



**Acoustics'08
Paris**
June 29-July 4, 2008

www.acoustics08-paris.org

euronoise

Modeling auditory attention focusing in multisource environments

Bert De Coensel^a and Dick Botteldooren^b

^aGhent University - Department of Information Technology, Sint-Pietersnieuwstraat 41, 9000 Ghent, Belgium

^bUniversity Ghent - Department Information Technology, Sint-Pietersnieuwstraat 41, 9000 Gent, Belgium
bert.decoensel@intec.ugent.be

In this paper, a mathematical submodel for auditory attention focusing is discussed in the framework of our ongoing research towards a unified model for soundscape perception. The submodel implements a balance between top-down focusing, in which higher level cognition guides attention towards expected sources, and bottom-up focusing, in which attention is triggered by the noticing of sound events. Attention elasticity – the ability to switch attention between different environmental sounds – depends on the current context (natural, urban...) as well as on the current activity of the modeled individual. The soundscape perception model is applied to case studies of simulated sound exposure in open area and in environments with a mixture of natural and man-made sounds. A comparison of the simulation results with survey results shows that introducing a submodel for auditory attention focusing into our model for soundscape perception enhances its ability to predict the emergence of annoyance.

1 Introduction

During the past few decades, a lot of research effort has been devoted to the development of methods for the characterization of outdoor sonic environments. Whereas the traditional aim of environmental noise research has been to specify the relationship between physical properties of environmental sound and its possible negative effects on the population as a whole (e.g. community noise annoyance), *soundscape research* takes a more positive and human-centered approach, thereby focusing on individual perception and expectation [1]. The increasing availability of computing power opens up the opportunity to use computational models for the perception of environmental sound by the individual listener in this research.

During the last decade, knowledge on (the neurobiological basis of) auditory perception has expanded enormously, for a large part thanks to the advent of several new brain imaging techniques (see e.g. [2] for a review). This evolution stimulated the development of very detailed computational models for human auditory processing, such as [3–5]. Most of these models, however, focus on speech processing; auditory processing of environmental sound has rarely been assessed with the same amount of detail [6].

Two important obstacles prevent that models such as those cited above are readily applied in soundscape research. Firstly, whereas processing of speech is usually a matter of seconds or minutes, environmental soundscape perception emerges over much longer stretches of time. For example, evaluating the sonic quality of the living environment is a process that could take months. Due to this huge difference in time scale, computational time and memory constraints make the use of detailed neurobiological models infeasible. Secondly, personal factors have a very strong effect on environmental sound perception (see e.g. [7–9]). For these reasons, studies on the effect of environmental sound are always based on averages over vast numbers of people (typically hundreds to thousands).

In earlier work [10–13], the authors have already presented an approach which tries to tackle the above described problems. The proposed methodology consists of simulating a large *synthetic population* of *modeled individuals*, each with its own personal characteristics and within its own context. The model for a single individual tries to achieve a balance between computational efficiency and psychoacoustical and psychological plausibility. Results are then analyzed statistically, exactly as one would analyze results of field studies with populations of biological agents.

This early work mainly focused on determining salient features of the soundscape (so called *notice-events*), but did not fully account for the influence of attention focusing. More in particular, noticing a single type of environmental sound was modeled to be independent of the presence of other types of sounds (other than masking effects). In this paper, the model for a single individual is extended with a simple but fast submodel for *attention focusing*. In Section 2, the layout of the modeling approach is briefly described, and the submodel for attention focusing is explained in detail. In Section 3, the model is applied to some test cases.

2 A model for environmental sound perception

2.1 Saliency detection

The input of the proposed model for a single individual is formed by a set of sound level time series, which simulate the various environmental sounds to which the individual is exposed. When a listener is exposed to an environment with multiple sound sources, the acoustic pattern at the ear will consist of the sum of all concurrent sounds. Nevertheless, the human auditory system is able to separate this mixture of sounds, and to form separate descriptions of each sound source. This mechanism is commonly referred to as *auditory scene analysis* (ASA) [14]. By simulating auditory streams for each separate environmental sound source, the non-trivial problems of modeling ASA and sound source recognition are effectively by-passed, making the model flexible and the numerical effort feasible¹.

In a first step (bottom half of Fig. 1), the model analyzes which environmental sounds are standing out of the ambient sound, and thus may be noticed during everyday activities and receive attention. It is believed that a large part of this process is pre-attentive [2]. *Saliency detection* is implemented as a peak detection algorithm, for each individual sound i in the multi-source environment, by comparing its signal-to-noise ratio with a time-varying noticing threshold $T_i(t)$:

$$L_{\text{sound},i}(t) - L_{\text{ambient},i}(t) > T_i(t) \quad (1)$$

The ambient level is hereby formed by all sounds other than sound i , present in the soundscape, and thus forms a physiological masking background. The physiological limitations of the ear are accounted for by A-weighting. Because the peak detection is implemented for each individual type of sound in the complex acoustic

¹ The problems of ASA and sound source recognition will have to be tackled, if the model is to be applied to soundscape recordings.

environment, deviant behavior on the frequency axis, which is normally included in the calculation of a saliency map [15], can be neglected in a first approximation. The process described in Eq. (1) marks *notice-events*: segments in time, during which the sound under study is noticed.

To include habituation in the saliency detection mechanism, the signal-to-noise ratio for a particular source – the left part of Eq. (1) – is reduced by an amount that depends linearly on the exponential average of past excitations:

$$C_h \int_{-\infty}^{t_0} [L_{sound,i}(t) - L_{ambient,i}(t)] \exp\left(\frac{t-t_0}{\tau_h}\right) dt \quad (2)$$

where τ_h is the time constant for habituation.

Very little is known on the dependence of the threshold for noticing a sound on attention. It is however safe to assume that this threshold monotonously decreases with increasing attention $A_{es,i}$ to the source considered; hence for simplicity, this relationship is approximated by a linear function,

$$T_i(t) = T_{base} - f \cdot A_{es,i}(t) \quad (3)$$

with f a general constant.

2.2 Attention focusing

It is widely acknowledged that attention focusing is directed using both bottom-up, sound based cues and top-down, activity dependent cues [2, 16]; similar mechanisms have been identified in visual attention focusing [17–18]. Top-down attention focusing is guided by higher level cognitive processing, in which meaning is attached to the sounds, within an *experienced* (including other senses than hearing) and *expected* context (based on prior experience). To date, no computer models are available that can handle this part of the complex problem. To provide the essential higher-level cognitive information on top-down attention $A_{topdown}$ for environmental sound, therefore, a very simplified model is introduced:

$$A_{topdown} = A_{es,0} + \alpha \int_{-\infty}^{t_0} \sum_i A_{es,i}(t) \exp\left(\frac{t-t_0}{\tau_{es}}\right) dt \quad (4)$$

where $A_{es,0}$ is the background attention for environmental sound, which depends among others on the current context, activity and intentions of the modeled individual. For simplicity, this background attention is kept constant during simulation, implying that the modeled individual only takes part in a single activity. The second term in Eq. (4) expresses the fact that bottom-up attention $A_{es,i}$ triggered by particular environmental sounds may eventually increase top-down attention as well. Focusing is thus based on earlier noticed sounds, and fades away with a time constant τ_{es} . The factor α weighs the importance of this top-down effect.

When a peak detect occurs for a particular sound, it tries to attract attention. This bottom-up attention process is introduced in the model in a simple and fast way by an attention switch process that distributes available attention over sounds reaching the detection threshold. Rather than strictly limiting total attention, an amount of attention is focused on the newly detected sound, decreasing as attention is already focused on other environmental sound,

$$\Delta A_{bottomup} = \beta \cdot \left(A_{es,d} - \sum_i A_{es,i}(t) \right) \quad (5)$$

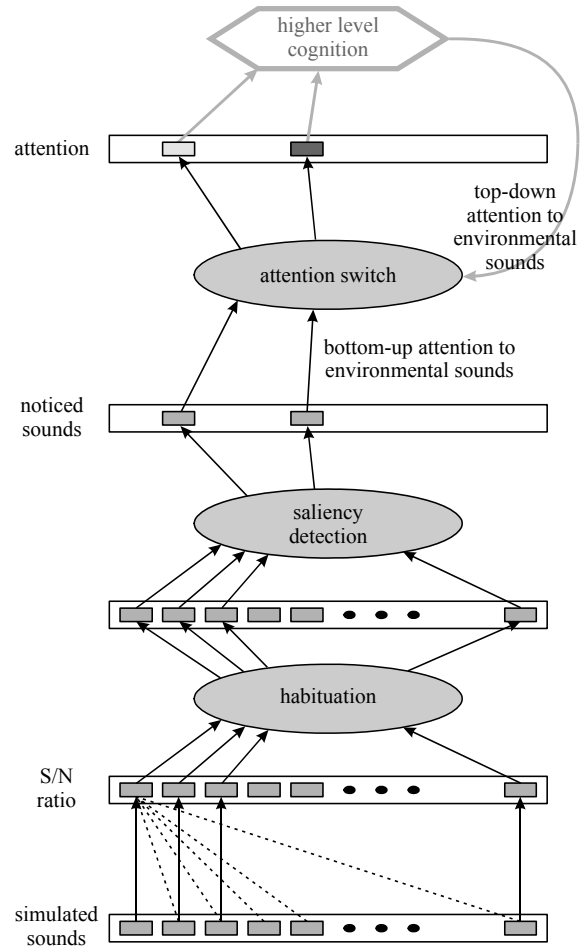


Fig. 1. General layout of the soundscape perception model.

where $A_{es,d}$ is the dynamic part of attention and β is a factor indicating which part of available attention is attributed to the newly detected sound. Together, $A_{es,d}$ and β determine the *attention elasticity*: the ability to switch attention between different environmental sounds. If the person-activity combination does not allow switching attention to environmental sounds easily, $A_{es,d}$ should be kept small. If attention can not easily be switched between environmental sounds, β should be kept small. The total level of attention attributed to the i th environmental sound finally becomes:

$$A_{es,i}(t) = A_{topdown} + \Delta A_{bottomup} \cdot \exp\left(\frac{t_i - t}{\tau_i}\right) \quad (6)$$

where t_i is the instance of the last increase of attention and τ_i is a suitable time constant. This model implicitly includes attentional gating. The top half of Fig. 1 shows the layout of the attention focusing module.

Attention and excitation level (S/N ratio) determine which environmental sounds are noticed. Note that the proposed model allows several environmental sounds to be noticed at the same time, and thus to share some of the listener's attention. This may seem surprising at first glimpse, but is solely determined by the smallest time steps of the order of 1s that are typically chosen for this type of simulation. During this time interval, attention may actually switch up and down between both sources at hand, leading to a non-zero average level of attention for each.

3 Simulation results

3.1 Open area soundscapes

In a first case study, we look at the influence of model parameters, linked to auditory perception, on the trends observed in simulations. Let us consider the perception of open area soundscapes, in which one is actively looking for recreation and psychological restoration. It can safely be assumed that the listener is attentive for the natural environment, including its soundscape. Hence attention for environmental sounds is relatively high.

In a typical but quite general situation for the open area soundscape, the listener is surrounded by natural sound and confronted with sounds from a highway, a road and a railway, located at respective distances d_{hw} , d_{rd} , and d_{rw} , and carrying traffic intensities N_{hw} , N_{rd} and N_{rw} (all sampled from distributions typical for the open area in northwest Europe). A typical noise emission for the European fleet is used [19]. Natural ambient sound levels also differ between situations, and are assumed to fluctuate following a $1/f$ characteristic – we refer to earlier work [10–13] for more details on this.

A synthetic population of 10000 individuals was simulated during 1 hour of exposure. Fig. 2 shows the average time that sounds are noticed as a function of ambient natural sound level, for different average levels of road traffic sound, for high and low top-down attention $A_{es,0}$ for environmental noise, respectively. The overall trend of these results corresponds to what could be expected. As

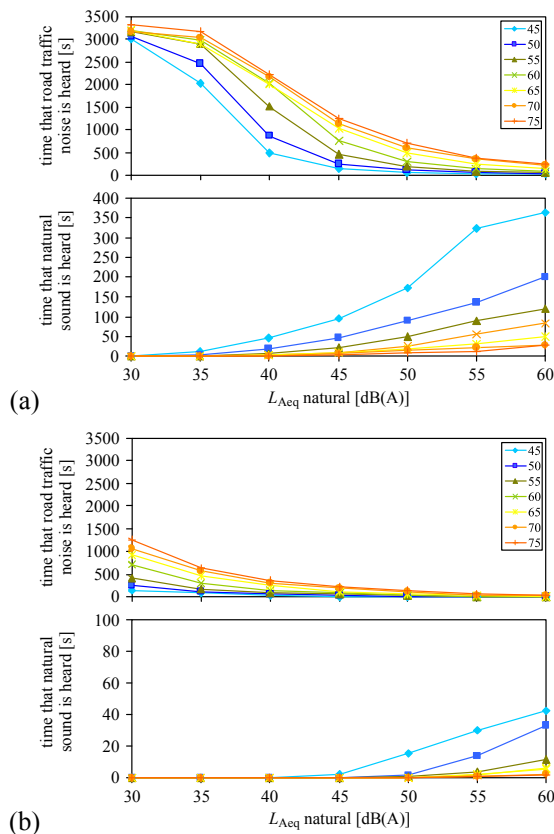


Fig. 2. Time that road traffic noise and natural sound is heard within one hour, as a function of average natural ambient sound level, for different average traffic noise levels. (a) *high* and (b) *low* top-down attention.

ambient noise level grows, traffic noise is less noticeable. The effect is less for higher traffic noise levels. Similarly, natural sound is more often noticed as natural ambient sound levels grow and as traffic noise levels are less. Since modeled attention can not be judged quantitatively, its value can only be deduced from observing phenomenological effects. From the results in Fig. 2(a) it could be assumed that this level of attention corresponds to a trained environmental scientist observing the rural soundscape; Fig. 2(b) would correspond to that of the recreating visitor. With decreasing attention for environmental sound, the curves seem to shift over the x -axes. As bottom-up attention is ignored, the main reason for the observed trends is physiological masking or at least reduction of signal-to-noise ratio.

Let us now increase the level of bottom-up attention to a level comparable to the difference between the two levels of top-down attention used above: $A_{es,d} \approx A_{topdown}^{(high)} - A_{topdown}^{(low)}$. In Fig. 3, results are shown for less and more volatile attention respectively. On average, the time that environmental sounds are noticed increases with added bottom-up attention, but the effect is different from the effect of increased top-down attention: the duration of noticing lower level traffic sound does not increase as much as the duration of hearing loud traffic sound. This can clearly be related to attracting more attention. Differences between Figs. 3(a) and (b) are most pronounced in the region where neither the traffic sound nor the natural sound dominates. Indeed, a different type of environmental sound needs to be present for switching of attention between sounds to occur, and thus for volatile attention to show its effect.

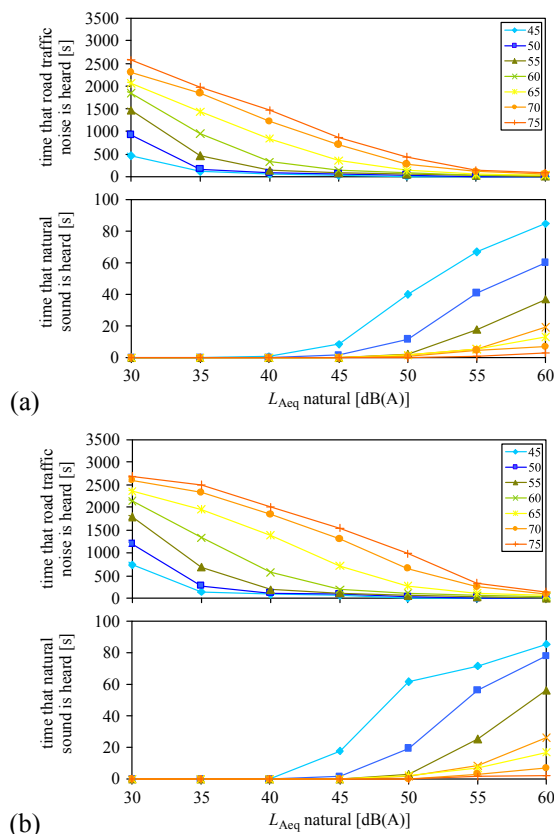


Fig. 3. Time that road traffic noise and natural sound is heard within one hour, as a function of average natural ambient sound level, for different average traffic noise levels, with low top-down attention and medium bottom-up attention. (a) $\beta=0.5$, (b) $\beta=1$.

3.2 Soundscapes with combined exposure

In a second case study, we look at the perception of the soundscape in an at-home context, in which individuals are exposed to a combination of road and railway traffic. A synthetic population was created, resembling the population of participants of a recent Belgian nationwide retrospective questionnaire survey on annoyance [20] as close as possible. Instead of sampling distances to roads and traffic intensities from random distributions, the addresses of the dwellings of the 7500 participants were used to calculate the distance of each dwelling to the nearest main road and railway track, and to determine the average hourly intensity of vehicles/trains during daytime on the road/track.

While at home, the modeled individuals may be engaged in various activities, and consequently may have a varying background attention to noise. Therefore, background attention was randomized between individuals. Dwelling acoustic insulation also differs between situations, but because of a lack of data on insulation of individual dwellings, this parameter was randomized between 0 (open window) and 30 dB(A). The sound produced by the home activity (radio, television, cooking, playing children...) can not be modelled accurately, so a similar approach was used as the approach used to create the natural ambient sound in outdoor situation.

In earlier work [10–13], we already showed that our model is able to give a qualitative explanation for the *railway bonus* – the difference in annoyance between road and railway traffic noise at the same average sound level – a phenomenon that has been observed in several field studies in the past [21]. Here, we will focus on railway noise annoyance more in particular. Fig. 4 shows the survey exposure-effect relationship for railway noise annoyance, for exposure to railway noise higher, lower and approximately equal (± 5 dB(A)) to road traffic noise exposure (L_{den} was calculated using standard noise mapping software). It can be seen that, for equal average railway noise level, railway noise annoyance is lower when road traffic noise exposure exceeds railway traffic noise exposure, at least for railway noise levels below 65 dB(A). Fig. 5 shows the average above-threshold exposure level² SEL_{thr} of the sound noticed by the simulated individuals, caused by railway traffic noise exposure, similarly split up for different relative levels of road traffic noise exposure. An even more pronounced trend can be spotted.

How can these observations be explained in light of the discussed model? Consider the sound event caused by a train passage. With increasing level of the event, the probability that it is noticed and draws attention increases gradually. When train noise is considered on its own, habituation gradually decreases the probability of noticing, while focusing attention raises this probability. When road traffic noise is added to this picture, it will act as an additional background noise for noticing train passages. The reduced signal-to-noise ratio will decrease the overall probability of noticing. More detailed analyses show that the crossing of curves at high exposure levels is accompanied by increased bottom-up attention for the (loud) train sounds.

² SEL_{thr} is defined as the integrated strength of the sounds noticed, where strength is defined as the S/N ratio suitably modified to account for habituation and attention.

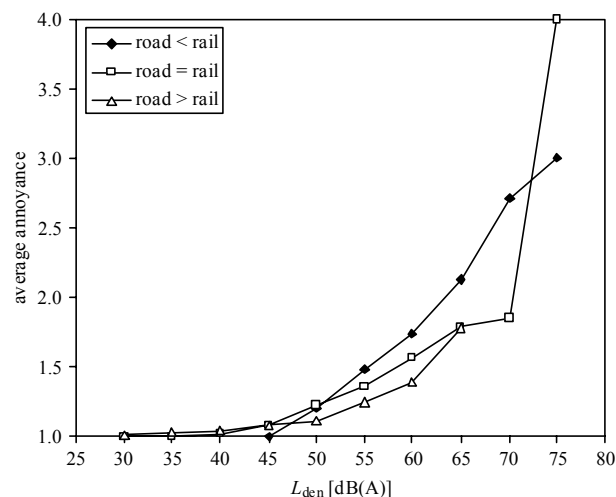


Fig. 4. Average annoyance of the survey population (5-point scale), caused by railway traffic noise exposure, split up for relative levels of road traffic noise exposure.

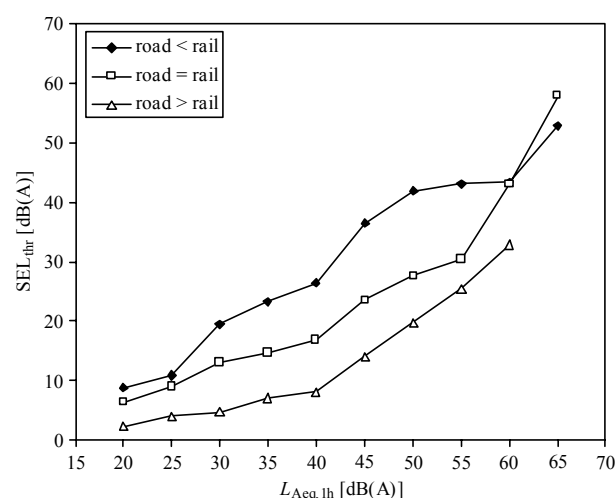


Fig. 5. Above-threshold exposure level of the sound noticed by the simulated individuals of the synthetic population, caused by railway traffic noise exposure, split up for relative levels of road traffic noise exposure.

4 Conclusion

In this paper, a submodel for auditory attention focusing was discussed in the framework of a computational model for human auditory perception of environmental sound. The influence of including bottom-up attention leads to less abrupt transitions in combined exposure situations than a model solely based on perceptual masking would. The application to perception of open area soundscapes shows expected results for the influence of natural ambient sound levels. The application in an at-home environment gives insight into the combined effect of road and rail noise on people. For the latter, qualitative agreement with the results of a large survey was found.

Acknowledgments

The authors would like to thank Mats E. Nilsson and Birgitta Berglund from Stockholm University, and Peter Lercher from the Medical University of Innsbruck, for the many fruitful discussions on the psychoacoustic and psychological mechanisms behind soundscape perception.

References

- [1] See e.g. the special issue on soundscapes, *Acta Acustica united with Acustica* 92(6) (2006).
- [2] J. B. Fritz, M. Elhilali, S. V. David, S. A. Shamma, "Auditory attention – focusing the searchlight on sound", *Current Opinion in Neurobiology* 17(4), 437-455 (2007).
- [3] D. Wang, G. J. Brown (eds.), *Computational auditory scene analysis: principles, algorithms, and applications*, John Wiley & Sons Inc., Hoboken, New Jersey, USA (2006).
- [4] B. Sagi, S. C. Nemat-Nasser, R. Kerr, R. Hayek, C. Downing, R. Hecht-Nielsen, "A biologically motivated solution to the cocktail party problem", *Neural Computation* 13(7), 1575-1602 (2001).
- [5] S. N. Wrigley, G. J. Brown, "A computational model of auditory selective attention", *IEEE Transactions on Neural Networks* 15(5), 1151-1163 (2004).
- [6] B. Gygi, V. Shafiro, "General functions and specific applications of environmental sound research", *Frontiers in Bioscience* 12, 3152-3166 (2007).
- [7] H. M. E. Miedema, H. Vos, "Noise sensitivity and reactions to noise and other environmental conditions", *Journal of the Acoustical Society of America* 113(3), 1492-1504 (2003).
- [8] R. F. S. Job, J. Hatfield, A. J. Hede, P. Peploe, N. L. Carter, R. Taylor, S. Morrell, "The role of attitudes and sensitivity in reactions to changing noise", *Proceedings of the 19th International Congress on Acoustics (ICA)*, Madrid, Spain (2007).
- [9] I. H. Flindell, P. J. M. Stillen, "Non-acoustical factors in environmental noise", *Noise and Health* 1(3), 11-16 (1999).
- [10] T. De Muer, D. Botteldooren, B. De Coensel, B. Berglund, M. E. Nilsson, P. Lercher, "A model for noise annoyance based on notice-events", *Proceedings of the 34th International Congress on Noise Control Engineering (Internoise)*, Rio de Janeiro, Brazil (2005).
- [11] B. De Coensel, T. De Muer, D. Botteldooren, "An agent based modeling approach to explain the perception of environmental stressors", *Proceedings of the 2007 International Conference on Artificial Intelligence (ICAI)*, Las Vegas, Nevada, USA (2007).
- [12] B. De Coensel, D. Botteldooren, "Models for soundscape perception and their use in planning", *Proceedings of the 2007 Congress and Exposition on Noise Control Engineering (Internoise)*, Istanbul, Turkey (2007).
- [13] D. Botteldooren, B. De Coensel, "A model for long-term environmental sound detection", *Proceedings of the 5th IEEE World Congress on Computational Intelligence (WCCI)*, Hong Kong (2008).
- [14] A. S. Bregman, *Auditory scene analysis: The perceptual organization of sound*, The MIT Press, Cambridge, MA, USA (1994).
- [15] C. Kayser, C. Petkov, M. Lippert, N. K. Logothetis, "Mechanisms for allocating auditory attention: an auditory saliency map", *Current Biology* 15(21), 1943-1947 (2005).
- [16] E. I. Knudsen, "Fundamental components of attention", *Annual Review of Neuroscience* 30, 57-78 (2007).
- [17] L. Itti, C. Koch, "Computational modelling of visual attention", *Nature Reviews Neuroscience* 2(3), 194-203 (2001).
- [18] H. Shi, Y. Yang, "A computational model of visual attention based on saliency maps", *Applied Mathematics and Computation* 188(2), 1671-1677 (2007).
- [19] H. Jonasson, U. Sandberg, G. van Blokland, J. Ejsmont, G. Watts, M. Luminari, *Source modelling of road vehicles*, Technical Report, Deliverable 9 of the Harmonoise project (2004).
- [20] *Schriftelijk leefomgevingsonderzoek SLO₀ & SLO₁* (Dutch), Environment, Nature and Energy Department (LNE), Flemish Government, Brussels, Belgium (2004).
- [21] H. M. E. Miedema, H. Vos, "Exposure-effect relationships for transportation noise", *Journal of the Acoustical Society of America* 104(6), 3432-3445 (1998).