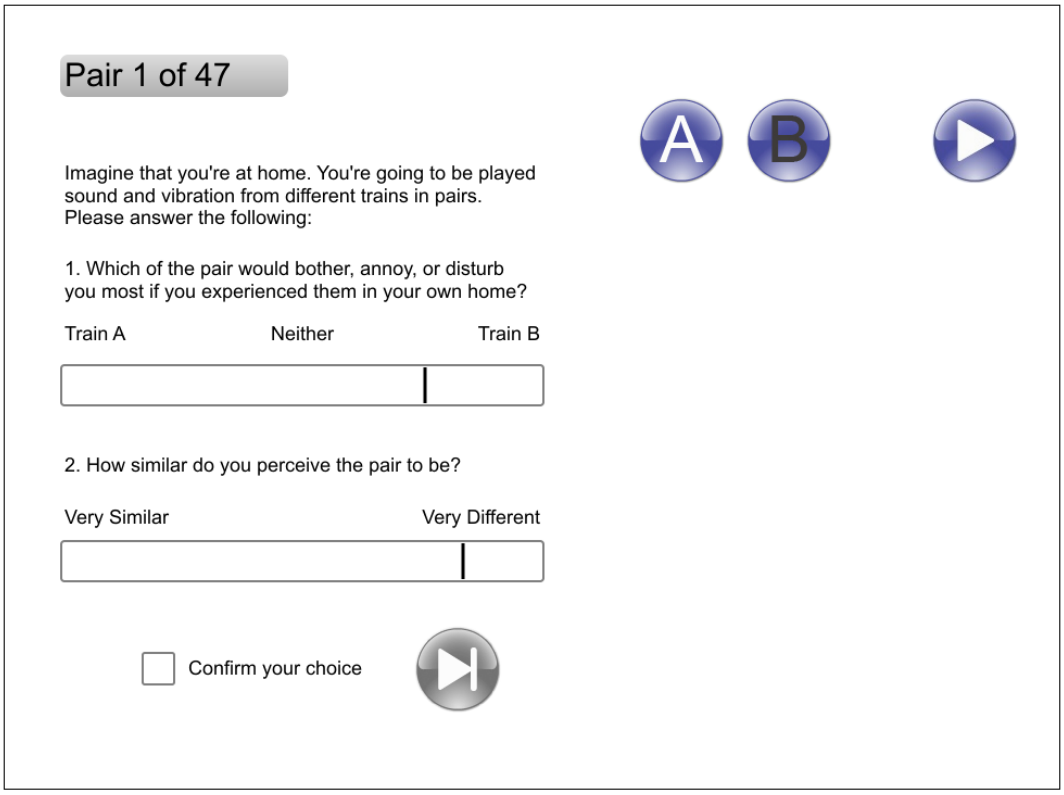

FIGURE 5.9: Graphical user interface used in the paired comparison test

5.4.5 Judgements of categorical annoyance

After completing the paired comparison section of the subjective test, the subjects were provided with an opportunity to take a break before moving on to the second and final part of the test, which required the subjects to make judgements of annoyance on a categorical scale. The paired comparisons in the first part of the subjective test result in single figure annoyance scores of an arbitrary scale and relate only to the set of stimuli upon which the judgements were made. In order to measure annoyance on an absolute scale and to allow the results of this

Figure 5.10: Photograph showing a subject using the touch screen graphical interface

subjective test to be comparable to the results of the field study by Waddington et al. (2014), it was necessary to determine absolute annoyance ratings for each stimulus. The graphical user interface for this part of the subjective test is shown in Figure 5.11. Each subject was allowed to play each of the 10 stimuli in a randomised order, and then asked to indicate on a five point semantic scale how bothered, annoyed or disturbed they would be by each stimulus. The five point scale ranged from "not at all" to "extremely" and is the same scale used in the field study of Waddington et al. (2014) and utilised in Chapter 4 to derive the exposure-response relationships for freight and passenger railway vibration. Once again, after making their judgements, the subjects were required to touch a box confirming their choice before being able to move on to the next stimulus.

5.5 Subjective test results

The data collected during the subjective test allows several forms of analyses to be performed. The paired comparison annoyance data can be used to determine single figure perceived annoyance ratings for each railway stimulus and the results

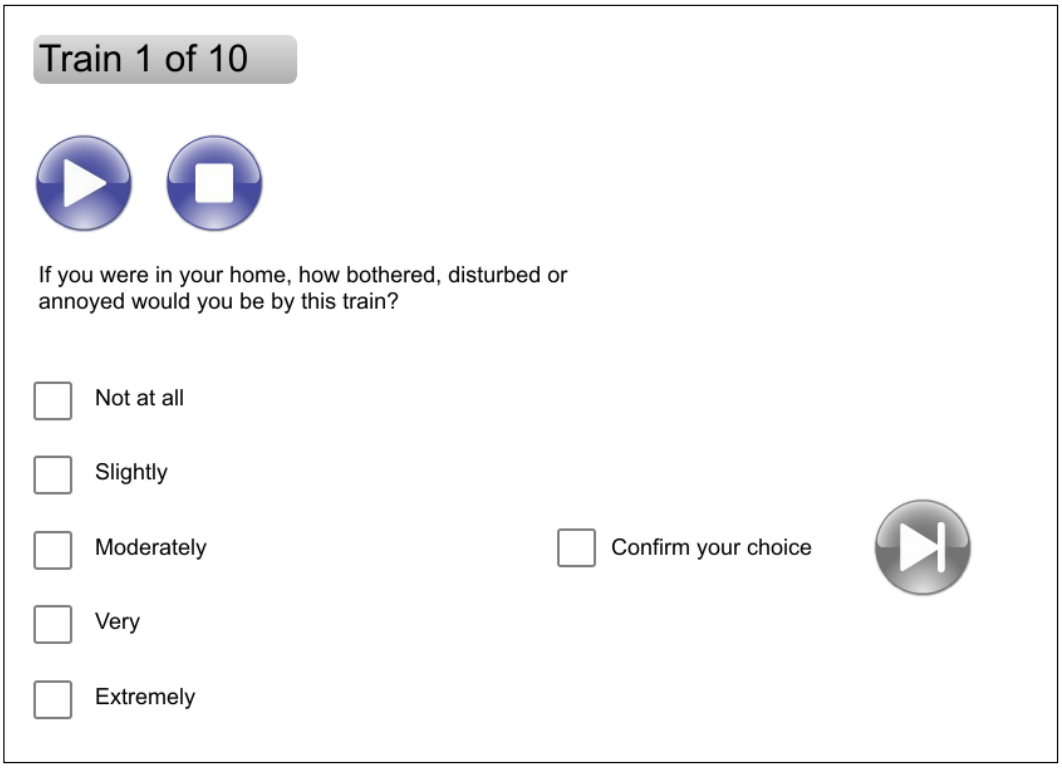

FIGURE 5.11: Graphical test interface used in the categorical annoyance test

of the categorical annoyance test can be used to determine absolute ratings of annoyance for each stimulus. This part of the analysis is presented in this chapter. In addition, the dissimilarity ratings can be used, through multidimensional scaling, to determine the perceptual dimensions associated with combined noise and vibration. A perceived annoyance model can then be determined by relating the perceived annoyance ratings to the positioning of each stimulus within the perceptual dimensions. This part of the analysis is presented in Chapter 6.

5.5.1 Circular error rates

An important measure of intra-subject consistency is the circular error rate. When presented with each possible pairing of three stimuli, i.e. (A, B), (A, C) and (B, C), it is possible for a subject to make an inconsistent judgement. This occurs when, for example, a subject judges stimulus A to be more annoying than stimulus B, stimulus B more annoying than stimulus C and stimulus C more annoying than stimulus A. Though the mechanism by which these inconsistencies can occur is not well defined, Weber (1999) suggests they can occur due to:

- An inaccuracy caused by the subject not paying attention during the test
- An alteration of assessment criteria during the test
- A misjudgement caused by stimuli being very perceptually similar, making the task of the subject more difficult

In the case of stimuli that are very perceptually similar, circular error rates can be reduced by allowing the subject the option of choosing "neither" when being asked which stimulus is more annoying (Parizet, 2002). A "neither" option was provided in the middle of the annoyance scale for this reason. Since the paired comparison user interface has a relatively high resolution (101 points on a sliding scale), it is difficult for a user to input the result that they find neither stimulus i or j more annoying (i.e. the stimuli are equally annoying, $A_{ij} = 0$) as this would require a very precise input on the touch screen interface. In reality, users attempting to input this result are likely to land slightly on either side of the exact middle of the scale, resulting in either i or j being randomly interpreted as more annoying by the count matrix. Not only can this lead to a misinterpretation of the results, but it can cause an increase in the circular error rates when small perceptual differences exist between stimuli (Parizet, 2002). To avoid this issue, an extended "neither zone" was created in the middle of the scale in which any paired comparison score within this range is recoded to $A_{ij} = 0$. The user interface with the neither zone highlighted in grey is shown in Figure 5.12. Note that the neither zone was not visible to the subjects and is only highlighted here for illustrative purposes. The neither zone is approximately equal in width to the text above stating "Neither" and covers the middle 10% of the scale leaving the remaining 90% available for the subjects to indicate a conscious preference of one stimulus over another.

Figure 5.13 shows the circular error rates calculated for each subject with and without the extended neither zone allowing easier selection of a "neither" response.

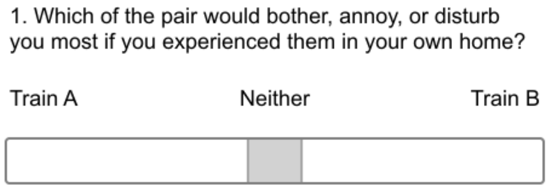

Figure 5.12: Paired comparison user interface with "neither zone" indicated in grey

In all but two cases, circular error rates are reduced with the inclusion of the neither zone and circular error rates are low across all subjects (the maximum circular error rate with the extended neither zone was 6.7% by subject 25). Overall, the circular error rates appear generally lower than those presented for similar paired comparison tests involving noise (Parizet, 2002) and vibration (Woodcock et al., 2014a), giving confidence that the subjective test was a relatively simple task to complete resulting in all subjects making relatively consistent judgements throughout the test. For this reason, no subject data was omitted from further analysis based on their circular error rates.

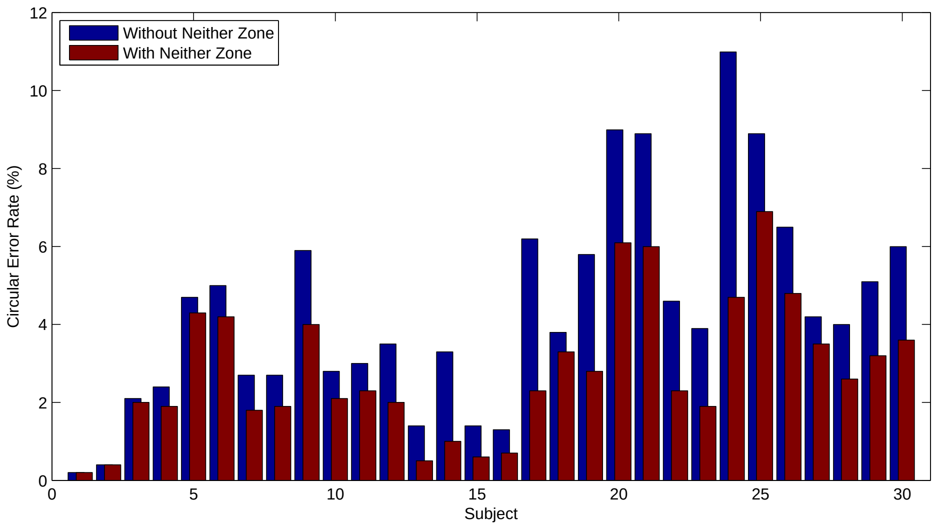

Figure 5.13: Circular error rates for each subject calculated with and without the extended "neither zone" allowing easier selection of a "neither" response

5.5.2 Averaged single figure annoyance score

Single figure perceived annoyance ratings can be obtained from paired comparison data. Each subject provided a paired comparison annoyance score for every possible pairing of the 10 combined noise and vibration stimuli they were exposed to. These annoyance scores can be summed across the subject group and linearly averaged to provide an average single figure annoyance score:

$$A_{av} = \frac{1}{N} \sum_{i \neq j} A_{ij} \tag{5.1}$$

where A_{av} is the averaged single figure annoyance score for stimulus i, N is the number of subjects and A_{ij} is the paired annoyance score for stimuli i and j summed across the subject group. This model was utilised by Parizet et al. (2008) to estimate a single figure metric for comparing the sound quality of car door closures and by Woodcock et al. (2014a) to estimate a single figure annoyance metric for comparing annoyance due to railway vibration. The validity of this model can be tested by measuring the correlation coefficient between the paired annoyance scores, A_{ij}, and paired annoyance scores estimated from the averaged single figure annoyance scores, $\tilde{A}_{ij}$:

$$\tilde{A}_{ij} = A_{av,i} - A_{av,j} \tag{5.2}$$

The correlation coefficient between the calculated and estimated paired annoyance scores is 0.98, suggesting that the paired annoyance scores are well represented by the single figure annoyance scores, A_{av}.

5.5.3 Thurstone's Case V Model

Another widely used model for the estimation of paired comparison single figure scores is Thurstone's Case V model. Thurstone's model (1927) pioneered psychometrics by utilising Gaussian distributions to analyse paired comparisons.

Assuming that an option's annoyance distribution is a Gaussian random variable, their probability density functions can be represented:

$$\begin{aligned} P(i) &= \tfrac{1}{\sigma_i}\phi\left(\tfrac{i-\mu_i}{\sigma_i}\right) \\ P(j) &= \tfrac{1}{\sigma_j}\phi\left(\tfrac{j-\mu_j}{\sigma_j}\right) \end{aligned} \tag{5.3}$$

where μ_i and σ_i are the mean and standard deviation of stimulus i's annoyance distribution and ϕ is the standard normal probability density function with zero mean and unit variance:

$$\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2} \tag{5.4}$$

The model states that when a subject makes a paired comparison judgement between stimulus i and stimulus j, they draw a perception from i's annoyance distribution and j's annoyance distribution and subsequently make a judgement of which stimulus they perceive to have a higher annoyance. This is equivalent to choosing stimulus i over stimulus j if their draw from the random variable $i - j$ exceeds 0:

$$P(i > j) = P((i - j) > 0) \tag{5.5}$$

Since i and j are Gaussian distributions, $i - j$ can be represented by a Gaussian random variable with the following properties:

$$\begin{aligned} \mu_{ij} &= \mu_i - \mu_j \\ \sigma_{ij}^2 &= \sigma_i^2 + \sigma_j^2 - 2\rho_{ij}\sigma_i\sigma_j \end{aligned} \tag{5.6}$$

where ρ_{ij} is the correlation between i and j. Equation 5.5, the probability of stimulus i being chosen over stimulus j then becomes:

$$\begin{aligned}
P(i > j) &= P((i-j) > 0) \\
&= \int_0^{\infty} \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} e^{-\frac{(x-\mu_{ij})^2}{2\sigma_{ij}^2}} dx \\
&= \int_{-\mu_{ij}}^{\infty} \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} e^{-\frac{x^2}{2\sigma_{ij}^2}} dx \\
&= \int_{-\infty}^{\mu_{ij}} \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} e^{-\frac{x^2}{2\sigma_{ij}^2}} dx \\
&= \int_{-\infty}^{\mu_{ij}} \frac{1}{\sigma_{ij}} \phi\left(\frac{x}{\sigma_{ij}}\right) dx \\
&= \Phi\left(\frac{\mu_{ij}}{\sigma_{ij}}\right) \qquad (5.7)
\end{aligned}$$

where Φ is the standard normal cumulative distribution function:

$$\begin{aligned}
\Phi(z) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}} dt \\
&= \int_{-\infty}^{z} \phi(t) dt \qquad (5.8)
\end{aligned}$$

Equation 5.7 can be rearranged to obtain the annoyance difference μ_{ij}:

$$\mu_{ij} = \sigma_{ij} \Phi^{-1}(P(i > j)) \qquad (5.9)$$

where Φ^{-1} is the inverse cumulative distribution function of the standard normal, also referred to as the z-score. Thurstone (1927) proposed estimating the probability $P(i > j)$ by the empirical proportion of subjects choosing i over j. A count matrix, C_{ij}, quantifying the number of times that each stimulus was preferred over each other stimulus can be defined:

$$C_{ij} = \begin{cases} \text{No. of times option } i \text{ preferred over option } j & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases} \tag{5.10}$$

By utilising the count matrix to estimate the probability $P(i > j)$, the annoyance difference estimate then becomes what is known as Thurstone's Law of Comparative Judgement:

$$\hat{\mu}_{ij} = \sigma_{ij}\Phi^{-1}\left(\frac{C_{ij}}{C_{ij} + C_{ji}}\right) \tag{5.11}$$

The use of Equation 5.11 requires that σ_{ij} be known or estimated, which in turn requires σ_i and σ_j to be known, as well as the correlation between i and j, ρ_{ij}, as shown in Equation 5.6. To avoid this, Thurstone (1927) provided a number of model simplifications, the simplest and most popular of which is known as the Case V model (Mosteller, 1951). This adaptation of the model assumes that all stimuli have equal variance and zero (or equal) correlations ($\sigma_i^2 = \sigma_j^2$ and $\rho_{ij} = 0$). Without loss of generality, the variances are set to one half so that the variance of $i - j$ is equal to 1 ($\sigma_{ij}^2 = \sigma_i^2 + \sigma_j^2 = \frac{1}{2} + \frac{1}{2} = 1$). Thurstone's law given in Equation 5.11 is then simplified to Thurstone's Case V model:

$$\hat{\mu}_{ij} = \Phi^{-1}\left(\frac{C_{ij}}{C_{ij} + C_{ji}}\right) \tag{5.12}$$

5.5.4 The Bradley-Terry-Luce model

Another popular paired comparison model is the Bradley-Terry-Luce (BTL) model (Bradley, 1954, 1955; Bradley and Terry, 1952). The BTL model was developed by giving each stimulus a rating π_i which satisfies the following:

$$P(i > j) = P((i - j) > 0 = \frac{\pi_i}{\pi_i + \pi_j} \tag{5.13}$$

Substituting $\pi_i = \exp(\mu_i/s)$, where s is a scale parameter:

$$\begin{aligned} P((i-j)>0) &= \frac{\exp(\mu_i/s)}{\exp(\mu_i/s)+\exp(\mu_j/s)} \\ &= \frac{1}{1+\exp\left(\frac{\mu_i-\mu_j}{s}\right)} \end{aligned} \qquad (5.14)$$

Since Equation 5.14 takes the form of a logistic function, it can be assumed that $i-j$ is a logistic random variable with mean $\mu_i - \mu_j$ and scale parameter s. The BTL model is similar to Thurstone's model except that it assumes the random annoyance difference $i-j$ has a logistic distribution whereas the Thurstone model assumes the random annoyance difference is Gaussian. Two proofs demonstrated by Block and Marshchak (1959) and Holman and Marley (presented in Luce and Suppes (1965)) show that if i and j have Gumbel distributions of annoyance then $i-j$ is indeed logistic.

The BTL model annoyance difference estimate $\hat{\mu}_{ij}$ can be determined by inverting Equation 5.14 and substituting the empirical count proportion $C_{ij}/(C_{ij}+C_{ji})$ for the probability $P(i>j)$, a substitution that was also applied in the development of the Thurstone model:

$$\hat{\mu}_{ij} = s\left(\ln\left(\frac{C_{ij}}{C_{ij}+C_{ji}}\right) - \ln\left(1-\frac{C_{ij}}{C_{ij}+C_{ji}}\right)\right) \qquad (5.15)$$

The inverse logistic cumulative distribution function shown in Equation 5.15 is also commonly known as the logit function. The BTL model scale differences can be compared with the Thurstone's Case V model scale differences by equating the variance using $s = \sqrt{3}/\pi$ (Tsukida and Gupta, 2011). Empirically, the logistic cumulative distribution function is very similar to the Gaussian cumulative distribution function, so Thurstone's model and the BTL model produce very similar results. However, the logistic cumulative distribution function has a fatter tail and has a slightly greater slope at the inflexion point than an equivalent Gaussian. This results in the BTL model estimating slightly smaller scale differences for proportions near 0.5 and slightly larger scale difference for proportions near 0 or 1 when compared with Thurstone's model. Historically, the BTL model was generally preferred for its reduced computation requirements as there is no need to compute the inverse Gaussian cumulative distribution function, though with modern computers this is no longer a significant issue (Tsukida and Gupta, 2011).

5.5.5 Single figure annoyance model fitting

Thurstone's Case V model provides a method for estimating the scale difference for a single pair of stimuli. Tsukida and Gupta (2011) have demonstrated that the Case V model, as shown in Equation 5.12, is the maximum likelihood solution for two stimuli. However, for several stimuli and hence multiple pairs there is no closed solution and instead a convex optimisation problem must be solved. The log likelihood function to be optimised is as follows (Tsukida and Gupta, 2011):

$$LL(\mu|C) \triangleq \ln P(\mu|C) = \sum_{i,j} C_{ij} \ln(\Phi(\mu_i - \mu_j)) \tag{5.16}$$

To determine the maximum likelihood annoyance scales, μ, the right hand side of Equation 5.16 must be maximised subject to $\sum_i \mu_i = 0$. The BTL model can likewise be extended to estimate the annoyance scales for multiple stimuli. A method to achieve this, proposed by Thurstone, was shown to be the solution of a least squares optimisation problem (Mosteller, 1951). The least squares estimate for the annoyance scores, μ, minimises the squared error between the annoyance scores and the BTL pairwise estimates:

$$\hat{\mu} = \arg\min_{\mu} \sum_{i,j} (\hat{\mu}_{ij} - (\mu_i - \mu_j))^2 \tag{5.17}$$

where $\hat{\mu}_{ij}$ is calculated using Equation 5.15. Assuming the mean of the annoyance scores is zero, the closed form solution of the least squares estimate is as follows (Tsukida and Gupta, 2011):

$$\mu_i = \sum_{j=1}^{m} \frac{\hat{\mu}_{ij}}{m} \tag{5.18}$$

where m is the number of stimuli. When estimating annoyance differences using the least squares method, a problem arises when C_{ij} is equal to 0 or the number of subjects, since this results in the empirical count proportion equalling 0 or 1. This will result in annoyance score estimates of $-\infty$ or 0 since $\ln(0) = -\infty$ and $\ln(1) = 0$. One solution to this problem is to ignore the 0/1 entries and use an incomplete matrix solution (Engeldrum, 2000; Gulliksen, 1956; Morrissey, 1955). However, Tsukida and Gupta (2011) argue that this is a heavy-handed solution and ignores important information that one stimulus is deemed as more annoying than another by all subjects. They therefore propose another solution which involves adding or subtracting a fractional count to the $C_{ij} = 0$ or 1 proportions. The modified count matrix then becomes:

$$\hat{C}_{ij} = \begin{cases} 0.5 & \text{if } C_{ij} = 0 \text{ and } i \neq j \\ C_{ij} - 0.5 & \text{if } C_{ij} = N \text{ and } i \neq j \\ C_{ij} & \text{otherwise} \end{cases} \tag{5.19}$$

where N is the number of subjects. The above solution changes the count matrix, though Tsukida and Gupta (2011) argue that it alters the data in a conservative way that biases the count towards less confidence, and that this solution is still preferable to ignoring the $C_{ij} = 0$ or 1 entries. However, they still recommend the use of maximum likelihood model fitting with Thurstone's Case V model (Equation 5.16) as it does not require ignoring or altering data.

Figure 5.14 shows the single figure perceived annoyance scores as determined by fitting Thurstone's Case V model using maximum likelihood. Figure 5.15 shows the averaged single figure annoyance scores (Equation 5.1) and the annoyance scores as determined by fitting the BTL model using the least squares method. Though the magnitudes of the annoyance scales are different, the three models of determining single figure annoyance scores show consistency in the ordering of the stimuli in terms of their annoyance and the relative difference between stimulus annoyance scores. The differences in magnitude between the models is not a significant issue as the point of interest is the relative differences between stimulus annoyance scores. Although a high annoyance score denotes a relatively high degree of annoyance, the overall magnitude of annoyance of each stimulus cannot be known. However, it can be stated that, for example, stimulus 8 is more annoying than stimulus 2.

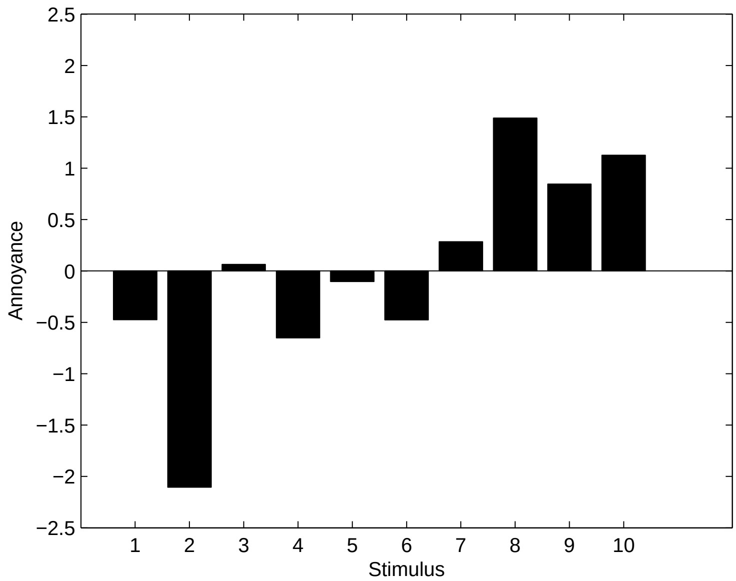

FIGURE 5.14: Single figure perceived annoyance scores for the 10 combined noise and vibration stimuli as determined by Thurstone's Case V model

Referring to Table 5.1, some preliminary judgements can be made about which types of trains cause relatively higher annoyance. Clearly, the freight trains (stimuli 8, 9 and 10) cause significantly higher annoyance than the passenger train stimuli. This may be due to their extended duration and their above average noise levels (in terms of L_{Aeq} and SEL_A) and vibration levels (in terms of VDV_b). This is in line with the exposure-response relationships developed in Chapter 4, which showed that the annoyance response due to exposure to freight railway vibration is higher than that due to exposure to passenger railway vibration.

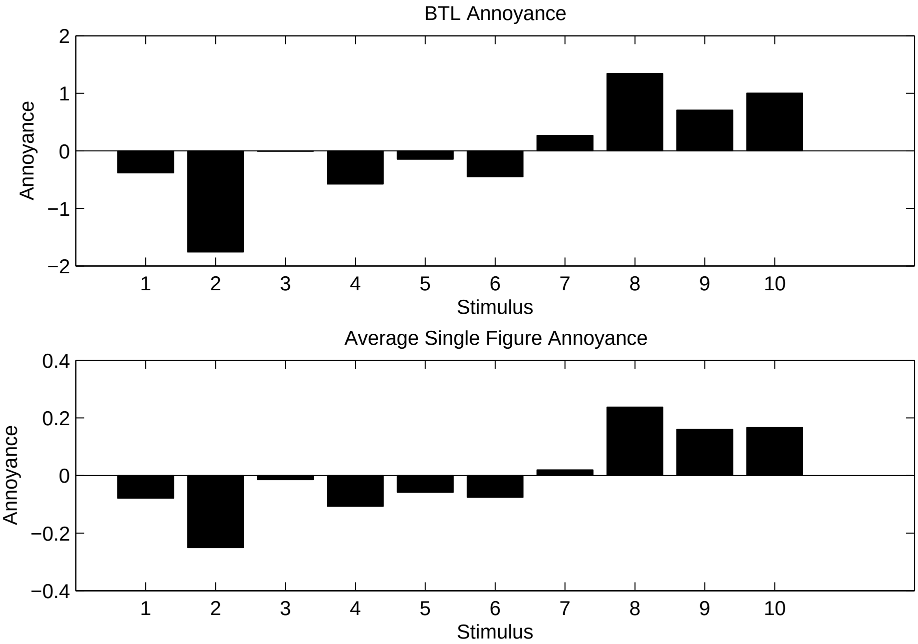

FIGURE 5.15: Single figure perceived annoyance scores for the 10 combined noise and vibration stimuli as determined by the BTL model and the averaged single figure annoyance score

5.5.6 Categorical annoyance model

The perceived single figure annoyance scores derived in Section 5.5.5 from the paired comparison subjective test are relative to the set of stimuli upon which the judgements are made, and are of an arbitrary scale. Though a higher single figure annoyance score denotes a greater level of annoyance, it is unknown how that annoyance may relate to an absolute scale of annoyance. In order to be able to relate the single figure annoyance scales to absolute categorical annoyance scales commonly used in field studies (for example, the field study by Waddington et al. (2014)), subjects were asked during the subjective test to rate the annoyance of each stimulus on a categorical scale (see Section 5.4.5). Figure 5.16 shows the proportion of subjects rating each of the combined noise and vibration stimuli in a given annoyance category during the categorical annoyance section of the subjective test. The figure shows a general trend of stimuli with higher single figure annoyance scores being rated in the higher annoyance categories more often.

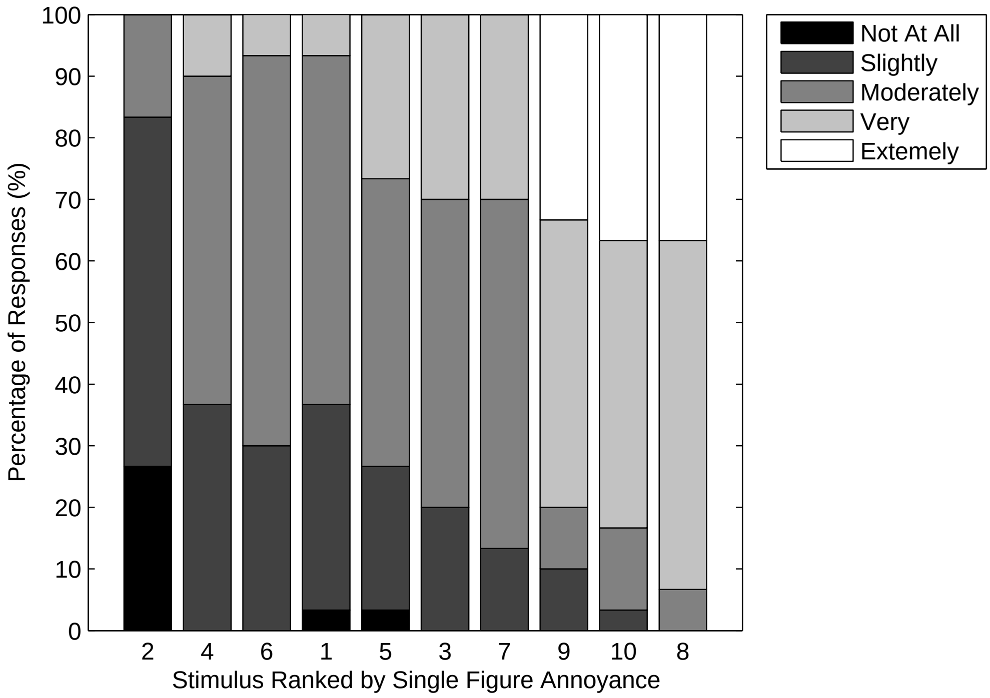

FIGURE 5.16: Proportion of respondents rating each stimulus in a certain annoyance category. Stimuli are ranked according to their Thurstone's Case V single figure annoyance scores.

Despite the clear general trend shown in Figure 5.16, there is still a degree of variance in the categorical ratings, with the ratings attributed to each stimulus spanning at least three, and sometimes four, annoyance categories. This illustrates an advantage of using paired comparison tests to determine annoyance ratings for a set of stimuli. In a paired comparison test, there is always a reference stimulus, resulting in relatively consistent inter-subject judgements. However, when judgements are made upon a single stimulus on an absolute scale, there is no common reference point, and the judgements are likely to be more affected by the subject's personal experience. As experience and perception varies a great deal between subjects, the subjective responses of absolute judgements vary also. This variance can also be seen in the exposure-response relationships derived in Chapter 4, since the responses used to develop the relationships were recorded on an absolute categorical scale.

However, the categorical annoyance scores can still be used to anchor the single figure annoyance scores, determined using paired comparison methods, to an absolute scale. This is a beneficial process, since the single figure annoyance scores are of an arbitrary scale. Though it can be said that a stimulus with a single figure annoyance score of 2 is perceived as more annoying than a stimulus with a single figure annoyance score of 1, it difficult to interpret what exactly an annoyance score of 1 or 2 represents without anchoring the scores to an absolute categorical scale. From the categorical annoyance ratings presented in Figure 5.16, a single figure categorical annoyance rating for each stimulus was calculated by taking the mode of the annoyance ratings for each stimulus and coding them as follows: 1 = not at all, 2 = slightly, 3 = moderately, 4 = very and 5 = extremely. These single figure categorical annoyance ratings are compared to the Thurstone's Case V single figure annoyance scores in Figure 5.17. Note that determining the single figure categorical annoyance ratings using the mean and the median of the categorical annoyance ratings gives the same result. The Spearman's correlation coefficient between the single figure categorical annoyance ratings and the Thurstone's Case V single figure annoyance scores is 0.87 ($p < 0.001$).

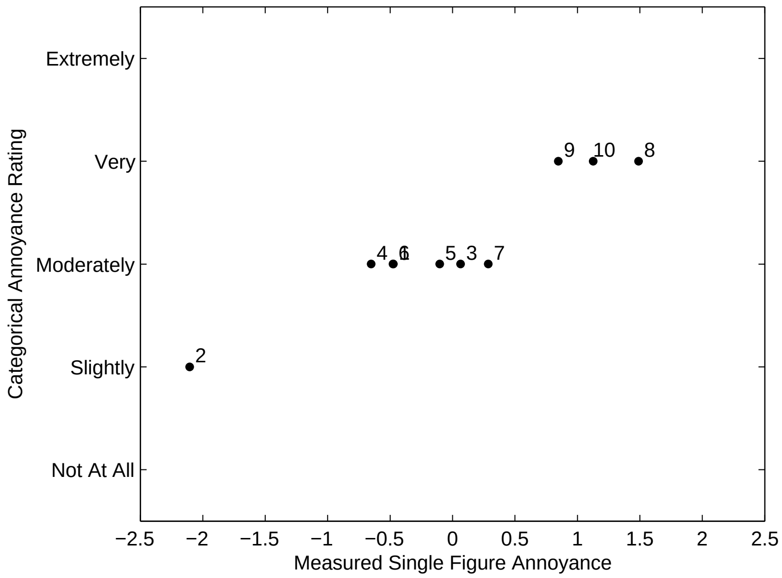

FIGURE 5.17: Relationship between measured Thurstone's Case V single figure annoyance and categorical annoyance ratings

A predictive relationship can be defined by relating the single figure annoyance scores to categorical annoyance ratings through ordinal regression. Ordinal regression is preferred to linear regression techniques in this case, due to the ordinal nature of the categorical annoyance scale. Ordinal regression can be computed with the continuous Thurstone's Case V single figure annoyance scores as the independent variable and the categorical single figure annoyance scores as the dependent variable using the the following equation:

$$\ln(-\ln(1 - p_{ij})) = \beta_{tj} + \beta A_T \tag{5.20}$$

where p_{ij} is the probability that stimulus i falls into the jth category, β_{tj} is the threshold coefficient for the jth category, β is a regression coefficient and A_T is the Thurstone's Case V single figure annoyance. The threshold coefficient and the regression coefficient are determined using maximum likelihood. The left hand side of Equation 5.20 is known as the link function and the particular link function in this equation is commonly referred to as the complementary log-log link function. Other link functions were investigated (logit, negative log-log, probit and cauchit), but the complementary log-log link function provided the best model fit. This is likely due to the fact that the this link function provides more successful predictions when stimuli are more likely to fall in higher categories, which is the case for the data set in question (McCullagh and Nelder, 1989). The result of the ordinal regression is presented in Figure 5.18 and details of the ordinal regression model are presented in Table 5.2. Note that, since the categorical single figure annoyance scores only fell in the categories of "Slightly", "Moderately" and "Very" annoying, the relationship shown in Figure 5.18 can only be used to predict the probability of a stimulus being rated within this range of annoyance categories. The p-value of the "Moderately/Very" threshold is not significant, meaning that the threshold value is not significantly different from zero. This could cast doubt over the model if the independent variable took only positive values. However, the independent variable, the single figure annoyance score, can be positive, negative or zero, so a threshold coefficient close to zero still has relevance.

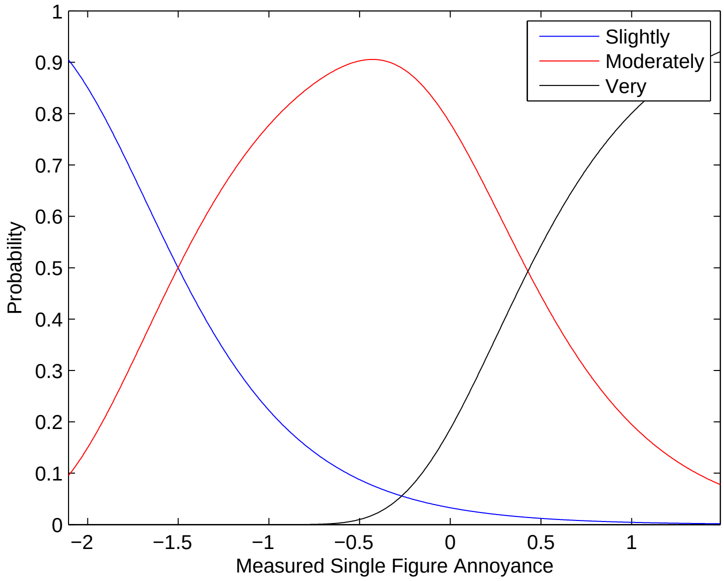

FIGURE 5.18: Probability of a combined noise and vibration stimulus with a given single figure annoyance score being rated in a certain category of absolute annoyance

Parameter	β Estimate	Standard Error	p-value	Overall Model	
Threshold				N	10
Slightly/Moderately	-3.40	1.56	< 0.050	p-value	< 0.001
Moderately/Very	0.52	0.59	n.s.	R^2_{pseudo}	1.00
Location					
A_T	2.02	0.84	< 0.050		

TABLE 5.2: Parameter estimates and other details of the ordinal regression model with the continuous Thurstone's Case V single figure annoyance scores as the independent variable and the categorical single figure annoyance scores as the dependent variable

5.6 Regression models for predicting annoyance due to combined noise and vibration

Several previous studies have looked at the effects of combined railway noise and vibration on annoyance. In two studies, Howarth and Griffin (1990; 1991) investigated the effects of combined railway noise and vibration, finding that the overall annoyance response depends on the magnitude of both stimuli, and that a reasonable approximation of the total annoyance caused can be determined from a summation of the effects of the individual stimuli. A study by Paulsen and Kastka (1995) found that, though assessment of combined noise and vibration stimuli from a passing tram and a hammermill was dominated by noise, it was also influenced by simultaneously perceivable vibration.

To investigate the effects of combined noise and vibration on overall annoyance, a multiple linear regression analysis was performed, firstly for noise and vibration alone, and then for noise and vibration combined, in a similar manner to that employed in the studies mentioned above. Multiple linear regression is a technique which can derive a model whereby several predictor variables are used to model a single response variable (Weisberg, 2005). The regression model takes the following form:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} \tag{5.21}$$

where $\mathbf{Y}$ is a vector of responses, $\mathbf{X}$ is a matrix of predictor variables and $\boldsymbol{\beta}$ is a vector of parameters to be estimated by the regression model. The parameters are estimated through the means of a least squares regression, by minimising the following function:

$$\mathrm{RSS}(\boldsymbol{\beta}) = \sum_i (y_i - x_i^{\mathrm{T}}\boldsymbol{\beta})^2 \tag{5.22}$$

where RSS is the residual sum of squares, y_i is the ith response, and x_i^{T} is the transpose of the ith row of $\mathbf{X}$.

Regression analyses were performed with the Thurstone's Case V single figure annoyance as the dependent variable, and the magnitude of noise alone, the magnitude of vibration alone and the magnitude of noise and vibration in combination as the independent variables. The vibration magnitude is quantified by the W_b weighted vibration dose value (VDV_b) and noise magnitude is quantified by the A-weighted equivalent continuous sound pressure level (L_{Aeq}). The results of these regression analyses are presented in Table 5.3. Though the relationship between noise magnitude and annoyance has a reasonable correlation, the correlation for the regression model with vibration magnitude alone is poor and the correlation is substantially higher for the regression model with noise and vibration magnitude combined.

Model	Regression Equation	R^2	p-value
Noise Only	$0.21L_{Aeq} - 12.0$	0.665	< 0.010
Vibration Only	$41.1\mathrm{VDV}_b - 2.76$	0.333	$n.s$
Noise & Vibration	$0.19L_{Aeq} + 30.7\mathrm{VDV}_b - 12.8$	0.843	< 0.010

Table 5.3: Regression models for annoyance predictions based on noise magnitude alone, vibration magnitude alone and combined noise and vibration magnitudes

The relationship between the measured single figure annoyance and the single figure annoyance predicted by the regression model with combined noise and vibration magnitudes is shown in Figure 5.19. Further details of this regression model are presented in Table 5.4. Though some scatter is apparent, the relationship shown in Figure 5.19 shows that a reasonable approximation of total annoyance caused by combined noise and vibration can be determined from a linear summation of the magnitudes of the individual stimuli. The standardised β coefficients in Table 5.4 give an indication of the relative weighting applied by subjects to the stimuli in this subjective test showing that, for this set of stimuli, variations in noise magnitude have a stronger effect on the total annoyance than variations in vibration magnitude.

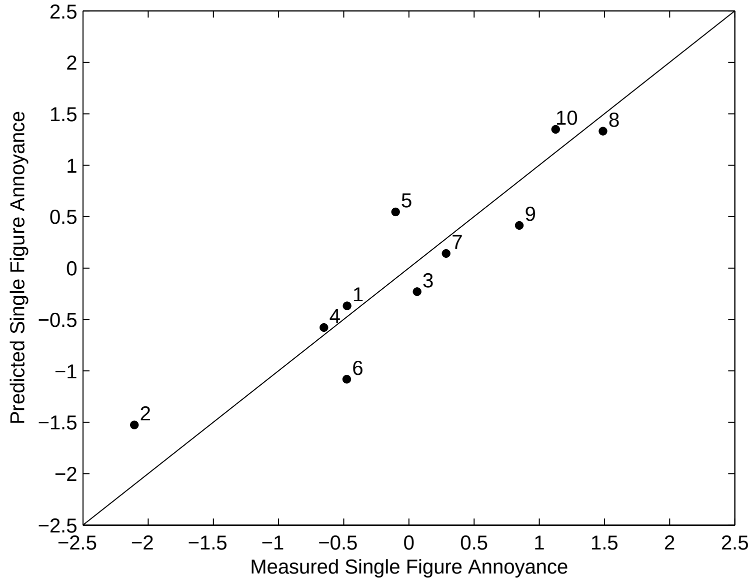

FIGURE 5.19: Comparison of single figure annoyance scores measured during the subjective test and those predicted using the regression model with combined noise and vibration magnitudes ($R^2 = 0.843$, $p < 0.010$)

Parameter	β Estimate	Standard Error	Standardised β Estimate	p-value	Overall Model	
Intercept	-12.8	2.23		< 0.010	N	10
L_{Aeq}	0.19	0.04	0.73	< 0.050	p-value	< 0.010
VDV_b	30.7	10.87	0.43	< 0.010	R^2	0.843

TABLE 5.4: Parameter estimates and other details of the multiple linear regression model with combined noise and vibration magnitudes

5.7 Summary

This chapter has presented the design and implementation of a subjective test on the perception of combined noise and vibration. In the subjective test, 30 subjects were asked to make a paired comparison judgement and a judgement of pairwise dissimilarity of all possible pairs of 10 stimuli composed of combined noise and vibration from railway passbys. An analysis of circular error rates of the subjects participating in the study suggests that the subjective test was a relatively simple task to complete resulting in all subjects making relatively consistent judgements throughout the test.

The results of the paired comparison judgements were used to model single figure annoyance scores using Thurstone's Case V model, the Bradley-Terry-Luce model and the averaged single figure annoyance scores. The freight train noise and vibration stimuli gave consistently higher perceived single figure annoyance scores. The single figure annoyance scores were also related to categorical annoyance ratings, with the highest recorded mode of category annoyance being "very" annoying, with only the freight train stimuli being assigned to this category.

Multiple linear regression models were derived for the prediction of single figure annoyance as a function of noise magnitude only, vibration magnitude only and noise and vibration magnitudes combined. The correlation coefficient for the regression model with combined noise and vibration magnitudes was higher than that for both regression models with noise magnitude alone and vibration magnitude alone, suggesting that overall annoyance due to combined railway noise and vibration stimuli is based on a summation of the effects of the individual stimuli.

Though a reasonable approximation of the total annoyance caused by combined railway noise and vibration can be determined from classic models of annoyance prediction involving summations of the noise and vibration magnitudes, there is still a substantial amount of scatter when comparing measured and predicted single figure annoyance scores (see Figure 5.19). Several studies have shown that the perception of noise is a multidimensional phenomenon that can be explained by multiple objective parameters that make up a perceptual space, using a technique known as multidimensional scaling (Grey, 1977; McAdams et al., 1995; Parizet et al., 2008; Schroeder et al., 1974; Trollé et al., 2014). Likewise, a study by Woodcock et al. (2014a) found that the perception of vibration can be explained by a multidimensional perceptual space using multidimensional scaling. However, no research has yet been performed on the perception of combined noise and vibration as a multidimensional phenomenon. Therefore, in Chapter 6, a multidimensional scaling analysis is performed on the results of the subjective test presented in this chapter, in the hope of determining the perceptual dimensions that underlay the perception of combined noise and vibration. This knowledge can then be used to develop models that are able to predict overall annoyance as a function of these perceptual dimensions, potentially with a greater degree of accuracy than models based on a summation of the magnitudes of the noise and vibration stimuli alone.

Chapter 6

The multidimensional perception of railway noise and vibration

6.1 Introduction

Chapter 5 details a laboratory study focussing on the subjective response to combined railway noise and vibration. During the test, as well as the subjects comparing each pair of stimulus in terms of their annoyance, each subject was asked to make a judgement on the dissimilarity of each pair of stimuli. This information can be used, in a method known as multidimensional scaling, to determine a multidimensional perceptual space upon which subjects make their judgements about a set of stimuli. These dimensions can then be related to objective parameters of the stimuli, allowing a perceptual annoyance model to be determined as a function of these objective parameters.

This chapter details the theory behind multidimensional scaling and the methods used to determine the perceptual dimensions used by the subjects in the laboratory test to make their dissimilarity judgements. An investigation into each perceptual dimension, and which objective parameters they may represent, is then presented. Finally, new models of annoyance due to combined railway noise and vibration, as determined by the perceptual dimension analysis, are presented.

6.2 Multidimensional scaling: theory and application

Multidimensional scaling (MDS) is a method by which measurements of similarity (or dissimilarity) among pairs of objects are represented as distances between coordinates on a multidimensional space. Borg and Groenen (2005) describe the main purposes of MDS as follows:

1. A method that represents (dis)similarity data as distances in a low-dimensional space in order to make these data accessible to visual inspection and exploration

2. A technique that allows one to test if and how certain criteria by which one can distinguish between different objects of interest are mirrored in corresponding empirical differences of these objects

3. A data-analytic approach that allows one to discover the dimensions that underlie judgements of (dis)similarity

4. A psychological model that explains judgements of (dis)similarity in terms of a rule that mimics a particular type of distance function

In this work, MDS will be used as a tool to determine the perceptual dimensions upon which subjects make their judgements of similarity for combined noise and vibration stimuli. These dimensions will then be related to objective properties of the noise and vibration stimuli and utilised to develop a perceptual model that explains judgements of dissimilarity and perceived annoyance.

6.2.1 Similarity judgements and multidimensional scaling

The methods of MDS rely on the related concepts of psychological similarities and psychological distances which are widely used in the field of cognitive psychology. Coombs' theory of data (Coombs, 1960) suggests that when subjects are presented with pairs of stimuli and asked to judge how similar they perceive the pair to be, the resulting judgement will take the form of a proximity relation. The quantity of similarity therefore represents a "distance" between two sets of coordinates in a psychological space. When judgements are made between all possible pairings of a set of stimuli, the resulting proximities between stimuli relate to coordinates in a multidimensional psychological space which describes the response of a subject to a set of stimuli. Analysing this multidimensional psychological space can therefore allow the understanding of the structure and dimensions of the psychological space on which perceptual judgements are made.

MDS is an exploratory technique that can be used to investigate these psychological dimensions that underlie the perception of a set of stimuli. It develops a configuration of a set of stimuli in an m-dimensional space to provide a representation of pairwise (dis)similarities and hence psychological distances between the stimuli. A commonly used example that is presented in many texts (Borg and Groenen, 2005; Cox and Cox, 2001; Mardia et al., 1979) is the use of MDS to analyse pairwise distances between cities. Figure 6.1 shows a two dimensional solution of such analysis of distances between cities in the United States, showing remarkable similarity to the same data represented on a geographical map, as shown in Figure 6.2 for comparison. Other historical uses of MDS, representing the great variety in its potential applications, include determining groups of similar whiskies (Lapointe and Legendre, 1994), investigating the proximity structure in a colony of Japanese monkeys (Corradino, 1990) and analysing the perceptual parameters used by bees to distinguish colour (Backhaus et al., 1987).

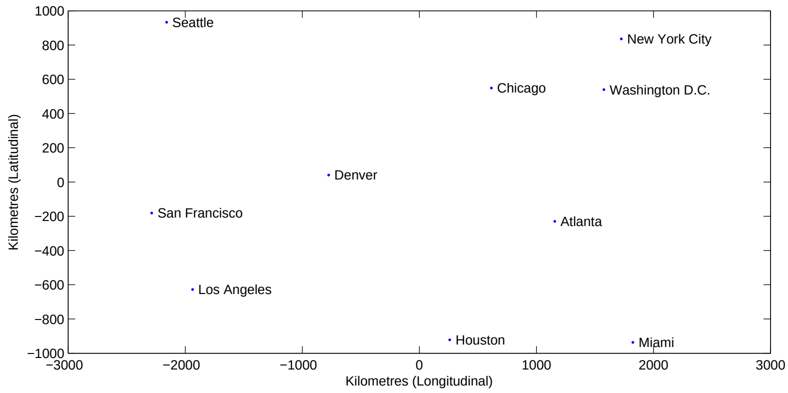

FIGURE 6.1: Two dimensional MDS solution of pairwise distances between cities in the United States

In a paired comparison subjective test, where subjects are presented with every possible pairing (i, j) of n stimuli and asked to judge how dissimilar they perceive the pair to be, a matrix of pairwise dissimilarities, δ_{ij}, can be constructed. This dissimilarity matrix can then be analysed using MDS methods to find the optimum representation of the stimuli in an m-dimensional space, with a large distance, d_{ij}, between stimuli in this space representing a large judged dissimilarity, δ_{ij}. Studying the configuration of the stimulus coordinates in the space, and the dimensions along which they lie, allows the identification of perceptual attributes used by subjects to make their judgements.

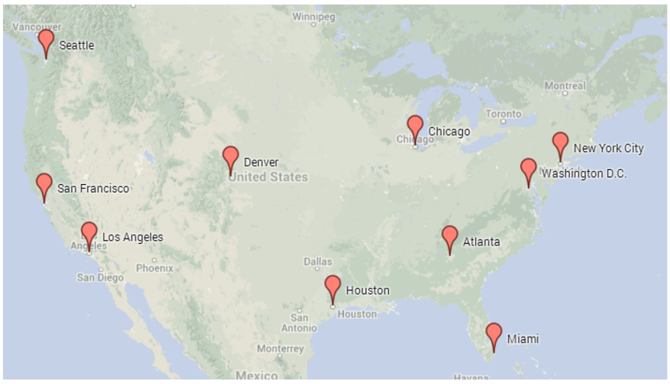

Figure 6.2: Geographical representation of distances between cities in the United States, created using Google Maps

Within the field of acoustics, MDS has been used in several studies to analyse the perception of noise and vibration. Grey (1977) studied the perception of musical timbres, discovering that the perception of timbre is related to: spectral energy distribution, the presence of synchronicity in the transients of the higher harmonics, spectral fluctuation and the presence of low amplitude high frequency energy in the initial attack sections of the musical tones. A similar study by McAdams et al. (1995) found the perception of musical timbre to be related to log-rise time, spectral centroid and the degree of spectral variation. Schroeder et al. (1974) studied the perception of concert hall sound quality, reporting correlations of the preference with objective parameters of the concert halls, including: volume, width, reverberation time, initial time-delay gap, "definition" and "internal coherence". Parizet et al. (2008) studied the perceived sound quality of car door closure sounds using MDS, discovering the particular importance of two timbral parameters which quantify the frequency balance of the sound and its "cleanness" (a clean sound is one in which only one temporal event is audible). Notably, the importance of the timbral parameters were found both when judgements were made in terms of similarity and when judgements were made in terms of preference, suggesting that the dimensional spaces for similarity and preference may be related. A study by Trollé et al. (2014) analysed the short term annoyance due to tram noise, determining that the perception of tram noise is related to the A-weighted equivalent sound pressure level, the variance of time-varying A-weighted pressure normalised by root mean square A-weighted pressure and a new index, the total energy of the tonal components within critical bands from 12 to 24 barks, which they term $TETC$. A precursor to this study, performed by Woodcock et al. (2014a), studied the perception of railway vibration using MDS, relating the perception of vibration to root mean square acceleration in the 16 Hz and 32 Hz octave bands, the duration defined by the 10 dB downpoints of the vibration envelope and the modulation frequency of the vibration envelope.

6.2.2 Metric and non-metric methods of scaling

In general, MDS attempts to represent perceptual similarities by distances between coordinates of an m-dimensional Euclidean configuration, $\mathbf{E}$, termed the MDS space. The optimal configuration is one such that the distances between points in the multidimensional space, d_{ij} are approximately equal to $f(\delta_{ij})$, where f is a transformation function whose parameters are to be estimated, and δ_{ij} are measured pairwise distances. A common method to achieve this is fitting the matrix of distances, d_{ij}, to the matrix of $f(\delta_{ij})$ by least squares regression or eigendecomposition. A squared error of representation between the distances in the MDS space, $d_{ij}(\mathbf{E})$ and the function $f(\delta_{ij})$ is defined as follows (Borg and Groenen, 2005):

$$e_{ij} = (f(\delta_{ij}) - d_{ij}(\mathbf{E}))^2 \tag{6.1}$$

Summing e_{ij} over all pairs (i, j) yields a measure of fit for the entire MDS representation, termed the raw stress:

$$\sigma_r = \sum_{(i,j)} (f(\delta_{ij}) - d_{ij}(\mathbf{E}))^2 \tag{6.2}$$

The raw stress, however, is not very informative as its value depends on the metric used to measure pairwise distances. Using the example of distances between cities (see Figure 6.1), measuring the distances in metres would give rise to a higher raw stress due to differences between $f(\delta_{ij})$ and $d_{ij}(\mathbf{E})$ being greater in terms of metres than in terms of kilometres. To avoid this dependency on scale, the stress is commonly normalised as follows:

$$\sigma_1^2 = \frac{\sigma_r}{\sum d_{ij}^2(\mathbf{E})} = \frac{\sum (f(\delta_{ij}) - d_{ij}(\mathbf{E}))^2}{\sum d_{ij}^2(\mathbf{E})} \tag{6.3}$$

Taking the square root of σ_1^2 (as in Equation 6.4) results in a value commonly known as “Stress-1” (Kruskal, 1964). The square root is taken since σ_1^2 values are almost always extremely small in practice and hence σ_1 values tend to be easier to discriminate. This is analogous to taking the standard deviation in place of the variance.

$$\text{Stress-1} = \sigma_1 = \sqrt{\frac{\sum (f(\delta_{ij}) - d_{ij}(\mathbf{E}))^2}{\sum d_{ij}^2(\mathbf{E})}} \tag{6.4}$$

The MDS model can then be optimised by minimising Stress-1, estimating the parameters of f and finding an optimal configuration $\mathbf{E}$ for a given dimensionality, m. The form taken by the transformation function f is largely dependent on the form of the measured distances δ_{ij}. If the data to be analysed are of an interval or ratio scale, metric multidimensional scaling can be used. For this method of scaling, a constraint is imposed such that the function f must be continuous and monotonic. For the ratio transformation, the transformed proximities are proportional to the original proximities. For the interval transformation the transformed proximities are proportional to the original proximities, plus an intercept term. For data of an ordinal or spline scale, however, non-metric multidimensional scaling is considered more appropriate (Borg and Groenen, 2005). In non-metric MDS, f may take the form of a non-parametric monotonic function. For the ordinal transformation, the rank order of the distances, d_{ij}, is preserved. For spline transformation, the transformed proximities are a smooth non-decreasing piecewise polynomial transformation of the original proximities. A summary of metric and non-metric transformations is presented in Table 6.1.

Transformation	Model
Absolute	$f(\delta_{ij}) = \delta_{ij}$
Ratio	$f(\delta_{ij}) = \beta\delta_{ij}$, $\beta > 0$
Interval	$f(\delta_{ij}) = \beta_0 + \beta\delta_{ij}$, $\beta_0 \geq 0$ and $\beta \geq 0$
Spline	A sum of polynomials of δ_{ij}
Ordinal	Preserve the rank order of δ_{ij} in $f(\delta_{ij})$

TABLE 6.1: Forms of the transformation function f for metric and non-metric MDS models (Borg and Groenen, 2005)

6.2.3 Individual difference scaling

Returning once again to the example of pairwise distances between cities in the United States: scaling this dataset requires minimising the stress function in Equation 6.4 using one of the metric or non-metric parameter models summarised in Table 6.1 to transform the measured distances. In this case, there is only one matrix of measured pairwise distances, δ_{ij}, since the distance between cities is a fixed value and should, in theory, be the same no matter how many times it is measured and regardless of the methods used for the measurement. This is not the case when considering paired comparison data from multiple subjects. Each subject will have a unique matrix of dissimilarity judgements and it is important to consider how such data should be aggregated. A simple method of aggregation would be to average responses across the subject group but this would lose important information about inter-subject variability. To retain information on the inter-subject variability, a method of MDS that retains individual subject differences was introduced by Horan (1969) and developed into an algorithm known as INDSCAL (Individual Difference Scaling) by Carroll and Chang (1970).

The INDSCAL algorithm preserves inter-subject differences by defining a "group space" MDS configuration that is common to all subjects and a "subject space" which represents the relative weightings that each subject attributes to each dimension that makes up the group space. A “private space” can therefore be defined for each subject which represents the group space MDS configuration modified by that subjects dimension weightings. A simple example of a subject space derived from an arbitrary data set with 3 stimuli, 2 dimensions and 4 subjects is shown in Figure 6.3. This subject space indicates the relative weightings that each of the 4 subjects attributes to the 2 dimensions, as well as an indication of how well each subject’s data is represented by the group space. The angle between the subject vectors and each dimension in the subject space relates to the relative weighting that the subject attributes to each of the dimensions. The magnitude of each vector from the origin relates to how well the subject’s data is represented by the aggregate group space. For example, subjects 1 and 2 place equal weighting on each dimension in the group space, represented by the equal angle between the subject vectors and each dimension. However, the dissimilarity judgements of subject 2 are better reproduced by the configuration of the group space than those of subject 1, represented by the vector of subject 2 having a greater magnitude than that of subject 1. In contrast, though the dissimilarity judgements of subject 3 and 4 appear well represented by the group space (represented by the equal magnitude of their vectors from the origin), they place greater weighting on dimensions 2 and 1 respectively.

The private space of each subject can be derived by scaling the dimensions of the group space with respect to the square root of the weightings as shown in the subject space:

$$y^n_{jm} = \sqrt{w^n_m} x_{jm} \tag{6.5}$$

where y^n_{jm} is the coordinate of the jth stimulus on the mth dimension in the private space of subject n, w^n_m is the weighting attributed by subject n on the mth dimension and x_{jm} is the coordinate of the jth stimulus on the mth dimension in the group space. The distance between stimuli i and j in the private space of subject n is therefore:

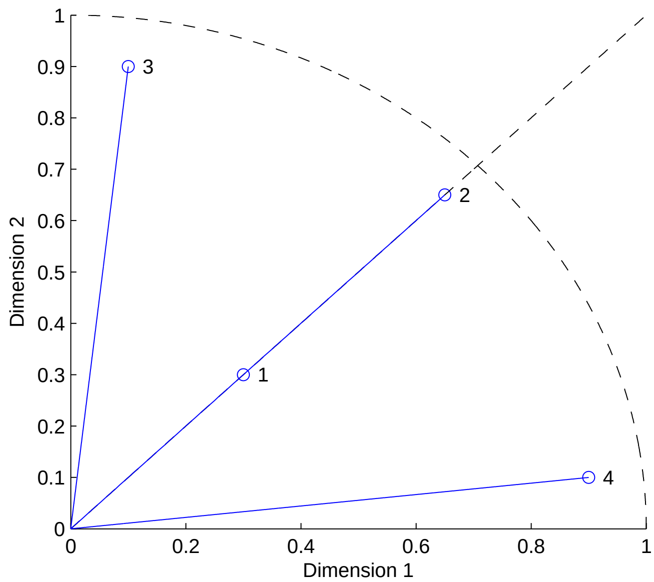

FIGURE 6.3: Example subject space of an arbitrary data set with two dimensions and 4 subjects

$$d_{ij}^n = \sqrt{\sum_m w_m^n (x_{jm} - x_{im})^2} \tag{6.6}$$

The concept of private spaces and weightings is illustrated in Figure 6.4, which shows the aggregated group space shared by all subjects and the private spaces of subjects 3 and 4 from the same data set as represented in the subject space shown in Figure 6.3. The stimulus coordinates in the private space for subject 3 is comparable to those of the group space in the 2nd dimension, due to the high weighting that the subject places on this dimension. The coordinates in the 1st dimension, however, have been scaled down a great deal due to the low weighting that the subject places on this dimension. The exact opposite is true for subject 4.

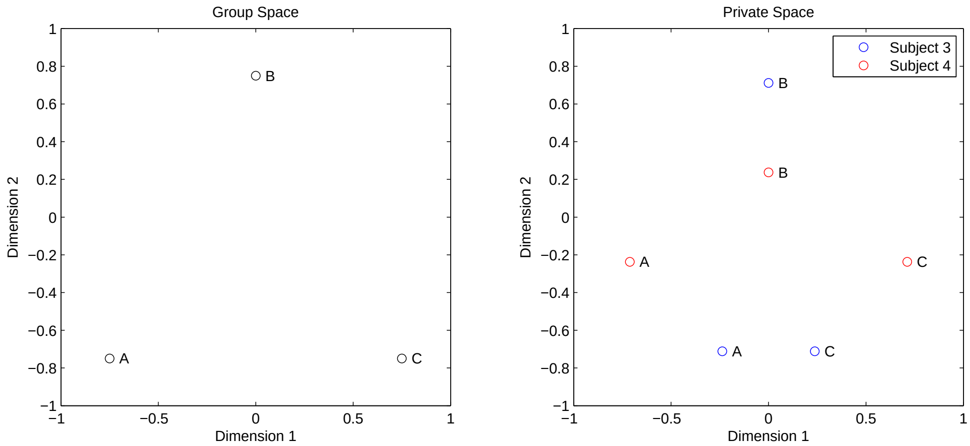

FIGURE 6.4: Example group space and private space of two subjects with different dimensional weightings

6.3 Multidimensional scaling of the dissimilarity judgements from the laboratory study

In Section 6.2, the theory of MDS, and its varied applications are discussed. In this section, the theory of MDS will be applied to the results of the laboratory study as described in Chapter 5. In the laboratory study, as well as being asked to make annoyance comparisons, subjects were asked for each pair of stimuli: “How similar do you perceive the pair to be?” (see the user interface in Figure 5.9). The subject’s responses were recorded on a 101 point scale ranging from 0, “very similar”, to 100, “very different”. The subjects were encouraged to make use of the full scale when making their judgements. These similarity judgements can be used to determine the perceptual dimensions used by the subject group to make their similarity judgements using the methods of MDS as described in Section 6.2.

6.3.1 Multidimensional scaling with PROXSCAL

The MDS models considered in this chapter were built and analysed using the PROXSCAL program in SPSS (IBM, 2011). PROXSCAL builds on the majorization algorithm developed by De Leeuw and Heiser (1980), which guarantees convergence of stress. PROXSCAL allows a great deal of customisation of the MDS model, with options for identity, weighted Euclidean, generalised Euclidean and reduced rank models. For the identity model, all subjects have the same configuration and no individual scaling is applied. The weighted Euclidean model is the individual difference scaled model (INDSCAL) as described in Section 6.2.3, where each subject has their own private space in which every dimension of the group space is weighted differently. The generalised Euclidean model is another form of individual scaling, where each subject has a private space that is equal to a rotation of the common space, combined with a differential weighting of the dimensions. The reduced rank model is a generalized Euclidean model for which the rank of the individual space can be specified. Proximity transformations of the function $f(\delta_{ij})$ can be of the ratio, interval, ordinal or spline form (see Table 6.1). Regardless of the model used, the PROXSCAL program derives its MDS configuration by minimising the normalised raw stress function (Equation 6.3), which represents the mean squared error between the transformed proximities and the pairwise measured distances of stimuli across all subjects. For the MDS analyses presented in this study, the weighted Euclidean individual scaling model will be used to preserve inter-subject variability. Ordinal proximity transformations will be applied. This could be interpreted as over cautious, since the responses could be interpreted on an interval scale, however, it was decided to use ordinal transformations in order to preserve the rank order of the dissimilarity judgements in the multidimensional space.

6.3.2 Considering dimensionality

When building an MDS model it is important to consider the number of dimensions that will make up the MDS configuration. Increasing the dimensionality of an MDS configuration reduces its stress at the cost of increased complexity. A convenient way to determine a reasonable number of dimensions is to create a visual representation of stress as a function of dimensionality in what is known as a "scree" plot. Figure 6.5 shows such a relationship. Kruskal (1964) suggests a reasonable choice of dimensionality is one such that the stress is acceptably small and for which a further increase in dimensionality does not significantly reduce stress. In some cases, an "elbow" in the data representing such a phenomenon can be visually identified. Such an elbow can possibly be identified in Figure 6.5 when the dimensionality is 4. Certainly the criteria that the stress is acceptably small is met with 4 dimensions, with a stress of 0.020, a value of fit which is quantified by Kruskal (1964) as "excellent" (see Table 6.2). In addition, previous research on the multidimensional perception of noise and vibration have utilised models with dimensionality of 3 (Grey, 1977; McAdams et al., 1995) and 4 (Schroeder et al., 1974; Woodcock et al., 2014a), giving further confidence that 4 perceptual dimensions is a sensible choice for the MDS configuration.

Stress	Goodness of Fit
0.200	Poor
0.100	Fair
0.050	Good
0.025	Excellent
0.000	Perfect

TABLE 6.2: Measures of goodness of fit as suggested by Kruskal (1964)

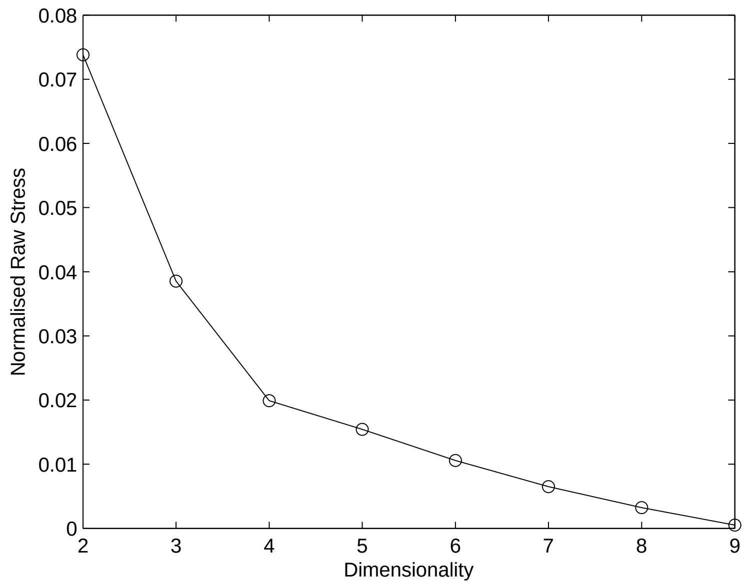

FIGURE 6.5: Normalised raw stress as a function of dimensionality

6.3.3 The four dimensional perceptual space

Figure 6.6 shows the aggregated group space for the 4 dimensional weighted Euclidean individual scaling MDS solution and Figure 6.7 shows the subject space solution. Each point in the group space represents one of the 10 combined noise and vibration stimuli and each point in the subject space represents one of the 30 subjects to take part in the study. The group space shows the stimuli are well distributed across all dimensions, suggesting that subjects are utilising the full continuum of each perceptual dimension in their dissimilarity judgements rather than simply categorising the stimuli. The subject vectors in the subject space are all clustered very close together (the subject labels are not shown for this reason) with approximately equal magnitude and angle, showing that the subjects are very consistent in their judgements, since the dissimilarity judgements of each subject are well represented by the group space, and each subject places approximately equal weighting on each perceptual dimension. Note that the data point in the bottom left of each panel in Figure 6.7 represents the origin of the subject space, and not a subject.

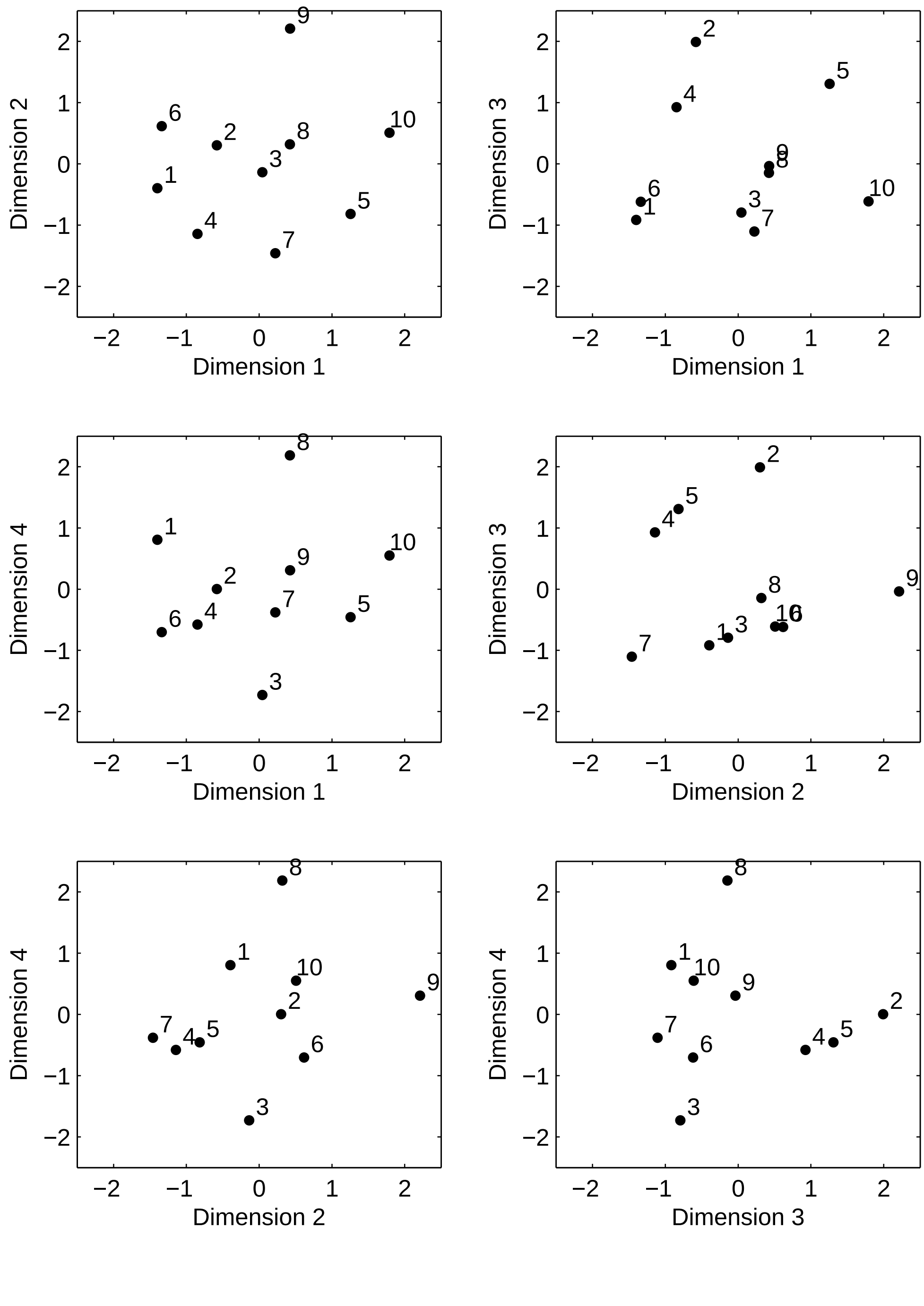

FIGURE 6.6: Aggregated group space for the 4 dimensional MDS solution

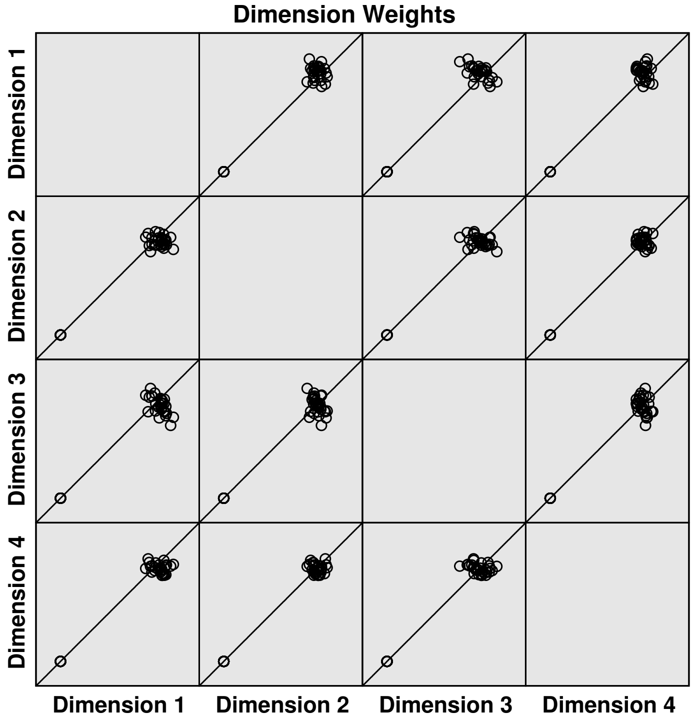

FIGURE 6.7: Subject space for the four dimensional MDS solution showing the relative weights attributed by each subject to each dimension

6.3.4 Relative annoyance model as a function of the perceptual dimensions

With the perceptual dimensions now defined, an annoyance model representing the relative annoyance scores (calculated in Section 5.5) as a function of the perceptual dimensions can be derived using multiple linear regression. The result of the multiple linear regression with Thurstone's Case V single figure annoyance scores as the response variable and the coordinates of the combined noise and vibration stimuli on the perceptual dimensions revealed through the multidimensional scaling as the predictor variables is shown in Equation 6.7, where A_p is the predicted single figure annoyance score and D_m is the position of the combined noise and vibration stimulus on the mth dimension. Figure 6.8 shows the relationship between the Thurstone Case V single figure annoyance scores measured through the subjective test and the predicted single figure annoyance scores calculated using Equation 6.7. Details of the regression model are presented in Table 6.3. Note that regression with the BTL annoyance score, and the averaged single figure annoyance score (see Section 5.5) provides models with commensurate R^2 values and significance ($R^2 = 0.848$, $p < 0.050$ and $R^2 = 0.857$, $p < 0.050$ respectively).

$$A_p = 0.56D_1 + 0.10D_2 - 0.56D_3 + 0.29D_4 \tag{6.7}$$

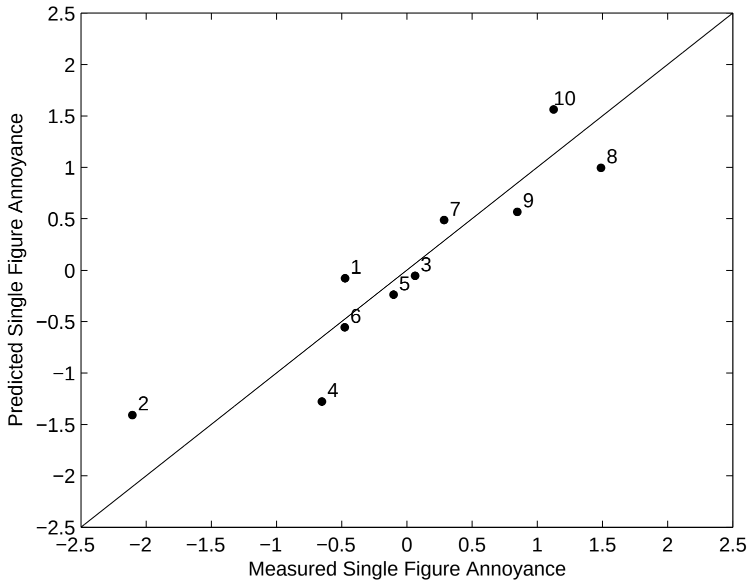

FIGURE 6.8: Comparison of single figure annoyance scores measured during the subjective test and those predicted using Equation 6.7 ($R^2 = 0.831$, $p < 0.050$)

Figure 6.8 shows a good agreement ($R^2 = 0.831$, $p < 0.050$) between the measured single figure annoyance scores and those predicted using the perceptual dimensions derived from the multidimensional scaling model. This is an important result of the MDS analysis as it shows that the perceptual dimensions of noise and vibration can be related to the overall annoyance caused by these stimuli. With the perception of noise and vibration described by this small number of perceptual dimensions, it should be possible to find objective parameters of noise and vibration that relate to these perceptual dimensions. If objective parameters can be found which correlate with these perceptual dimensions, a model can then be derived to predict

Parameter	β Estimate	Standard Error	p-value	Overall Model	
Intercept	0.000	0.180	n.s.	N	10
D_1	0.564	0.183	< 0.050	p-value	< 0.050
D_2	0.098	0.188	n.s.	R^2	0.831
D_3	-0.559	0.181	< 0.050		
D_4	0.294	0.189	n.s.		

Table 6.3: Parameter estimates and other details of the multiple linear regression model for single figure annoyance as a function of the perceptual dimensions

self reported annoyance due to noise and whole-body vibration exposure based on objective features of the noise and vibration stimuli.

6.4 Interpretations of the perceptual space

In Section 6.3, it was demonstrated that a multidimensional scaling configuration can be derived from the dissimilarity judgements from the laboratory study described in Chapter 5, and that self reported annoyance due to the noise and vibration stimuli presented in the laboratory study can be modelled using the perceptual dimensions derived from the MDS configuration. An investigation into correlations between objective descriptors of the noise and vibration stimuli and the perceptual dimensions used to predict self reported annoyance should therefore allow the development of a model that can predict annoyance due to objective parameters of the noise and vibration stimuli. The first step into such an investigation is to calculate objective features of each stimulus. Since MDS is an exploratory technique, no restriction was imposed on the objective features to be calculated and as such a wide range of features which quantify magnitude, frequency content, duration and envelope modulation are considered.

6.4.1 Correlations with objective parameters of noise

Objective parameters of noise were calculated from pressure time histories recorded using a 01dB Symphonie measurement system with an MCE212 free field microphone (frequency range = 3.15 Hz to 20 kHz, dynamic range = 15 to 146 dB, sensitivity = 50 mV/Pa). Calculated measures of the magnitude of the noise signal include the peak level (L_{max}), the root mean square pressure (rms), the equivalent continuous sound pressure level (L_{eq}), the sound exposure level (SEL) and the sound pressure level exceeded for 10% of the signal duration (L_{10}). These properties were calculated on both the un-weighted and A-weighted noise signal.

Calculated psychoacoustic parameters include the overall specific loudness, the peak specific loudness, the overall sharpness, the peak sharpness, the fluctuation strength, and the roughness. For more information on these psychoacoustic parameters and how they are calculated, see Fastl and Zwicker (2007). The psychoacoustic parameters that are found to correlate with the perceptual dimensions, namely the loudness, the sharpness and the roughness, are discussed in further detail in the following sections.

Statistical parameters of the noise signal include the skewness, the kurtosis and the crest factor. Spectral parameters of the signal include the spectral centroid and the dominant frequency of the power spectral density (f_{max}). Parameters relating to the duration of the signal include the duration defined by the 3 dB and 10 dB downpoints of the signal, and the duration exceeded by the top 1/5th, 2/5ths, 3/5ths and 4/5ths of the signal's dynamic range.

Additionally, two parameters which were found by Trollé et al. (2014) to correlate with instantaneous annoyance of tram noise were calculated, namely the variance of time-varying A-weighted pressure normalised by root mean square A-weighted pressure (VAP) and a new psychoacoustic index, the total energy of the tonal components within critical bands from 12 to 24 barks ($TETC$).

All of these parameters combined cover the magnitude, psychoacoustic, statistical, spectral and temporal characteristics of the stimuli in terms of noise and should therefore sufficiently cover potential perceptual dimensions upon which subjects in the test make their dissimilarity judgements. Table 6.4 shows the Pearson's correlation coefficient, and its significance, between the above mentioned noise parameters and the four perceptual dimensions revealed through the multidimensional scaling analysis. Only parameters that have a significant correlation with at least one dimension are shown, any parameters not appearing in the table have been found to have no significant correlation with any of the perceptual dimensions. Only the roughness parameter has a significant correlation with more than one dimension, giving confidence that each dimension represents an independent component of an overall psychological perception. These correlations will be discussed further in the following sections.

6.4.2 Correlations with objective parameters of vibration

As well as objective parameters of the noise signals of the combined stimuli, objective parameters of the vibration signals were calculated in a similar fashion. Objective parameters of vibration were calculated from acceleration time histories measured at the interface between a 70 kg subject and the seat cushion with a Svantek SV38 seat pad accelerometer (frequency range = 0.1 to 100 Hz, dynamic range = 0.01 to 50 m s^{-2}, sensitivity = 100 mV/(m s^{-2})). All parameters were calculated for vibration in the vertical axis, and parameters were calculated on un-weighted signals as well as W_b weighted and W_k weighted signals where appropriate. Parameters which quantify the magnitude of the signal include the vibration dose value (VDV), the root mean square acceleration (rms), the equivalent continuous vibration level (V_{eq}), the vibration exposure level (VEL) and the peak acceleration (a_{max}). The modulation depth is defined as the average difference between the maxima and minima of the signal envelope.

Statistical parameters of the vibration signals include the skewness, the kurtosis and the crest factor. Spectral parameters of the vibration signals include spectral centroid and the dominant frequency of the power spectral density (f_{max}) and the rms acceleration in the 4 Hz, 8 Hz, 16 Hz, 32 Hz and 64 Hz octave bands, expressed

Parameter	Dim 1	Dim 2	Dim 3	Dim 4
rms_A	0.952**	–	–	–
L_{Aeq}	0.933**	–	–	–
SEL_A	0.836**	–	–	–
L_{10}	–	–	-0.759*	–
L_{A10}	0.908**	–	–	–
L_{max}	0.853**	–	–	–
Overall Loudness	0.727*	–	–	–
Peak Loudness	0.945**	–	–	–
Overall Sharpness	–	-0.799**	–	–
Roughness	–	0.671*	–	0.650*
Kurtosis	0.754*	–	–	–
Crest Factor	0.836**	–	–	–
Spectral Centroid	–	-0.680*	–	–
f_{max}	–	–	0.713*	–
3 dB Downpoints Duration	–	–	-0.640*	–
10 dB Downpoints Duration	–	0.806**	–	–
1/5th Dynamic Range Duration	–	–	-0.649*	–
2/5th Dynamic Range Duration	–	0.792**	–	–
3/5th Dynamic Range Duration	–	0.792**	–	–
4/5th Dynamic Range Duration	–	0.801**	–	–
VAP	-0.663*	–	–	–
$TETC$	0.778**	–	–	–

Table 6.4: Pearson's correlation coefficients between objective parameters of the noise signals and the perceptual dimensions. – not significant, * $p < 0.05$, ** $p < 0.01$

both as an absolute value and as a proportion of the overall rms acceleration in the signal. The modulation frequency is defined as the inverse of the average period between the maxima of the signal envelope. Temporal parameters of the vibration signals include the duration defined by the 3 dB and 10 dB downpoints of the signal and the duration exceeded by the top 1/5th, 2/5ths, 3/5ths and 4/5ths of the signal's dynamic range.

As is the case with the objective noise parameters, these vibration parameters cover the magnitude, statistical, spectral and temporal characteristics of the vibration signals and should therefore sufficiently cover potential perceptual dimensions upon which subjects in the laboratory test make their dissimilarity judgements. Table 6.5 shows the Pearson's correlation coefficient, and its significance, between the above mentioned vibration parameters and the four perceptual dimensions revealed through the multidimensional scaling analysis. As with Table 6.4, only parameters that have a significant correlation with at least one perceptual dimension are shown, and any parameters not appearing in the table have been found to have no significant correlation with any of the perceptual dimensions. Note that no vibration parameters have been found to correlate with the first perceptual dimension, suggesting that this dimension is related to the noise component of the combined exposure. The correlations of objective parameters with each dimension in turn will be discussed in detail in the following sections.

6.4.3 Dimension 1

The correlations presented in Tables 6.4 and 6.5 suggest that the first perceptual dimension is related to objective parameters of the noise signal, since no vibration parameters show a significant correlation with this dimension. Figure 6.9 shows the relationship between the 1st perceptual dimension and the five parameters that exhibit the strongest correlation with this dimension. Though some degree of scatter is apparent, the figure gives confidence that the 1st perceptual dimension can be related to these objective parameters of noise.

Many of the strongest correlations with the 1st perceptual dimension exist with parameters that quantify the magnitude of the signal: the A-weighted rms pressure, the A-weighted equivalent continuous sound pressure level, the A-weighted sound pressure level exceeded for 10% of the time, the peak sound pressure level, and the sound exposure level. The A-weighted rms pressure, rms_A, is calculated as follows:

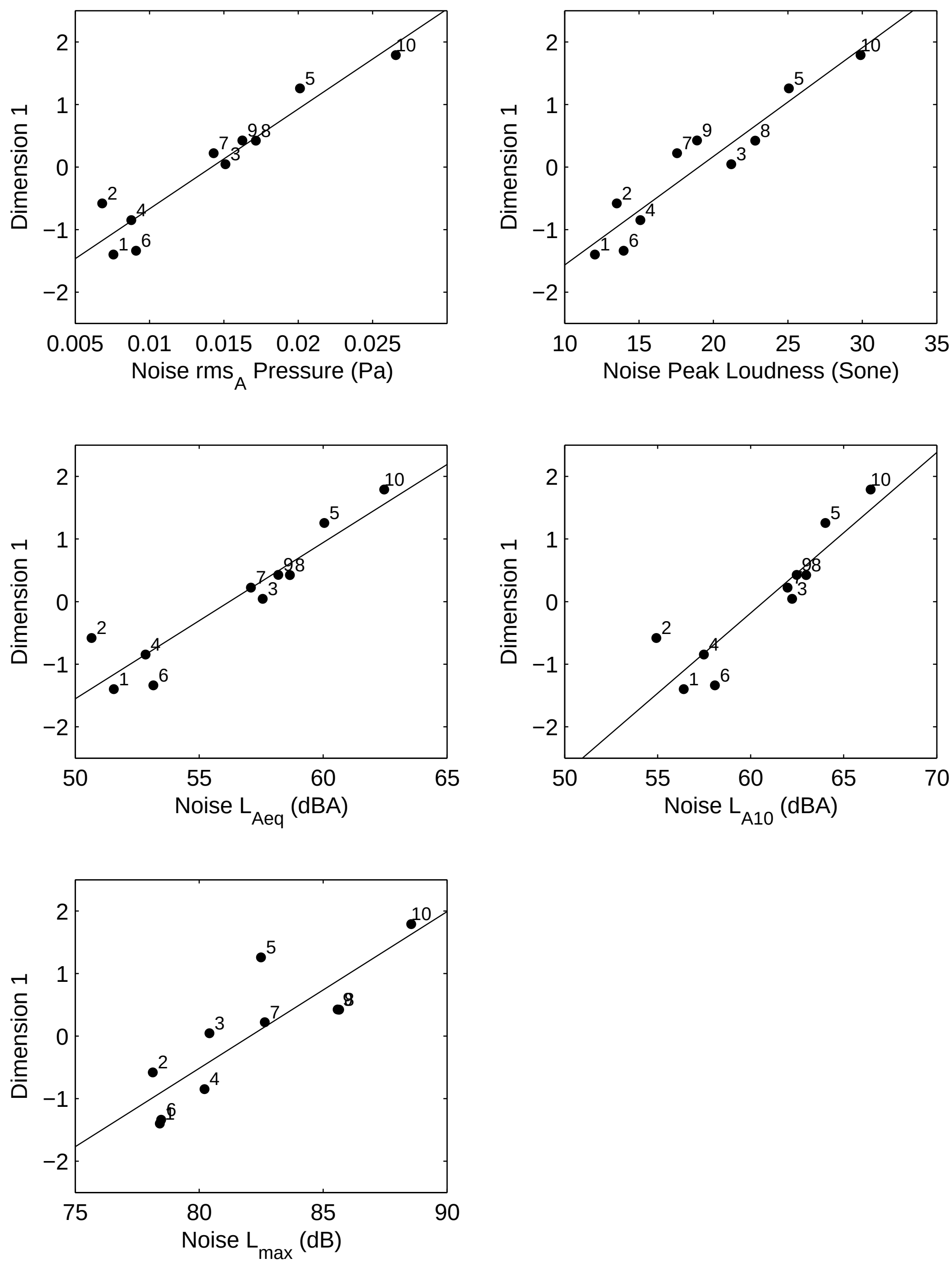

FIGURE 6.9: Relationship between the 1st perceptual dimension and the five parameters that exhibit the strongest correlation with this dimension

Parameter	Dim 1	Dim 2	Dim 3	Dim 4
VDV	–	–	–	0.699*
VDV_b	–	–	–	0.865**
rms	–	-0.773**	–	–
rms_k	–	-0.858**	–	–
V_{eq}	–	-0.807**	–	–
VEL	–	–	-0.634*	0.679*
Peak Acceleration	–	–	–	0.858**
Kurtosis	–	–	–	0.725*
Crest Factor	–	–	–	0.784**
64 Hz rms	–	-0.719*	–	–
4 Hz rms (proportional)	–	0.748*	–	–
8 Hz rms (proportional)	–	0.814**	–	–
3 dB Downpoints Duration	–	0.784**	–	–
10 dB Downpoints Duration	–	0.786**	–	–
1/5th Dynamic Range Duration	–	–	–	0.676*
2/5th Dynamic Range Duration	–	0.788**	–	–
3/5th Dynamic Range Duration	–	0.782**	–	–
4/5th Dynamic Range Duration	–	0.782**	–	–
Modulation Frequency	–	-0.636*	–	-0.641*

TABLE 6.5: Pearson's correlation coefficients between objective parameters of the vibration signals and the perceptual dimensions. – not significant, * $p < 0.05$, ** $p < 0.01$

$$rms_A = \sqrt{\frac{1}{T}\sum_{t=0}^{T} p_A(t)^2} \tag{6.8}$$

where $p(t)$ is a pressure time history of T seconds. The A-weighted continuous sound pressure level, L_{Aeq}, is calculated as follows:

$$L_{Aeq} = 10\log_{10}\left(\frac{1}{T}\int_0^T 10^{L_{pA}(t)/10}dt\right) \tag{6.9}$$

where $L_{pA}(t)$ is the time-varying A-weighted sound pressure level. The A-weighted sound exposure level, SEL_A, is calculated as follows:

$$SEL_A = L_{Aeq} + 10\log(T) \tag{6.10}$$

The sound exposure level is the sound level of duration 1 second that would have the same energy content as the whole event. It is therefore a function of not only the energy of the signal but its duration. The sound exposure level is used in the measurement and assessment of railway noise (Department of Transport, 1995).

Correlations also exist with the measures of overall loudness, and the peak loudness of the pressure-time signal. Loudness is a psychoacoustic metric used to quantify the intensity of a signal as perceived by the ear (Fastl and Zwicker, 2007). Loudness is expressed in sones, where 1 sone is the level given by a pure 1 kHz tone at 40 dB. The quantities of loudness and peak loudness of the time varying signal were calculated using the dBFA32 software of the 01 dB Symphonie measurement system.

Strong correlations also exist with the statistical measures of kurtosis and the crest factor, both of which quantify the "peakiness" of the noise signal. Kurtosis of the noise signal is calculated as follows:

$$K_t = \frac{1}{T\sigma^4}\sum_{t=0}^{T}[p(t) - \bar{p}]^4 \tag{6.11}$$

where σ is the standard deviation of a time-pressure signal with mean pressure $\bar{p}$. The crest factor is calculated as follows:

$$C_r = \frac{p_{max}}{rms} \tag{6.12}$$

where p_{max} is the peak pressure of the noise signal. Finally the 1st perceptual dimension has a significant correlation with two parameters found by Trollé et al. (2014) to relate to annoyance due to tram noise: VAP and $TETC$. VAP is the variance of of the time-varying A-weighted pressure signal, p_A, normalised by the A-weighted rms pressure as follows:

$$VAP = \frac{var(p_A)}{rms_A} \tag{6.13}$$

where *var* is the finite sample variance operator. Trollé et al. (2014) suggest that VAP may quantify the irregular/continuous character, or the overall pressure rise, of a noise signal. $TETC$ is defined by Trollé et al. (2014) as follows:

$$TETC = 10\log_{10}\left(\int_{12}^{24} 10^{L(z)/10} dz\right) \tag{6.14}$$

where $L(z)$ represents the maximal (across time) level of the tonal components as a function of the critical-band rate, z. $TETC$ was specifically proposed by Trollé et al. (2014) to capture squeal noise in trams, and hence emphasises tonal components at high frequencies. They found that $TETC$, combined with the equivalent continuous A-weighted sound pressure level, L_{Aeq}, provides a satisfactory model for predicting short-term annoyance due to tram noise.

In summary, all of the parameters that show a significant correlation with the 1st dimension are descriptors of the noise component of the combined stimuli. The correlating parameters quantify the magnitude, loudness and "peakiness" of the signal, as well as the psychoacoustic parameters of VAP, which quantifies the irregular/continuous character of the signal, and $TETC$ which emphasises high frequency components, such as wheel squeal, of a noise signal.

6.4.4 Dimension 2

The correlations presented in Tables 6.4 and 6.5 suggest the second perceptual dimension could be related to objective characteristics of either the noise or the vibration signal signal. Figure 6.10 shows the relationship between the second perceptual dimension and the five parameters which have the strongest significant correlation with this dimension. As with relationships with the 1st dimension, though some scatter is apparent, the figure gives confidence that the 2nd perceptual dimension can be related to these objective parameters.

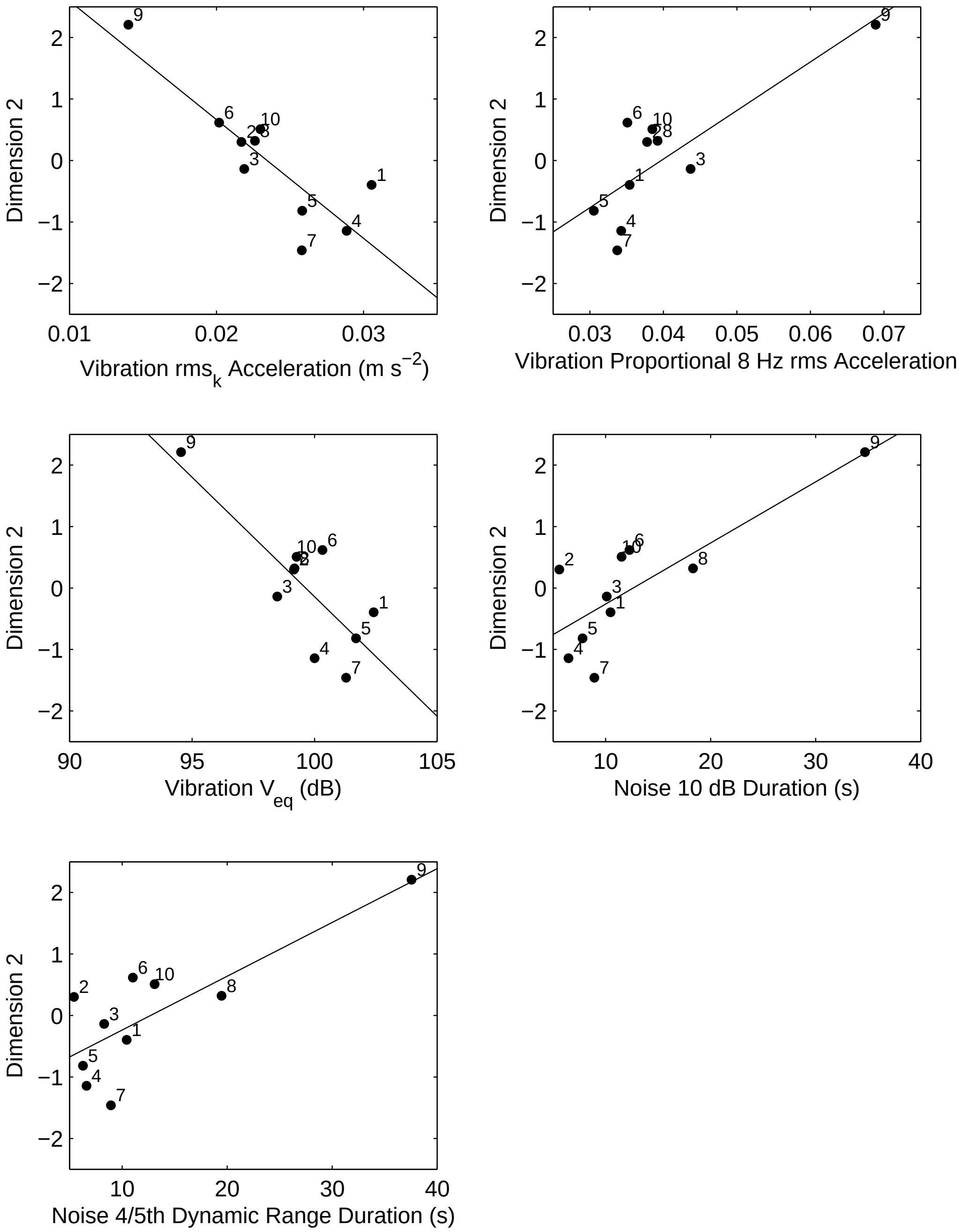

FIGURE 6.10: Relationship between the 2nd perceptual dimension and the five parameters that exhibit the strongest correlation with this dimension

Many of the strongest correlations exist with parameters that quantify the vibration exposure: the un-weighted and W_k weighted *rms* acceleration (see Equation 3.15), the *rms* acceleration contained within the 64 Hz octave band, the proportional *rms* acceleration contained within the 4 and 8 Hz octave bands and the equivalent vibration level (see Equation 3.16). Strong correlations also exist with parameters that quantify the duration of the noise signal: the duration defined by the 10 dB downpoints of the noise signal and the duration defined by 2/5ths, 3/5ths and 4/5ths of the dynamic range of the noise signal. Similarly, correlations exist with parameters that quantify the duration of the vibration signal: the duration defined by the 3 dB and 10 dB downpoints of the signal and the duration defined by 2/5ths, 3/5ths and 4/5ths of the dynamic range of the vibration signal.

Correlations also exist with two psychoacoustic parameters of the noise signal: sharpness and roughness. Fastl and Zwicker (2007) provide the following model for calculating sharpness, in the form of a weighted function of the specific loudness (N'):

$$S = 0.11 \frac{\int_0^{24} N' g(z) z dz}{\int_0^{24} N' dz} \tag{6.15}$$

where $g(z)$ is a weighting function that increases with the critical band rate z. The denominator in Equation 6.15 is the total loudness. Sharpness is a measure of the high frequency content of a sound and can also be related to the timbral brightness and the sensory pleasantness of a sound (Fastl and Zwicker, 2007; Von Bismarck, 1974). A significant correlation also exists with roughness. Roughness is a metric which quantifies the perception of rapid (15 to 300 Hz, with maximum roughness at around 70 Hz) amplitude modulation of a sound. Roughness, as well as sharpness, is related to the sensory pleasantness of a sound. Fastl and Zwicker (2007) provide an approximation for determining roughness based on the product of the perceived temporal masking depth, $\triangle L$ and the modulation frequency, f_{mod}:

$$R \approx f_{mod} \triangle L \tag{6.16}$$

Further correlation exists with the spectral centroid of the noise signal, which is defined as the weighted mean of the spectral energy in the signal:

$$f_{sc} = \frac{\sum f(n) c_{mf}(n)}{\sum c_{mf}(n)} \tag{6.17}$$

where $f(n)$ is the central frequency of the nth spectral bin and $c_{mf}(n)$ is the magnitude Fourier coefficient of the nth spectral bin. The spectral centroid has been found to influence the perception of the timbral brightness of a particular sound (Grey and Gordon, 1978; Schubert and Wolfe, 2006). Finally, a correlation exists between the 2nd perceptual dimension and the modulation frequency of the vibration signal, defined as the inverse of the average period between peaks of the vibration signal envelope. In a similar MDS study looking at the perception of railway vibration, Woodcock et al. (2014a) found that one of the perceptual dimensions upon which subjects were making their judgements of dissimilarity was related to the modulation frequency of the vibration envelope.

In summary, descriptors of both the noise and vibration components of the combined stimuli show a significant correlation with the second perceptual dimension. Many of the correlating descriptors relate to the duration of either the noise or vibration signal, with other correlations existing with descriptors of the magnitude of the vibration signal and the modulation frequency of its envelope. Other correlations exist with the psychoacoustic parameters of roughness, sharpness and the spectral centroid of the noise signal.

6.4.5 Dimension 3

The correlations shown in Tables 6.4 and 6.5 suggest that the 3rd perceptual dimension could be related to objective parameters of either the noise or the vibration signal. Figure 6.11 shows the relationship between the 3rd perceptual dimension and all five objective parameters of noise and vibration which exhibit a correlation with this dimension. Once again, though some scatter is apparent, this figure gives confidence that there is a relationship between these parameters and the 3rd perceptual dimension.

The strongest correlation exists with the sound pressure level exceeded for 10% of the time, L_{10}, a property that quantifies the magnitude of the signal and that is used in the measurement and assessment of road traffic noise (Department of Transport, 1988). The next strongest correlation exists with the dominant frequency of the noise signal's power spectral density, f_{max}.

The next two strongest correlations exist with duration descriptors of the noise signal: the duration defined by 1/5th of the dynamic range, and the duration defined by the 3 dB downpoints. Note that these duration descriptors are somewhat different than the noise duration descriptors which were found to correlate with the 2nd perceptual dimension. Whilst the 10 dB downpoints duration, and the 2/5ths, 3/5ths and 4/5ths dynamic range duration tend to represent the duration of the whole signal, the 3 dB downpoints and 1/5th dynamic range duration only capture the duration of the loudest part of the signal, which may not correlate with the overall duration of the signal, particularly if the signal has a significant peak. For example, the freight train stimulus 9 has the longest duration in terms of the 10 dB downpoints (see Figure 6.10), yet has one of the shortest durations in terms of the 3 dB downpoints (see Figure 6.11). This is due to a large peak near the beginning of the signal, which can clearly be identified in Figure 5.5. To investigate the independence of these duration descriptors for the studied stimulus set, Pearson's correlation coefficients were calculated between all noise duration descriptors which have been shown to correlate with the 2nd and 3rd perceptual dimensions. These correlations are presented in Table 6.6. Though strong correlations exist between duration parameters that correlate with the 2nd and 3rd perceptual dimensions respectively, there are no correlations between these two groups. The fact that there is no correlations between these groups, and that the two groups of duration descriptors correlate with two different perceptual parameters, suggests that subjects perceive these two sets of duration descriptors differently in the stimuli used in this test.

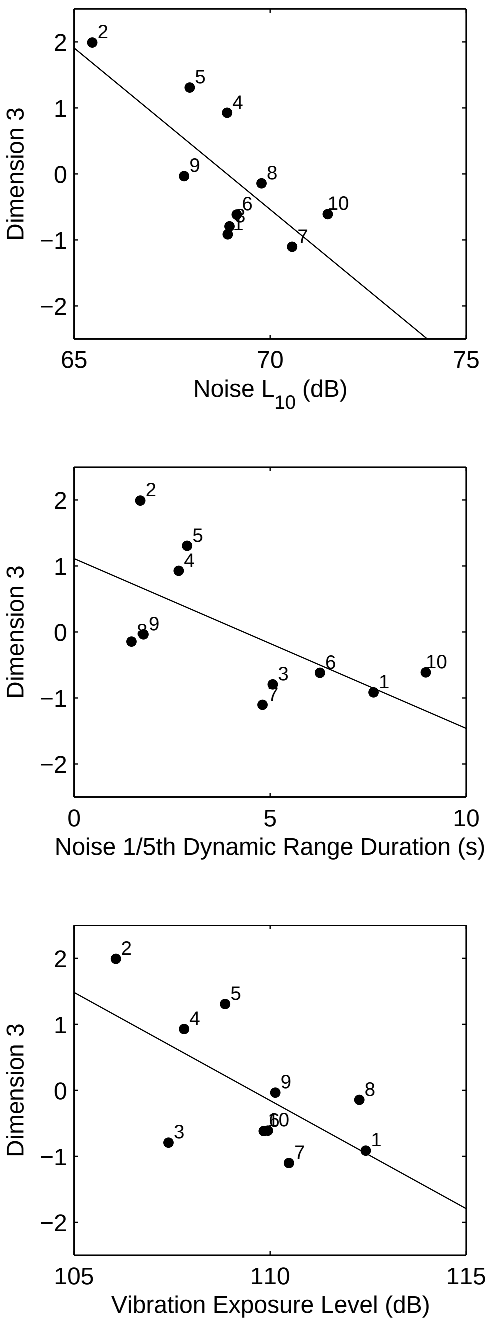

FIGURE 6.11: Relationship between the 3rd perceptual dimension and the five parameters that have a significant correlation with this dimension

		Dimension 2				Dimension 3	
		10 dB	2/5ths	3/5ths	4/5ths	3 dB	1/5th
Dim 2	10 dB	1.000**	0.992**	0.991**	0.995**	–	–
	2/5ths	0.992**	1.000**	0.998**	0.998**	–	–
	3/5ths	0.991**	0.998**	1.000**	0.998**	–	–
	4/5ths	0.994**	0.998**	0.998**	1.000**	–	–
Dim 3	3 dB	–	–	–	–	1.000**	0.999**
	1/5th	–	–	–	–	0.999**	1.000**

Table 6.6: Pearson's correlation coefficients between duration descriptors of the noise signal that have been shown to correlate with the 2nd and 3rd perceptual dimensions. – not significant, * $p < 0.05$, ** $p < 0.01$

The only vibration parameter which shows a significant correlation with the 3rd perceptual dimension is the vibration exposure level, VEL (see Equation 3.17). The vibration exposure level is analogous to the sound exposure level and is a function of the energy of the signal as well as its duration. It is defined as the vibration level of a signal of duration 1 second that would contain the same energy content as the whole signal in question.

In summary, the third perceptual dimension appears to correlate mainly with descriptors of the noise signal, with only one vibration descriptor, the vibration exposure level, showing a significant, but also the weakest, correlation with this dimension. Correlating parameters of the noise signal include those which quantify the magnitude, duration and dominant frequency of the signal.

6.4.6 Dimension 4

The correlations presented in Tables 6.4 and 6.5 suggest that the 4th perceptual dimension could be related to objective descriptors of either the noise or the vibration signals. Figure 6.12 shows the relationship between the 4th perceptual dimensions and the 5 parameters that exhibit the strongest correlation with this dimension.

Eight out of the nine parameters that exhibit a significant correlation with the 4th perceptual dimension relate to the vibration signal, with the only noise parameter that shows a significant correlation being the psychoacoustic measure of roughness. Several of the correlating parameters relate to the magnitude of the vibration signal: the un-weighted and W_b weighted vibration dose value, VDV (see Equation 3.14), the peak acceleration, and the vibration exposure level, VEL (see Equation 3.17), which is also a function of the vibration signal duration.

Other significant correlations exist with the kurtosis (see Equation 3.20) and crest factor (see Equation 3.19) of the vibration signal, two parameters which quantify the "peakiness" of a signal, and the modulation frequency, which quantifies the frequency at which peaks occur in the signal.

A correlation also exists with the duration defined by 1/5th of the dynamic range of the vibrations signal. As was demonstrated for the noise duration descriptors, this duration descriptor quantifies a different property of the vibration signal than other duration descriptors, since it captures the duration of the "loudest" part of the signal. However, unlike for the noise duration descriptors, the duration defined by the 3 dB downpoints does correlate with other duration descriptors that tend to capture the whole event duration, since the dB levels scale differently for vibration and noise due to the different units involved. To demonstrate the difference in the vibration duration descriptors that correlate with the 2nd and 4th dimensions, correlations between these descriptors are shown in Table 6.7. Although strong correlations exist between vibration duration descriptors that correlate with the 2nd perceptual dimension, no significant correlation exists between the duration

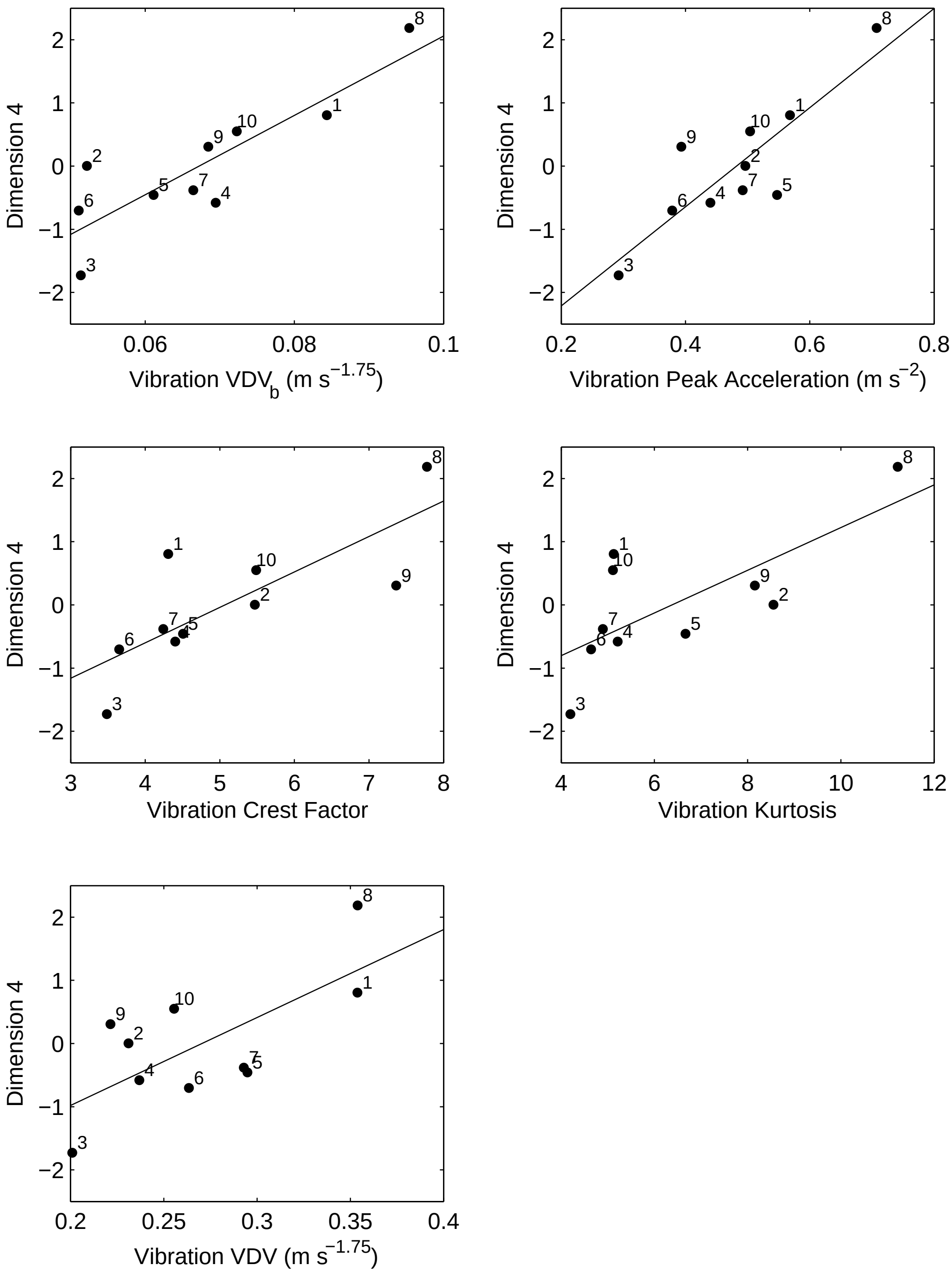

FIGURE 6.12: Relationship between the 4th perceptual dimension and the five parameters that exhibit the strongest correlation with this dimension

descriptor which correlates with the 4th perceptual dimension, the duration defined by the top 1/5th of the vibration dynamic range, and any of the vibration duration descriptors which correlate with the 2nd perceptual dimension. As is the case with the noise duration descriptors, this gives confidence that these two sets of duration descriptors quantify difference aspects of the vibration signal, which are perceived differently by subjects in the test.

		Dimension 2					Dimension 4
		3 dB	10 dB	2/5ths	3/5ths	4/5ths	1/5th
Dim 2	3 dB	1.000**	0.995**	0.990**	0.994**	0.995**	–
	10 dB	0.995**	1.000**	0.998**	0.999**	0.999**	–
	2/5ths	0.990**	0.998**	1.000**	0.998**	0.997**	–
	3/5ths	0.994**	0.999**	0.997**	1.000**	0.999**	–
	4/5ths	0.995**	0.999**	0.997**	0.999**	1.000**	–
Dim 4	1/5th	–	–	–	–	–	1.000**

TABLE 6.7: Pearson's correlation coefficients between duration descriptors of the vibration signal that have been shown to correlate with the 2nd and 4th perceptual dimensions. – not significant, * $p < 0.05$, ** $p < 0.01$

In summary, the 4th perceptual dimension appears to mainly correlate with objective parameters of the vibration component of the combined stimuli, with the only noise parameter which shows a significant correlation being the psychoacoustic roughness. Correlations with objective descriptors of vibration include those which quantify the magnitude of the signal, the duration of the signal, the "peakiness" of the signal and the frequency at which peaks occur in the signal.

6.5 New models for predicting annoyance due to combined noise and vibration

In the previous sections, for each perceptual dimension in turn, the correlations between the dimensions and objective parameters of the noise and vibration stimuli are examined. Based on these correlations, the following hypothesis is introduced:

1. The first perceptual dimension is related to aspects of the noise signal: its magnitude, loudness, "peakiness" and psychoacoustic properties of its irregular/continuous character and high frequency content.

2. The second perceptual dimension is related to the duration of the combined stimulus, the magnitude and modulation frequency of the vibration signal as well as psychoacoustic parameters of roughness, sharpness and the spectral centroid of the noise signal.

3. The third perceptual dimension is related to aspects of the noise signal: its magnitude, duration and dominant frequency.

4. The fourth perceptual dimension is related to aspects of the vibration signal: its magnitude, duration, "peakiness" and the frequency at which peaks occur

This hypothesis suggests that subjects make their dissimilarity judgements based on separate aspects of the combined noise and vibration stimuli, and in particular that they make their judgements based not only on the magnitude of the stimuli but also on spectral and temporal characteristics of the stimuli.

In order to develop new annoyance models for combined railway noise and vibration, multiple linear regression models can be created based on the above hypothesis. One method to achieve this is to derive a regression model with the Thurstone's Case V annoyance scores as the dependent variables, and the parameters which exhibit the strongest correlation with each perceptual dimension as the independent variables. The result of this regression is described in the following equation:

$$A_p = -20.3 + 53.0 rms_A - 91.8 rms_k + 0.3 L_{10} + 32.2 \mathrm{VDV}_b \tag{6.18}$$

where A_p is the predicted single figure annoyance score, rms_A is the A-weighted rms noise pressure, rms_k is the W_k weighted rms vibration acceleration, L_{10} is the noise level exceeded for 10% of the signal duration and VDV_b is the W_b weighted vibration dose value. A comparison between the Thurstone's Case V single figure annoyance scores measured during the subjective test and the single figure annoyance scores predicted using Equation 6.18 is shown in Figure 6.13. Details of the multiple linear regression are presented in Table 6.8.

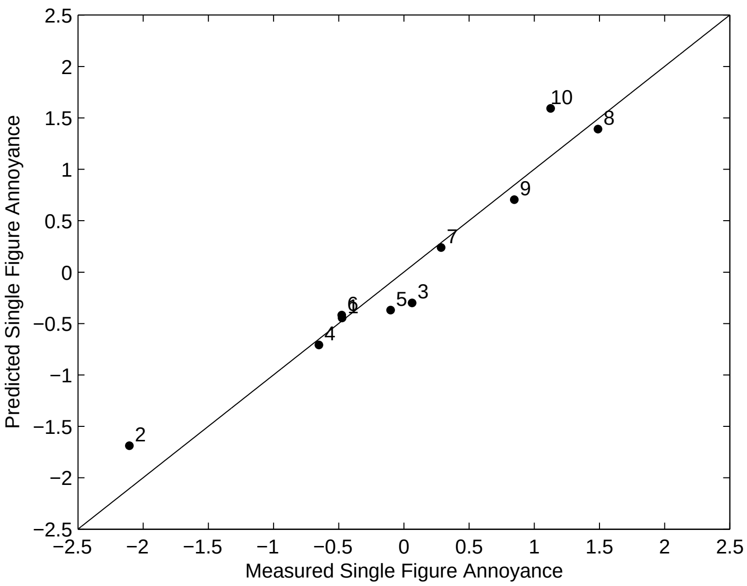

FIGURE 6.13: Comparison of single figure annoyance scores measured during the subjective test and those predicted using Equation 6.18 ($R^2 = 0.934, p < 0.010$)

Figure 6.13 shows a good agreement between measured single figure annoyance scores and those predicted using Equation 6.18 ($R^2 = 0.934, p < 0.010$). However, this model is based entirely on the energy magnitudes of the combined noise and vibration stimuli and neglects aspects of the combined stimuli which have shown to correlate significantly with the perceptual dimensions such as duration, frequency content and aspects of the peaks of the signals. It is perhaps not surprising,

Parameter	β Estimate	Standard Error	Standardised β Estimate	p-value	Overall Model	
Intercept	-20.3	6.28		< 0.050	N	10
rms_A	53.0	25.1	0.322	n.s.	p-value	< 0.010
rms_k	-91.8	29.8	-0.415	< 0.050	R^2	0.934
L_{10}	0.284	0.099	0.450	< 0.050		
VDV_b	32.2	9.26	0.452	< 0.050		

Table 6.8: Parameter estimates and other details of the multiple linear regression model for single figure annoyance as a function of the objective parameters that have the highest correlation with the perceptual dimensions

therefore, that the model shows only a slight improvement on the model derived in Chapter 5 as a linear summation of the magnitudes of the noise and vibration stimuli (see Figure 5.19).

Though choosing the parameters that have the highest correlation with the perceptual dimensions as the independent variables for the multiple linear regression is a sensible choice, it is worth considering regressions with other parameters that have shown significant correlation with the dimensions, since many of the parameters quantify the same property of the signal (i.e. there are several different quantifies of duration). Additionally, though the parameters in Equation 6.18 individually exhibit the highest correlations with each perceptual dimension, they may not necessarily provide the best model in combination. Therefore, multiple linear regressions were computed with all possible combination of parameters that have a significant correlation with the perceptual dimensions in an attempt to find the optimal model for predicting single figure annoyance. All possible combinations that result in a significant regression model ($p < 0.050$) have R^2 values ranging between 0.806 and 0.993. The model with the highest R^2 value is presented in the following equation:

$$A_p = -69.7 + 4.16TETC + 0.054T_{4/5,v} + 0.316L_{10} - 0.273f_{mod} \tag{6.19}$$

where $TETC$ is the total energy of the tonal components within critical bands from 12 to 24 barks, $T_{4/5,v}$ is the stimulus duration defined by the duration for which the vibration signal exceeds the top 4/5ths of its dynamic range, L_{10} is the sound pressure level exceeded for 10% of the stimulus duration and f_{mod} is the vibration modulation frequency. This model captures a greater degree of complexity of the combined stimulus than the model presented in Equation 6.18, since it takes into account the duration of the stimulus ($T_{4/5,v}$), the spectral distribution of the noise signal ($TETC$) and the frequency at which peaks occur in the vibration signal (f_{mod}) as well as a measure of the pressure magnitude of the noise signal (L_{10}). A previous MDS analysis of the perception of railway vibration by Woodcock et al. (2014a) found two parameters in common with this model, the vibration duration and vibration modulation frequency, to correlate with the perceptual dimensions discovered in their perceptual test. They also found these two parameters, combined with the rms acceleration in the 16 Hz and 32 Hz octave bands, to provide a successful annoyance prediction model. Though Woodcock et al. (2014a) defined the duration by the 10 dB downpoints of the vibration signal, this duration descriptor is very highly correlated with the duration defined by 4/5ths of the dynamic range (see Table 6.7).

A comparison between the Thurstone's Case V single figure annoyance scores measured during the subjective test and those predicted using Equation 6.19 is shown in Figure 6.14. Details of the multiple linear regression model are presented in Table 6.9.

Comparing Figures 6.13 and 6.14 it can be seen that the optimal multiple linear regression model (Equation 6.19) provides a better agreement between the measured and predicted single figure annoyance scores. The R^2 value is higher, the model is significant to a higher level and each regression parameter is significant to a higher level. Regression models with the single figure annoyance scores predicted by the BTL model and the averaged single figure annoyance scores also provide

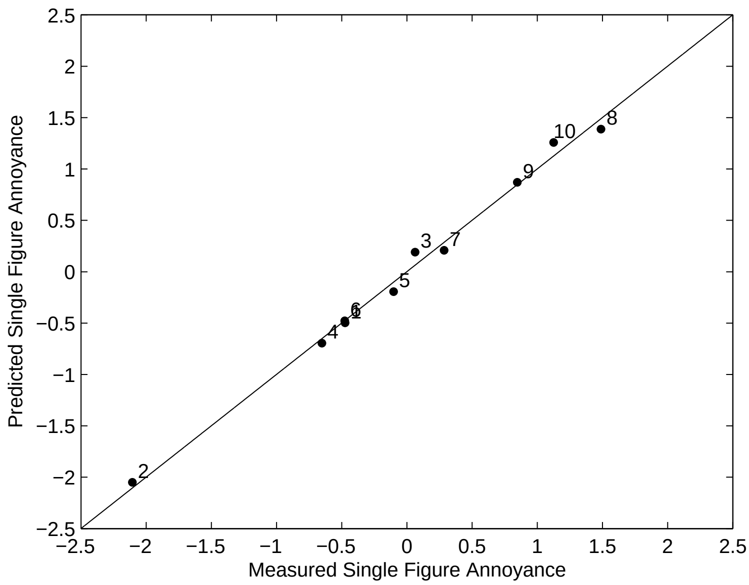

FIGURE 6.14: Comparison of single figure annoyance scores measured during the subjective test and those predicted using Equation 6.19 ($R^2 = 0.993$, $p < 0.001$)

Parameter	β Estimate	Standard Error	Standardised β Estimate	p-value	Overall Model	
Intercept	-69.7	4.36		< 0.001	N	10
$TETC$	4.16	0.464	0.427	< 0.001	p-value	< 0.001
$T_{4/5,v}$	0.054	0.006	0.490	< 0.001	R^2	0.993
L_{10}	0.316	0.032	0.501	< 0.001		
f_{mod}	-0.273	0.064	-0.221	< 0.010		

TABLE 6.9: Parameter estimates and other details of the multiple linear regression model for single figure annoyance as a function of the objective parameters that give the optimal multiple linear regression

good agreement with the measured annoyance scores, with similar R^2 values and significance ($R^2 = 0.991$, $p < 0.001$ and $R^2 = 0.989$, $p < 0.001$ respectively).

The standardised β coefficients in Table 6.9 give an indication of the relative weight carried by each parameter, with the sound pressure level exceeded for 10% of the signal duration providing the greatest influence on the annoyance response, followed by the stimulus duration, then the measure of the high frequency content of the noise signal and finally the vibration modulation frequency.

To test the trade off between the number of predictor variables included in the annoyance model and the amount of variance explained by the model, a stepwise linear regression was conducted, with a constraint implied that for a predictor variable to be included in the regression model it must have an estimated β coefficient with a p-value of less than 0.05. The steps taken by the stepwise regression are shown below and the results of this stepwise regression are presented in Table 6.10. The fact that all predictor variables are included in the model by the stepwise regression gives confidence that an optimal model has been found and that the perception of combined noise and vibration is well represented by four perceptual dimensions.

- Initial predictors included: none
- Step 1: added L_{10} ($R^2 = 0.633$, $p < 0.050$)
- Step 2: added $T_{4/5,v}$ ($R^2 = 0.927$, $p < 0.010$)
- Step 3: added $TETC$ ($R^2 = 0.968$, $p < 0.010$)
- Step 4: added f_{mod} ($R^2 = 0.993$, $p < 0.010$)
- Final predictors included: L_{10}, $T_{4/5,v}$, $TETC$ and f_{mod}

The relationship described in Equation 6.19 and shown in Figure 6.14 is a key result of the multidimensional scaling analysis. Comparing the annoyance predictions described by the multidimensional model (Figure 6.14) and the regression model

Parameter	β Coefficient	Status	p-value
Intercept	-8.44	In	< 0.001
L_{10}	0.046	In	< 0.010
$T_{4/5,v}$	0.008	In	< 0.010
$TETC$	4.63	In	< 0.010
f_{mod}	-0.57	In	< 0.050

Table 6.10: Stepwise regression results

where annoyance is predicted by magnitudes of noise and vibration (Figure 5.19) shows that the new multidimensional model provides a substantial improvement in the accuracy of the prediction of overall annoyance, accounting for approximately 15% more variance in annoyance scores. Equation 6.19, and the high accuracy of annoyance prediction, suggests that the perception of combined noise and vibration is a complex multidimensional phenomenon that takes into account not only the energy magnitude of the stimuli but also the duration, frequency content and envelope modulation of the combined stimulus.

6.6 The multidimensional perception of freight and passenger railway noise and vibration

In the subjective test presented in Chapter 5, the three freight train stimuli were shown to be perceived as more annoying than the passenger train stimuli. This can be seen in Figures 5.14 and 5.15, where the freight train stimuli (stimuli number 8, 9 and 10) have higher single figure annoyance scores than the passenger train stimuli, and in Figure 5.17, where the freight train stimuli are the only stimuli with the categorical annoyance rating of "very" annoying, with the passenger trains have lower ratings of either "slightly" or "moderately" annoying. In this section, the differences in the response to freight and passenger trains will be investigated in terms of multidimensional perceptual models.

6.6.1 Multidimensional analysis of freight and passenger noise and vibration stimuli

The annoyance model derived in this chapter predicts the freight train stimuli to have a higher single figure annoyance rating (see stimuli 8,9 and 10 in Figure 6.14) based on the high frequency content of the noise signal, the duration of the stimulus, the sound pressure level exceeded for 10% of the signal duration and the frequency modulation of the vibration envelope (see Equation 6.19). To test whether these parameters are significantly different for the freight and passenger train signals used in the subjective test, a Kruskal-Wallis one-way analysis of variance test was performed for each of the four parameters in the annoyance model. According to this test, the only parameter that has significantly different distributions ($p < 0.05$) for the freight and passenger train stimuli is the duration descriptive parameter, $T_{4/5,v}$. Though it is difficult to make strong conclusions based on the very small sample size of stimuli (7 passenger and 3 freight trains), these tests suggest that the reason for the increased annoyance response from freight trains may be primarily due to their increased duration.

6.6.2 Multidimensional analysis of field study freight and passenger vibration signals

Though the stimuli set of freight and passenger trains used in the subjective test is small, there is a large set of freight and passenger vibration stimuli (14 874 freight trains and 85 103 passenger trains) extracted from the field study data of Waddington et al. (2014) using the machine learning methods described in Chapter 3. The annoyance model derived in this chapter cannot be applied to this data set, as it is a model based on combined noise and vibration, and no measurements of noise were made during the field study of Waddington et al. (2014). However, Woodcock et al. (2014a) developed a multidimensional annoyance model based on the vibration signals of the field study database, using a similar methodology as that applied in the subjective test for this research. Their annoyance model takes the following form:

$$A_p = -0.40 + 4.57rms_{16\text{Hz}} + 3.18rms_{32\text{Hz}} + 0.02T_{10dB} + 0.02f_{mod} \qquad (6.20)$$

where $rms_{16\text{Hz}}$ and $rms_{32\text{Hz}}$ is the rms acceleration contained in the 16 Hz and 32 Hz octave bands respectively, T_{10dB} is the duration defined by the 10 dB downpoints of the vibration signal and f_{mod} is the modulation frequency of the vibration envelope. Note that two of the parameters in Equation 6.20 are equivalent to those in the annoyance model derived in this chapter (Equation 6.19), namely the duration defined by the 10 dB downpoints of the vibration signal and the modulation frequency of the vibration signal. It should also be noted, however, that these two parameters were removed from a model in a stepwise regression performed by Woodcock et al. (2014a), who stated that further work should be performed to determine objective correlates which better describe the third and fourth perceptual dimensions in Equation 6.20. Although Equation 6.19 cannot be successfully applied to the separated freight and passenger vibration signals from the field study since it was derived from a subjective test involving combined noise and vibration, Equation 6.20 can be, since it was derived from vibration stimuli obtained from the field study measurements.

To test whether the parameters in Equation 6.20 are significantly different for the freight and passenger signals measured during the field study, these parameters were calculated for all freight and passenger vibration signals extracted from the database of field study measurements, a substantially larger data set than the noise and vibration signals used in the subjective test.

A Kruskal-Wallis one way analysis of variance test was performed for each parameter in Equation 6.20 calculated for both the freight and passenger vibration signals. The results of this test indicates that each parameter has a significantly different distribution ($p < 0.01$) for freight and passenger vibration signals. However, a visual representation of the distributions of these parameters revealed some differences in the parameter distributions for these two sources. Figure 6.15 shows the probability density estimates of the parameter distributions for freight and passenger vibration signals. There is a significant degree of overlap and similarity between the distributions, especially for the $rms_{16\text{Hz}}$ and $rms_{32\text{Hz}}$ parameters. The T_{10dB} parameter, however, shows quite a different distribution for freight and passenger vibration signals, with the freight signals tending to have a significantly higher duration. Again this suggests that freight trains may be perceived as more annoying because of their greater passby duration. Note that the tail of the probability density curve going below zero in some cases is due to the smoothing function of the probability density estimate; none of the parameters have a negative value.

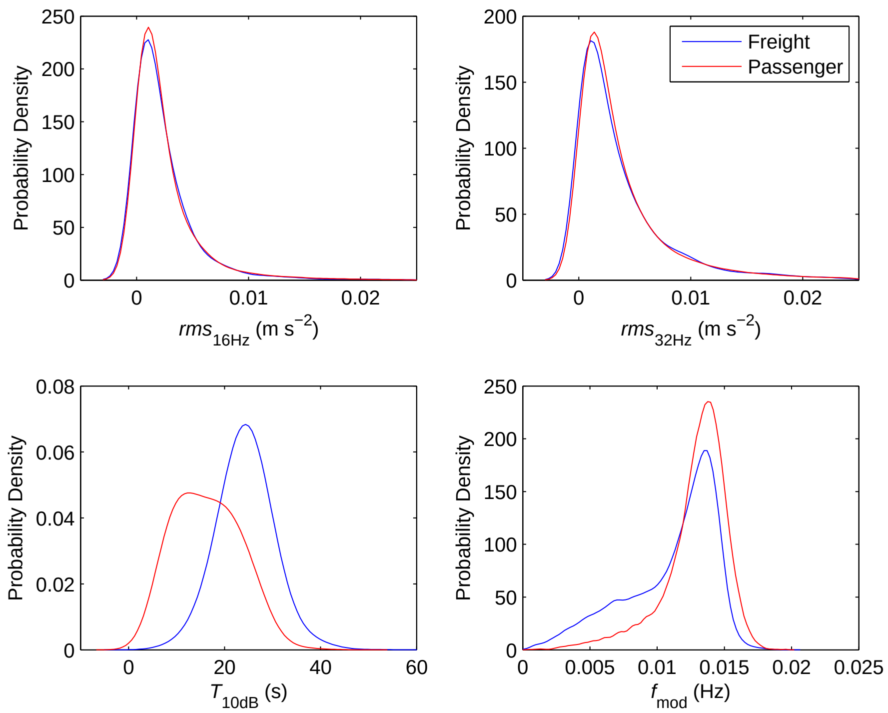

FIGURE 6.15: Probability density estimates of the four parameters in the annoyance model of Woodcock et al. (2014a) for freight and passenger vibration signals

6.6.3 Perceived annoyance of field study freight and passenger vibration signals

Predicted perceived annoyance scores were calculated for each of the freight and passenger vibration signals using Equation 6.20. Once again, a Kruskal-Wallis one way analysis of variance test was performed on the resulting distributions of freight and passenger perceived annoyance scores, the result of which indicates that the distributions of perceived annoyance are significantly different for these two sources ($p < 0.01$). The probability density estimate of the distributions of perceived annoyance scores for freight and passenger vibration signals is shown in Figure 6.16. This figure shows a significantly different distributions for the perceived annoyance of these two sources, with freight trains being consistently perceived as more annoying than passenger trains. The mean single figure annoyance scores for freight and passenger vibration signals are 0.11 and -0.04 respectively. Note that these annoyance scores are on a different scale to others presented in this chapter, since the annoyance models derived in this chapter were based on the Thurstone's Case V annoyance, whereas the model derived by Woodcock et al. (2014a) was derived using the average single figure annoyance score (see Equation 5.1). Notably, the distributions of perceived annoyance due to freight and passenger vibration signals are very similar to the distributions of the duration parameter T_{10dB} for these two sources, as shown in Figure 6.15, giving further confidence that the increased annoyance due to freight trains may be primarily attributed to their greater passby duration.

6.6.4 Categorical annoyance of field study freight and passenger vibration signals

Although the distributions shown in Figure 6.16 provide the information that freight train vibration signals tend to be perceived as more annoying than passenger train vibration signals, it is difficult to relate the single figure annoyance

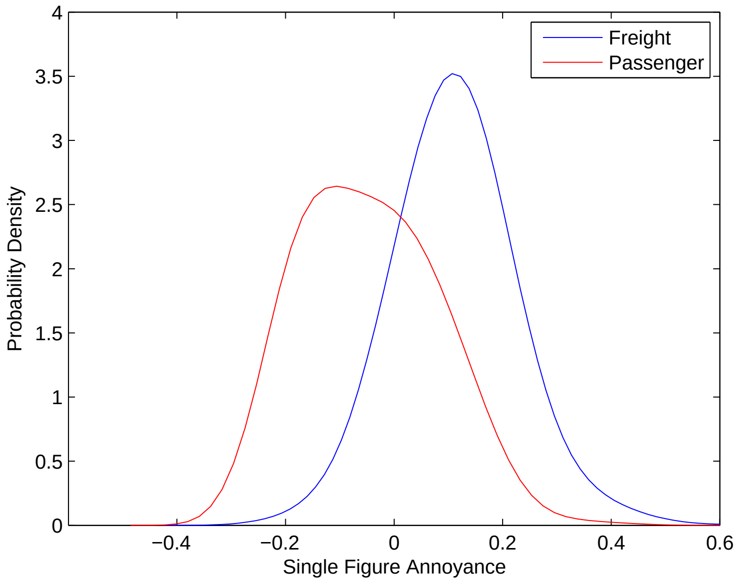

FIGURE 6.16: Probability density estimates of perceived single figure annoyance of freight and passenger vibration signals

scores to absolute levels of annoyance without anchoring the single figure annoyance scale to categorical ratings of annoyance. Woodcock et al. (2014a) related their average single figure annoyance scores to categorical annoyance scores using the same methods described in Section 5.5.6. They also used ordinal regression for their relationship, but used the logit link function, which is more appropriate for their evenly distributed data. The ordinal regression model with the logit link function is as follows:

$$\ln\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_{tj} + \beta A_{av} \tag{6.21}$$

where A_{av} is the average single figure annoyance (see Equation 5.1). The parameters of the ordinal regression model derived by Woodcock et al. (2014a) were kindly supplied by the authors for the purpose of further analysis in this research and are presented in Table 6.11. The ordinal regression model with these parameters is presented in Figure 6.17.

Categorical annoyance ratings can be determined for all the freight and passenger

Parameter	β Estimate	Standard Error	p-value	Overall Model	
Threshold				N	14
Slightly/Moderately	-12.8	7.53	n.s.	p-value	< 0.001
Moderately/Very	-1.75	1.62	n.s.	R^2_{pseudo}	0.823
Very/Extremely	12.6	8.50	n.s.		
Location					
A_{av}	76.3	43.3	n.s.		

TABLE 6.11: Parameter estimates and other details of the ordinal regression model derived by Woodcock et al. (2014a)

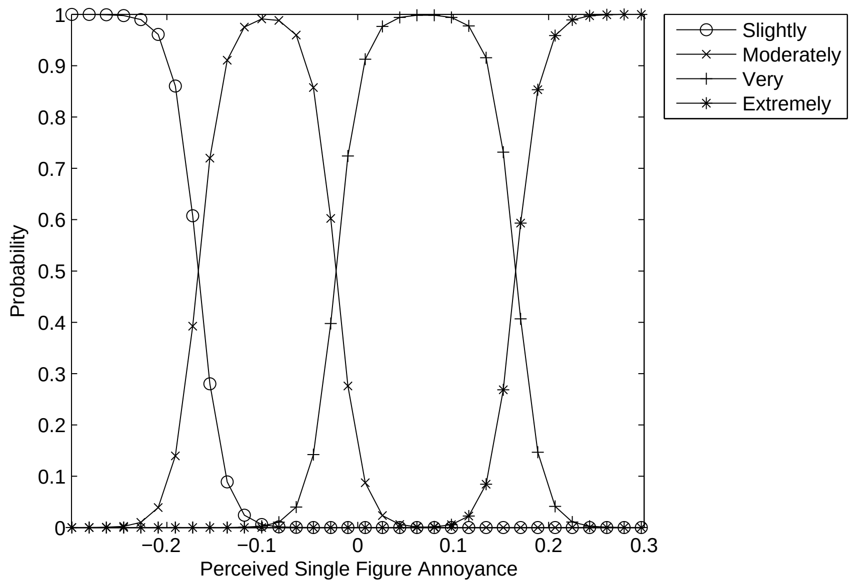

FIGURE 6.17: Probability of a vibration signal with a given perceived annoyance rating being rated in a certain category of absolute annoyance

vibration signals measured in the field study of Waddington et al. (2014), by first calculating their predicted perceived annoyance rating using Equation 6.20 and then assigning their categorical annoyance to the category with the highest probability for that perceived annoyance rating, according to the model presented in Figure 6.17. The results of this procedure is presented in Figure 6.18, which shows the percentage of all freight and passenger vibration signals that are assigned to each category of annoyance. A clear trend can be seen for freight vibration signals, with a greater proportion of signals being assigned to higher annoyance categories. A significant majority, approximately 75%, of all vibration signals measured in the field study have been predicted to fall in the "extremely" annoying category. In contrast, the majority of passenger vibration signals are assigned to the category of "moderately" annoying or lower.

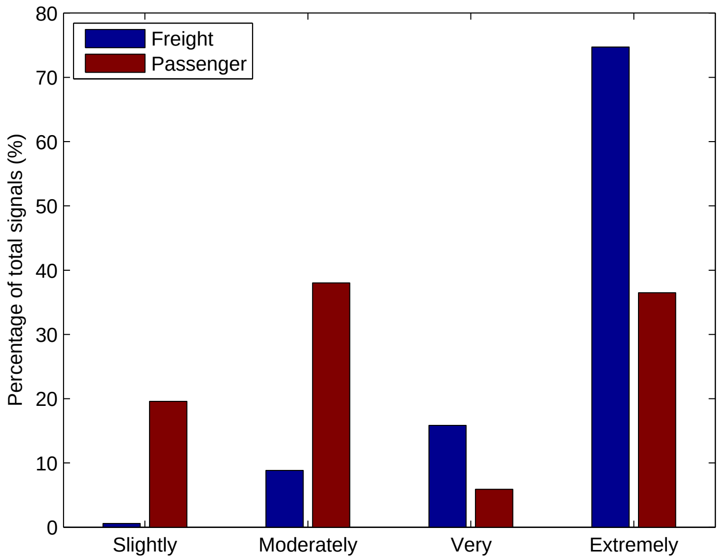

FIGURE 6.18: Proportion of freight and passenger signals that have been assigned to each annoyance category

Note that the categorical annoyance model was derived from a dataset with no stimuli assigned to the "not at all" annoyed category. The model therefore only predicts the probability of signals being placed in the range of annoyance categories from "slightly" to "extremely" and will therefore over emphasise the proportion of signals assigned to these categories.

6.7 Summary

This chapter presents further analysis of the results of the subjective test described in Chapter 5. Using the methods of multidimensional scaling, a multidimensional perceptual model was created from the perceived dissimilarity, measured during the subjective test, of a set of noise and vibration railway stimuli. This multidimensional scaling model revealed four perceptual dimensions upon which subjects made their judgements of dissimilarity, and which can be used to successfully predict the perceived single figure annoyance of the noise and vibration stimuli. These perceptual dimensions can be related to several objective properties of the noise and vibration stimuli which quantify the magnitude of the noise or vibration signals, the duration of the signals, the spectral distribution of the signal, aspects of the "peakiness" of the signals and psychoacoustic properties of the noise signal.

Multiple linear regression was used to model the predicted single figure annoyance as a function of the objective parameters of the noise and vibration signals that have been shown to correlate with the perceptual dimensions of the multidimensional scaling model. The model with the greatest accuracy of prediction of the perceived single figure annoyance was a function of the spectral distribution of the noise signal, the duration of the stimulus, the magnitude of the noise signal and the modulation frequency of the vibration signal envelope. Considering the parameters of this model, an analysis of the freight and passenger noise and vibration stimuli suggests that the freight train stimuli in the subjective test may have been perceived as more annoying based primarily on their increased duration.

The freight and passenger vibration signals measured during the field study of Waddington et al. (2014) were analysed using a model of perceived annoyance derived by Woodcock et al. (2014a) using similar methods to those outlined in this chapter, but for vibration stimuli in the absence of noise. Using this model, and the freight and passenger vibration signals identified using the methods outlined in Chapter 3, single figure annoyance scores and categorical annoyance ratings were calculated for each of the vibration signals. Analysis of these signals shows that the annoyance distributions of the freight and passenger vibration signals are significantly different, with freight trains having consistently higher estimated annoyance ratings than passenger trains. In terms of categorical annoyance ratings, a significant majority of freight trains were assigned to the highest category of "extremely" annoying, whereas the majority of passenger trains were assigned to the category of "moderately" annoying or lower. Once again, an analysis of the distribution of the parameters in the annoyance model derived by Woodcock et al. (2014a) suggests that freight trains may be perceived as more annoying primarily due to their increased duration.

Chapter 7

Summary, conclusions and further work

7.1 Summary and conclusions

With the proposed increase in freight transport carried by rail, there will be an associated increase in noise and vibration from freight railway traffic. Though the human response to railway noise is well documented, the human response to railway vibration has been less researched, and in particular studies into the human response to freight railway noise and vibration are scarce. The primary aim of this research, therefore, is to further the understanding of the human response to freight railway noise and vibration, in an attempt to understand the potential impacts that an increase in freight railway traffic may have on residents living close to railway lines. In this chapter, the main outcomes and conclusions of the research are summarised, and potential avenues of further research are presented.

7.1.1 Classification of unknown railway vibration signals

Waddington et al. (2014) performed one of the few field studies that have been carried out worldwide on the human response to railway vibration in residential environments. They were successful in developing exposure-response relationships for annoyance due to exposure to railway vibration in residential environments. However, these exposure-response relationships were derived for all railway traffic, and no attempt was made to distinguish between freight and passenger railway vibration. In this research, a machine learning logistic regression algorithm was developed in order to classify unknown railway vibration signals in the database of measurements collected by Waddington et al. (2014) as freight or passenger vibration events, for further analysis. The logistic regression model was optimised using a combination of univariate and multivariate significance testing and classification accuracy testing. The optimised logistic regression model is a function of only two objective vibration parameters: the event signal duration, defined as the normalised duration for which the vibration signal exceeds the top third of its dynamic range, and the proportional *rms* acceleration contained within the 5 Hz 1/3rd octave band. The logistic regression model is able to classify unknown railway vibration signals as freight or passenger vibration signals with an accuracy of 96%, and a validation of the model on new measurements resulted in an even higher classification accuracy of 98%. The high classification accuracy of this model gives confidence that it can be used to classify unknown signals measured during the field study of Waddington et al. (2014), allowing further analysis of vibration exposure resulting from freight and passenger railway traffic.

7.1.2 Exposure-response relationships for annoyance due to exposure to freight and passenger railway vibration

With the logistic regression model applied to the measurement database, it was possible to determine 24 hour exposures of freight and passenger railway vibration for all of the 752 case studies collected during the field study. These exposures were then paired with the annoyance responses to freight and passenger railway vibration which were also collected during the field study, in order to investigate the exposure-response relationships for these two sources of environmental vibration. Initially, a pooled exposure-response relationship with a dummy variable for source type was developed, using an ordinal probit grouped regression model. The regression coefficient for the source type dummy variable was positive and significant, giving confidence that it is valid to derived separate exposure-response relationships for freight and passenger railway vibration.

Separate exposure-response relationships were then developed, showing the annoyance response due to exposure to freight railway vibration, and the annoyance response due to exposure to passenger railway vibration. These exposure-response relationships show that the annoyance response to freight railway vibration is significantly higher than that due to passenger railway vibration, even for equal levels of vibration exposure magnitude. This difference in response can be quantified in several ways. For example, for a 24 hour vibration exposure of 0.01 m s^{-2}, 4% of the population studied is likely to be highly annoyed if the source is passenger railway vibration, whereas 13% of the studied population is likely to be highly annoyed if the source is freight railway vibration.

The difference in response is further quantified by deriving exposure-response relationships with a source type dummy variable using an ordinal logit model. The regression coefficient for the source type dummy variable can then be used to compute the odds ratio of a respondent responding in a higher annoyance category for a given magnitude of vibration exposure for freight railway vibration rather than passenger railway vibration. The odds ratio suggests that respondents, for a given 24 hour rms acceleration, are 1.6 times more likely to respond in a higher annoyance category due to freight railway vibration than they are due to passenger railway vibration. For a given 24 hour VDV, the odds ratio of responding in a higher annoyance category increases to 1.8.

7.1.3 A subjective test on the perception of combined railway noise and vibration

The disparate annoyance responses to freight and passenger railway vibration, even for equal levels of vibration exposure magnitude, suggest that current metrics for quantifying vibration exposure may not accurately represent the human perception of railway vibration. In addition, railway vibration very often exists in concert with noise. Therefore, a subjective test was designed in order to investigate the human response to railway noise and vibration as a multidimensional phenomenon, in the hopes of identifying objective parameters of the vibration signals, and accompanying noise, that may lead to the increased annoyance response due to freight railway vibration.

The subjective test took the form of a paired comparison test, with subjects making judgements on pairwise annoyance and dissimilarity across all possible pairings of 10 freight and passenger combined noise and vibration stimuli. The results of the paired comparison judgements of annoyance were used to model single figure annoyance scores. The freight railway noise and vibration stimuli gave consistently higher perceived single figure scores than the passenger train stimuli. When related to categorical annoyance ratings, freight trains were the only stimuli assigned to the highest recorded categorical rating of "very" annoying.

Multiple linear regression models were derived for the prediction of single figure annoyance as a function of noise magnitude, vibration magnitude and combined noise and vibration magnitudes. The single figure annoyance predicted by the regression model with combined noise and vibration magnitudes showed the highest correlation coefficient with measured single figure annoyance, suggesting that overall annoyance due to combined railway noise and vibration stimuli is based on a summation of the effects of the individual stimuli.

Though a reasonable approximation of the total annoyance caused by combined railway noise and vibration can be determined from classic models of annoyance predictions, involving summations of the noise and vibration magnitudes, there is still a substantial degree of scatter when comparing measured single figure annoyance scores and those predicted by the magnitude summation model. This scatter, and the disparity in the exposure-response relationships derived from vibration exposure magnitude, suggests that the human response to vibration and combined noise is not sufficiently quantified by existing exposure metrics. Therefore, multidimensional scaling analysis was performed on the results of the subjective test, in the hopes of determining the perceptual dimensions that underlay the perception of combined noise and vibration. This may reveal aspects of the noise and vibration signals that account for the difference in the human response, even for equal levels of noise and vibration magnitude.

7.1.4 The multidimensional perception of railway noise and vibration

The multidimensional scaling of the paired comparison judgements of dissimilarity revealed a four dimensional perceptual space upon which subjects made their judgements of dissimilarity, and which can be used to successfully predict the perceived single figure annoyance of the noise and vibration stimuli. The perceptual dimensions can be related to several objective properties of the noise and vibration stimuli which quantify the magnitude of the noise and vibration signals, the duration of the signals, the spectral distribution of the signals, aspects of the "peakiness" of the signals and psychoacoustic properties of the noise signal.

Multiple linear regression models were used to predict single figure annoyance scores as a function of the parameters of the noise and vibration signals that were shown to have a significant correlation with the perceptual dimensions revealed by the multidimensional scaling. The model with the greatest accuracy of prediction of the single figure annoyance scores is a function of the spectral distribution of the noise signal, the duration of the combined stimulus, the magnitude of the noise signal and the modulation frequency of the vibration envelope. This model has a greater prediction accuracy than classic models of annoyance prediction based only on the combined magnitudes of the noise and vibration stimuli. Considering the parameters of the model, an analysis of the noise and vibration stimuli used in the subjective test suggests that the freight train stimuli may have been perceived as more annoying based primarily on their increased duration.

The freight and passenger vibration signals measured during the field study of Waddington et al. (2014) could not be analysed with the newly derived model, since only measurements of vibration were made. Instead, a multidimensional model previously derived by Woodcock et al. (2014a) was used to analyse the freight and passenger vibration signals measured during the field study. Using this model, single figure annoyance scores and categorical annoyance ratings were calculated for each of the vibration signals. The annoyance distributions of the freight and passenger vibration signals were significantly different, with freight trains having consistently higher perceived annoyance ratings than passenger trains. A significant majority of freight trains were assigned to the highest categorical annoyance category of "extremely" annoying, whereas the majority of passenger trains were assigned to the category of "moderately" annoying or lower. Again, an analysis of the distributions of the parameters in the annoyance model derived by Woodcock et al. (2014a) for freight and passenger train signals, suggests that freight trains may be perceived as more annoying primarily due to their increased duration.

7.2 Further Work

The work in this thesis has demonstrated that there is a significant difference in the human response to freight and passenger railway vibration. It has been suggested that the difference in response is not well quantified by currently used metrics of vibration exposure, leading to higher annoyance responses due to freight railway vibration, even for equal levels of vibration exposure magnitude. A multidimensional analysis of the perception of railway noise and vibration has helped to highlight some of the aspects of the noise and vibration signals that may lead to this difference in the annoyance response, and a new model of annoyance prediction has been presented. However, there is a great deal of further work that can be performed in order to further understand the differences in the human response to different sources of environmental vibration, and how these differences can be used to understand potential impacts that increases in railway noise and vibration may have on residents living close to railways. Some potential avenues of further research are suggested in this section.

7.2.1 Relating individual annoyance ratings to overall annoyance response

The multidimensional annoyance model developed by Woodcock et al. (2014a) has been applied to all of the freight and passenger vibration events in the measurement database of the field study of Waddington et al. (2014). This has resulted in each vibration event being assigned a single figure annoyance score and a categorical annoyance rating. However, it is unclear how these individual event annoyance ratings are related to an overall annoyance response. With further work, an investigation could be performed on how these individual event annoyance ratings can be related to the overall annoyance response for each of the 752 case studies in the field study, allowing new and perhaps more accurate predictions of exposure-response relationships for annoyance due to exposure to freight and passenger railway vibration. Exposure-response relationships developed in this work have either needed to be developed separately for freight and passenger railway vibration, or with a source type penalty for freight railway vibration. If the multidimensional models of vibration perception can be utilised in the development of new exposure-response relationships, then it should be possible to take into account aspects of the vibration signals that cause the increased annoyance response to freight vibration, such as the increased duration. This could allow the development of robust exposure-response relationships that are able to more accurately predict the human response to all railway vibration.

7.2.2 Development of new vibration exposure metrics

Related to the above avenue of further work is the possibility of developing a new metric for quantifying human exposure to vibration. The multidimensional scaling analyses presented in this research, and in the work of Woodcock et al. (2014a), have shown that the perception of vibration, and vibration and noise in combination, is a multidimensional phenomenon, and can be modelled using not only the magnitude of the vibration exposure, but also the duration, spectral distribution and envelope modulation of the signal. With further work, it may be possible to develop new vibration exposure metrics which take these aspects of the vibration signal into account and which are perhaps better quantifiers of vibration exposure in terms of potential human response. These new metrics could then be used to accurately predict annoyance responses due to environmental vibration, or to set limits and guidance for appropriate levels of vibration.

7.2.3 Time of day analysis

Previous research by Peris et al. (2012) has shown that the annoyance response due to vibration is higher during the night-time (23:00 to 07:00) than during evening time (19:00 to 23:00) and higher during evening time than during daytime (07:00 to 19:00). Freight railway traffic is generally more prevalent evening and night-time hours. For example, for the 752 case studies analysed in this research, the proportion of freight traffic during the daytime, evening time and night-time is 10%, 18% and 22% respectively. The increased annoyance response due to freight railway vibration, therefore, may be partially due to the prevalence of freight traffic to occur more frequently during the evening time and during the night-time, where sleep disturbance also becomes an issue. The time of day analysis performed by Peris et al. (2012) utilised response data collected during the field study of Waddington et al. (2014), in which questions were asked of respondents about how annoyed they were by railway vibration at different times of the day. However, these time of day responses were only collected for all railway traffic, and not freight and passenger railway traffic separately. This means that it is not possible to derive exposure-response relationships for annoyance due to freight and passenger railway traffic at different times of day. With further work, and possibly further field studies, it may be possible to develop these relationships and investigate how freight traffic during the night relates to overall annoyance due to freight railway vibration.

7.2.4 Situational, attitudinal and demographic factors

In exposure-response relationships, there is always a significant degree of scatter, due to inherent differences in the population studied. Some of the variation is due to different sensitivities to noise and vibration, whilst other differences may not be necessarily associated with the vibration or noise exposure but with situational, attitudinal and demographic factors. Peris et al. (2014) studied the effects of these situational, attitudinal and demographic factors on annoyance due to railway vibration, using data collected during the field study of Waddington et al. (2014). In particular, they found that annoyance responses were strongly influenced by two attitudinal factors: concern of property damage and expectations about future levels of vibration. It is possible that these two attitudinal factors are also related to the increased annoyance response due to freight railway vibration. Freight trains, with their longer duration and greater low frequency vibration energy could be perceived by respondents as more likely to damage their property. In addition, the fact that freight railway transportation is increasing could be related to increased annoyance, due to expectations about future levels of vibration increasing. Further work could focus on investigating how these non exposure factors may relate to freight railway traffic in particular, perhaps helping to further explain the differences in the annoyance response to freight and passenger railway vibration.

7.3 Final summary

This research has demonstrated that the human response to freight and passenger railway vibration is significantly different, and in particular that the annoyance response due to freight railway vibration is higher than that due to passenger railway vibration, even for equal levels of vibration exposure magnitude. This suggests that current vibration exposure metrics do not sufficiently quantify aspects of freight vibration signals that may account for the higher annoyance response, such as their increased duration. Indeed, a subjective study has shown that the human response to railway noise and vibration is not simply a function of the combined noise and vibration magnitudes, but also takes into account spectral distributions of the noise signal, the duration of the stimulus and the modulation frequency of the vibration envelope. Further work is required to utilise these results in the development of new exposure-response relationships and new metrics to quantify vibration exposure that can more accurately account for the difference in the response to these sources of environmental vibration. Further work may also help to determine other factors that may account for the difference in response to freight and passenger railway vibration, such as time of day effects and the effects of situational, attitudinal and demographic factors.

The findings that the human response to freight and passenger railway vibration is significantly different, and that the human response to combined railway noise and vibration is a complex multidimensional phenomenon, are important results, particularly in light of increasing freight transportation on rail. These results may have important implications for the expansion of freight traffic on rail, or for policy that aims to promote passenger railway traffic.

References

Abercrombie, C. and Braasch, J. (2009). A method for multimodal auralization of audio-tactile stimuli from acoustic and structural measurements. In *Proceedings of the Audio Engineering Society Convention*.

Agresti, A. (2002). *Categorical data analysis*. John Wiley & Sons, Hoboken, NJ, 2nd edition.

Alpaydin, E. (2010). *Introduction to machine learning*. MIT Press, Cambridge, MA.

Andersen, T. V., Kühl, K., and Relster, E. (1983). Reactions to railway noise in Denmark. *Journal of Sound and Vibration*, 87(2):311–314.

Arnberg, P. W., Bennerhult, O., and Eberhardt, J. L. (1990). Sleep disturbances caused by vibrations from heavy road traffic. *The Journal of the Acoustical Society of America*, 88(3):1486–1493.

Association of Noise Consultants (2012). *Red book: measurement and assessment of groundborne noise and vibration*. Fresco, 2nd edition.

Backhaus, W., Menzel, R., and Kreiß l, S. (1987). Multidimensional scaling of color similarity in bees. *Biological Cybernetics*, 56(5-6):293–304.

Basner, M., Babisch, W., Davis, A., Brink, M., Clark, C., Janssen, S., and Stansfeld, S. (2014). Auditory and non-auditory effects of noise on health. *Lancet*, 383:1325–1332.

Bech, S. and Zacharov, N. (2006). *Perceptual audio evaluation - Theory, method and application*. John Wiley & Sons, Chichester, UK.

Bellmann, M. A. (2002). *Perception of whole-body vibrations: from basic experiments to effects of seat and steering-wheel vibrations on the passenger's comfort inside vehicles.* PhD thesis, University of Oldenburg.

Benson, A. J. and Dilnot, S. (1981). Perception of whole-body linear oscillation. In *Proceedings of the United Kingdom Informal Group Meeting on Human Response to Vibration*, Edinburgh, UK.

Block, H. D. and Marshchak, J. (1959). Random orderings and stochastic theories of responses. *Cowles Foundation Discussion Papers*, 66.

Blumenthal, T. D. (1988). The startle response to acoustic stimuli near startle threshold: effects of stimulus rise and fall time, duration, and intensity. *Psychophysiology*, 25(5):607–611.

Blumenthal, T. D. and Berg, W. K. (1986). The startle response as an indicator of temporal summation. *Perception & Psychophysics*, 40(1):62–68.

Blumenthal, T. D., Cuthbert, B. N., Filion, D. L., Hackley, S., Lipp, O. V., and van Boxtel, A. (2005). Committee report: guidelines for human startle eyeblink electromyographic studies. *Psychophysiology*, 42(1):1–15.

Blumenthal, T. D. and Goode, C. T. (1991). The startle eyeblink response to low intensity acoustic stimuli. *Psychophysiology*, 28(3):296–306.

Borg, I. and Groenen, P. J. F. (2005). *Modern multidimensional scaling: Theory and applications.* Springer, New York, NY, 2nd edition.

Bradley, R. A. (1954). Rank analysis of incomplete block designs: II. Additional tables for the method of paired comparisons. *Biometrika*, 41(3):502–537.

Bradley, R. A. (1955). Rank analysis of incomplete block designs: III. Some large-sample results on estimation and power for a method of paired comparisons. *Biometrika*, 42(3):450–470.

Bradley, R. A. and Terry, M. E. (1952). Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika*, 39(3/4):324–345.

British Standards Institution (1997). BS ISO 2631-1:1997 Mechanical vibration and shock — evaluation of human exposure to whole-body vibration - part 1: general requirements.

British Standards Institution (1998). BS 6840-13:1998, IEC 60268-13:1998 Sound system equipment - part 13: listening tests on loudspeakers.

British Standards Institution (2005). BS EN ISO 8041:2005 Human response to vibration – measuring instrumentation.

British Standards Institution (2008). BS 6472-1:2008 Guide to evaluation of human exposure to vibration in buildings - part 1: vibration sources other than blasting.

Brown, J. C. (1999). Computer identification of musical instruments using pattern recognition with cepstral coefficients as features. *The Journal of the Acoustical Society of America*, 105(3):1933–1941.

Brown, J. C. and Smaragdis, P. (2009). Hidden Markov and Gaussian mixture models for automatic call classification. *The Journal of the Acoustical Society of America*, 125(6):EL221–EL224.

Brown, J. C., Smaragdis, P., and Nousek-McGregor, A. (2010). Automatic identification of individual killer whales. *The Journal of the Acoustical Society of America*, 128(3):EL93–EL98.

Carroll, J. D. and Chang, J.-J. (1970). Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition. *Psychometrika*, 35(3):283–319.

Clark, C. and Stansfeld, S. A. (2007). The effect of transportation noise on health and cognitive development: A review of recent evidence. *International Journal of Comparative Psychology*, 20:145–158.

Commission of the European Communities (2001). White Paper - European transport policy for 2010: time to decide.

Coombs, C. H. (1960). A theory of data. *The Pyschological Review*, 67(3):143–159.

Corbridge, C. and Griffin, M. J. (1986). Vibration and comfort: vertical and lateral motion in the range 0.5 to 5.0 Hz. *Ergonomics*, 29:249–272.

Corradino, C. (1990). Proximity structure in a captive colony of Japanese monkeys (Macaca fuscata fuscata): an application of multidimensional scaling. *Primates*, 31:351–362.

Cox, T. F. and Cox, M. A. A. (2001). *Multidimensional scaling.* Chapman & Hall/CRC, Boca Raton, FL, 2nd edition.

Cycling '74 (2013). Max, version 6.1.6. San Francisco, CA: Cycling '74.

David, H. A. (1988). *The method of paired comparisons.* C. Griffin, London, UK, 2nd edition.

De Jong, R. G. (1979). A Dutch study on railroad traffic noise. *Journal of Sound and Vibration*, 66(3):497–502.

De Jong, R. G. and Miedema, H. M. E. (1996). Is freight traffic noise more annoying than passenger traffic noise? *Journal of Sound and Vibration*, 193(1):35–38.

De Leeuw, J. and Heiser, W. J. (1980). Multidimensional scaling with restrictions on the configurations. In Krishnaiah, P. R., editor, *Multivariate analysis*, pages 501–522. North-Holland Publishing Company, Amsterdam, Netherlands.

Department of Transport (1988). Calculation of road traffic noise.

Department of Transport (1995). Calculation of railway noise.

Deutsches Institut für Normung (1999). DIN 4150-2 Structural vibration part 2: human exposure to vibration in buildings.

Donati, P., Grosjean, A., Mistrot, P., and Roure, L. (1983). The subjective equivalence of sinsusoidal and random whole-body vibration in the sitting position (an experimental study using the "floating reference vibration" method). *Ergonomics*, 26:251–273.

Dupuis, H., Hartung, E., and Louda, L. (1972a). Random vibrations of a limited frequency range compared with sinusoidal vibrations with regard to its effects on man. SAM TT G-115. School of Aerospace Medicine, Brooks Air Force Base, TX.

Dupuis, H., Hartung, E., and Louda, L. (1972b). The effect of random vibrations of a limited frequeny band compared with sinusoidal vibrations, on human beings. Library translation 1603. Royal Aircraft Establishment, Farnborough, UK.

Dupuis, H., Hartung, E., and Louda, L. (1972c). Vergleich regelloser schwingungen eines begrenzten frequenzbereiches mit sinusförmigen schwingungen hinsichtlich der einwirkung auf den menschen. *Ergonomics*, 15:237–265.

Eldick Thieme, H. C. A. V. (1961). Passenger riding comfort criteria and methods of analysing ride and vibration data. In *Proceedings of the Society of Automotive Engineers International Congress*, Detroit, MI.

Elmenhorst, E.-M., Pennig, S., Rolny, V., Quehl, J., Mueller, U., Maaß, H., and Basner, M. (2012). Examining nocturnal railway noise and aircraft noise in the field: sleep, psychomotor performance, and annoyance. *The Science of the total environment*, 424:48–56.

Engeldrum, P. G. (2000). *Psychometric scaling: A toolkit for imaging systems development.* Imcotek Press, Winchester, MA.

European Broadcasting Union (1998). EBU Tech. 3276 Listening conditions for the assessment of sound programme material: monophonic and two-channel stereophonic.

Fairley, T. E. and Griffin, M. J. (1990). The apparent mass of the seated human body in the fore-and-aft and lateral directions. *Journal of Sound and Vibration*, 139(2):299–306.

Fastl, H. and Zwicker, E. (2007). *Psychoacoustics: facts and models.* Springer, New York, NY, 3rd edition.

Federal Transit Administration (2006). FTA-VA-90-1003-06 Transit noise and vibraton impact assessment.

Fidell, S., Barber, D. S., and Schultz, T. J. (1991). Updating a dosage-effect relationship for the prevalence of annoyance due to general transportation noise. *The Journal of the Acoustical Society of America*, 89(1):221–233.

Fidell, S., Mestre, V., Schomer, P., Berry, B., Gjestland, T., Vallet, M., and Reid, T. (2011). A first-principles model for estimating the prevalence of annoyance with aircraft noise exposure. *The Journal of the Acoustical Society of America*, 130(2):791–806.

Fields, J. M. (1979). Railway noise and vibration annoyance in residential areas. *Journal of Sound and Vibration*, 66(3):445–458.

Fields, J. M., De Jong, R. G., Gjestland, T., Flindell, I. H., Job, R. F. S., Kurra, S., Lercher, P., Vallet, M., Yano, T., Guski, R., Felscher-Suhr, U., and Schumer, R. (2001). Standardized general-purpose noise reaction questions for community noise surveys: research and a recommendation. *Journal of Sound and Vibration*, 242(4):641–679.

Fields, J. M. and Walker, J. G. (1982). The response to railway noise in residential areas in Great Britain. *Journal of Sound and Vibration*, 85(2):177–255.

Gidlöf-Gunnarsson, A., Ögren, M., Jerson, T., and Öhrström, E. (2012). Railway noise annoyance and the importance of number of trains, ground vibration, and building situational factors. *Noise & health*, 14(59):190–201.

Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *The Journal of the Acoustical Society of America*, 61(5):1270–1277.

Grey, J. M. and Gordon, J. W. (1978). Perceptual effects of spectral modifications on musical timbres. *The Journal of the Acoustical Society of America*, 63(5):1493–1500.

Griffin, M. J. (1976). Subjective equivalence of sinusoidal and random whole-body vibration. *The Journal of the Acoustical Society of America*, 60(5):1140–1145.

Griffin, M. J. (1996). *Handbook of human vibration*. 988 p. Academic Press, London, UK.

Griffin, M. J. and Whitham, E. M. (1980a). Discomfort produced by impulsive whole-body vibration. *The Journal of the Acoustical Society of America*, 68(5):1277–1284.

Griffin, M. J. and Whitham, E. M. (1980b). Time dependency of whole-body vibration discomfort. *The Journal of the Acoustical Society of America*, 68(5):1522–1523.

Griffin, M. J., Whitham, E. M., and Parsons, K. C. (1982). Vibration and comfort: I. Translational seat vibration. *Ergonomics*, 25:603–630.

Groothuis-Oudshoorn, C. G. M. and Miedema, H. M. E. (2006). Multilevel grouped regression for analyzing self-reported health in relation to environmental factors: the model and its application. *Biometrical Journal*, 48(1):67–82.

Gulliksen, H. (1956). A least squares solution for paired comparisons with incomplete data. *Psychometrika*, 21:125–134.

Guski, R., Felscher-Suhr, U., and Schuemer, R. (1999). The concept of noise annoyance: how international experts see it. *Journal of Sound and Vibration*, 223(4):513–527.

Hiramatsu, K. and Griffin, M. J. (1984). Predicting the subjective response to nonsteady vibration based on the summation of subjective magnitude. *The Journal of the Acoustical Society of America*, 76(4):1080–1089.

Horan, C. B. (1969). Multidimensional scaling: combining observations when individuals have different perceptual structures. *Psychometrika*, 34(2):139–165.

Hosmer, D. W. and Lemeshow, S. (1989). *Applied logistic regression (Wiley series in probability and statistics)*. John Wiley & Sons, New York, NY.

Howarth, H. V. C. and Griffin, M. J. (1988a). Human response to simulated intermittent railway-induced building vibration. *Journal of Sound and Vibration*, 120(2):413–420.

Howarth, H. V. C. and Griffin, M. J. (1988b). The frequency dependence of subjective reaction to vertical and horizontal whole-body vibration at low magnitudes. *The Journal of the Acoustical Society of America*, 83(4):1406–1413.

Howarth, H. V. C. and Griffin, M. J. (1990). Subjective response to combined noise and vibration: summation and interaction effects. *Journal of Sound and Vibration*, 143(3):443–454.

Howarth, H. V. C. and Griffin, M. J. (1991). The annoyance caused by simultaneous noise and vibration from railways. *The Journal of the Acoustical Society of America*, 89(5):2317–2323.

Huang, Y. and Griffin, M. J. (2014). The relative discomfort of noise and vibration: effects of stimulus duration. *Ergonomics*, 57(8):1244–1255.

IBM (2011). IBM SPSS Statistics for Windows, version 20.0. Armonk, NY: IBM Corp.

International Organization for Standardization (2003a). ISO 2631-2:2003 Mechanical vibration and shock - evaluation of human exposure to whole-body vibration - part 2: vibration in buildings (1 Hz to 80 Hz).

International Organization for Standardization (2003b). ISO/TS 15666:2003 Acoustics — Assessment of noise annoyance by means of social and socio-acoustic surveys.

International Telecommunication Union (2014). ITU-R BS 1116-2 Methods for the subjective assessment of small impairments in audio systems.

Italian Organization for Standardization (1990). UNI 9641:1990 Misura delle vibrazioni negli edifici e criteri di valutazione del disturbo.

Jik Lee, P. and Griffin, M. J. (2013). Combined effect of noise and vibration produced by high-speed trains on annoyance in buildings. *The Journal of the Acoustical Society of America*, 133(4):2126–35.

Job, R. F. S. (1988). Community response to noise: a review of factors influencing the relationship between noise exposure and reaction. *The Journal of the Acoustical Society of America*, 83(3):991–1001.

Jones, A. J. and Saunders, D. J. (1972). Equal comfort contours for whole body vertical, pulsed sinusoidal vibration. *Journal of Sound and Vibration*, 23(1):1–14.

Jones, A. J. and Saunders, D. J. (1974). A scale of human reaction to whole body, vertical, sinusoidal vibration. *Journal of Sound and Vibration*, 35:503–520.

Kandel, E. R., Schwartz, J. H., and Jessell, T. M. (2000). *Principles of neural science*. McGraw-Hill, New York, NY, 4th edition.

Klæboe, R., Öhrström, E., Turunen-Rise, I. H., Bendtsen, H., and Nykänen, H. (2003a). Vibration in dwellings from road and rail traffic — part III : towards a common methodology for socio-vibrational surveys. *Applied Acoustics*, 64:111–120.

Klæboe, R., Turunen-Rise, I. H., Hårvik, L., and Madshus, C. (2003b). Vibration in dwellings from road and rail traffic — part II: exposure-effect relationships based on ordinal logit and logistic regression models. *Applied Acoustics*, 64:89–109.

Knall, V. (1996). Railway noise and vibration: effects and criteria. *Journal of Sound and Vibration*, 193(1):9–20.

Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1):1–27.

Kryter, K. D. (1982). Community annoyance from aircraft and ground vehicle noise. *The Journal of the Acoustical Society of America*, 72(4):1222–1242.

Kryter, K. D. (1983). Response of K. D. Kryter to modified comments by T. J. Schultz on K. D. Kryter's paper, "Community annoyance from aircraft and ground vehicle noise" [J. Acoust. Soc. Am. 72, 1243-1252 (1982)]. *The Journal of the Acoustical Society of America*, 73(3):1066–1068.

Landström, U., Lundström, R., and Strandberg, U.-K. (1983a). Perceptionför horisontella vibrationer i staende stallning (Perception of horizontal vibrations in standing posture). Undersokningsrapport 1983:40. Umea: Arbetarskyddsstyrelsen.

Landström, U., Lundström, R., and Strandberg, U.-K. (1983b). Perceptionforhallanden avseende helkroppsvibrationer i staende stallning (Perception of whole-body vibration in standing posture). Undersokningsrapport 1983:17. Stockholm: Arbetarskyddsstyrelsen.

Lapointe, F.-J. and Legendre, P. (1994). A classification of pure malt scotch whiskies. *Applied Statistics*, 43(1):237–257.

Leatherwood, J. D. and Dempsey, T. K. (1976). Psychophysical relationships characterizing human response to whole-body sinusoidal vertical vibration. NASA Technical Note TN D-8188.

Long, J. S. (1997). *Regression models for categorical and limited dependent variables.* Sage Publications, Inc, Thousand Oaks, CA.

Luce, R. D. and Suppes, P. (1965). Preference, utility, and subjective probability. In Luce, R. D., Bush, R. R., and Galanter, E. H., editors, *Handbook of mathematical psychology, vol. 3*, pages 249–410. Wiley, New York, NY.

Madshus, C., Bessason, B., and Hårvik, L. (1996). Prediction model for low frequency vibration from high speed railways on soft ground. *Journal of Sound and Vibration*, 193(1):195–203.

Magid, E. B., Coermann, R. R., and Ziegenruecker, G. H. (1960). Human tolerance to whole body sinusoidal vibration: short-time, one-minute and three-minute studies. *Aerospace Medicine*, 31:915–924.

Mansfield, N. J. (2005). *Human response to vibration.* CRC Press, Boca Raton, FL.

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979). *Multivariate analysis.* Academic Press, London, UK.

Matsumoto, Y. and Griffin, M. J. (1998). Dynamic response of the standing human body exposed to vertical vibration: influence of posture and vibration magnitude. *Journal of Sound and Vibration*, 212(1):85–107.

Matsumoto, Y. and Griffin, M. J. (2002). Effect of muscle tension on non-linearities in the apparent masses of seated subjects exposed to vertical whole-body vibration. *Journal of Sound and Vibration*, 253(1):77–92.

McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., and Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: common dimensions, specificities, and latent subject classes. *Psychological Research*, 58(3):177–192.

McCullagh, P. and Nelder, J. A. (1989). *Generalized linear models*. Springer, London, UK, 2nd edition.

McKay, J. R. (1971). Human perception of whole body vibration. Memorandum No. 435. Institute of Sound and Vibration Research, University of Southampton.

Miedema, H. M. E. and Oudshoorn, C. G. M. (2001). Annoyance from transportation noise: relationships with exposure metrics DNL and DENL and their confidence intervals. *Environmental Health Perspectives*, 109(4):409–416.

Miedema, H. M. E. and Vos, H. (1998). Exposure-response relationships for transportation noise. *The Journal of the Acoustical Society of America*, 104(6):3432–3445.

Miwa, T. (1967a). Evaluation methods for vibration effect. Part 1. Measurements of threshold and equal sensation contours of whole body for vertical and horizontal vibrations. *Industrial Health*, 5:183–205.

Miwa, T. (1967b). Evaluation methods for vibration effect. Part 3. Measurements of threshold and equal sensation contours on hand for vertical and horizontal sinusoidal vibrations. *Industrial Health*, 5:213–220.

Miwa, T. (1967c). Evaluation methods for vibration effects. Part 2. Measurement of equal sensation level for whole body between vertical and horizontal sinusoidal vibrations. *Industrial Health*, 5:206–212.

Miwa, T. (1968a). Evaluation methods for vibration effect. Part 4. Measurements of vibration greatness for whole body and hand in vertical and horizontal vibrations. *Industrial Health*, 6:1–10.

Miwa, T. (1968b). Evaluation methods for vibration effect. Part 7. The vibration greatness of the pulses. *Industrial Health*, 6:143–164.

Miwa, T., Yonekawa, Y., and Kojima-Sudo, A. (1973). Measurement and evaluation of environmental vibration. Part 3. Vibration exposure criterion. *Industrial Health*, 11:185–196.

Moehler, U. (1988). Community response to railway noise: a review of social surveys. *Journal of Sound and Vibration*, 120(2):321–332.

Moehler, U., Liepert, M., Schuemer, R., and Griefahn, B. (2000). Differences between railway and road traffic noise. *Journal of Sound and Vibration*, 231(3):853–864.

Morioka, M. and Griffin, M. J. (2006). Magnitude-dependence of equivalent comfort contours for fore-and-aft, lateral and vertical whole-body vibration. *Journal of Sound and Vibration*, 298(3):755–772.

Morioka, M. and Griffin, M. J. (2010). Magnitude-dependence of equivalent comfort contours for fore-and-aft, lateral, and vertical vibration at the foot for seated persons. *Journal of Sound and Vibration*, 329(14):2939–2952.

Morrissey, J. H. (1955). New method for the assignment of psychometric scale values from incomplete paired comparisons. *Journal of the Optical Society of America*, 45(5):373–378.

Mosteller, F. (1951). Remarks on the method of paired comparisons: I. The least squares solution assuming equal standard deviations and equal correlations. *Psychometrika*, 16(1):3–9.

Nawayseh, N. and Griffin, M. (2003). Non-linear dual-axis biodynamic response to vertical whole-body vibration. *Journal of Sound and Vibration*, 268(3):503–523.

Nawayseh, N. and Griffin, M. J. (2005). Non-linear dual-axis biodynamic response to fore-and-aft whole-body vibration. *Journal of Sound and Vibration*, 282:831–862.

Nordtest Tekniikantie (2001). NT ACOU 106 Acoustics: assesment of annoyance caused by vibrations in dwellings from road and rail traffic by means of socio-vibrational and social surveys.

Notess, C. B. (1963). Flexible airplanes, gusts, crew, a triangle. Full-scale Division Memorandum No. 343. Cornell Aeronautical Laboratory Inc., Ithaca, NY.

Oborne, D. J. and Boarer, P. A. (1982). Subjective response to whole body vibration: the effects of posture. *Ergonomics*, 25:673–681.

ÖENORM (2010). ÖENORM S 9012:2010 Beurteilung der einwirkung von schwingungsimmissionen des landgebundenen verkehrs auf den menschen in gebäuden - schwingungen und sekundärer luftschall.

Ögren, M., Öhrström, E., and Gidlöf-Gunnarsson, A. (2009). Effects of railway noise and vibrations on sleep – experimental studies within the Swedish research program TVANE. In *Proceedings of Euronoise*, Edinburgh, UK.

Öhrström, E. (1997). Effects of exposure to railway noise — a comparison between areas with and without vibration. *Journal of Sound and Vibration*, 205(4):555–560.

Öhrström, E., Gidlöf-Gunnarsson, A., Ögren, M., and Jerson, T. (2009). Effects of railway noise and vibration in combination: field and laboratory studies. In *Proceedings of Euronoise*, Edinburgh, UK.

Öhrström, E. and Skånberg, A. B. (1996). A field survey on effects of exposure to noise and vibration from railway traffic, part I: annoyance and activity disturbance effects. *Journal of Sound and Vibration*, 193(1):39–47.

Parizet, E. (2002). Paired comparison listening tests and circular error rates. *Acta Acustica United with Acustica*, 88:594–598.

Parizet, E., Guyader, E., and Nosulenko, V. (2008). Analysis of car door closing sound quality. *Applied Acoustics*, 69(1):12–22.

Parsons, K. C. and Griffin, M. J. (1988). Whole-body vibration perception thresholds. *Journal of Sound and Vibration*, 121:237–258.

Parsons, K. C., Griffin, M. J., and Whitham, E. M. (1982). Vibration and comfort. III. Translational vibration of the feet and back. *Ergonomics*, 25:705–719.

Paulsen, R. and Kastka, J. (1995). Effects of combined noise and vibration on annoyance. *Journal of Sound and Vibration*, 181(2):295–314.

Pennig, S., Quehl, J., Mueller, U., Rolny, V., Maass, H., Basner, M., and Elmenhorst, E.-M. (2012). Annoyance and self-reported sleep disturbance due to night-time railway noise examined in the field. *The Journal of the Acoustical Society of America*, 132(5):3109–3117.

Peris, E., Woodcock, J., Sica, G., Moorhouse, A. T., and Waddington, D. C. (2012). Annoyance due to railway vibration at different times of the day. *The Journal of the Acoustical Society of America*, 131(2):EL191–EL196.

Peris, E., Woodcock, J., Sica, G., Sharp, C., Moorhouse, A. T., and Waddington, D. C. (2014). Effect of situational, attitudinal and demographic factors on railway vibration annoyance in residential areas. *The Journal of the Acoustical Society of America*, 135(1):194–204.

Purves, D., Augustine, G. J., Fitzpatrick, D., Katz, L. C., LaMantia, A. S., McNamara, J. O., and Williams, S. M., editors (2001). *Neuroscience*. Sinauer Associates, Sunderland, MA, 2nd edition.

Putnam, L. E. and Roth, W. T. (1990). Effects of stimulus repetition, duration, and rise time on startle blink and automatically elicited P300. *Psychophysiology*, 27(3):275–97.

Real Decreto (2007). Real Decreto 1367/2007 de 19 de octubre, por el que se desarrolla la Ley 37/2003, de 17 de noviembre, del Ruido, en lo referente a zonificación acústica, objetivos de calidad y emisiones acústicas.

Reiher, H. and Meister, F. J. (1931). The senstiveness of the human body to vibrations (Empfindlichkeit des menschen gegen erschutterung). Forschung (VDI) 2: 381-386. Translation, Report No. F-TS-616-RE (1946). Headquarters Air Material Command, Wright Field, Dayton, OH.

Rice, C. G. and Izumi, K. (1984). Annoyance due to combinations of noises. In *Proceedings of the Institute of Acoustics Spring Conference.*

RIVAS (2011). Deliverable D1.4: Review of existing standards, regulations and guidelines, as well as laboratory and field studies concerning human exposure to vibration.

Ross, R. T. (1934). Optimum orders for the presentation of pairs in the method of paired comparisons. *Journal of Educational Psychology*, 25:375–382.

Saremi, M., Grenèche, J., Bonnefond, A., Rohmer, O., Eschenlauer, A., and Tassi, P. (2008). Effects of nocturnal railway noise on sleep fragmentation in young and middle-aged subjects as a function of type of train and sound level. *International Journal of Psychophysiology*, 70:184–191.

SBR (2006). SBR Richtlijn – Deel B: Hinder voor personen in gebouwen.

Schomer, P., Mestre, V., Fidell, S., Berry, B., Gjestland, T., Vallet, M., and Reid, T. (2012). Role of community tolerance level (CTL) in predicting the prevalence of the annoyance of road and rail noise. *The Journal of the Acoustical Society of America*, 131(4):2772–2786.

Schroeder, M. R., Gottlob, D., and Siebrasse, K. F. (1974). Comparative study of European concert halls: correlation of subjective preference with geometric and acoustic parameters. *The Journal of the Acoustical Society of America*, 56(4):1195–1201.

Schubert, E. and Wolfe, J. (2006). Does timbral brightness scale with frequency and spectral centroid? *Acta Acustica United with Acustica*, 92:820–825.

Schultz, T. J. (1978). Synthesis of social surveys on noise annoyance. *The Journal of the Acoustical Society of America*, 64(2):377–405.

Schultz, T. J. (1982). Comments on K. D. Kryter's paper, "Community annoyance from aircraft and ground vehicle noise". *The Journal of the Acoustical Society of America*, 72(4):1243–1252.

Shamir, L., Yerby, C., Simpson, R., von Benda-Beckmann, A. M., Tyack, P., Samarra, F., Miller, P., and Wallin, J. (2014). Classification of large acoustic datasets using machine learning and crowdsourcing: application to whale calls. *The Journal of the Acoustical Society of America*, 135(2):953–962.

Shoenberger, R. W. (1975). Subjective response to very low frequency vibration. *Aviation, Space and Environmental Medicine*, 46:785–790.

Shoenberger, R. W. and Harris, C. S. (1971). Psychophysical assessment of whole-body vibration. *Human Factors*, 13:41–50.

Sica, G., Peris, E., Woodcock, J. S., Moorhouse, A. T., and Waddington, D. C. (2014). Design of measurement methodology for the evaluation of human exposure to vibration in residential environments. *Science of the Total Environment*, 482-483:461–471.

Sica, G., Woodcock, J., Peris, E., Koziel, Z., Moorhouse, A. T., and Waddington, D. C. (2011). Human response to vibration in residential environments (NANR209), technical report 3: calculation of vibration exposure.

Simic, D. (1974). Contributions to the optimisation of the oscillatory properties of a vehicle: physiological foundations of comfort during oscillations. Library Translation No. 1707. Royal Aircraft Establishment, Farnborough, UK.

Smith, M. G., Croy, I., Ögren, M., and Persson Waye, K. (2013). On the influence of freight trains on humans: a laboratory investigation of the impact of nocturnal low frequency vibration and noise on sleep and heart rate. *PloS one*, 8(2):e55829.

Standard Norge (2005). NS 8176:2005 Vibration and shock - measurement of vibration in buildings from landbased transport and guidance to evaluation of effects on human beings.

Stevens, S. S. (1975). *Psychophysics: introduction to its perceptual, neural, and social prospects*. John Wiley & Sons, New York, NY.

Swedish Standards Institute (1992). SS 460 48 61:1992 Vibration oct stöt – mätning och riktvärden för bedömning av komfort i byggnader.

Taylor, S. M. (1982). A comparison of models to predict annoyance reactions to noise from mixed sources. *Journal of Sound and Vibration*, 81(1):123–138.

Thompson, D. (2009). *Railway noise and vibration: mechanisms, modelling and means of control*. Elsevier, Oxford, UK.

Thurstone, L. L. (1927). A law of comparative judgement. *Psychology Review*, 34:273–276.

TRL, Temple, ISVR, and Arup Acoustics (2007). NANR172 Human response to vibration in residential environments.

Trollé, A., Marquis-Favre, C., and Klein, A. (2014). Short-term annoyance due to tramway noise: determination of an acoustical indicator of annoyance via multilevel regression analysis. *Acta Acustica United with Acustica*, 100(1):34–45.

Truett, J., Cornfield, J., and Kannel, W. (1967). A multivariate analysis of the risk of coronary heart disease in Framingham. *Journal of Chronic Diseases*, 20(7):511–24.

Tsukida, K. and Gupta, M. R. (2011). How to Analyze Paired Comparison Data. Technical Report UWEETR–2011-0004, University of Washington.

Turunen-Rise, I., Brekke, A., Hårvik, L., Madshus, C., and Klæboe, R. (2003). Vibration in dwellings from road and rail traffic — part I: a new Norwegian measurement standard and classification system. *Applied Acoustics*, 64(1):71–87.

Von Bismarck, G. (1974). Sharpness as an attribute of the timbre of steady sounds. *Acustica*, 30:150–172.

Waddington, D. C., Woodcock, J., Peris, E., Condie, J., Sica, G., Moorhouse, A. T., and Steele, A. (2014). Human response to vibration in residential environments. *The Journal of the Acoustical Society of America*, 135(1):182–193.

Watts, G. R. (1984). Vibration nuisance from road traffic - results of a 50 site survey. TRRL Laboratory Report 1119.

Watts, G. R. (1987). Traffic-induced ground-borne vibrations in dwellings. TRRL Research Report 102.

Watts, G. R. (1990). Traffic induced vibrations in buildings. TRRL Research Report 246.

Weber, R. (1999). Interior sound quality - assesment of acceleration noises. In *Proceedings of Forum Acusticum*, Berlin, Germany.

Weisberg, S. (2005). *Applied linear regression.* John Wiley & Sons, Hoboken, NJ, 3rd edition.

Whitham, E. M. and Griffin, M. J. (1978). The effects of vibration frequency and direction on the location of areas of discomfort caused by whole-body vibration. *Applied Ergonomics*, 9:231–239.

Whittle, N., Peris, E., Condie, J., Woodcock, J., Brown, P., Moorhouse, A. T., Waddington, D. C., and Steele, A. (2015). Development of a social survey for the study of vibration annoyance in residential environments: good practice guidance. *Applied Acoustics*, 87:83–93.

Wiebe, E., Sandor, J., Cheron, C., and Haas, S. (2011). ERRAC roadmap WP 01 - the greening of surface transport, towards 2030 – noise and vibrations roadmap for the European railway sector.

Woodcock, J. (2013). *Field and laboratory studies into the human response to groundborne vibration: exposure-response relationships, perceptual dimensions, and models of annoyance.* PhD thesis, University of Salford.

Woodcock, J., Moorhouse, A. T., and Waddington, D. C. (2014a). A multidimensional evaluation of the perception and annoyance caused by railway induced groundborne vibration. *Acta Acustica United with Acustica*, 100(4):614–627.

Woodcock, J. S., Peris, E., Moorhouse, A. T., and Waddington, D. C. (2014b). Guidance document for the evaluation of railway vibration. CargoVibes Deliverable D1.5.

Woodruff, H. J. and Griffin, M. J. (1987). A survey of the effect of railway-induced vibration on the community. ISVR Technical Report 160.

World Health Organization (2000). Guidelines for community noise.

Wyllie, I. H. and Griffin, M. J. (2007). Discomfort from sinusoidal oscillation in the roll and lateral axes at frequencies between 0.2 and 1.6 Hz. *The Journal of the Acoustical Society of America*, 121(5):2644–2654.

Yano, T., Morihara, T., and Sato, T. (2005). Community response to Shinkansen noise and vibration: a survey in areas along the Sanyo Shinkansen Line. In *Proceedings of Forum Acusticum*, Budapest, Hungary.

Yokoshima, S., Morihara, T., Sano, Y., Ota, A., and Tamura, A. (2011). Community response to Shinkansen Railway vibration. In *Proceedings of Internoise*, Osaka, Japan.

Yokoshima, S. and Tamura, T. (2005). Combined annoyance due to the Shinkansen Railway noise and vibration. In *Proceedings of Internoise*, Rio de Janeiro, Brazil.

Yonekawa, Y. and Miwa, T. (1972). Sensational responses of sinusoidal whole body vibrations with ultra-low frequencies. *Industrial Health*, 10:63–76.

Zapfe, J. A., Saurenman, H., and Fidell, S. (2009). Ground-borne noise and vibration in buildings caused by rail transit. Contractor's final report for TCRP Project D-12.