Auralization of truck engine sound – preliminary results using a granular approach

Jens Forssén[1], Patrik Andersson[1], Penny Bergman[1], Krister Fredriksson[2], Peter Zimmerman[2]

[1]*Division of Applied Acoustics, Chalmers University of Technology, Göteborg, Sweden,*
Email: jens.forssen@chalmers.se
[2]*Volvo GTT, Göteborg, Sweden*

Introduction

There is an interest in night-time delivery using heavy duty vehicles, motivated by a potential alleviation in traffic flow and congestion during daytime. The current project focuses on urban environments and the sound indoors due to delivery vehicles. Recent work within auralization of road traffic has shown significant progress (e.g. [1-6]). However, to the knowledge of the authors, the source modelling of heavy duty road vehicles are currently lacking in auralization models.

Here, a granular approach is used in order to try and capture the characteristics of a Volvo truck diesel engine, which can be characterised as a rather rich sound in low to mid frequencies. In this first step, constant engine speeds are studied and the objective is an auralization that can function as a source signal to simulate drive-by indoor scenarios, using outdoor and indoor acoustically calculated transfer functions.

Method and results

Recordings of truck engine sound

The sound of a Volvo FM truck (13 litre diesel engine, 460 hp, emission class EEV) was recorded in the Truck Noise Chamber (semi-anechoic laboratory) at Volvo GTT, Lundby, Gothenburg on April 23, 2012 (Figure 1). The laboratory is equipped with a rolling dynamometer for the driving wheels of the truck. It was set to simulate a load of 20 000 kg. The driving conditions considered here are constant speed of 20, 50, 60, and 70 km/h and low idling (about 950, 1400, 1050, 1250, and 600 rpm, respectively). The speed of the truck was manually controlled by a driver and therefore within ±3 km/h, and sometimes varying slightly with time. There is no significant speeds variation during the 4 seconds of recordings used in the sound synthesis.

Microphones (Brüel & Kjær Type 4189, 1/2-inch free-field) were distributed around the truck (fulfilling ISO 3744 for sound power measurements and with additional microphones for detailed investigations). One position in front of the truck and one position on the left hand side of the truck have been used to evaluate the granular approach. The front position is labelled F5 and is 2.00 m from the front at 1.20 m height. The left-hand side position is labelled V1 and is 2.00 m from the truck driver's door at 2.93 m height. On the left-hand side of the truck also the exhaust pipe termination is located. The sound was recorded using a sampling frequency of 44.1 kHz.

Figure 1: The sound of the Volvo FM truck was recorded in a semi-anechoic laboratory. Microphone positions F5 and V1 are indicated by circles.

Granular synthesis

In the granular synthesis approach used here, short time pieces of a laboratory recorded pressure signal are stored and later combined to synthesize an engine source sound. Grains are captured on basis of best correlation with 6-period long sinusoid with variable length. At synthesis, grains are picked at random from a bank and added synchronously using 124 samples overlap and a Hann windowing, as suggested in Ref. [6].

With the five driving conditions and the two microphone positions, we have ten cases, which we have studied in listening tests. The synthesis was made using ten grains for each case. All sound signals (synthesized and recorded) were prepared to be 4 s long, with a 10 ms long tuning in and out. To summarize, the ten pairs for the evaluation are listed in Table 1.

Micro-phone position	Driving condition				
	Idle 600 rpm	20 km/h 950 rpm	50 km/h 1400 rpm	60 km/h 1050 rpm	70 km/h 1250 rpm
F5	AB	AB	AB	AB	AB
V1	AB	AB	AB	AB	AB

Table 1: Set of pairs from synthesis (A) and recordings (B) using ten grains.

Evaluation of the granular approach

To evaluate how well the granular approach captured the characteristics of the truck, the syntheses (A) were compared with the recordings (B) in a two-part experiment. In the first part of the experiment (Part I) the participants performed an AX discrimination task (Same-Different). In a same-different task the participants are presented with AA, AB, BB and BA combinations of the set of pairs and the task to determine whether the presented sounds are the same or different. The hypothesis is that the synthesized sounds should not be possible to discriminate from the recordings, and thus the proportion of correct responses should be .5. In the second part (Part II) the participants rated each sound individually, on a set of five semantic differential scales. Part I, the discrimination task provides a stronger validation of the granular approach method, whereas Part II, the semantic differential, may provide additional information if the participants are able to differentiate between the synthesized and the recorded stimuli.

In the listening test, 15 participants participated, whereof 6 women and 9 men. The average reported age was 26.8 years old (standard deviation 4.5 years). All participants reported normal hearing. The participants were paid for their participation and gave their informed consent prior to their inclusion in the study.

The stimuli were presented as a monaural signal through dynamic headphones (Sennheiser HD650). All stimuli were normalized in loudness to avoid the same-different judgement to be coloured by differences in loudness between the different set of pairs.

In Part I, each set of pair was tested for the two microphone positions and the five driving conditions: idle (0 km/h), 20 km/h, 50 km/h, 60 km/h, and 70 km/h. Each pair was combined in the four different combinations AA, AB, BB and BA, in random order. For each participant the four combinations were tested twice.

In Part II, each stimuli was presented individually and the participants were asked to answer five 9-point semantic differential scales measuring the emotional responses (valence and activation) and how demanding, realistic, and annoying each sound was perceived. Each sound was here tested once.

The listening test took approximately 45 minutes to complete and each participant made the test individually in a sound-attenuated room.

Concerning the results of Part I, each response was coded to either a correct response (1) or an incorrect response (0) and analysed by a binomial test to determine if the participants could differentiate between the recordings and the synthesis at an above-chance level. The results showed that in nine of the ten sets of pairs the participants could not differentiate whether it was the same or different sounds that were played, at a 95 % confidence interval (see Figure 2). For the front microphone position at a speed of 20 km/h the participants could discriminate between the synthesized and the recordings at an above-chance performance level (significant at $p<.001$).

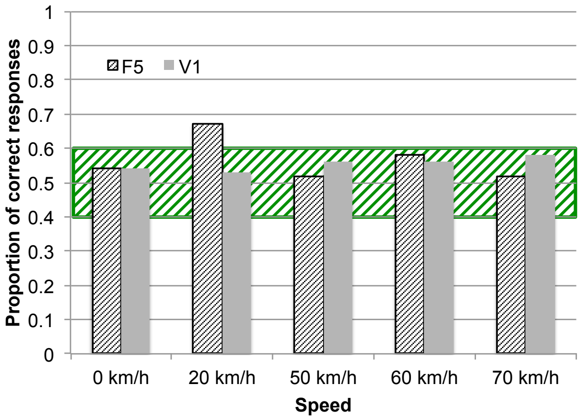

Figure 2: Proportion of correct responses in the same-different experiment, for the five driving speeds and the two microphone positions (denoted F5 and V1). The dashed range marks the 95 % confidence interval.

Overall, the incorrect responses were mainly due to that different stimuli (i.e. AB or BA) were perceived as being the same. The proportions of 'false alarms' (i.e. two stimuli of the same origin, AA or BB, was perceived as being different from each other) varied between 0 and 0.1 (median proportion 'false alarms'=.025).

The results of Part II could be used to determine the origin of the differences between the recordings and the synthesized signal for the deviating case, i.e. for the front microphone position at constant speed of 20 km/h. The semantic differential was analysed by a repeated measures analysis of variance and Fisher's LSD (Least Significant Difference) post-hoc test. This revealed that the recorded signal was perceived as more realistic than the synthesized one. There was however no significant difference in the emotional responses, level of annoyance, or how demanding the two sounds were perceived.

Conclusion

For steady-state propulsion sounds of a Volvo truck, an auralization model has been developed using a granular approach.

It has been shown by listening tests that the participants could not differentiate between synthesized and recorded sounds, for nine out of the ten studied sound pairs. For the deviating case, the recorded sound was perceived as more realistic than the synthesized sound. There was however no significant difference between the two sounds in the emotional responses, level of annoyance, or how demanding the two sounds were perceived. Therefore we find it a reasonable conclusion that the granular approach developed here gives a high enough quality for using as a basis for further auralization. Further work includes synthesis of pass-by as well as variable engine speed.

Acknowledgement

The work presented here is part of the Swedish national project *Quieter Transport for More Efficient Distribution* and is financed by the VINNOVA program FFI Transport Efficiency. Project partners are Volvo GTT, Volvo GTT/AT&R, Chalmers University of Technology, YKI and JABA Group.

References

[1] Kaczmarek, T. Road-vehicle simulation for psycho-acoustic studies. ICA 2007, MADRID, Spain, 2007.

[2] Maillard, J., Martin, J. A simulation and restitution technique for the perceptive evaluation of road traffic noise. Proc. Euronoise, Paris, 2008.

[3] Forssén, J., Kaczmarek, T. Alvarsson, J., Lundén, P., Nilsson, M.E. Auralization of traffic noise within the LISTEN project – Preliminary results for passenger car pass-by. Euronoise 2009, Edinburgh, UK EAA, 2009.

[4] Nilsson, M.E., Rådsten-Ekman, M., Alvarsson, J., Lundén, P., Forssén, J. Perceptual validation of auralized road traffic noise. Proc. of Inter-Noise 2011, Osaka, Japan, 2011.

[5] Maillard, J. and Jagla, J. Auralization of non-stationary traffic noise using sample based synthesis - Comparison with pass-by recordings. Proc. of Inter-Noise 2012, New York City, New York, 2012.

[6] Jagla, J., Maillard, J. & Martin, N. Sample-based engine noise synthesis using an enhanced pitch-synchronous overlap-and-add method. J. Acoust. Soc. Am.132 (5), 2012.