A Framework for Road Traffic Noise Auralisation

Alex Southern[1&2]
[1] AECOM Acoustics, 225 Bath Street, Glasgow, G2 4GZ

Damian Murphy[2]
[2] AudioLab, Department of Electronics, University of York, Heslington, York, UK, YO10 5DD

Summary
Auralisation of road traffic noise has gained an increased amount of attention in the literature in recent years. Tools are required that can form the basis for the proper evaluation of the subjective aural impact of planned or existing road transport developments in residential areas. This paper presents a new framework for the capture, analysis and synthesis of moving road vehicles within a virtual auditory environment rendered using spatial sound techniques for headphone or loudspeaker based systems. The new framework is presented in the context of the spatial sound rendering of a single car pass-by sourced from a roadside recording. The analysis and synthesis processing is performed in the time-frequency domain and incorporates changes at the listening position based on an approximation of the vehicle's directional noise radiation pattern. The presented framework is then used to generate single vehicle pass-bys that simulate the original vehicle recording. Blind listening tests indicate that the processing framework can reliably synthesize vehicle pass-by noise in a manner that listeners deem subjectively plausible.

PACS no. xx.xx.Nn, xx.xx.Nn

1. Introduction

In June 2002 the European Parliament and Council adopted the The Environmental Noise Directive (END) (2009/49/EC) relating to the assessment and management of environmental noise [1]. The END is applicable to sources of noise to which the public are exposed, including built up areas and quiet areas in open country, as well as noise sensitive buildings, rather than localised noise from domestic activities. Road traffic noise in open country is of particular relevance to this work.

Since the adoption of the END, there has been an increasing amount of interest in the affects and abatement of road traffic noise. The World Health Organisation has also recognised it as a serious problem for public health with annoyance being recognised as the most widespread of these adverse effects [2]. This is not surprising when considering that environmental noise, in urban environments in particular, is often dominated by road traffic noise, for example Calixto *et al.* [3] reported that around 70% of total noise in the urban environment is due to vehicles. More recently the WHO published the *Night Noise Guidelines for Europe* (NNG) and as an example showed that in the Netherlands the percentage of the population highly disturbed by noise during sleep in 1998 and 2003 had increased by around 8% due to all noise sources, with road traffic noise accounting for around 50% of total sleep disturbances in both 1998 and 2003 [4]. Furthermore the NNG also gathered evidence and reported on the adverse affects of noise on human health and well-being.

In the UK, the Office for National Statistics predicts that by 2033 the population will increase to 71.6 million from 61.3 million in 2008 [5]. The increasing and ageing population is in common with other European countries and it suggests that more people are likely to be contributing to and be affected by road traffic noise in the future.

It is therefore important to sustainably manage noise pollution from road traffic vehicles and for planners and consultants to be able to optimise designs at the planning stage. This is primarily achieved based on predicted noise level exposures, however these are an objective measure of noise whereas the auditory system is sensitive to, and subjectively interprets, the varying time-frequency dependent characteristics of a soundfield rather than the time-averaged noise level alone. Furthermore it is also recognised that quantitative assessments conducted based on noise level exposures are difficult to communicate with non-specialists.

Auralisation techniques offer the ability to present a predicted/modelled or recorded soundfield of a given environment to a listener. Therefore, auralisation techniques can potentially provide a means to predict the perceived sound quality, in addition to sound quantity, for a proposed road scheme development. Previous authors have considered the auralisation of road traffic noise in urban areas and have presented different methods for synthesising road vehicle soundfields and these are discussed further in Section 2 in the context of this work [6][7].

Modern road vehicles typically have two primary sources of noise - the engine/exhaust system and rolling tyre noise. For fast moving vehicles, particularly on major (motorway/freeway) roads through otherwise open countryside, the rolling tyre noise is the dominant noise source and will generally mask any engine/exhaust noise when the vehicle is cruising e.g. not accelerating, particularly when listening over large open distances. This paper presents a framework and method for auralising road traffic noise, where the focus is to synthesise vehicles moving at a relatively high speed compared to those in built-up urban environments, and hence engine/exhaust related noise can be discounted. Furthermore, single car pass-by recordings are analysed and synthesised and compared under listening test conditions to assess their plausibility as a representation of the original recordings.

In Section 2 an overview of a general analysis/synthesis framework for road traffic noise auralisation is presented and discussed in relation to similar works. In Section 3 a general method for encoding recorded vehicle pass-bys for use in this auralisation framework is described. This is in addition to a preliminary investigation into an appropriate data compression scheme for reducing the size of the analysis data for each vehicle.

2. Auralisation Framework Overview

An overview of an auralisation framework for road traffic noise is presented in Figure 1 and similar generalised frameworks have been presented previously i.e. [8][6][9]. In [9] a real-time overlap-add algorithm is used to synthesise vehicle engine sound components using granular synthesis and road-tyre interaction using an asynchronous overlap-add scheme, where random length samples are added at random locations. This differs from [6] in that it uses a constant overlap-add spectral modelling synthesis method often applied in musical instrument synthesis. The trajectories of periodic components are analysed and subtracted from the recording to leave a residual signal that represents the noise-like components of the analysed sound. The primary concern in [8] is to model hybrid vehicles and assess annoyance compared to regular combustion engines. An overview of an auralisation system is presented along with a brief overview

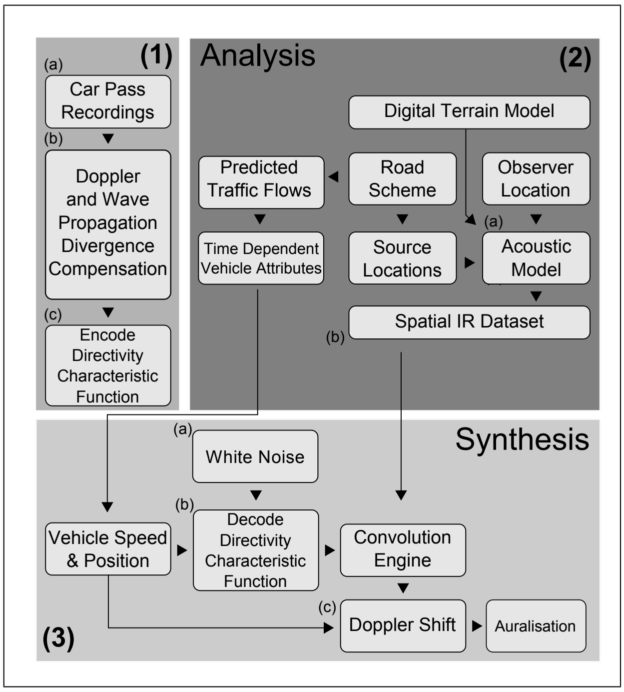

Figure 1. An overview of a processing framework for the auralisation of road traffic noise. The analysis of single road vehicle sound recordings is considered separately from the synthesis of moving road traffic vehicles.

of the analysis and synthesis techniques. In particular, the non-periodic components of the vehicle noise are smoothed in the frequency domain which is likely to have a perceptually similar effect to the smoothing applied in this work.

While the engine noise is not considered in this work, the road-tyre interaction is analysed/synthesised differently to that presented in any of the aforementioned works.

The auralisation of road traffic noise can be considered as a product of analysis and synthesis processes. With reference to (1) in Figure 1, there is an initial data acquisition phase (1a).

Obtaining freefield recordings of even a small sample of typical road vehicles moving at high speed presents difficulties and can be very time consuming, especially if different road surface types are required. This work has adopted a passive roadside measurement methodology that allows plausible freefield recording representations of cars to be synthesised. This is followed by some post-processing (1b) discussed in section 2.1. In addition, a directional noise radiation characteristic profile (1c) for each vehicle is estimated from the roadside recording and used to synthesise a moving car, which is of primary interest in this work.

In Figure1 (2), A 3D virtual world representation of an existing/proposed road scheme can be constructed using digital terrain and planning data. Acoustic modelling (2a) can then calculate the sound propagation paths between a vehicle on a section of road to the listening position, a relationship which is defined by an impulse response. In addition, a spatial impulse response can be modelled if directional information about incoming wave fronts at the listening position is also captured, see (2b). There are different types of acoustic modelling methods that can calculate the spatial impulse response in different ways e.g. [10].

With reference to (3) in Figure 1, to produce an auralisation of road traffic noise the spatial impulse response dataset is dynamically convolved with the sound of every vehicle in the scene. In the simplest implementation every vehicle is assumed to possess a directionally uniform noise radiation pattern and it is necessary to store only a freefield recording of every vehicle for convolution with the spatial impulse response dataset. A more considered implementation would also account for the directional noise radiation pattern of each vehicle at the acoustic modelling stage, and this is the motivation behind the approach presented in this paper, indicated by blocks (1c) and (2b).

In Figure 1 (3) synthesis of freefield rolling car tyre noise is achieved within an overlap-add framework by applying a frequency and direction dependent energy weighting (3b) for an individual vehicle to windowed white noise (3a). As the direction of the vehicle with the respect to the listener is constantly changing over time, (3b) will change from one instant to the next. Different techniques for encoding a vehicle's directional characteristic are discussed in section 3.

2.1. Assumptions in this work

This paper is primarily concerned with validating the method of vehicle characteristic analysis and synthesis for use within a road traffic noise auralisation framework. Specifically it is the incorporation of a directivity characteristic function into the synthesis framework that is of interest in this work. To allow a plausibility comparison of the synthesised pass-bys with their corresponding recordings, it is appropriate to synthesise the same setup used in the recordings. This choice simplifies a number of other aspects in the auralisation framework which are explained below. Note that under this condition of single vehicle pass-bys on flat terrain there is no need for a digital terrain model, road scheme design or traffic flow predictions.

2.1.1. Acoustic Modelling

As recordings were conducted in an open soft-ground environment free from substantial reflective surfaces, see section 3, the acoustic modelling of the setup reduces to the calculation of the length of direct sound path to compute the amplitude changes due to wave propagation divergence. However it is explained in the next section that even this length calculation is not necessary. Consequently the any propagation path calculation and convolution of the spatial impulse response dataset is not considered in this work.

2.1.2. Wave Propagation Divergence

As the synthesised listening position is the same as the recording position for the purpose of this work, the average amplitude changes in the recordings are used to both compensate for the wave propagation divergence in the analysis phase and also represent the amplitude changes due to wave propagation divergence in the synthesis phase. In practice the synthesised amplitude changes would be represented in the spatial impulse dataset and be specific to the desired vehicle movement as defined by the *time-dependent vehicle attributes* in Figure 1 .

2.1.3. Doppler Shift

Generally, the speed of each vehicle recording being analysed is required to be known or accurately estimated in order to compensate for the effects of doppler frequency shift. However in this work, accurate speed estimation or doppler shift compensation was not of great importance as the synthesis simply applied the inverse of the doppler compensation filter used in the analysis, and consequently it is not discussed further.

The remainder of this paper discusses validating the plausibility of the synthesis methodology in general and the analysis and synthesis of single vehicle pass-bys using the vehicle directivity characteristic encoding functions.

3. Directivity Encoding Methods

Recordings of single vehicle pass-bys were taken in open country-side on a 60 mph road, with a measurement microphone (with environmental windshield) at 2.6m from the nearside edge of a dry single carriageway road. The effects of amplitude changes due to the distance of the vehicle from the microphone (wave propagation divergence) and to doppler shift were compensated in the recordings. Therefore the angle and frequency dependent noise emission characteristic of the vehicle are present in the resulting signal, but must be captured and stored as efficiently as possible to facilitate an optimal synthesis implementation. The angle dependent response is stored in the frequency domain and is obtained by calculating the FFT of a 1024 sample sliding window with step size of 256. This results in 513 samples per angle. The corresponding angle of each frame is computed trivially by estimating the distance travelled between the centre of the sliding window to the closest position to the microphone.

Below is a description of the methods of encoding the directivity characteristic, namely the *encoding schemes*, that are considered in this paper. Note that simply storing the magnitude response in each time frame is not considered. This is because it produces notable *metallic* sounding artefacts. Generally, smoothing the magnitude response in each angular frame produces a more natural sounding vehicle pass-by but whether or not this, or any other encoding scheme, is perceived as a plausible representation of the real vehicle pass-by will be considered in section 4. The goal of an encoding scheme in this paper is to reduce the number of samples used to represent the magnitude response of each noise radiation angle, while ideally still allowing a perceptually similar synthesis of the pass-by.

No Encoding
Store the smoothed magnitude response for every vehicle noise emission direction. The smoothing was performed using the Savitsky-Golay filter of polynomial order 5 and frame size of 51 samples.

Downsampling
Store a downsampled version of the full magnitude response for every vehicle noise emission direction.

Hybrid Downsampling
Store a downsampled version of the full magnitude response for every vehicle noise emission direction where the downsampling rate is different for the loudest parts of the spectrum compared to the quietest. In this work, the crossover point is determined by the frequency at which a 2^{nd} order polynomial fit of the magnitude characteristics over all angles meets the point that is 10dB below the loudest part of the polynomial fit. This crossover point usually corresponds to around 2 kHz and most of the vehicle's energy is around or below this frequency, although this could change with different recording conditions.

Table I details the parameters of the encoding schemes used in this work and a typical example of how each scheme reduces the number of points in a given magnitude response is presented in Figure 2. Note that the -10dB crossover frequency between the LF and HF downsampling factors for (H1 - H4) is highlighted and that most of the vehicle noise energy is in the lower part of the spectrum.

Later, in the synthesis phase, the downsampled angle dependent magnitude characteristic for the current synthesis frame is upsampled to the original 513 samples and, if the hybrid encoding schemes have been successful, there should be no perceivable difference in the synthesised audio when compared to the no-encoding (NE) scheme.

Table I. A listing of the different downsampling encoding rates used to represent the LF and HF parts of the magnitude response. The typical number of samples used to represent the magnitude response after downsampling.

	Encoding Scheme						
	NE	**H1**	**H2**	**H3**	**H4**	**D1**	**D2**
LF Factor	1	4	8	16	24	32	64
HF Factor	1	32	32	32	32	32	64
LF Samples	55	14	7	4	3	2	1
HF Samples	458	16	16	16	16	16	8
Total	513	30	23	20	19	18	9

4. Listening Test Comparison

In [9] the authors validated the synthesis method using a paired comparison test that asked participants to identify two equal size groups of recorded and synthesised pass-bys from a dataset. This approach is not appropriate for this work as, although the intention is to create auralisation where the vehicle is at large distances from the vehicle, the practicality of recording freefield noise characteristic of vehicles dictates that the best location for the microphone is roadside to avoid noise from unwanted noise sources. As such some low level engine/exhaust noise is sometimes audible causing participants to easily differentiate the difference.

The aim is for the synthesised pass-by events to sound like a similar and plausible representation of the recorded pass-by rather than to synthesise a pass-by that is indistinguishable from the original recording. The listening tests are therefore designed to find out two things:

- Do listeners agree that the synthesised car pass-by is more similar to the recorded pass-by when compared to other car pass-bys?
- Do any of the hybrid encoding schemes produce similar sounding synthesised pass-bys when compared to the NE scheme?

Two listening tests were performed; 10 recorded car pass-bys were selected ensuring that the vehicle was moving at speed but not accelerating. At the recording distance of 2.6m some very low level engine/exhaust noise could be heard on some vehicles and some of these pass-by recordings were used in the test to ensure that a robust encoding method is eventually developed. In the first test, 13 participants were asked to identify one pass-by from 7 options that best matched a given reference. The reference was always a real recording and the 7 other stimuli were made up of the NE equivalent of the reference and 6 other random pass-bys which could be either a real recording, NE, H1 or H2 for a different vehicle. The results are presented in Table II and it could be generally noted that reference stimuli 4, 9 and 10 in particular are more difficult to correctly match. It is unclear why this is occurring, although it could be related to listener fatigue or subtle synthesis artefacts. However it is also important to note that while participants were only 54% successful with stimulus 4 as the worst case, this is still significantly more than chance, i.e. assuming for any given stimulus the probability of 6/13 participants correctly matching the hidden stimulus is calculated using an exact binomial test (one-tailed) $= \mathrm{P}(\mathrm{Y} \geq 6 | n = 13, p = 1/7) = 0.0059$.

Table II. The percentage of correctly matched (M) hidden NE stimuli with the reference.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
M	92	85	77	54	77	92	100	92	62	62

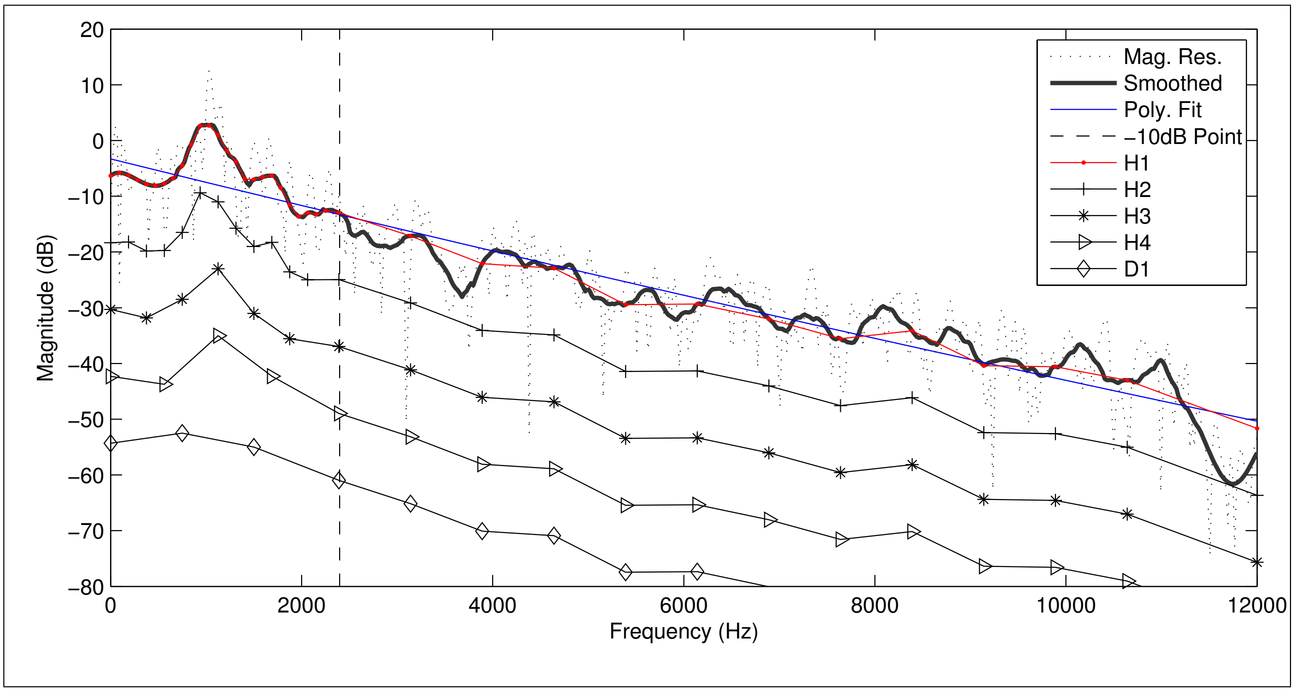

Figure 2. An example of how a typical directional magnitude characteristic is encoded using the downsampling schemes given in TableI. Note that all schemes apart from H1 are offset for graphical clarity.

In the second test, each participant was presented with 10 questions and asked to rate the similarity of 7 stimuli to a reference stimulus on a continuous scale of 0 - 100. The reference was always a synthesised pass-by using the NE scheme and the 7 other stimuli consisted of all encoding scheme, including NE which was therefore a hidden reference. The D2 scheme purposely produces clearly degraded sound quality and encoding artefacts and therefore acts as a hidden anchor which should be easily identifiable as the least similar to the reference. Part 2 is a MUSHRA inspired listening test where similarity instead of quality was the parameter under test [11]. The results of the test for 16 participants are provided in Figure 3 which shows the mean scores for each encoding scheme for each stimulus question along with the 95% confidence interval.

The participants were able to rank the hidden reference stimuli as the most similar to the reference stimuli 68% of the time. The remaining 32%, the hidden reference stimuli was confused with another processing technique, mainly H1 (20%) or H2 (11%). This suggests in some situations that either H1 or H2 can be indistinguishable from the reference under critical listening conditions and therefore they can be accepted as perceptually identical to the non-compressed reference in some practical situations. This is not too surprising when comparing the confidence interval overlap for the different sample numbers in Figure 3, which suggests that the true mean score could lie in a similar region and as such confusions due to chance are possible.

5. Conclusions

This paper has discussed some preliminary results of the similarity of synthesised vehicle pass-by noise to real recordings with application to open field auralisations of road traffic noise. Furthermore, encoding/decoding strategies associated with the vehicle directivity characteristic function have been presented and evaluated using a subjective listening test.

The purpose of the test was to evaluate if the synthesised pass-by events sound like similar and plausible representations of the corresponding recorded pass-by, as opposed to attempting to synthesise a pass-by that is indistinguishable from the original recording under critical listening conditions. Firstly, the listening tests were designed to find if listeners agree that the synthesised car pass-by is more similar to the recorded pass-by when hidden amongst other random pass-by events; this was shown to be true. Secondly, do any of the hybrid encoding schemes produce similar sounding synthesised pass-by when compared to the NE scheme? This was less conclusive, although encouraging as some of the time the H1 encoding scheme was confused with the hidden reference but this did not occur often enough to conclude that participants were in-fact guessing.

Further statistics from part 2 of the listening test could reveal correlations between objective measures of signal differences between the NE and H1-H2 encoding scheme and reveal methodologies of objectively determining the subjective quality of a given vehicle tyre synthesis sample.

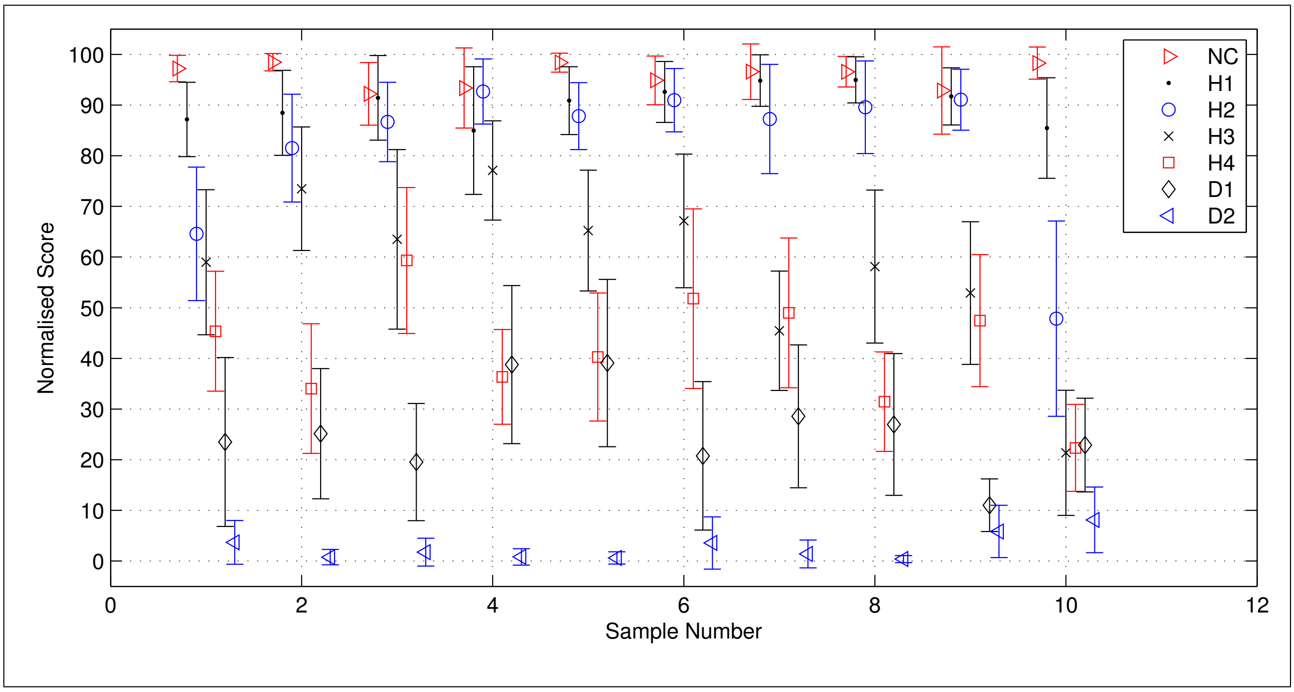

Figure 3. The results of part two of the listening test. The mean score and 95% confidence interval are provided for each encoding scheme for each of 10 questions. Note the mean values and associated confidence intervals have been offset on the x-axis for readability.

Overall the results of this paper suggest more work is needed to improve the robustness of the hybrid encoding scheme to work well with more vehicle recordings, however the general technique using the vehicle directivity characteristic function is a success as the NE encoding scheme produces synthesised vehicle pass-by events that represent the characteristics of the original recordings.

Acknowledgement

This project has been funded by the Royal Society Industry Fellowship Scheme.

References

[1] European Commision Evironment. *The Environmental Noise Directive (2002/49/EC)*. 2015. URL: `http://ec.europa.eu/environment/noise/directive.htm` (visited on 02/15/2015).

[2] den Boer L.C. and A. Schroten. *Traffic noise reduction in Europe: Health effects, social costs and technical and policy options to reduce road and rail traffic noise*. CE Delft, March 2007. URL: `http://www.transportenvironment.org/sites/te/files/media/2008-02_traffic_noise_ce_delft_report.pdf`.

[3] A. Calixto, F.B. Diniz, and P. H. T. Zannin. "The statistical modeling of road traffic noise in an urban setting". In: *Cities* 20 (2003), pp. 23–29.

[4] World Health Organisation Europe. *Night Noise Guidelines for Europe*. 2009. URL: `http://www.euro.who.int/__data/assets/pdf_file/0017/43316/E92845.pdf` (visited on 02/15/2015).

[5] Office for National Statistics. *National Projections: UK population to exceed 65m by 2018*. 2009. URL: `http://www.ons.gov.uk/ons/rel/npp/national-population-projections/2008-based-projections/sum-2008-based-national-population-projections.html` (visited on 02/15/2015).

[6] C. Pendharkar. "Auralization of road vehicles using spectral modeling synthesis." MA thesis. Chalmers, 2012.

[7] J. Maillard and J. Jagla. "Real-Time Auralization of Non-Stationary Traffic Noise - Quantative and Perceptual Validation in an Urban Street". In: *AIA-DAGA Merano*.

[8] A. Telle. *Objective and psychoacoustic noise evaluation of pass by of quiet hybrid vehicles and comparison with standard vehicles*. 2012. URL: `http://www.cityhush.eu/downloads/280213/D030102_HAC_M36.pdf` (visited on 02/15/2015).

[9] J. Maillard and Jagla. "Auralization of non-stationary traffic noise using sample based synthesis - Comparison with pass-by recordings". In: *Internoise 2012, New York City, New York.*

[10] A. Southern et al. "Room Impulse Response Synthesis and Validation Using a Hybrid Acoustic Model". In: *Audio, Speech, and Language Processing, IEEE Transactions on* 21.9 (2013), pp. 1940–1952. ISSN: 1558-7916. DOI: `10.1109/TASL.2013.2263139`.

[11] T. Sporer, J. Liebetrau, and S. Schneider. "Statistics of MUSHRA Revisited". In: *Audio Engineering Society Convention 127*. 2009. URL: `http://www.aes.org/e-lib/browse.cfm?elib=15021`.