Modeling the role of attention in the assessment of environmental noise annoyance

Dick Botteldooren[1*], Bert De Coensel[1], Birgitta Berglund[2], Mats E. Nilsson[2], Peter Lercher[3]

[1] Acoustics Group, Department of Information Technology, Ghent University
St.-Pietersnieuwstraat 41, B-9000 Ghent, Belgium

[2] Gösta Ekman Laboratory, Karolinska Institutet and Stockholm University, Sweden

[3] Division of Social Medicine, Medical University Innsbruck, Austria

* corresponding author (e-mail: dick.botteldooren@intec.ugent.be)

ABSTRACT

Community noise effects in general and noise annoyance in particular are mostly studied by relating them to exposure through blind statistical analyses of large datasets. This paper reports on a specific part of a quite different approach. Using mathematical simulation of basic perception and psychophysical mechanisms for a large synthetic population, insight is sought into the mechanisms underlying the emergence of noise effects. This is achieved by comparing – in a phenomenological way – the statistics of the data gathered from the simulated synthetic population to that of the real population. This paper focuses on modeling the role of attention. Attention could play a role in two distinct aspects of the process: firstly, attention can be drawn away from other tasks by the environmental sound or tasks requiring sustained attention can suppress the noticing of the environmental sound; secondly, attention can jump between sounds in multisource sonic environments. In modeling this dual role of attention, care must be taken to simplify existing knowledge on these aspects of perception in such a way that the model can be used to study long exposure times and large populations. Such modeling may support the assessment of real life situations where multiple environmental sounds interfere and cause noise annoyance. Example simulations involving exposure to railway noise, road traffic noise, natural sound and sound produced by the individual's own activity show the influence of attention on the model outcome.

INTRODUCTION

Noise plays an important role in the perception of the quality of the living environment, and there is a growing public concern about the disruptive effect of an adverse living environment on health. Noise annoyance is commonly considered as the most widespread effect of environmental noise, and community noise annoyance is found to be a reasonable indicator to assess the impact of environmental noise pollution on man. Community noise effects in general, and noise annoyance in particular, are usually studied in relation to exposure through blind statistical analyses of large datasets.

In earlier work (De Muer et al. 2005; De Coensel & Botteldooren 2007, 2008; Botteldooren & De Coensel 2008), the authors have followed a quite different approach. 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 psychoacoustic and psychological plausibility. Results are analyzed statistically on a population basis, exactly as one would analyze results of field studies with human participants. This paper reports on a specific process within the model for a single individual: the role of attention.

Attention could play a dual role in the perception of environmental noise, and consequently in the emergence of noise annoyance. Firstly, attention can be drawn away from other tasks by the environmental sound, or tasks requiring sustained attention can suppress the noticing of environmental sound. Secondly, attention can jump between sounds in multisource sonic environments. Knowledge on the neurological basis of auditory attention has recently expanded enormously (Fritz et al. 2007). This evolution stimulated the development of very detailed computational models, such as that of Wrigley & Brown (2004). However, they mainly focus on speech processing. In modeling the dual role of attention in environmental noise perception, care must be taken to simplify existing knowledge on auditory perception in such a way that the model can be used to study long exposure times and large populations.

In the next section, the layout of such a model for a single individual is explained in detail. Subsequently, results of simulations involving a large number of individuals are given. The influence of attention on the model outcome is illustrated with simulations in environments with road traffic noise, railway noise, natural sound and sound produced by the individual's own activity.

METHODOLOGY

The proposed framework for including attention mechanism in modeling perception of and annoyance caused by environmental sound is shown in Figure 1. Simulated time series of the sound levels caused by various environmental sound sources form the input of the model. In a pre-attentive phase, salient parts of the sonic environment are detected. Inspired by available neurobiological knowledge on attention (Fritz et al. 2007; Knudsen 2007), the model implements a balance between top-down and bottom-up focusing; similar mechanisms have been identified in visual attention focusing (Itti & Koch 2001; Shi & Yang 2007). In the following subsections, we will briefly describe how each sub-mechanism is implemented in the model.

Simulating the sonic environment

Typical simulations consider the sound produced by vehicular traffic, railway traffic and natural ambient sources (wind in trees, birds, etc.) in the vicinity of the modeled individual. Vehicular and railway traffic is accounted for by simulating the emission of each vehicle/train individually. However, the model treats natural ambient sounds (including the sound produced by the modeled individual itself) as a whole, rather than to consider the sounds produced by each bird, each tree, etc. separately. Natural ambient sound is assumed to fluctuate according to a 1/*f* characteristic, because this characteristic was found in many recordings of environmental sounds (De Coensel et al. 2003). Using a simple sound propagation model which considers only attenuation caused by geometrical divergence, the time-varying sonic environment ($L_{Aeq,1s}$ time series for each source) at the location of the modeled individual is simulated.

Pre-attentive processes: masking, stream regrouping and saliency detection

A sound presented at the ear will be observed only when it is not completely energetically masked by other parts of the sonic environment. Therefore, the first step in the processing of the environmental sound reaching the ear of the modeled individual will account for masking. Temporal effects involved have time constants of a few

100 ms at most. Therefore, within the scope of a long-term analysis and given the time step of 1 s that was chosen, temporal effects of masking are safely ignored.

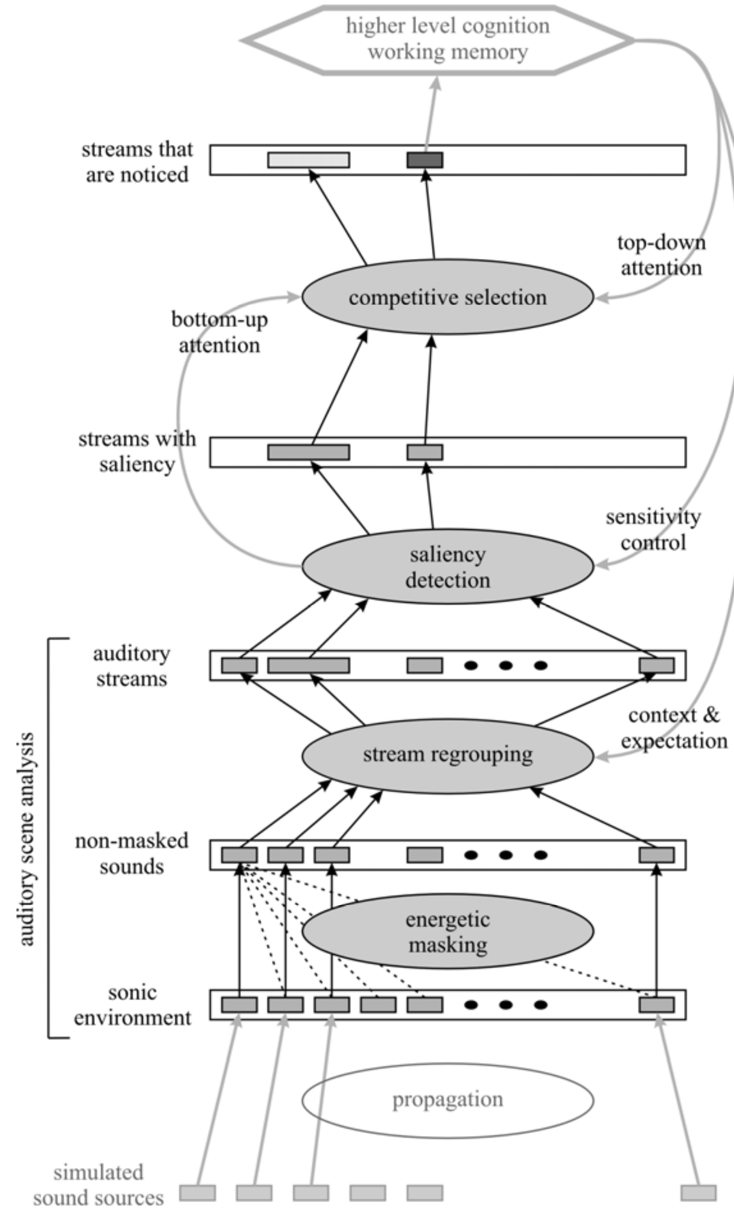


Figure 1: Layout of the proposed model

In the proposed model, masking thresholds for each sound in the sonic environment are based on the overall sound of all other sources. A comprehensive model for (partial) masking should be based on the specific loudness of each sound. However, because the proposed model is aimed at simulating large populations of listeners for long durations, the use of a detailed model such as that of Moore et al. (1997) is computationally infeasible. Therefore, the total A-weighted sound level of all other sounds is used as a proxy for the masking threshold. In other words, sounds (signals) are not energetically masked by all other sound (noise) if their signal-to-noise ratio (S/N) is positive.

The model could also allow for regrouping of auditory streams. 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) (Bregman 1994). Stream separation is performed on the basis of a variety of acoustic cues (bottom-up) and on the basis of acquired expectations from prior experience or knowledge (Fritz et al. 2007). By simulating auditory streams for each separate environmental sound source, the non-trivial problems of modeling ASA and sound source identification are effectively by-passed. However, this approach imposes an a priori stream segregation, which may not coincide with actual perception. Examples are the sound of a fountain making road traffic noise being perceived as part of the fountain sound, whether individual cars are heard or the sound of a street, whether individual bird songs are heard or the morning chorus of birds etc. Therefore, at least conceptually, a stream regrouping process is allowed in the model.

Saliency detection is probably the most important feature of the biological auditory system. Novelty may trigger bottom-up attention, but this is not a necessary consequence. Saliency may be detected based on various temporal and spectral cues. Saliency detection is implemented in the model, for each auditory stream separately, by comparing the S/N of the stream to a predicted value. For simplicity, a linear predictor based on the exponential average of earlier S/N is used. A future improvement could be to include a neural network, dynamically trained for optimal prediction, possibly based on remembered sounds retrieved from memory. Because the peak detection is implemented for each individual type of sound, deviant behavior on the frequency axis, which is usually included in the calculation of a saliency map (Kayser et al. 2005), can be neglected in a first approximation.

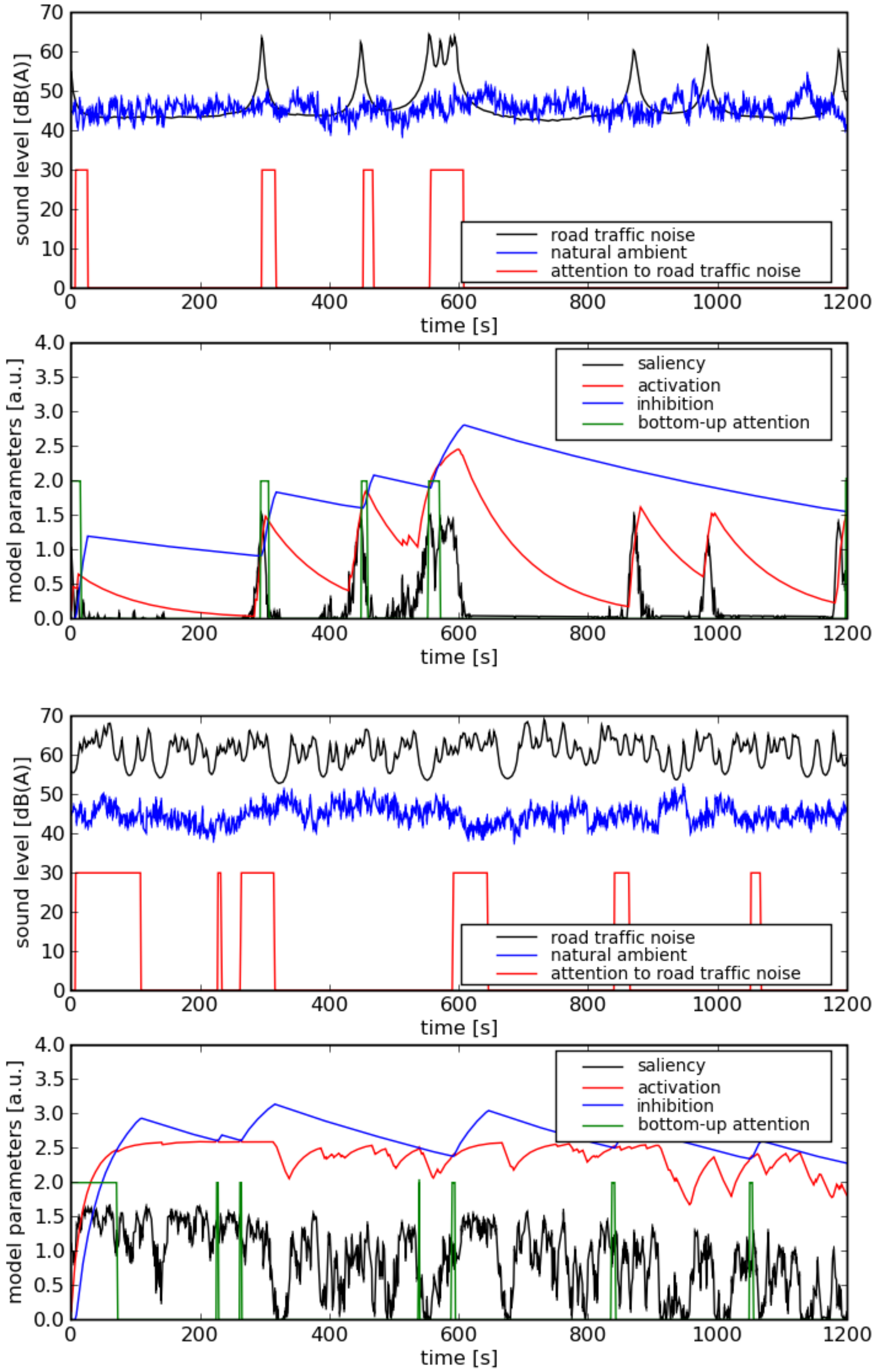
Figure 2 shows two examples of calculated saliency. Close to the road where only few vehicles an hour come by, saliency roughly follows the individual car passage events. Since remembered sounds of car passages are not used in this first approximation and the time between passages is unpredictable, the prediction is poor and saliency keeps popping up at each individual passage. Close to the road carrying more traffic, the predictor gradually improves and saliency decreases.

Bottom-up attention mechanism

High levels of saliency attract attention (bottom-up). However, the perceptual system does not stay focused on this salient feature continuously. The mechanism playing a crucial role here is inhibition of return (IOR) (Itti & Koch 2001), which prohibits attention to come back to the same salient streams over and over again.

This mechanism is implemented in a simplified way as a competition between an activator, AA, and inhibitor, AI. Activation is triggered by saliency. It has a very short rise time τ_{Ar} and a somewhat longer decay time τ_{Ad}. Because the rise time probably is shorter than the 1 s resolution of the simulation, AA jumps up almost instantaneously in the model. Decay starts immediately after this jump. The strength of activation is assumed to be a saturation process. If activation is higher than inhibition for a particular auditory stream, this stream spikes for attention. As soon as attention gets focused on the particular stream by the mechanism explained in the next section, inhibition comes into play. It is modeled as a slow process with a short rise time τ_{Ir} ($>\tau_{Ar}$) and a very long decay time τ_{Id} ($>\tau_{Ad}$). AI continues increasing until eventually it exceeds AA at which time the stream stops spiking for attention. Attraction of bottom-up attention is thus implemented as a spiking mechanism: the process will fire a couple of times, until either saliency decreases or attention is obtained, at which time the IOR mechanism will stop the spiking.

In Figure 2 activation and inhibition levels are shown together with the spikes (green line) resulting from it. For the low intensity traffic stream, (bottom-up) attention is asked for at almost every car passage although occasionally a noise event is overlooked. The higher traffic intensity situation shown in the lower graphs results in slightly more attention requests, but after the initial adaptation period of a few hundred seconds, a quite different regime sets in: the time since the last granted request is as least as important as the instantaneous saliency.



a)

b)

Figure 2: Road traffic noise level and background level in dBA together with intervals of attention for road traffic noise (upper graphs); saliency, activation, inhibition and bottom-up attention triggers (lower graphs): a) close to a road with few vehicles/h, b) close to a road with about 500 vehicles/h

Top-down attention mechanism and attention switching

Top-down attention focusing is guided by higher cognitive processes. It results in a change in sensitivity for a particular auditory stream, which implies that it has no effect as long as no stimulus is present. Top-down attention for environmental sounds may depend on

- the task that the person is performing, which in turn is related to the current activity;
- emotional state such as anxiety, arousal etc.
- personal treats such as noise sensitivity etc.
- the information content of the stream to which attention is currently directed.

The bottom-up mechanism makes attention switch between auditory streams, whereas the top-down mechanism tries to focus attention on a single stream. Sporadically attention stays caught by the bottom-up mechanism, resulting in remembered noticing of the environmental sound. It could be assumed that limited availability of resources mainly plays a role in sustained attention.

The attention switching mechanism can be seen as a gating mechanism that switches off all but a couple of auditory streams at a single instance in time. The sensitivity for non-attended sources is decreased. Parallel processing is changed to serial processing: the individual listens to a single stream at a time.

The implementation of the top-down mechanism consists of an activity related part DA_a, a personal related part DA_p, and a part related to the information content of the attended sound DA_i. The first two parts are independent from the sound. They could be modeled using time-activity patterns of the studied population, but currently are fixed for every modeled individual in the population. The last part could depend on the meaning that people give to the sound, but currently is modeled solely on the basis of the overall saliency of the attended signal, which could be seen as a measure of interesting variation in the sound. If this variation is lost, attention drops.

The attention switch that is mathematically implemented is a winner-take-all mechanism. The selection is based on the sum of the overall DA and an activation of the bottom-up attention. The activation of the bottom-up attention is an integrator that receives an additional activation δUA every time the bottom-up mechanism described in the previous section spikes for attention, and decays rather fast with a time constant τ_{UA} afterwards. The gate is opened for the winning stream, resulting in a passing of the auditory stream to the cognitive process of attaching meaning (currently only the calculation of DA_i based on saliency). This situation is kept until another stream gets attention.

In order for this numerical model to work properly, non-auditory processes should be allowed to draw attention away from the auditory streams. Indeed, particular tasks could prevent noticing of sounds even though they do not involve listening at all. To account for this, an additional non-auditory stream is added that periodically fires for attention.

Figure 2 also shows the actual attention paid to the road traffic (red line in upper curves). It can be seen that it responds roughly to the triggers received from the bottom-up mechanism but not always since the modeled individual may be involved in other activities requiring attention. The duration of actual attention is also longer than the duration of bottom-up attention request. This prolonged attention is in this simple approximation fully determined by saliency, be it in an indirect and complex way.

SIMULATION RESULTS

The main purpose of the numerical simulation is to observe emergence of features that have not been modeled explicitly and thus to relate them to underlying basic mechanisms. For the example given in this paper, simulations are performed for a simplified setting: people seeking recreation in mildly disturbed rural areas. The sonic environment of this setting is modeled using the simplified layout of the environment shown in Figure 3 and with parameters ranges defined in Table 1. In total 20,000 combinations of people and environment were simulated for a one hour visit and the results for this synthetic population are analyses statistically in a manner very similar to the usual analyses of survey data.

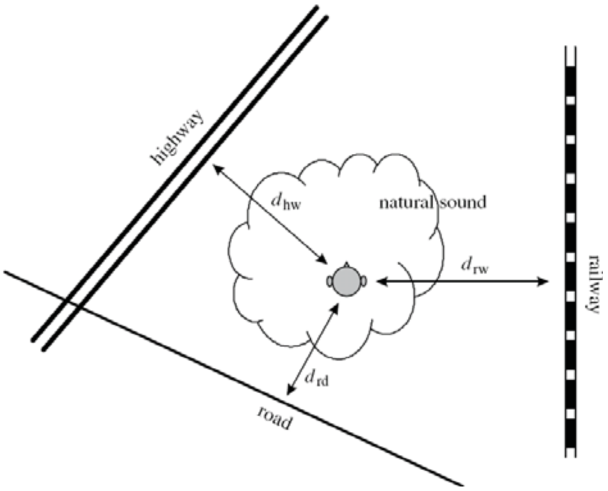


Table 1: Range of distances and traffic intensities used in the numerical experiment

Quantity	Average	Minimum	Maximum
d_{hw}	3.9 km	1 km	10 km
N_{hw}	2000 cars/h	1500 cars/h	2500 cars/h
	400 trucks/h	300 trucks/h	500 trucks/h
d_{rd}	185 m	5 m	1 km
N_{rd}	250 cars/h	5 cars/h	500 cars/h
	10 trucks/h	1 truck/h	20 trucks/h
d_{rw}	1950 m	500 m	5 km
N_{rw}	1 train/h	No trains	3 trains/h

Figure 3 (left): Simplified layout of the rural setting

In Figure 4 we focus on the effect of activities of the modeled individual seeking recreation. In the left column figures, the modeled individual is assumed to be focused on activities not related to observing environmental sound such as conversation, reading, active sporting, and thus not paying much (top down) attention to the sound. In this case, the average modeled individual in the synthetic population observes road traffic for some amount of time but hardly any other sources. The saturation with increasing $L_{Aeq,\ road}$ (upper graphs) could be linked to the saliency detection gradually decreasing, which could be called habituation or adaptation to the sonic environment. Within the range of modeled railway L_{Aeq}, there is no effect of railway noise on the time that road traffic sound is observed. With increasing level of natural sound, this time drops down very strongly but interestingly enough, the time that natural sound is observed hardly rises.

The picture changes when it is assumed that the modeled individual is actively listening to natural sounds, e.g. while watching birds (right column of graphs in Figure 4). Due to the fact that relatively strong top down attention is used, the modeled individual succeeds in its goal and observes natural sounds most of the time, except when the level of natural sounds drops very low. The interference with this listening task caused by road traffic is observable but less than expected, even at road traffic noise levels above 70 dBA. More detailed analyses showed that this is due to the fact that a rural area is modeled and high levels of road traffic noise correspond to close roads carrying moderate traffic intensity. Hence energetic masking happens only infrequently and the attention mechanism is still able to pick up on the natural sound.

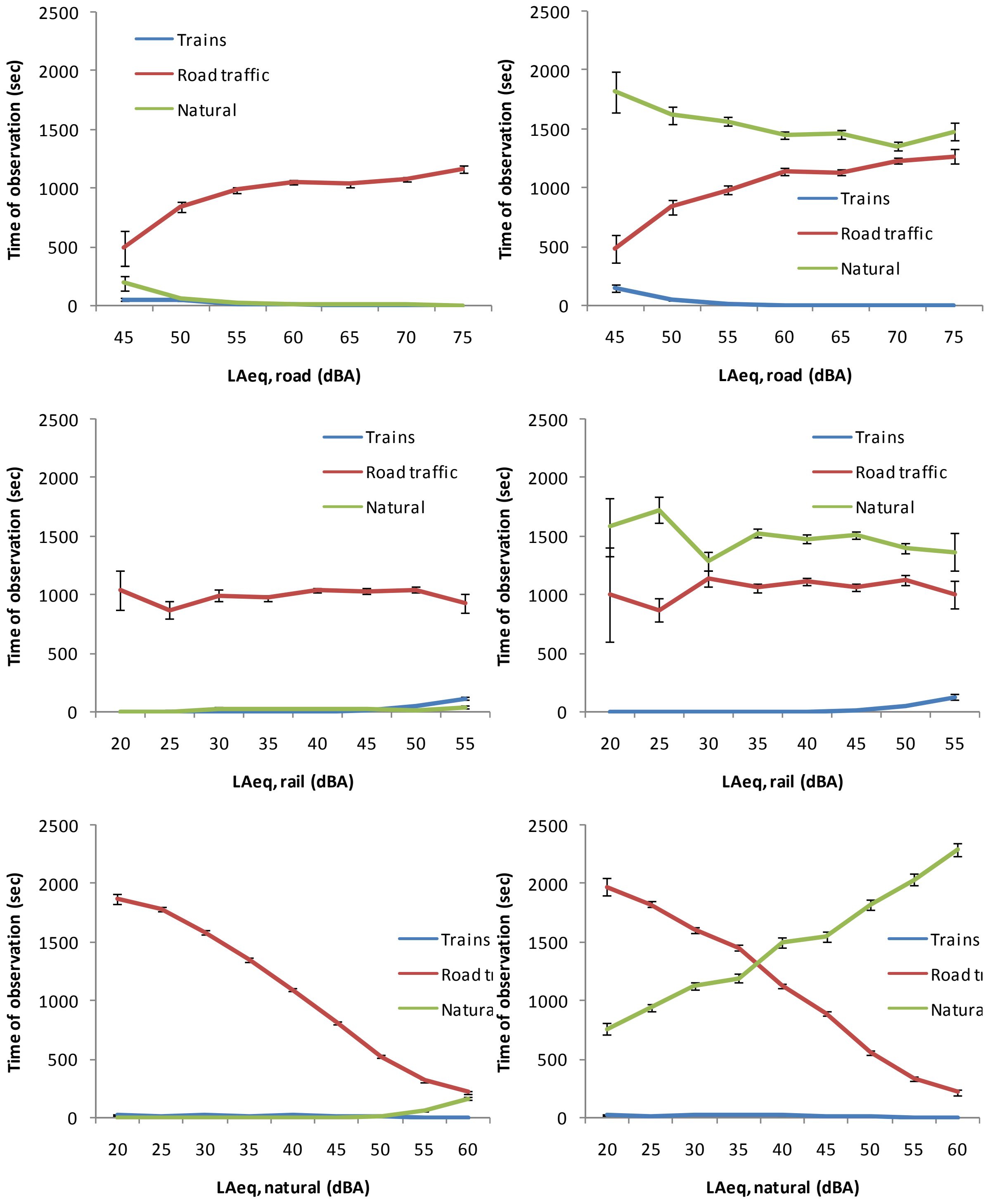


Figure 4: Time of observation of different types of sound as a function of equivalent noise levels: for top down attention focussed on other activities in left column; dop down attention focussed on listening to natural sounds

CONCLUSIONS

In this paper we discussed how attention mechanisms can be included in a numerical model for individual observation of environmental sounds. Care is taken to keep computational complexity of the proposed model limited in order to allow simulating the effect on large synthetic population for considerable observation times. On the one hand saliency detection and bottom up attention triggering were modeled in detail, on the other hand top down attention was included in a more general way.

Simulation of a synthetic population recreating in a mildly disturbed rural area allowed investigating how such a model can generate emergent features for the population as a whole. At least some of these results can be related in a qualitative way to earlier field studies such as (De Coensel & Botteldooren 2006) where we noted that the equivalent noise level of road traffic noise did not accurately predict the disturbance of silence. This could be explained by the saturating time of observation of road traffic sound with L_{Aeq} and the limited effect on the observation of natural sound (when focusing attention on it) with increasing L_{Aeq}. The better correlation with LA50 observed in this field work could not be checked in the modeling at the time that this paper was written.

This paper presents another step forward in an ongoing effort to model the effect of environmental sound on everyday life numerically based on basic knowledge on human perception. An effort that aims at bridging the gap between environmental noise specialists, medical researchers, and psychologists.

REFERENCES

Botteldooren D, De Coensel B (2008). A model for long-term environmental sound detection. In: Proceedings of the IEEE World Congress on Computational Intelligence, Hong Kong.

Bregman AS (1994). Auditory scene analysis: the perceptual organization of sound. Cambridge, MA: MIT Press.

De Coensel B, Botteldooren D (2006). The quiet rural soundscape and how to characterize it. Acta Acust Acust 92: 887–897.

De Coensel B, Botteldooren D (2007). Models for soundscape perception and their use in planning. In: Proceedings of Internoise, Istanbul, Turkey.

De Coensel B, Botteldooren D (2008). Modeling auditory attention focusing in multisource environments. In: Proceedings of Acoustics'08, Paris, France.

De Coensel B, Botteldooren D, De Muer T (2003). 1/*f* noise in rural and urban soundscapes. Act Acust Acust 89: 287-295.

De Muer T, Botteldooren D, De Coensel B, Berglund B, Nilsson ME, Lercher P (2005). A model for noise annoyance based on notice-events. In: Proceedings of Internoise, Rio de Janeiro, Brazil.

Fritz JB, Elhilali M, David SV, Shamma SA (2007). Auditory attention – focusing the searchlight on sound. Curr Opin Neurobiol 17: 437-455.

Gygi B, Shafiro V (2007). General functions and specific applications of environmental sound research. Front Biosci 12: 3152-3166.

Itti L, Koch C (2001). Computational modelling of visual attention. Nat Rev Neurosci 2: 194-203.

Kayser C, Petkov C, Lippert M, Logothetis NK (2005). Mechanisms for allocating auditory attention: an auditory saliency map. Curr Biol 15: 1943-1947.

Knudsen EI (2007). Fundamental components of attention. Annu Rev Neurosci 30: 57-78.

Moore BCJ, Glasberg BR, Baer T (1997). A model for the prediction of thresholds, loudness, and partial loudness. J Audio Eng Soc 45: 224-240.

Sagi B, Nemat-Nasser SC, Kerr R, Hayek R, Downing C, Hecht-Nielsen R (2001). A biologically motivated solution to the cocktail party problem. Neural Comput 13: 1575-1602.

Shi H, Yang Y (2007). A computational model of visual attention based on saliency maps. Appl Math Comput 188: 1671-1677.

Wrigley SN, Brown GJ (2004). A computational model of auditory selective attention. IEEE Trans Neural Netw 15: 1151-1163.