Perceptual validation of auralized heavy-duty vehicles

Penny Bergman, Astrid Pieringer, Jens Forssén, Patrik Höstmad
Division of Applied Acoustics, Chalmers University of Technology, Gothenburg, Sweden.

Summary
Auralization is a valuable tool when evaluating the effect of traffic noise on people. The present study focuses on the validation of auralization of heavy-duty vehicles with a diesel engine. To capture the characteristics of the diesel engine a granular approach has been used. The granular approach has proven to be successful in a previous validation test examining two microphone positions around a still-standing truck. In the present study a granular approach was used to achieve pass-by noise at an artificial listening position alongside a Volvo truck (experiment 1) and pass-by noise inside an apartment (experiment 2). The aim of experiment 1 was to determine the number of interpolated sets of grains needed, in order to create a perceptually valid auralized signal. The results were used in the auralization of pass-by noise in an apartment in experiment 2. 20 and 15 participants respectively rated original recordings and auralized signals on four different attributes: realism, annoyance, and emotional response measured by valence and arousal. The results of both experiments suggest that auralizations of heavy-duty vehicles are successful and usable. It further indicates that what distinguish the auralized signals from the original recordings is mostly the arousal responses.

PACS no. 43.50.Qp, 43.66.Lj

1. Introduction

Traffic noise is associated with several negative effects on health as well as well being (see e.g., [1], [6], [8], [10]). In a current project the focus has been to determine the possibility to use heavy-duty vehicles for night-time delivery in an urban area without reducing the well being of the habitants. Auralization is a valuable tool when evaluating how different parameters influence well being. The present study focuses on the possibility to use auralization in evaluations of heavy-duty vehicles. Auralization has been validated in several studies using road traffic, especially cars (see e.g., [5]) however there exists few auralization models for heavy-duty vehicles. This study is a continuation of a previous study where the possibility to use a granular approach was examined and proved successful for auralization of a single microphone recording [4]. In this study two experiments are conducted. In the first experiment the aim is to validate auralization models for an artificial position alongside a heavy-duty truck to determine the number of interpolated sets of grains needed. This is followed by a second experiment where the results of the first auralization are utilized in order to create a pass-by noise of a heavy-duty vehicle inside an apartment.

There are different methods to validate an auralization. The highest level of validation is an auralization that is indistinguishable from the real recording. This can be tested in a discrimination test (e.g., an AX discrimination task), where the capability to discriminate between recordings and auralization signals are evaluated. In several cases it is however not needed to be completely indistinguishable. Instead it is enough if the auralized signals render the same perceptual responses as real recordings. Dependent on the aim of the auralization the perceptual responses can be measured by several different attributes, e.g., the perceived realism of the sound, whether it renders the same responses to psychoacoustic parameters, emotional responses and attributes relating to a particular situation, e.g., the perceived speed of the vehicle.

In the first experiment of the present study both these two approaches were utilized. First the participants conducted a discrimination test, where pairwise presented sounds were evaluated. Further the sounds were evaluated on a set of attributes. The aim of the auralizations is to evaluate well being in habitants, hence the sounds were evaluated on how they were perceived in terms of the emotional responses. This was measured by the two dimensions of valence (pleasantness) and arousal (activation). Valence and arousal together provide a good understanding of the full emotional response to the sound, making it suitable to evaluate the well-being of the listeners (for evaluations of measures of emotions see e.g., [3], [9], [11], [12]). Further the participants were asked to rate the level of annoyance to the sound as well as perceived realism.

In the second experiment the participants evaluated a pass-by of a heavy-duty vehicle when being inside an apartment. These sounds were approximately 10 seconds long, and deemed to long to evaluate in a discrimination test. Thus, the second experiment evaluated whether the auralized signals rendered the same perceptual response using the same attributes as in the first experiment.

The experiments reported here were conducted in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki. The participants were paid for their participation and gave their informed consent to participate in the study prior to the experiment.

2. Experiment 1

2.1. Method

2.1.1. Recordings of truck engine sounds

The sound of a Volvo FM truck (13 litre diesel engine, 460 hp, emission class EEV) was recorded in a semi-anechoic laboratory. 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 20km/h ($\approx$ 950 rpm) and 50 km/h ($\approx$1400 rpm). The speed of the truck was manually controlled by a driver and might therefore vary with $\pm$3 km/h. Microphones (Brüel & Kjær Type 4189, 1/2-inch free-field) were distributed in front, behind and on the right-hand side of the truck. Six microphones are poistioned to simulate pass-by recordings, i.e. at 1.2 m height, 7.5 m from the center line of the truck starting at 3.9 m in front of the truck and backwards, separated 2 m apart. One set of 14 microphones are distruted at surface of a virtual box at a distance of 2 m from the truck according to ISO 3744 for sound power measurements. Eight microphones placed at 2 m distance from the vehicle on the floor in front (2), behind (1) and on the righthand side (5) of the vehicle. The sound was recorded using a sampling frequency of 44.1 kHz. Method 1–3 described below uses the pass-by microphones and method 4 uses all the microphones. The evaluated listening positions are at the pass-by microphones placed 1.9 m in front of and 0.1 m and 2.1 m behind the front of the truck.

2.1.2. Auralization technique

In an earlier study, a granular approach has been used to capture the characteristics of a diesel engine [4]. The granular approach uses short time pieces

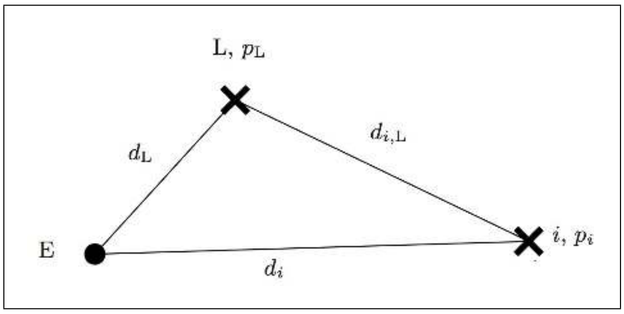

Figure 1. Schematic drawing: engine E, listening position L and microphone position i.

of recorded signals that are stored in a data base. At synthesis, the grains are picked at random from the data base and added synchronously using 124 samples overlap and Hann windowing. The approach was validated in a study examining two microphone positions around a still-standing truck [4].

In the present study, the same approach is used to achieve a pass-by noise at an artificial listening position alongside a Volvo FM truck. Grains are extracted from the recordings in all the microphone positions described in the previous subsection and stored in a data base. To generate the sound p_L in a listening position L the sound signals p_i from a set of N different microphone positions are weighted and superposed. In detail, the following steps are carried out:

1. The individual sound signals p_i, $i = 1, 2, \dots N$, are synthesized from the grain data base.
2. Assuming that the diesel engine is the only sound source and behaves as a point source, the source distance is accounted for according to

$$p_{i,\mathrm{L}} = p_i \frac{d_i}{d_\mathrm{L}} , \tag{1}$$

where d_i and d_L are the source distances of positions i and L, respectively, see Figure 1.

3. The weights w_i are calculated from Shepard's method, which provides an inverse distance weighting:

$$w_i = \frac{d_{i,\mathrm{L}}^{-1}}{\sum_{j=1}^{N} d_{j,\mathrm{L}}^{-1}} . \tag{2}$$

The distance between L and i is denoted $d_{i,\mathrm{L}}$.

4. The signals $p_{i,\mathrm{L}}$ are weighted and superposed to obtain the sound in the listening position:

$$p_\mathrm{L} = \sum_{i=1}^{N} w_i p_{i,\mathrm{L}} . \tag{3}$$

Four different sets of microphones have been used to construct the sound in the listening position L.

Method 1:
The set of microphones includes only the two microphones closest to L, which are at the same distance

from the truck and height above the ground as L, $N = 2$.

Method 2:
The set of microphones includes all microphones which are at the same distance from the truck and height above the ground as L, except the two closest ones, $N = 3$.

Method 3:
The set of microphones includes all microphones which are at the same distance from the truck and height above the ground as L, $N = 5$.

Method 4:
The set of microphones includes all microphones on the right side of the truck which are at varying distance from the truck and height above the ground, $N = 27$.

In the listening test, the synthesized sound p_{L} is compared to the original measured sound in position L. In order to avoid effects due to loudness, the sound pressure level of the synthesized sound is adjusted equal to the sound pressure level of the original sound. To achieve this, the weights w_i in Eq.(3) are multiplied with a factor f_i in a simple engineering approach:

$$f_i = 1 + (1 - w_i)\alpha\,. \quad (4)$$

The same constant α is used for all weights.

Applying the described procedure for the three listening positions, four methods and two truck speeds, 24 auralized signals are obtained with a stimuli length of 6 seconds.

2.1.3. The experiment

Twenty participants (M=27.3 years; SD=3.8 years; 9 females) took part. All were naïve as to the purpose of the experiment. All participants reported normal hearing.

The stimuli were presented as monaural signals through dynamic headphones (Sennheiser HD650). The sound level for the original sounds were set as to be as similar as possible (app. 65 dB(A)), to avoid a result dependent on similarity in loudness level.

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. Each auralized sound was tested against the original recording resulting in 96 pairs of sounds.

In the second part (Part II) the participants rated each sound individually (in total 30 sounds, 24 auralized signals and 6 original recordings), on a set of four attributes: perceived realism, perceived annoyance, and ratings of their emotional responses measured by valence and arousal. All the ratings were done on a scale between 1-9.

2.2. Results

2.2.1. Discrimination test

The responses in the discrimination test were coded to either a correct response (1) or an incorrect response (0) and analyzed by a binomial test to determine if the participants could differentiate between the recordings and the synthesis at an above-chance level. In all four different synthesis the participants could differentiate between the recordings and the synthesis methods well above-chance level, the correct responses varied between .86 and .99.

2.2.2. Attribute test

Each tested attribute was analyzed by a repeated measure analysis of variance (ANOVA) with speed (20 and 50 km/hour), microphone position (pos. 1, 2, and 3) and method (original, synth. method 1, method 2, method 3 and method 4) as within-participants factors. The interest in the present study is whether the different synthesis methods differs from the original recordings, hence only the results of the method factor are reported. The results showed that there was a main effect of method in the ratings of realism ($F(4, 76) = 3.37, p < .05, \eta_p^2 = .15$), subsequent post hoc comparisons using the Bonferroni correction did however not reveal any significant differences between the different auralization methods and the original recordings. There was also a main effect of method in the ratings of annoyance ($F(4, 76) = 7.88, p < .001, \eta_p^2 = .29$), subsequent post hoc comparisons using the Bonferroni correction suggest that the difference was driven by the different auralization methods. For the emotional responses there was a main effect of method in valence ($F(4, 76) = 6.76, p < .001, \eta_p^2 = .26$) as well as arousal ($F(4, 76) = 16.43, p < .001, \eta_p^2 = .46$). The post hoc comparisons using the Bonferroni correction of the valence ratings did however not reveal any significant differences between the different auralization methods and the original recordings. In the arousal ratings the post hoc comparison indicated that the original sound ($M = 5.02, SE = .27$) was significantly different from two of the synthesis method (synth. method 2: $M = 5.47, SE = .30$) and synth. method 4: $M = 4.3, SE = .33$)), the other two synthesis methods did not differ significantly from the original sound (synth method 1: $M = 4.92, SE = .27$); synth. method 3: $M = 4.76, SE = .33$)).

2.3. Discussion

The results of the discrimination test showed that the auralized and the original recordings were clearly distinguishable from each other. The results varied between .86 and .99 in correct responses. The second method using grains from all microphones at the same distance from the truck and height from the ground, except the two closest were the easiest to distinguish from the original recording for all three microphone positions. The results of the attribute test indicated that the main differences comparing the three auralization methods and the original recordings were mainly due to differences between the different auralization methods. The only attribute that was significantly different between the original recordings and synthesis methods were the ratings of arousal. The second method together with the fourth method were both significantly different from the recordings, the second method was more arousing whereas the fourth method was less arousing than the original sounds. The results are not univocal but for the further auralization of the pass-by the results suggest that using the microphones on the same height might provide a better result than using all microphones on one side. There is further an indication that the closest microphones are of importance in the auralizations.

3. Experiment 2

For the second experiment the results from Experiment 1 was utilized in order to create pass-by signals. The same type of recordings as in Experiment 1 was used in order to create the grains employed for the auralization. To do an initial validation of the pass-by auralizations the same attributes were evaluated in the second experiment: perceived realism, annoyance ratings and the emotional responses measured by valence and arousal. The recordings were conducted in an apartment at a central location in Gothenburg [7]. Two trucks were used and two speeds (20 km/h and 30km/h).

3.1. Method

3.1.1. Auralization technique

In order to estimate the transfer function from a monopole source within a street canyon, to a receiver on a building's façade, a ray-model was built up where the street canyon is modelled as two parallel walls to which absorption and diffusion properties could be ascribed. (The effect of a hard ground plane assumed to be included in the grains.) In the street canyon modelling, the height of the façade walls were given an infinite extent, after numeric tests for a finite height of 10 m showing no significant difference. The absorption is modelled as a reduction in amplitude for each order of reflection, as function of frequency. For the diffusion, an alternative approach was implemented with assigned diffusing elements randomly distributed along both sides of the street, on the façades. Here, in this two-dimensional diffusion model, an average of one diffusive element per ten metres along a façade was used. The angular variation in diffusion was according to Lambert's cosine law, i.e. where the outgoing intensity varies as cosine of the angle and is independent on the incoming angle, also as function of frequency.

The canyon is modelled as 200 m long in total, with specular reflections up to order 20, and with first order diffusion for each reflection order and for each diffusion element. For the frequency dependence of both absorption and diffusion, third-octave band value inputs were used. The implementation was made using a mixed time and frequency-domain approach, where the direct sound pulse and the resulting pulses are collected concerning amplitude and timing, and the frequency domain input data for absorption and diffusion is multiplied on the Fourier transform of the time-domain signal (in narrow band, after frequency interpolation). The final time-domain impulse-response is then obtained via an inverse Fourier transform (using the FFT algorithm). Input values of absorption and diffusion were chosen as reasonable values based on the resulting gain in relation to that of a measured impulse-response in a real street canyon. It should be noted that, as a result of this approach, the modelled gain is more even as a function of frequency compared with the measured gain, which showed higher peaks as function of frequency.

The impulse response, and from that the gain, is then calculated for a fixed receiver position and for a set of source positions, which are located on a straight line parallel to the façades and to the ground surface, simulating the line of travel of the acoustic source. Along the line, a discretisation of 0.25 m were used for the source positions, along the whole length of the canyon. For a receiver placed midway, symmetry of the two halves is used. The modelled width of the canyon is 11 m, the receiver is at 10 m height and the source line is at 1 m height, located 5 m from the receiver façade.

In order to model the variation in character as the direction from the truck to the listener varies, a weighting between the grains from the different microphone positions was used. It should be noted that such an approach gives a richer directivity information than the usually applied gain as function of angle (in e.g. third-octave bands). In order to calculate the weights for any chosen angle, the set of microphone positions is first described on a convex hull based on Delaunay triangulation (using the delaunay function in Matlab) assuming the engine as the acoustic centre. Then, for a given angle, the point of intersection between the source-to-receiver line and the convex hull is identified in terms of local coordinates on the intersected triangular surface element. Thereby the distances from the point of intersection to the three corner nodes (i.e. three microphone positions) of the intersected triangle can be calculated, as well as the distances to all other microphone positions. (To calculate metric distances to all other microphones, the nodes of the convex hull were projected on a sphere with volume equated to that of the truck, and where the distances are calculated as the shortest distance along the surface of that sphere.) As the source passes by a given receiver position, the source-to-receiver line, updated at discrete time steps, gives rise to an updated set of microphone position weights, corresponding to the weights of the database grains. For the auralizations, two different approaches to calculate the weights were used: (i) Shepard's method, as presented above, but now involving microphone positions surrounding the truck, auralization model 1 and (ii) a weighting including only the three corner node microphone positions corresponding of the currently intersected triangular surface element, using so-called barycentric weights, auralization model 2.

For auralization of a vehicle pass-by, the weights are pre-calculated as function position with a prescribed resolution that can be much lower than the corresponding sampling of the sound; here a resolution corresponding to 1° is used. The pre-calculated weights are then interpolated to fit the pass-by time behaviour of the wanted auralization case, and the corresponding grains are weighted and mixed to a mono signal. This source signal is then subject to the transfer modelling of the path between source and receiver. This includes, in addition to the gain of the street canyon, as described above, the distance decay, air attenuation and Doppler effect, which are inferred using a previously developed methodology [5]. For the indoor sounds, the façade reduction indices in third-octaves were used as an equalizing filter, based on measurements within the project for a façade and room at 3rd floor [7]. The final indoor sounds were then given by convolving with synthetic room impulse responses corresponding to measured third-octave band reverberation times.

3.1.2. The experiment

15 participants (M=26.9years; SD=3.6 years; 6 females) took part. All were naïve as to the purpose of the experiment. All participants reported normal hearing. 1 participant reported tinnitus at higher frequencies (>15kHz), this was deemed to not affect the results and the participant was included.

The sounds tested were from two trucks, two speeds (20km/h and 30km/h) and compared using the two auralization models and the recording on site, in total 12 different stimuli. The stimuli were presented as monaural signals through dynamic headphones (Sennheiser HD650). Due to a problem with the sound levels for the auralization the sounds were loudness equalized (Zwicker ISO, 532B) and presented at 45 dB(A).

The participant rated each sound individually (in total 12 sounds repeated twice) on a set of four attributes: perceived realism, perceived annoyance, and ratings of their emotional responses measured by valence and arousal. All the ratings were done on a scale between 1-9.

3.2. Results

Each tested attribute was analyzed by a repeated measure analysis of variance (ANOVA) with speed (20 and 30 km/hour), truck type (truck 1 and truck 2) and origin (original, auralization method 1, auralization method 2) as within-participants factors. As the interest in the present study is the different synthesis methods only the results of origin are reported. The average responses together with the standard deviation is reported in figure 2. The results showed that there was no main effect of origin in the ratings of realism ($F(2, 28) = 0.61$, $p = n.s.$, $\eta_p^2 = .04$). There was neither an effect of origin in the ratings of annoyance ($F(2, 28) = .19$, $p = n.s.$,$\eta_p^2 = .01$). For the emotional responses there was a main effect of origin in valence ($F(2, 26) = 3.36, p < .05, \eta_p^2 = .21$). The post hoc comparisons using the Bonferroni correction did not reveal any significant differences, the first auralization method rendered lowest valence ratings in comparison to the original recordings ($M_{orig.} = 5.1$, $SE_{orig.} = .34$; $M_{aural.1} = 4.3$, $SE_{aural.1} = .22$; $M_{aural.2} = 4.9$, $SE_{aural.2} = .28$). There was also a main effect of origin in arousal ($F(2, 28) = 7.98$, $p < .01$, $\eta_p^2 = .36$). The post hoc comparisons using the Bonferroni correction indicated that the original sound ($M = 5.68$, $SE = .25$) was significantly different from the first auralization method ($M = 4.75$, $SE = .24$), there was no significant difference of the second auralization method ($M = 5.27, SE = .26$). There was no interaction effects between origin and truck type or speed in any of the attributes.

3.3. Discussion

The results of the second experiment showed that the participants could not differentiate between the original sound and the two auralized sounds in terms of realism and level of annoyance. There was however a significant difference in the emotional ratings, where the first auralization method rendered significantly lower level of arousal than the original recordings. This indicate that a barycentric weighting including only the three corner node microphone positions might be a more preferable method.

4. Conclusion

The results of both experiments suggest that the possibility for auralizations of heavy-duty vehicles are good and gives high enough quality. Both experiments indicate that it is slightly better to use microphones closely gathered, synthesis 1 in experiment 1 with the two closest microphones at the same height and distance and the barycentric weighting (auralization

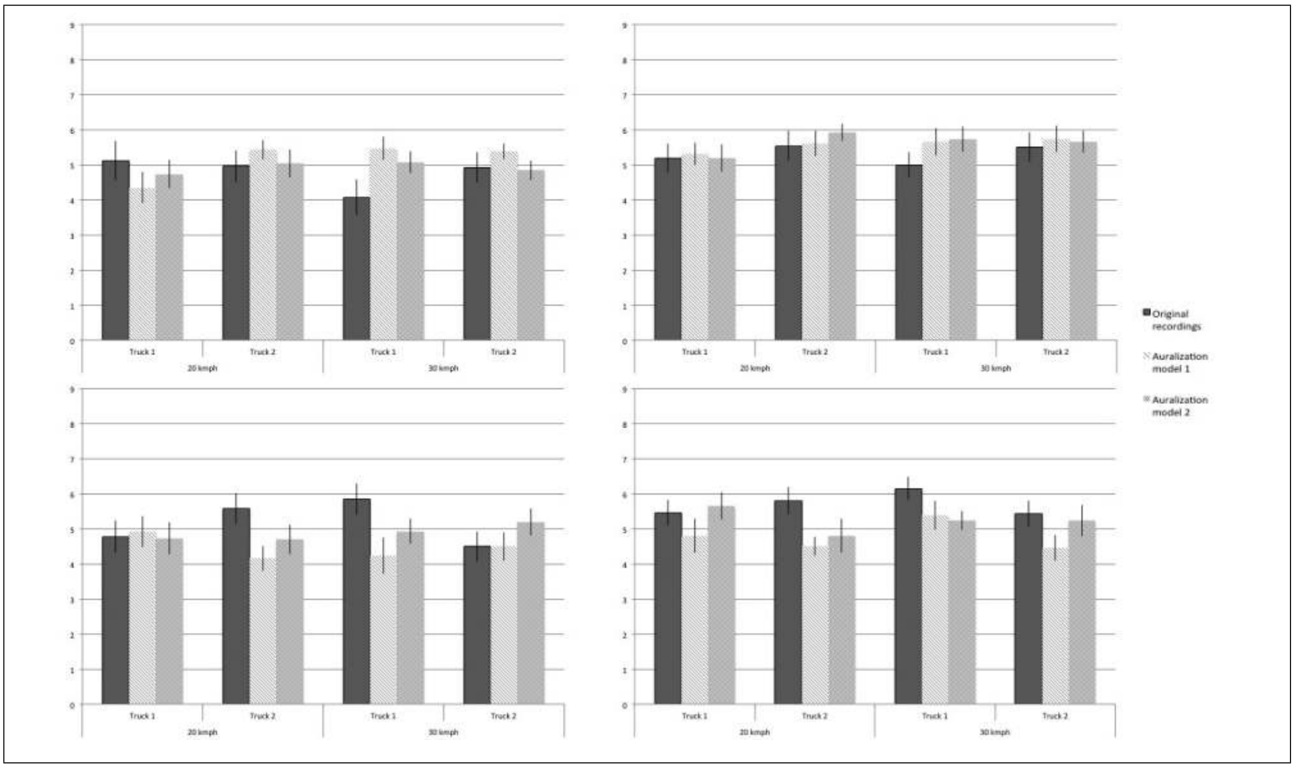

Figure 2. The average responses to the four different attributes in experiment 2, error bars indicate the standard error of the means, upper left: Perceived realism; upper right: Annoyance levels; lower left: Valence level; lower right: Arousal levels.

method 2) in experiment 2. The validation experiments included two different speeds and in the second experiment two different heavy-duty vehicles, however there was no interaction effects with method or truck type or speed, indicating that the auralization models also work for varying speed and different heavy-duty vehicles.

In both experiments the listeners appear to be more sensitive to changes in arousal levels. This is in accordance with several studies on emotional reactions to sound (see [2] for a discussion on the topic). This suggest that the arousal level is of importance in studies of heavy-duty vehicles and of higher importance than valence levels for measures of well-being to this type of sounds.

References

[1] W. Babisch: Road traffic noise and cardiovascular risk. Noise Health 10 (2008) 27-33.

[2] P. Bergman: The Role of Embodied Emotions in Perceptual Decisions and Categorization of Sounds. PhD-thesis, Applied Acoustics, Dep. of Civil and Env. Eng., Chalmers University of Technology, Göteborg, Sweden, 2014.

[3] M. Bradley, P. Lang: Affective reactions to acoustic stimuli. Psychophysiology 37 (2000) 204-215.

[4] J. Forssén, P. Andersson, P. Bergman, K. Fredriksson, P. Zimmerman: Auralization of truck engine sound: A preliminary results using a granular approach. Proc. AIA-DAGA 2013, Merano, Italy.

[5] J. Forssén, T Kaczmarek, J. Alvarsson, P. LundŐn, M.E. Nilsson: Auralization of traffic noise within the LISTEN project Ð Preliminary results for passenger car pass- by. Proc. Euronoise 2009, Edinburgh, UK.

[6] A. Gidlöf-Gunnarsson, E. Öhrström: Noise and well-being in urban residential environments: The potential role of perceived availability to nearby green areas. Landscape and Urban Planning 83 (2007) 115-126.

[7] P. Höstmad: Volvo FM-499 and FM-490 Passby Measurements at Landsvägsgatan, Göteborg, June 11, 2014, Report S15-02, 2015, Applied Acoustics, Dep. of Civil and Env. Eng., Chalmers University of Technology, Göteborg, Sweden, 2015.

[8] H. Ising, B. Kruppa: Health effects caused by noise: evidence in the literature from the past 25 years. Noise Health 6 (2004) 5-13.

[9] R. Larsen, B. Fredrickson: Measurement issues in emotion research. In D. Kahneman, E. Diener, N. Schwarz (Eds.), Well-being: Foundations of hedonic psychology, New York, USA, 1999.

[10] HM. Miedema, H. Vos: Associations between self-reported sleep disturbance and environmental noise based on reanalyses of pooled data from 24 studies. Behav Sleep Med 5 (2007) 1-20.

[11] J.A. Russell: A circumplex model of affect. Journal of Personality and Social Psychology 39 (1980) 1161-1178.

[12] M. Yik, J.A. Russell, L. Barrett: Structure of self-reported current affect: Integration and beyond. Journal of Personality and Social Psychology 77 (1999) 600-619.