LISTEN Auralization of Urban Soundscapes

Final report to the Knowledge Foundation

Mats E. Nilsson
Department of Psychology, Stockholm University, Stockholm, Sweden

Jens Forssén
Applied Acoustics, Chalmers University of Technology, Gothenburg, Sweden

Peter Lundén,
Sonic Studio, the Interactive Institute, Stockholm, Sweden

Andrew Peplow,
KTH Royal Institute of Technology, Stockholm, Sweden

Björn Hellström
University College of Art, Crafts and Design, Stockholm, Sweden

Preface and Acknowledgements

This is the final report of the 3-year research project “LISTEN: Auralization of Urban Soundscapes” that started in September 2008. The project was financed by the Visualization-in-Sweden call for projects, launched by the Knowledge Foundation, Invest in Sweden Agency, Swedish Governmental Agency for Innovation Systems (VINNOVA), Swedish Foundation for Strategic Research, and the Vårdal Foundation. The project was co-financed by its six end-user partners: Swedish Road Administration, Swedish Railway Administration, Stockholm City, WSP Acoustics, ÅF Consulting, and Rambøll Denmark A/S. The project’s research partners were Stockholm University (coordinator), Chalmers University of Technology, the Interactive Institute, KTH Royal Institute of Technology and University College of Art, Crafts and Design.

We would like to thank the following representatives of the end-user partners for their dedication and contribution to the project: Allan Jensen from Rambøll Denmark; Klas Hagberg, Crispin Dickson and Andreas Novak from WSP Acoustics; Björn Hellström, Philip Zalyaletdinov and Leif Åkerlöf from ÅF-consulting; Kjell Strömmer and Lars Dahlbom from the Swedish Road Administration (now the Swedish Transport Administration); Gisela Köthnig and Karin Blidberg from the Swedish Railway Administration (now the Swedish Transport Administration); Magnus Lindqvist from Stockholm City.

We would also like to thank the following researchers and students for their help in the project: Maria Rådsten-Ekman and Jesper Alvarsson from Stockholm University; Chinmay Prafulla Pendharkar from Chalmers University of Technology; Marja Gustlin from KTH Royal Institute of Technology, Peter Becker from the Interactive Institute; Dr. Ricardo Avienca from University College of Art, Crafts and Design. Finally, we would like to thank Dr Tomasz Kaczmarek from Institute of Acoustics, Adam Mickiewiz University, Poznan, Poland for his contribution to the development of auralization methodology in the early stages of the project.

Content

Summary

Noise pollution is an increasing environmental problem in urban areas. Consequently, there is an increasing need for city planners and architects to consider the acoustic qualities of new residential and public areas. Auralization, that is the simulation of sounds using mathematical models, may be used for this purpose. Currently there are auralization tools available for architects or engineers which can evaluate the acoustic quality of rooms in new buildings. However, the corresponding tools for design of outdoor environments are lacking.

Therefore, the main objective of the project "LISTEN: Auralizing Urban Soundscapes" was to develop a demonstrator (the LISTEN-Demonstrator), to illustrate the potential of auralization as a tool for evaluating future sound environments (or soundscapes). With such a tool, auditory effects of architectural, noise control and design solutions may be evaluated at the planning stage, by simply listening to their effect on the perceived soundscape. Specifically, the LISTEN-Demonstrator should illustrate the potential of auralization for three scenarios in a typical urban environment exposed to road-traffic or railway noise: (1) outdoor soundscapes at a noise exposed side of an apartment building; (2) indoor soundscapes in an apartment room with noise exposed windows; (3) outdoor soundscapes at the shielded side ("quiet side") of a noise exposed apartment building.

The project's main result is the LISTEN-Demonstrator. This software auralizes the effect of a number of environmental factors on traffic noise. The factors can be varied independently, which makes it possible to create a limitless number of soundscapes. For road-traffic noise, the LISTEN-Demonstrator allows real-time listening determined by the following environmental factors:

(i) Distance from listener to road.

(ii) Traffic density.

(iii) Mean speed of vehicles.

(iv) Type of ground between listener and road (asphalt or grass).

(v) Height of noise barrier between listener and road. The following heights were included: 0 m (= no barrier), 2 m and 4 m height. The model included diffraction at the top of the barrier, but not at its sides (i.e., barrier assumed to be of infinite length).

(vi) Type of window (indoor scenario).

(vii) Degree of openness of windows (indoor scenario).

In addition, methods were developed for auralizing road-traffic noise behind barriers of finite length and behind a building (the shielded side scenario). The model of the finite barrier included diffraction both at the top of the barrier and at its sides. The model of the shielding building included diffraction both at building roof and building sides, as well as, reflections from other buildings. These situations were found to be too complex to achieve real-time auralizations. Therefore a smaller number of pre-calculated situations were included in the LISTEN-Demonstrator. This was also true for railway noise, which was more difficult to auralize than road-traffic noise, mainly due to the larger number of sources that had to be modeled. For railway noise, precalculated auralizations were included for different distances to the rail and with barriers of different heights close to the railway.

The development of the LISTEN-Demonstrator was based on results from the project's research on acoustic modeling, auralization and perceptual evaluation. The main research results are summarized below:

1. A general methodology was developed for auralizing sound from road-traffic and railway vehicles.

2. A method was developed for auralizing the effect of noise barriers on road-traffic and railway noise, including effects of diffraction both at the top of the barrier and at barrier sides.

3. Real-time implementations of the acoustic models were accomplished for the majority of environmental factors included in the project. The computational demands of very complex situations involving multiple diffraction and reflection paths were found to exceed the performance of today's personal computers. However, it will be possible in the near future with the expected increase in performance of personal computers.

4. A methodology was developed for perceptual validation of auralizations, which was used to test the auralizations of single road-traffic vehicles. The results showed high agreement between auralized sound and real recordings, supporting the validity of the LISTEN approach to auralization.

In conclusion, the project has provided a demonstration of how auralization may be used for evaluating urban soundscapes, specifically the effect of various environmental factors on the perception on traffic noise. Further developments of this work should include (a) more complex scenarios, which would require effective use of parallel processing in multi-core computers, (b) more detailed source modeling to make auralization of decelerating and accelerating vehicles possible, (c) more advanced traffic flow models that may allow modeling of situations with traffic congestion, and (d) visual presentations, linking auralizations to visual images of the environment.

The project was financed by the Visualization-in-Sweden call for projects, launched by the Knowledge Foundation, Invest in Sweden Agency, Swedish Governmental Agency for Innovation Systems (VINNOVA), Swedish Foundation for Strategic Research, and the Vårdal Foundation. The project was co-financed by its six end-user partners: Swedish Road Administration, Swedish Railway Administration, Stockholm City, WSP Acoustics, ÅF Consulting, and Rambøll Denmark A/S. The project's research partners were Stockholm University (coordinator), Chalmers University of Technology, the Interactive Institute, KTH Royal Institute of Technology and University College of Art, Crafts and Design

1. Introduction

Traffic noise is a great and increasing environmental problem in urban areas [1, 2]. The increasing urban population and the limited space available in urban areas means that new residential and public areas will be built in close proximity to major noise sources, such as main roads and railways. Reduction measures at the source alone will not solve the noise problem, because drastic traffic-volume reductions or legislation for quieter cars and tires are difficult to achieve for political and economical reasons [3]. Therefore, healthy outdoor sound environments, in accordance with existing guideline values for traffic noise, require methods for soundscape improvements at the receiver end. This includes noise control measures (e.g., noise barriers), architectural solutions (e.g., creating quiet court-yards) and city-planning measures (e.g., directing traffic in order to maximize shielding effects of large buildings) [4].

Evaluation of such mitigation methods solely based on predicted reductions in noise decibels (dB) are insufficient for predicting the associated perceptual effects [5]. This because any method for improving the soundscape[1] will cause a number of perceptual changes that are unrelated to the decibel reduction. For example, a noise barrier will change the spectral and temporal structure of traffic noise, as well as, the relative prominence of other sound sources in the soundscape. All these changes are perceptually relevant, since they may influence traffic noise annoyance and perceived quality of the soundscape [6]. Therefore, better decisions regarding methods for soundscape improvement may be obtained if calculations of dB-values are complemented with perceptual evaluation of the expected soundscape. This requires auralizations[2], by which all perceptually relevant changes of the sound field can be modeled and reproduced in real-time (or interactive time).

1.1 Objectives of the project

The main objective of the LISTEN project was to develop a demonstrator of an auralization tool, by which architectural, noise control and design solutions for improving urban soundscapes can be auralized at the planning stage. Various solutions for soundscape improvement may thus be evaluated by simply listening to their effect on the perceived soundscape.

The demonstrator developed in the project (henceforth called the LISTEN-Demonstrator) was designed to illustrate the potential and feasibility of soundscape auralization, by demonstrating the application for three scenarios in a typical urban environment exposed to road-traffic or railway noise (Figure 1.1):

1) Outdoor soundscapes at a noise exposed side of an apartment building.
2) Indoor soundscapes in an apartment room with noise exposed windows.
3) Outdoor soundscapes at the shielded side (“quiet side”) of a noise exposed apartment building.

All scenarios included the effects of noise barriers of various heights close to the source. The soundscapes presented in the LISTEN-Demonstrator consisted of auralized traffic noise combined with recordings of background sounds (soundscapes undisturbed by traffic noise). The perceptual validity of the auralization methodology that forms the basis of the LISTEN-Demonstrator was evaluated in psychoacoustic listening experiments.

---

[1] We use the term ”soundscape” synonymous with ”sound environment”, referring to the totally of sounds in a given environment. The soundscape may include both wanted sounds, such as pleasant bird song, and unwanted sounds, such as annoying road traffic noise. “Soundscape quality” refers to the overall perception of the soundscape on a scale from poor to good.

[2] Kleiner *et al.* [7] defines auralization as “… the process of rendering audible, by physical or mathematical modeling, the sound field of a source in a space, in such a way as to simulate the binaural listening experience at a given position in the modeled space.”

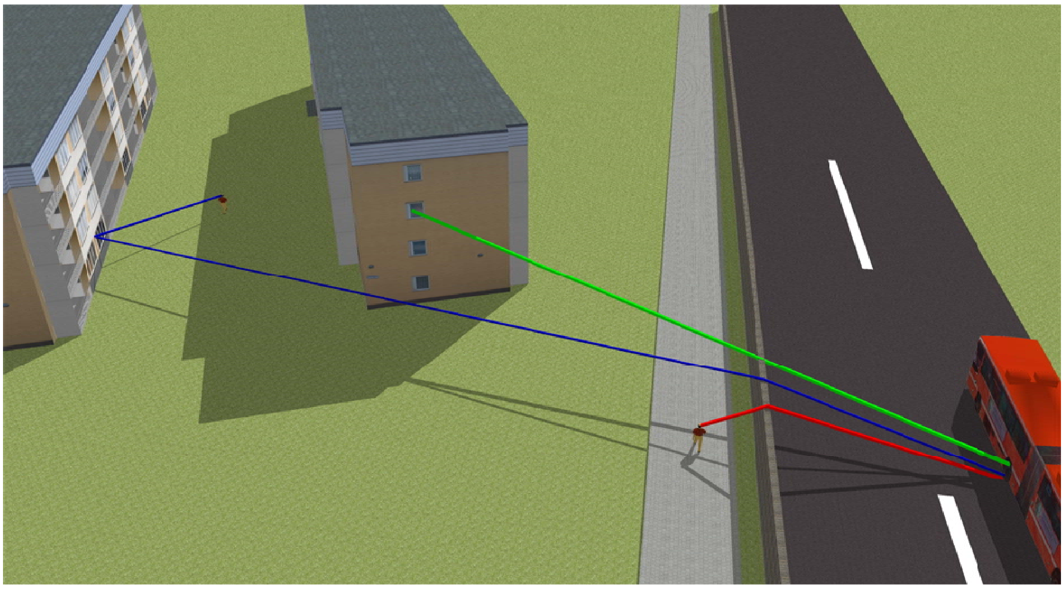

Figure 1.1 Illustration of the three LISTEN-scenarios. (Note that the exact geometry of the auralized scenarios was different than shown in this figure, see Figure 2.3.)

1.2 Plan of the report

The organization of this report relates to the project's work plan as outlined in the application, see Figure 1.2. Section 2 describes the user-driven nature of the work in the project, where the specific purpose and design of the LISTEN-Demonstrator was decided in dialogue between researchers and end-user partners. Section 3 describes the results from the Acoustic modeling work package, in which existing or new models for sound propagation were developed and applied to obtain traffic-noise auralizations. The main part of this work was conducted by Chalmers University of Technology and KTH Royal Institute of Technology. Section 4 describes the Auralization work package, in which the acoustic models were implemented in the software Pure Data (PD, http://puredata.info) to obtain the LISTEN-Demonstrator. This work was conducted by the Interactive Institute. Section 5 describes the work with perceptual validation and optimization of the auralizations. This work was led by Stockholm University and University College of Art, Crafts and Design. Finally, Section 6 summarizes the main conclusions of the project and discusses future perspectives, including a SWOT analysis.

2. User-driven Specification of Demonstrator Purpose and Design

The general purpose and design of the LISTEN-Demonstrator was specified through dialogue between research partners and end-user partners (WSP Acoustics, ÅF Consulting, Rambøll Denmark, Swedish Transport Agency, Stockholm City). The end-user partners had a large impact on which aspects that should be auralized in the LISTEN-Demonstrator, for example effects of hard and soft ground, effects of various window types, and effects of finite barriers. Effects of different road surfaces, of various proportions of electric cars, of mixed soft and hard ground, and of barriers close to the receiver (e.g., to protect outdoor sitting places) were also requested by the end-user partners. Although not included in the final LISTEN-Demonstrator, the methodology developed in the project was directed to allow such additions in the future. In addition, the end-users influenced the development of the software for the LISTEN-Demonstrator, to assure that it would be possible to evaluate it within the end-user organizations. The main forum for the end-user and research partner interaction was the strategic meetings of the project. In between these meetings, informal discussion and feedback from end-users on developments in the project guaranteed the end-user driven development of the LISTEN-Demonstrator.

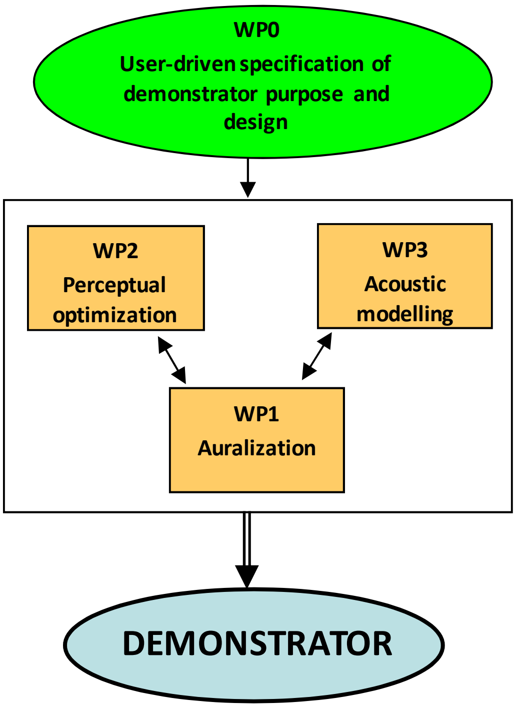

Figure 1.2. Schematic work plan of the LISTEN project.

3. Acoustic Modeling

Section 3.1 briefly describes the basic methodology of auralization and the approaches used for creating the traffic sound environments of the project. Sections 3.2 and 3.3 deal with the approaches to model the acoustic sources for road and rail vehicles, respectively. Section 3.4 describe the general sound propagation modeling, followed by sections describing the specific modeling of the different scenarios (Sections 3.5–3.7). Section 3.8 lists the conclusions draws from the project's research work on acoustic modeling.

3.1 General approach

At the heart of the auralization is the separation of the source signal and the propagation effects. For a time signal of the source, $s(t)$, the received sound pressure, $p(t)$, can be calculated as

$$p(t) = \int_0^\infty s(t-\tau)h(\tau)\,\mathrm{d}\tau\,,$$

where $h(\tau)$ is the impulse response due to the propagation.

As an example, the impulse response of free-field propagation in three-dimensional space is the free-field Green function $h(\tau) = \delta(\tau - R/c)/4\pi R$, where $\delta(\tau)$ is the Dirac delta function, R is the distance and c is the speed of sound. This results in a received sound signal according to

$$p(t) = \frac{s(t-R/c)}{4\pi R}\,.$$

This can be described as a delay in time, of the source signal, by an amount equal to R/c and a decrease in amplitude inversely proportional to R. For a moving source the distance becomes a function of time, $R(t)$. The retarded time, $t - R(t)/c$, is used here to model the Doppler shift, via a resampling of the time signal.

In the processes of auralizing road or rail traffic noise, for a person standing still at a position exterior to the vehicle, a number of parameters need to be considered. The sound path that reaches the listener from a vehicle can be described as starting from a set of sources that each has its own properties concerning directivity and spectral content, including noisy and tonal characteristics.

While propagating from source to receiver, the sound will be influenced by decaying amplitude due to spherical spreading, air attenuation, which leads to a larger reduction of the higher frequencies compared to the lower frequencies, and the reflection in the ground surface, leads to an interference pattern over frequency. The theoretical interference pattern for a point source showing significantly deep dips. The real interference pattern is however usually weakened due to smearing effects like decorrelation caused by random ground roughness and air turbulence as well as effects of the finite sized sources, which has been considered here. The acoustic ground impedance can be calculated for different types of ground (e.g. [7]), which enables calculation of the spherical reflection factor. Here, asphalt or grass has been modeled. For the moving source, relative to the receiver, the Doppler effect needs to be modeled as described above. Note we have not considered the effects of refraction, i.e. the curved sound paths due to wind or height varying temperature, which mainly influence the sound at longer distance between source and receiver than studied in the present project.

Finally, the sound entering our ears is influenced by our head and torso, which here is modeled using head related transfer functions, HRTFs (from the CIPIC database, subject 165, a KEMAR manikin [8]). The whole auralization process can, in our methodology, be summarized as starting with the creation of the noises and tones of the sources followed by the modeling of the air attenuation, the ground effect, the directivity, the Doppler effect, the spherical spreading and ending with the HRTFs. Modeling in 1/3 octave bands is undertaken for the air attenuation, the ground effect and the directivity, whereas the remaining effects are modeled directly on the time signal, i.e. the Doppler effect, the spherical spreading and the HRTFs. The further modeling of the sound propagation for the different scenarios is described below.

3.2 Modeling of the acoustic sources of road vehicles

For the auralization of the road vehicles, the source signals are modeled as steady-state noises and steady-state tones. To define the noises and tones, an inverted process is applied where a recorded time signal in mono is used as input. The process going from the mono recording to the source signal can be described as inverting the effect of spherical spreading, the Doppler effect, the ground effect, and the air-attenuation. Resulting from the inversion process is a signal that is seen as a steady-state source signal shaped by the source directivity. The slowly varying amplitudes (envelopes) of each 1/3 octave band are stored as directivity polynomials for later use in the auralization. After separation into forward and backward direction, a second-degree polynomial fitting is used with an adjustment to also fit the centre level in order to give a continuous curve. After removing the variation in amplitude modeled as directivity, a period shortly after point of passage is used as an estimate of the source signal. As default, a 1 s long signal was used, starting 1 s after point of passage.

For the road vehicles, two sources are used to determine the vehicle response, (*i*) a propulsion source, which models engine, air intake, air exhaust, fans, compressors, etc., and (*ii*) a road/tire source, which models the noise generated by the contact between tire and road surface. This set is modeled according to current engineering methods (Harmonoise and Nord2000 [8, 9], with the sources for passenger cars located on a vertical line at heights (*i*) 0.3 m and (*ii*) 0.01 m. For medium heavy and heavy vehicles the height of the propulsion source is at 0.75 m. The transition frequency, where the road/tire source starts to dominate, is estimated from the Harmonoise source model [8].

Initially in the project, two different approaches were used for the modeling of the tonal characteristics and for the synthesis of the auralization sounds, as described by Forssén et al. [10]. The methodology described above is referred as the Time-Domain approach (TD) since some propagation effects are inferred on the time signal (like the Doppler effect). Connected to the TD approach, the tonal characteristics were modeled from peaks of a narrow band spectrum of the source signal using one tone per 1/3 octave band. The other approach is based on Additive synthesis (AS), as previously described by Kaczmarek [11]. In the additive synthesis, sinusoids with 0.5 Hz resolution and random initial phase are added to produce the noise-like character of the pass-by sound. For each single sinusoidal component the propagation effects on the amplitude and frequency change over the time are calculated as due to the geometrical spreading, the Doppler effect, the ground effect and the air attenuation. The prominent tonal components were identified from cepstral analysis and added as separate sinusoids. Both approaches were concluded to perform acceptably well [10], whereas it was decided to develop Time-Domain approach (TD) for the future development of the project, partially due to the larger numerical cost of the additive synthesis approach (AS).

Later on within the project work, the modeling of the tonal character of the passenger cars was further developed, based on the idea of so-called granular synthesis [9]. The main idea is to find the shortest time pattern that describes the propulsion sound, which here was made via auto correlation analysis. The resulting auralizations were considered to be an improvement, as shown by a new set of listening tests, see Section 5.4.

For the heavier road vehicles, the tonal character of the engine sound gets both more prominent and more complex. This is also the case for diesel engine passenger cars, however to a lesser extent. To model sounds with such rich tonal character imbedded in noise, another methodology was investigated, called spectral modeling synthesis (SMS). With that approach, the amplitudes of noise and tones as well as the frequency of the tones are allowed to vary with time [12]. The rate of variation may be much higher than e.g. that of the directivity. Thereby the complex sound character can be better described and re-synthesized. In the SMS re-synthesis process it was possible to take into account the propagation effects using the same methodology as for the passenger cars, as previously developed within the project, and thereby auralize passages also for the road vehicles with the more complex propulsion sounds. In a master thesis work connected to the LISTEN project, the SMS approach was investigated. Pass-by sounds of a diesel engine passenger car, a medium heavy truck and a bus were synthesized for different driving speeds and the results were evaluated in comparison with recordings in a listening test showing that the SMS approach works well [13].

3.3 Modeling of the acoustic sources of railway vehicles

Railway noise consists of noise from the engine and fans (traction), noise from the friction of wheels over the rails (rolling), noise from turbulence of airflow over the structure of the railway (dominated at high speeds), impact noise and wheel noise (wheels).

For the train auralizations, a similar methodology to car auralization was adopted, whereas a directivity factor has been included in the methodology according to the work by Bongini and Bonnet [14] for the traction noise and the rolling noise. The switch from monopole to a dipole-like source was implemented for rolling noise whereas for traction sources which dominate low-frequency ranges any modifications to directivity did not influence the global pass-by signature.

In adopting the synthesis method described above one has to define the rolling stock, traction and track as a set of noise point sources. The synthesis method determines each noise source pass-by into a time-frequency environment, taking into account, its time-evolution (depending on pass-by scenario), the Doppler effect and the ground effect. The time-frequency plans, each of them corresponding to an equivalent point source contribution towards the pass-by noise, are summed. The global time-frequency plan is therefore transposed into the time domain. The main consequence of the synthesis method on the source model is that the point source model is considered for equivalent source. For passenger car realizations this adoption of point source models seems reasonable upon listening panel analysis. However, for train auralizations it is still unclear whether a global point-source methodology is viable for all passenger train types.

It must be pointed out that most of trains in urban rolling conditions are composed of about twenty real noise sources. Moreover, since this method has been developed to carry out parametric studies on the influence of each noise source on the global pass-by noise, the equivalent sources have to be easily handled according to the modifications managed on real sources. Consequently, it is recommended to represent one physical noise source by very few point sources preferably only one. An RC6 intercity train shown in Figure 3.1 was modeled by 17 point sources, whereas a smaller commuter train Regina was modeled according to a seven equivalent point source scheme.

Figure 3.1. Illustration showing location of equivalent point sources in the train model for RC6 inter-city train.

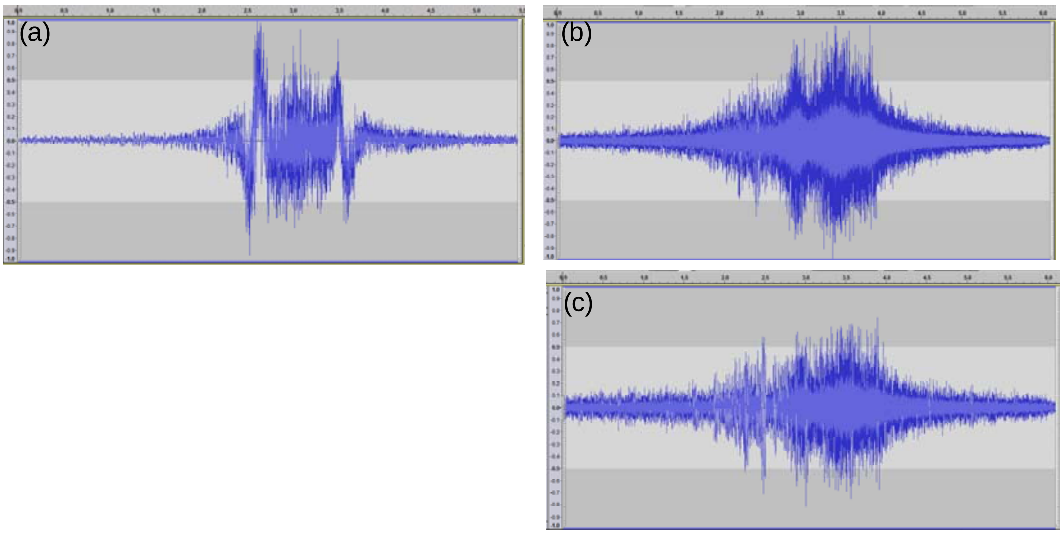

Figure 3.2. (a) Time signal recording for Regina commuter train, (b) auralization of the same recording, (c) auralization with barrier (0.5 m height, 1.5 from track side).

Measurements undertaken by WSP Sweden correspond to rail measurements adopted in the ISO 3095 standard for the pass-by measurement. This method consists on recording pass-by noise with a microphone located at 7.5m from the track. It allows a characterization of an isolated source level spectrum. However, as soon as several sources are close together, the contribution of each physical source cannot be separated. Since it does not locate equivalent point source it does not provide directivity patterns. These short-comings are partially resolved in the auralization de-dopplerization methodology. Results showing time signals for an auralization of an original recording and a scenario including a small screen at track-side are presented in the following figures. The train length is 55 m and estimated to travel at 90 km/h. The noise level at the receiver is estimated at L_{AFmax} = 95 dB(A).

Figure 3.2(a) shows the time signal derived from a Regina pass-by recording. Inspection of the figure showing the onset of a pulse-type wave, as the first wagon passes by the microphone, is clearly evident. This is an aerodynamic effect, and evidence of the difficulty in modeling this as a source may be seen in Figure 3.2(b), where an auralization for this situation has been performed.

A small rigid noise screen 0.5 m high and 1.5 m from the track-side was included in the auralization incorporating the Hadden-Pierce diffraction model. A reduced noise level compared to Figure 3.2(a) calculated at 7.5 m from the source, around L_{AFmax} = 85 dB(A), is clearly displayed in Figure 3.2(c). Although not visible here, the high frequency squeal noise has been attenuated significantly.

3.4 Modeling of sound propagation

For the sound propagation modeling, mainly the results from the Nord2000 project and the EU project Harmonoise have been used [7, 15], which give a solid physical basis for sound propagation modeling. Also, effects of refraction can be modeled to some degree of accuracy (i.e. curving of ray paths due to wind or varying temperature with height), whereas within the LISTEN project we have only used models assuming no significant effect of refraction, due to the shorter ranges of propagation of interest here. For these cases, the models of the two projects Harmonoise and Nord2000 are very similar.

For most parts of the sound propagation modeling it is straightforward to implement the reported models in computer code: decay due to distance, air attenuation, ground effect and diffraction by barriers of infinite length [7]. This can be calculated in decibels according to the following formula.

$$L_p = L_W - A_{div} - A_{atm} - A_{excess},$$

where L_p is the received sound pressure level, L_W the power level of the acoustic source (including the directivity) and where the three attenuation terms used here, A_{div}, A_{atm} and A_{excess}, respectively are for distance, air attenuation and excess attenuation, caused by ground, noise barriers, etc.

However, since the reporting of the Harmonoise model is quite concise and focuses on modeling two-dimensional propagation situations, the details for modeling typical three-dimensional situations are not described, e.g. for a noise barrier of finite length. For such a case, sound paths above the barrier can be seen to coexist with those via the side edges.

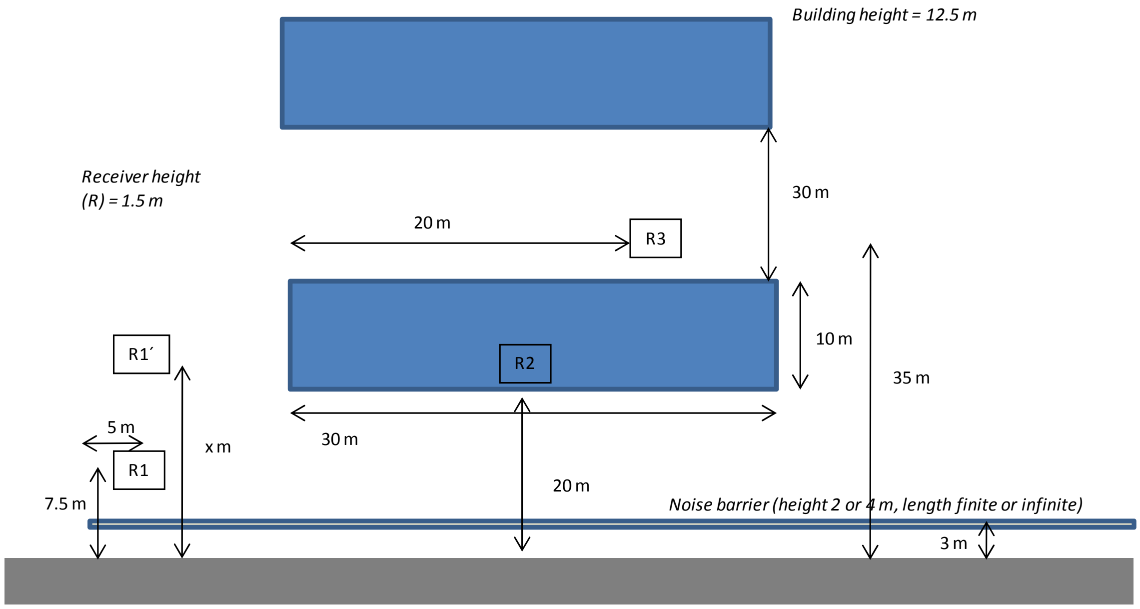

Figure 3.3 Geometry of three scenarios auralized in the LISTEN-project. R1, R2 and R3 refer to receiver positions at Scenario 1, 2 and 3, respectively

For the air attenuation, we have used input values according to: 40 % relative humidity, temperature 24°C, and a static atmospheric pressure of 101.325 hPa. The modeling of dense asphalt ground uses a value of the effective flow resistivity of $2 \cdot 10^8$ Nsm^{-4}. For grass ground, a less simplified approach has been used; according to the two-parameter model of Attenborough for hard warn lawn [16]. For the combined case of two ground types, the asphalt of the road and the grass lawn, the engineering approach of Fresnel weighting is used, where the Fresnel weights depend on the sound frequency and on the point of reflection of sound rays in relation to the position of the ground impedance jump [7]. The ground effect (in dB) is first calculated for both ground types individually and then combined as a weighted average, using the Fresnel weights. Concerning the frequency range, the Harmonoise model uses third-octave bands from 25 Hz to 10 kHz. In our modeling we have extended the range to include also higher audible frequencies, by using third-octave bands from 25 Hz to 20 kHz.

3.5 Modeling Scenario 1

Scenario 1 involves flat ground and an optional noise barrier that may be of finite length (see Figure 3.3). The basics of the ground modeling are described above, including the modeling for the case of mixed ground (asphalt and grass). The effect of the noise barrier may be calculated using a more precise model than the one of the Harmonoise method.

To maximize the acoustic performance of noise barriers, it is important to have accurate prediction schemes to calculate sound reduction by barriers. In fact, theoretical and experimental studies of wave diffraction by a thin plane have been a subject of interest for more than two centuries. There has been continual interest in the past few decades in studying the attenuation of sound by a barrier. Many theoretical results for the diffraction problem of a spherical wave by a half plane or a wedge have been confirmed by experimental results. The theoretical results are often based on some analytical or empirical formulae. For diffraction based models, the pressure at the receiver is determined by summing the contribution from each diffracted path as it propagates from the source. Figure 3.4 illustrate the different diffraction paths around a barrier present in both two dimensional and three dimensional cases.

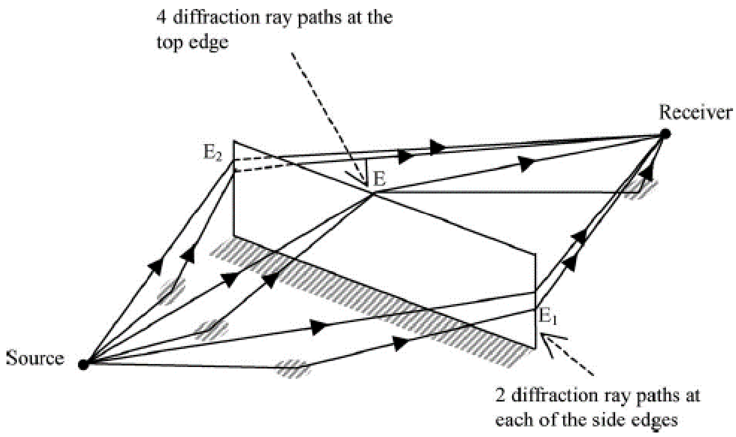

Figure 3.4: Schematic diagram showing the different paths associated with diffraction around a finite screen.

The pressure due to the diffracted path *i* is considered. The function, *G*, is the geometric spreading for the source being considered. The total pressure at the receiver is the sum of the individual paths,

$$p_R = \sum_{i=1}^{n} p_i = \sum_{i=1}^{n} A_i e^{j\psi_i} G(kd_i)$$

where A_i, ψ_i and d_i represent amplitude change, phase shift and path-length from source to receiver of the diffracted wave respectively.

A simple test concerning the effectiveness of a finite screen model in regard to a 3D point source placed symmetrically and a 20 m finite length (red line), 400 m finite length (blue line) and infinite barrier (black line) shown in Figure 3.5. In all cases, the source is 0.5 m from the ground and 7.5 m from the barrier. The receiver point 30 m away from the source and 3 m off the ground. For this test, stark differences are observed between the three curves. The close comparison of the 3D point source in front of an infinite barrier and a finite length (400 m) barrier is not surprising owing to the fact that the contribution of the side diffraction paths is small for the long barrier considered.

The final important aspect of this comparison is evident in the plot related to reduction factors determined by two long “barriers” and the short finite barrier model. The highly oscillating insertion loss values, up to 10 dB, for the short barrier are not realistic. The dips appearing at short frequency intervals are due to phase-differences realized as strong path interference effects. In reality these are smoothed by decorrelation of the reflected source and diffracted paths and the ground reflection smearing factors in real situations. This phenomena has been overcome by frequency averaging and decorrelation described earlier although other aspects related to finite length barriers appear later in Scenario 3. A coherence loss due frequency band averaging is used together with that of uncertainties in source and receiver positions (using 10 % relative variance in the uncertainty of the height of sources, receivers and screens). In addition, a coherence loss due to turbulence is applied (assuming a Gaussian model for temperature fluctuations with a correlation length of 1 m and a normalized temperature variance of $8 \cdot 10^{-5}$, e.g. [17]).

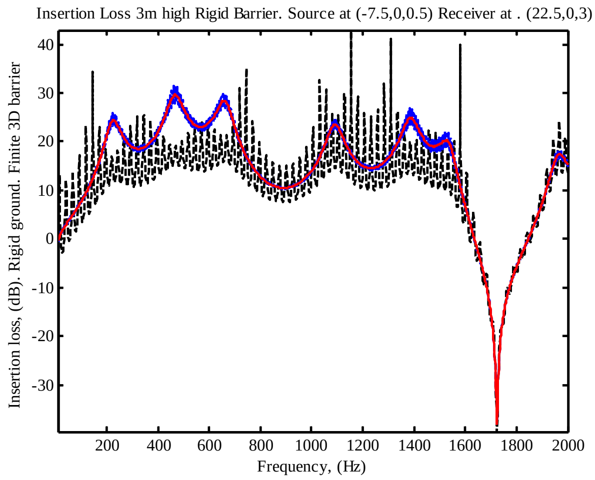

Figure 3.5 Sound pressure insertion loss as a function of frequency, for three different noise barrier lengths. Black line: Infinite length; blue line: 400 m length; red line: 20 m length. Barrier height 3 m, source at 0.5 m height and 7.5 m distance from the barrier, receiver at 3 m height and 40 m distance.

Table 3.1. Reduction indices in third-octave bands for various window configurations.

	2 x 4 mm glass, spacing 6-16 mm, no frame	2 x 3 mm glass, spacing ca 40 mm	2 x 3 mm glass, spacing ca 40 mm	Old 1+2 (isolating unit), leakage in sealant	Reconditioned 1+2. Exchanged to new isolating unit, 2 x 4 mm glass, 12 mm spacing	DOMUS ALU MSEL, VTT/1858, 4+4+6	Window with partial opening (opening of 0.1 m2)
Size (ca.) [m2]	1.6	1.1	3.2	1.1	1.1	1.5	1.0
Rw [dB]	29.0	35.0	34.0	19.0	36.0	48.0	11.0
f [Hz]	R [dB]	R [dB]	R [dB]	R [dB]	R [dB]	R [dB]	R [dB]
25.0	**13.0**	**17.0**	**12.0**	**20.0**	**20.0**	**21.0**	**12.0**
31.5	**14.0**	**17.0**	**13.0**	**20.0**	**20.0**	**21.0**	**12.0**
40.0	**15.0**	**17.0**	**14.0**	**20.0**	**20.0**	**21.0**	**12.0**
50.0	16.0	17.0	**15.0**	20.0	20.0	**22.0**	**12.0**
63.0	17.0	16.0	**16.0**	18.0	18.0	**23.0**	12.0
80.0	18.0	15.0	**17.0**	17.0	18.0	**24.0**	10.0
100.0	19.0	14.0	17.7	16.0	17.0	25.4	8.0
125.0	21.0	12.0	15.8	15.0	16.0	37.0	6.0
160.0	19.0	10.0	17.8	17.0	18.0	32.6	6.0
200.0	18.0	20.0	18.6	19.0	21.0	34.0	7.0
250.0	17.0	26.0	24.5	20.0	24.0	37.3	8.0
315.0	19.0	29.0	26.6	18.0	27.0	42.2	8.0
400.0	22.0	32.0	28.9	17.0	30.0	43.4	8.0
500.0	25.0	35.0	31.7	15.0	34.0	45.4	8.0
630.0	27.0	37.0	33.4	15.0	38.0	47.2	7.0
800.0	32.0	39.0	35.7	18.0	42.0	48.7	6.0
1000.0	35.0	41.0	38.4	20.0	46.0	50.1	5.0
1250.0	35.0	43.0	40.6	20.0	50.0	50.6	6.0
1600.0	36.0	45.0	42.1	20.0	54.0	51.3	8.0
2000.0	37.0	47.0	40.4	20.0	55.0	50.9	10.0
2500.0	35.0	47.0	40.5	21.0	56.0	51.2	12.0
3150.0	32.0	45.0	40.7	22.0	50.0	49.9	14.0
4000.0	31.0	42.0	**39.0**	24.0	45.0	52.7	15.0
5000.0	**30.0**	41.0	**38.0**	26.0	48.0	53.2	**15.0**
6300.0	**29.0**	**40.0**	**37.0**	**26.0**	**51.0**	**54.0**	**15.0**
8000.0	**28.0**	**39.0**	**36.0**	**26.0**	**54.0**	**54.0**	**15.0**
10000.0	**27.0**	**38.0**	**35.0**	**26.0**	**57.0**	**54.0**	**15.0**
12500.0	**26.0**	**37.0**	**34.0**	**26.0**	**57.0**	**54.0**	**15.0**
16000.0	**25.0**	**36.0**	**33.0**	**26.0**	**57.0**	**54.0**	**15.0**
20000.0	**24.0**	**35.0**	**32.0**	**26.0**	**57.0**	**54.0**	**15.0**

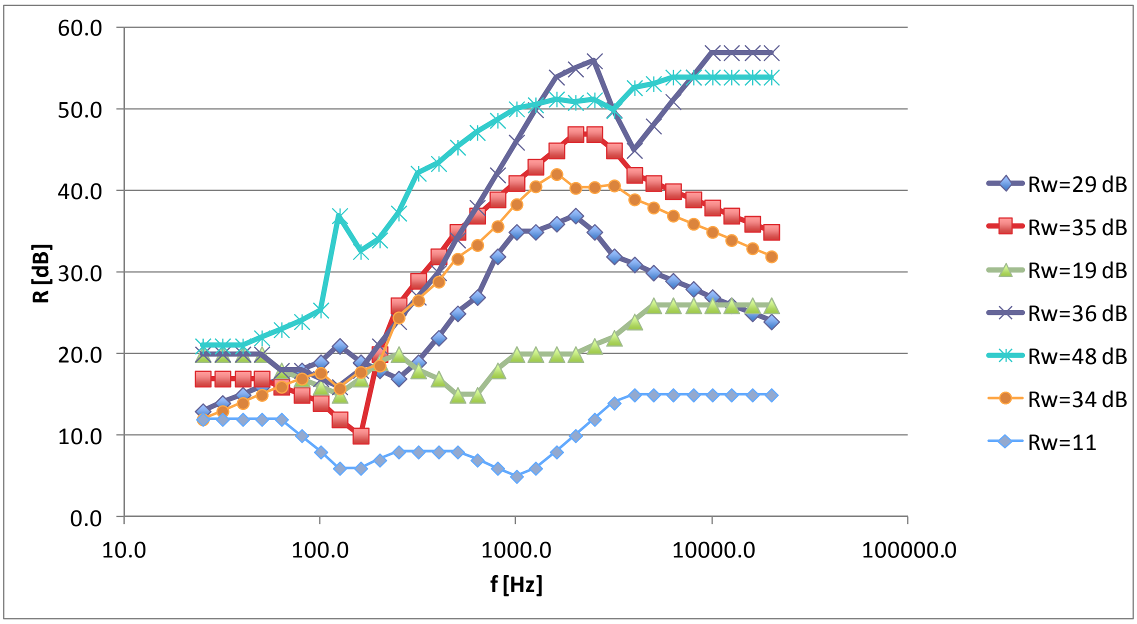

Figure 3.6. Plot of reduction indices corresponding to Table 3.1

3.6 Modeling Scenario 2

Scenario 2 also assumes a flat ground, but here the indoor acoustic environment is auralized. As a first step the same methods as for Scenario 1 are used to calculate the sound propagated toward a window, here however for a higher receiver position and without a noise barrier. As a second step, the indoor sound is calculated using the sound reduction of the window together with a room acoustic model, based on a recorded room impulse response. The reduction index of the window can be chosen by the user of the demonstrator from a small database (see Table 3.1 and Figure 3.6).

3.7 Modeling Scenario 3

In Scenario 3 the basis of the situation is similar to Scenario 1. However, instead of the thin noise barrier, sound shielding is caused by an apartment building of finite length together with reflected sound due to the location of a rear second building (see Figure 3.3, where receiver point R3 represents Scenario 3). The house thus strongly shields the direct path to the receiver from the nearest points on the road. However, further out on the road there is line-of-sight propagation from the vehicles to the receiver, which enhances the overall sound level. An additional contribution is caused by the reflected sound, due to the second building façade.

The sound without involving the reflected paths can be seen to consist of three possible paths: to the left of the building, to the right of the building and via the top (see Figure 3.7). If there is a direct sound contribution, i.e. line-of-sight propagation to the left (or to the right) of the building, the left (or right) sound path is seen to be along that line. The sound path on the other side of the building is seen to go via the side edge of the building. The path of the top contribution is seen to go via a point on the roof closest to the straight line between source and receiver. These three directions are then considered for the head related transfer functions (HRTFs). For the first order reflections (reflected once in the second building, see dashed lines in Figure 3.7), a similar approach is made, however considering an image receiver, located on the other side of the façade of the second building. To simplify, higher order reflections were not modeled. The largest effect of including also higher order reflections would be for shielded source positions, as known from previous work (e.g. [18]). In total, including paths up to first order reflection, gives six different paths.

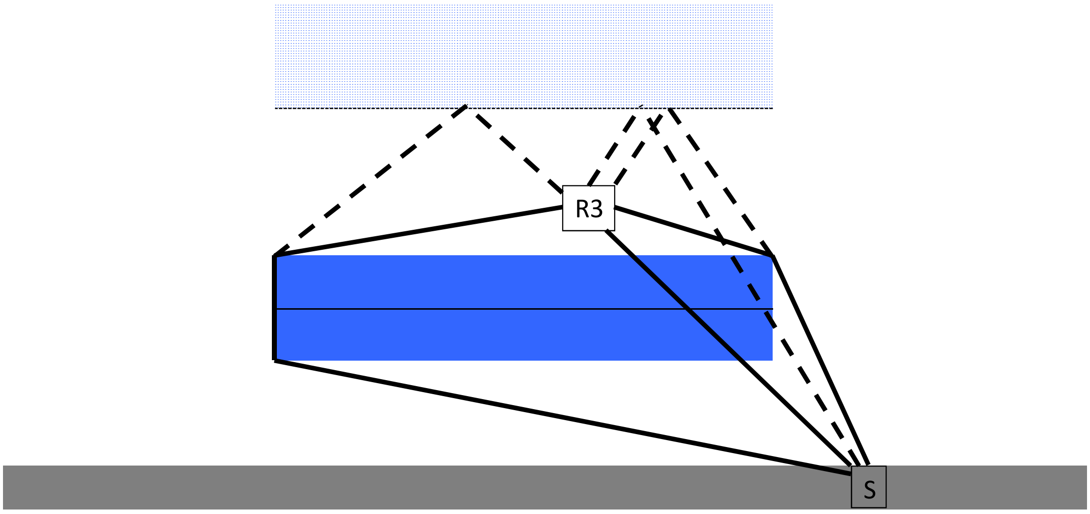

Figure 3.7. Sketch of the sound propagation paths of Scenario 3. S = source, R3 = receiver in Scenario 3. Buildings in blue, road in grey color.

For the modeling of the strengths of the different paths of Scenario 3, the Harmonoise method has been used also for the diffraction. The contributions of the diffraction from the finite edges have been reduced in strength according to Fresnel-zone weighting. All paths are added as uncorrelated contributions, where the reflected paths have been reduced in energy by 5 % to model absorption of typical façades.

3.8 Summary of results acoustic modeling

1) A methodology has been developed to auralize road and rail traffic situations. The methodology includes different ways of modeling the acoustic sources of the vehicles followed by the modeling of the sound propagation effects of air attenuation, ground effect, spherical spreading, Doppler effect and finally the effect of the listener's head and torso in order to create a binaural signal.
2) For the road vehicles the sources are split into tire/road sources and propulsion sources. At first, for the passenger cars (i.e. the light road vehicles), the sound sources were modeled as containing steady state noises and, for the propulsion, steady state tones, both given slowly varying directivity patterns. Later on, the modeling of the tonal character was further developed to approach that of granular synthesis. For the vehicles with more prominent and also more complex sound of the propulsion, a model based on SMS (spectral modeling synthesis) was developed. Tests were made on a diesel passenger car, a medium heavy truck and a bus, showing good results.
3) The methodology developed for road vehicles was extended to railway vehicles. This seemed a natural extension. Source files were provided by industrial partners. The general method following sound propagation effects and inclusion of tonal components determined by the nature of the source (diesel engine) and the contribution of each source type using Harmonoise tabulations. Generally, the RC6 locomotive auralization comprised sixteen source signals and proved to be problematic example for informal and formal listening tests. More favorable results proved to be generated from Regina auralizations although further fundamental research is required.
4) For the sound propagation modeling, a large frequency range was used based on the third-octave bands from 25 Hz to 20 kHz. In the modeling of the ground both hard and soft ground types were allowed (dense asphalt and grass) as well as their combination, for which a Fresnel zone modeling was implemented. In addition, decorrelation effects were implemented, necessary to make the interference pattern more realistic, i.e. how the direct waves interact with the waves reflected in the ground surface. In general, well-based and validated modeling

approaches of large previous projects (Nord2000 and Harmonoise) have been followed in order give a solid theoretical and largely validated basis to the LISTEN approach.

5) Specifically, the developments within the different scenarios, auralization for a noise barrier of finite length is enabled in Scenario 1, transmission through building façade elements and indoor reverberation is developed for Scenario 2 and the effects of a screening building and a second reflecting building façade has been modeled for Scenario 3.

4. Auralization

This section describes the implementation of the LISTEN-demonstrator. The implementation is based on the results from the research on acoustical modeling described in Section 3. The acoustic models were implemented in Matlab which serves to document, test and verify the models.

It was obvious that we could not use the Matlab implementations directly in the demonstrator as its performance were far from real-time. To achieve real-time performance the LISTEN-demonstrator was implemented in Pure Data (PD [19]). It is an open source data-flow based programming environment for audio processing. PD is extendible though a plug-in mechanism which allows new primitives to be added to the language. The implementation language of PD is C and it is used to implement the extensions of PD that where developed in this project. PD was chosen because it is open source and it allows very fast implementation of prototypes which can easily be modified and experimented with. The Graphical User Interface (GUI) of the LISTEN-demonstrator is based on the tool-kit available in PD which is somewhat clinical in presentation. The GUI can easily be replaced in a future version.

4.1 Head related transfer functions

Head Related Transfer Function (HRTF) technology was used to model the acoustic of the receiver. A HRTF describes the effect of the receiver‘s head, ears and torso on the sound [20, 21]. It is the complex frequency response measured at the entrance of the ear-canal (other measurement points are sometimes used) of a sound source at a given direction. HRTF is a frequency-domain representation, head related impulse response (HRIR) is its dual time-domain representation.

An HRIR can be recorded by inserting microphones in a human subject's or artificial head's ear-canals, often at the entrance and with a blocked canal. A loudspeaker at a given position is emitting a measurement signal that is recorded via the microphones. The HRIR can then be extracted from the recorded signal. A large number of directions are sampled where each pair of HRIRs, from left respectively right ear, are representing the direction where the loudspeaker was positioned when the signal was recorded.

To simulate a sound source from a given direction, a pair of HRIRs are selected which represents the direction, the HRIRs are convoluted (filtered) with a mono signal representing the sound of the source [22]. The resulting signal represents the sound pressure at the entrance of the listener’s left and right ear-canals, respectively.

To implement HRTF filtering, a software library was implemented in this project and adapted for the CIPIC HRTF database [23]. The CIPIC database was chosen because of its high quality, dense spatial sampling space and large number of measured subjects including two sets of measurements made on the KEMAR artificial heads with small respectively large pinnae. The CIPIC database contains pairs of HRTFs for 1250 directions, 2500 responses in total. This project used the HRTFs of subject 165 as the standard HRTF set. Subject 165 is the KEMAR artificial head with small pinnae.

As the processing of the HRTFs is very computation intensive it is important that the implementation is very efficient. There are two parts where a careful implementation is most important, the interpolation (explained below) and the convolution with the source signal.

Humans can detect small differences in the direction of a sound source [24]. It would not be practical to sample every detectable direction. Bi-linear interpolation on a sphere centered at the listener's head was therefore used to approximate the intermediate directions that were not contained in the database. For a given direction, the closest four HRTFs on the sphere were selected and a bi-linear interpolation was used to approximate the given direction. There are several ways to implement this [25]. This project used a method where the interpolation was performed in the frequency domain, which assumes that the HRTFs can be approximated by a minimum-phase signal [26].

To be able to apply HRTF filtering on single vehicle pass-bys signal generated in MATLAB, a separate application in PD was developed in the project, and was used in project's listening tests (see Section 5). This application was based on the HRTF-library, and a large part of this code was later reused in the implementation of the direct synthesis method (see below).

4.2 Traffic flow simulation

To create road-traffic noise, a number of single vehicle pass-bys are mixed together. A simple model of traffic flow based on a Poisson process was used to determine the start time of each individual pass-by. The distance in time between vehicles was assumed to be randomly distributed with a modified exponential distribution, where a minimum distance is forced to avoid vehicle collisions. A limitation of this method is that it assumes that vehicles travel independently of each other, which is not the case in certain situations, such as in traffic congestion. We used this simplified approach, because informal listening tests suggested that the simplified method worked well for the scenarios covered by the project plan (free flow traffic). However, the LISTEN-Demonstrator was designed to easy allow future implementations of more advanced traffic flow simulations.

4.3 First approach: Pre-computed synthesis method

Our first approach to traffic flow auralization was based on pre-calculated single vehicle pass-bys. A number of such single pass-bys were mixed together to create the sound of a traffic flow. The rationale for this approach was to save computational power and the fact that we already had the means to compute single vehicle pass-bys by using the Matlab code developed in the acoustic modeling work package (Section 3).

A database consisting of the sound signals of pre-calculated single vehicle pass-bys was created. All the source characteristics and propagation effects were included in the signals such as the Doppler shift, ground reflection, air-attenuation and, if present, the effect of diffraction over a noise barrier. The signals were depending on a set of parameters such as the speed of the vehicle, the perpendicular distance to the path of vehicle and the height of a noise barrier if present.

The database needed to include one pass-by signal for each combination of the parameters. The most important parameters were velocity in the range of 20-140 km/h in steps of 2 km/h which gives 61 different velocities, two distances 7.5 m and 25 m and the height of the noise barrier which is 0 m (no barrier) or 3 m. In total there were 1464 samples in the database which now had reached a total size of 5.6 Gbyte.

4.4 Second approach: Direct synthesis method

We found the first approach to be too limited. The parameter space of the simulations was severely restricted by the size of the database which grew exponentially with the size of the parameter space. Another drawback was that the simulation had to be stopped before allowing any change of parameters. This because new pass-bys signals had to be loaded into the simulator and PD does not allow loading of sound-files while running, it may causes unacceptable interruptions of the audio output stream.

If increasing levels of computation could be done in real-time then the size of the database would decrease accordingly. The obvious solution would be to conduct all computation in real-time. Unfortunately, this is beyond reach with the current computer technology.

Except for the Doppler shift, spherical spreading and HRTF filtering, the acoustic modeling methods used in the project simulated propagation effects using a time-varying 1/3-octave filter-bank. These effects can be combined by simply multiplying the frequency response of each effect together. One separate filter-bank is needed for each distinct source as they may have different propagation paths and source characteristic, for example, a passenger car would need two filter-banks one for the rolling noise source and one for the propulsion source.

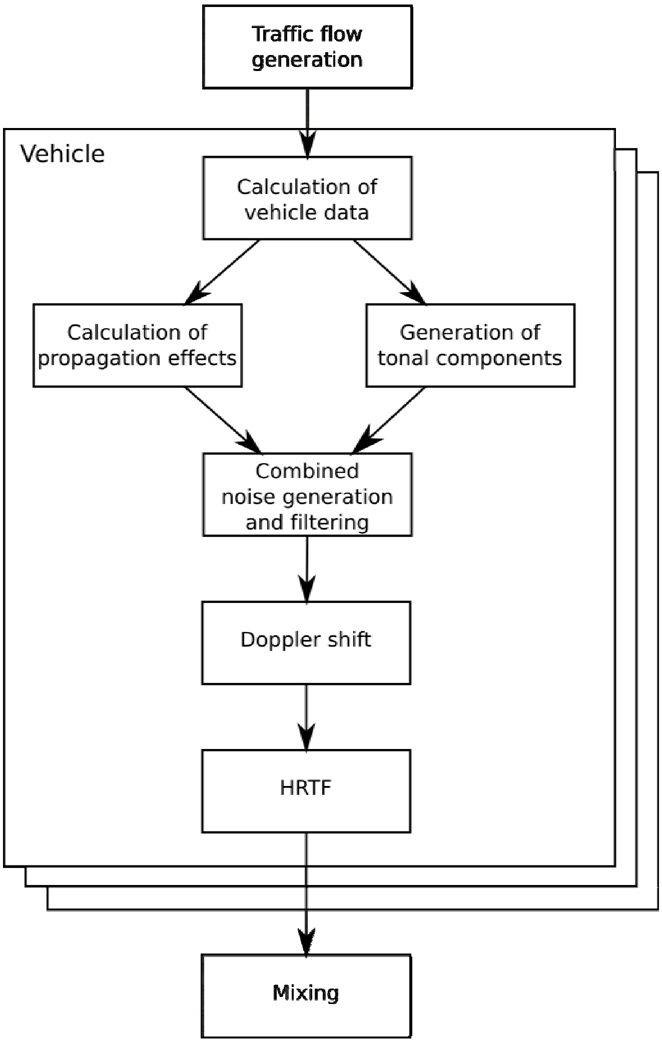

Figure 4.1 Overview of the vehicle synthesis process

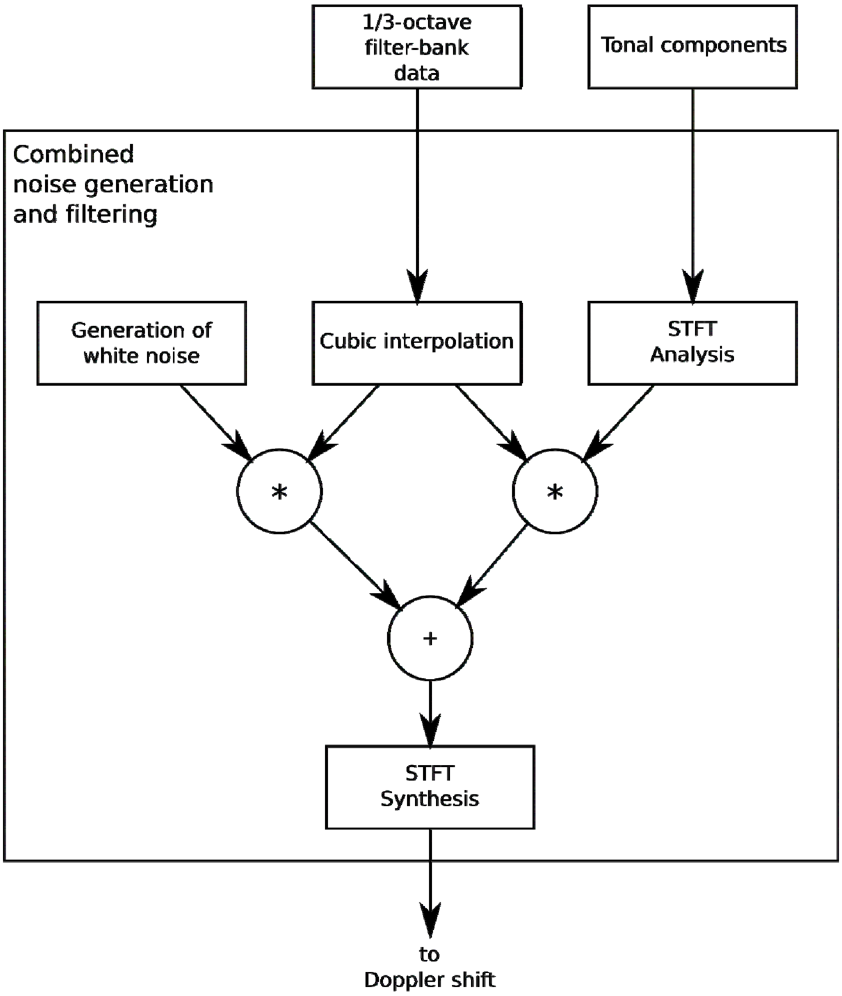

Figure 4.2 The combined noise generation and filtering process

To avoid the large database of signals of vehicle pass-bys from the first approach, the source signals had to be synthesized and filtered in real-time using pre-calculated filter-bank data. The computations of propagation effects were expensive to compute. It is hard to compute both the acoustics of the propagation and to synthesize the source signals at the same time. However the size of the data representing the effects as a time-varying frequency response in 1/3-octave bands was very small. Propagation effects was pre-calculated and stored in tables. It is computationally cheap to recall and combined the effects in real-time and it will not impose the server limitations on the parameter space as in the first approach using the pre-computed synthesis method.

The syntheses process may conceptually be described as follows (see Figure 4.1). As described in Section 3, the source signals of each vehicle was modeled as a stochastic part consisting of colored noise and a deterministic part containing the tonal components. The stochastic part was generated by filtering a white noise signal. The deterministic part, which is assumed to be periodic, was generated by looping a single period of the sound. The combined frequency response of the propagation effects and the source characteristics where applied to the signals by filtering the signals through a time-varying 1/3-octave filter-bank. The Doppler shift was handled separately and was simulated using a time-varying delay-line. In the last step the signals were process by the HRTF filters which was implemented as a block convolution in the frequency domain [27]. At the end, all signals from the vehicle synthesis were mixed together to create the sound of a traffic flow.

Implementing a time-varying filter in the time-domain is difficult and computational expensive as new filter coefficients has to be computed each time the filter is changing. There can also be problems with the stability when changing the filter coefficients. A better and more efficient solution is to generate the noise signal in the frequency-domain. A colored noise signal with a time varying spectrum can be efficiently generated in the frequency domain using a short-time Fourier transform (STFT) synthesis method, in this case the overlap add method (OLA [28]).

The noise generation and the filtering of the stochastic part can be efficiently combined by adding the two components in the frequency domain and using one single STFT process instead of two and thereby saving one STFT synthesis process (see Figure 4.2). The first step was to generate white noise directly in the frequency domain which by definition has constant magnitude and random phase. The magnitude was set to give unit power of the resulting signal and the random phase was generated using a pseudo random number sequence with rectangular distribution, $-\pi <= X(n) < \pi$. In PD, this could be generated by the *noise~* primitive. However, we found that the noise it generated was of unsatisfactory quality. Artifacts could clearly be heard, caused by too high correlation between the numbers of the underlying pseudo random number sequence. A new noise primitive was therefore implemented based on the Mersenne Twister (MT) algorithm [29]. The MT algorithm is able to generate very long pseudo random number sequence of good statistical quality and the risk of artifacts was thereby eliminated. In addition, the new noise primitive was found to be significantly faster than the original PD primitive.

The tonal components were generated in the time-domain using a looped wave-table in PD. There were five different wave-tables, where each one is used for a certain range of velocity. The signal was transformed to the frequency-domain by a STFT analysis section.

The calculation of the propagation effects was performed off-line and store in a database. The instant effect of the sound propagation was computed by recalling data from the database and combined the appropriate propagation effects. The result was given as coefficients of a 1/3-octave filter bank. The STFT may be viewed as a filter bank and thereby used to implement the propagation effect. However, two problems had first to be overcome. First, the STFT uses linear frequency spacing, whereas as a 1/3-octave filter bank uses logarithmic frequency spacing. We therefore used a STFT large enough to handle the resolution of the 1/3-octave filter-bank at low frequencies where the bandwidth is smallest. The frequency response of the STFT was then interpolated to make it smooth using cubic Hermitian splines.

Second, the size of the STFT was too large which caused the time domain signal to get blurred. The solution was to (a) divide the computation in two frequency bands lowering the sampling rate of the low frequency bands and perform separate STFTs (at different sampling rate) and then (b) up-sample the low frequency parts to the original sampling rate and, finally, (c) to combine the results. In the final application, we used a decimating factor of 16, a STFT size of 256 and a sampling rate of 44.1 kHz, which gave a bandwidth of less than 20 Hz at the low frequency end.

The traffic-intensity auralization was limited by the number of vehicles that could be synthesized simultaneously. This limit was found to be too restrictive. To increase it to an acceptable level a memorization mechanism was implemented. The vehicles were classified by type and velocity. The velocity range was divided into slots with a step of 5 km/h. This was found to be a satisfactory resolution in the listening tests performed in the project (see Section 5.3). The first time a vehicle of a particular class was needed it was synthesized and the sound signal was stored in a wave-table. The next time a vehicle of the same class was needed the sound in the wave-table was played instead of been re-computed. Playing a sound from a wave-table is computationally very cheap compared to synthesizing it. The traffic intensity could be increased with at least a factor 3 by using the memorization mechanism. As s consequence of using this method, the parameters of a given vehicle cannot be changed once it has started. The reaction in the sounding result on the user changing a parameter would be delayed. The delay is approximately 8-10 seconds which is the time it takes for a vehicle to run from its start position to the point it passes the listener.

A drawback of choosing PD, is that it is a single threaded application. This means that it cannot use the full power of a modern multi-core computer. However we partly overcome this problem. The interpolation of the HRTFs was one of the most computationally resource consuming process in the LISTEN-demonstrator. By implementing it using multi-threading techniques we could make use of multiple cores and thereby significantly improve the performance of the system. The load balancing is not perfect but the system behaves quite well on the dual-core system (DELL XPS M1730 with Intel Core2 Duo T9300) that the system was implemented on .

4.5 Implementation of Scenario 1-3

Scenario 1 was implemented using the direct synthesis method. The user can control the simulation by a set of parameters: traffic intensity specified as the average number of vehicles per hour, mean velocity, standard deviation of the velocity, distance to the road, screen height in three steps 0 (no screen), 2 and 4, soft or hard ground, four different background sounds and the level of the background. The user can change the parameters while the simulation is running but it take 10-15 seconds before the changes has full effect because the changes are only affects new vehicle pass-bys and not the ones that are already running.

Scenario 2 was implemented using the direct synthesis method. The acoustics of the room was simulated using a recorded room impulse response. All the above parameters was available but with some additional restriction on the ranges. In addition the user could select a window from a set of windows with different reduction index and the degree of openness of the window could be specified.

Scenario 3 was implemented using the pre-computed synthesis method. The geometry of the scenario is fixed. The only parameters the user can control are the traffic-flow parameters, the intensity and the velocity.

4.6 Summary of results auralization

Two different approaches to implementation of the LISTEN-demonstrator were developed and applied in this project, pre-computed synthesis method and a direct synthesis method. The idea behind the first method was to save computational effort at the cost of a large database of the signals from pre-computed synthesis of single vehicle pass-bys. The parameter space would be limited to a number of fixed values defined by the set of signals in the database. The second method was more computationally demanding but it did not suffer from the limitations forced by the size of the database, the parameters could be continuous and could take any (reasonable) value.

The pre-computed synthesis method was found to be unsatisfactory for the Scenario 1 where the parameter space was to limited by the size of database and by restriction on when parameter changes can be effected. It was therefore replaced by the direct synthesis method. This method could also without problem be used in for Scenario 2.

Scenario 3 included many more signal paths from different directions than in Scenario 1 and 2. It was not possible to compute this is real-time using the direct synthesis method. Therefore, the LISTEN-Demonstrator used the pre-computation method for Scenario 3. In the future, it will be possible to use the direct synthesis method also in this scenario. However, this requires personal computers of better performance than today.

5. Perceptual Validation and Optimization

This section describes the perception evaluation part of the project, which included acoustic analyses of auralized sounds as well as listening experiments with subjects. The work was led by Stockholm University, which conducted the work together with the Interactive institute and University College of Art, Crafts, and Design.

Perceptual validation and perceptual optimization are two types of perceptual evaluation of auralizations. Perceptual *validation* involves comparisons of real and auralized sounds[3], to assure that auralized sounds are perceptually correct. This means that they should be perceptually indistinguishable from real sounds or, at least, similar with respect to perceptual factors crucial for correct decisions [30, 31]. Perceptual *optimization* involves comparisons of auralizations of different degree of complexity. The goal is to find short-cuts that simplify the calculations without compromising the perceived realism of auralized sounds [32].

The lack of previous research on perceptual validation and optimization of auralizations meant that methodology had to be developed within the project. In the early part of the project, before any auralizations had been produced, we tested different methods for perceptual validation using room-acoustic auralizations (see Section 5.1). The methods were then refined and used in listening tests evaluating different auralization approaches (see Section 5.2). The methods were further refined in the last series of validation experiments, where the project's final auralization approach was evaluated (see below Section 5.4). Methods for perceptual optimization were also developed and tested in a series of experiments (see Section 5.3). Altogether, 14 listening experiments were conducted in the project, involving totally 196 listeners (113 unique individuals, some of which participated in more than one experiment).

5.1 Method Development

A procedure for perceptual validation of auralization was developed in the beginning of the project. At that time, the project had not yet produced auralizations of traffic noise. Therefore, we first tested the procedure on auralizations of room acoustics developed by the Interactive Institute [31]. The procedure were then refined in the following test series with auralizations produced in the project [10, 33]. The general outline of the validation procedure is described below.

The first and obvious step of the validation procedure was an informal listening test, in which the researchers themselves listen to and compared real and auralized sounds. This indicated whether the quality was good enough to proceed to more advanced evaluations (if not, the auralization procedure had to be improved). In the second step, real and auralized sounds were analyzed acoustically, using models of loudness perception. Specifically, we calculated spectrograms displaying specific loudness per frequency region as a function of time. Specific loudness was calculated using Zwicker's method [34]. This method is based on a model of loudness perception that takes into account several factors in the auditory periphery affecting loudness of complex sounds, such as the variation in sensitivity at different frequency regions and masking of sound components in adjacent frequency regions [35]. If spectrograms of real and auralized sounds differed markedly, this would indicate that the auralization procedure needed to be improved. A descent agreement between spectrogram was thus considered a necessary, but not sufficient condition for successful auralization. This because the auditory system may detect details at a much finer scale than captured by the spectrograms. Therefore, the last and most important step of the procedure was a series of listening experiments, focusing on different aspects of perceptual validity.

[3] In this report, "real sound" refers to a binaural signal created from a mono recording of a real sound. The binaural signal was created using the HRTF method developed in the project, which also was applied to the auralized sounds (see Section 4.3).

We identified four aspects of perceptual validity relevant for the auralizations developed in the project, discussed in turn below:

(i) *Discrimination.* A discrimination test is the hardest test of an auralization, because any detectable difference may be used to discriminate between the sounds. It does not require that the listener can hear which sound is real and which is auralized, only that they differ [33]. Note that auralizations that fail the discrimination test may still be perfectly useful. Our project is an example of this, because its goal was to auralize traffic noise that sounded like typical traffic noise, not to create identical copies of specific traffic noise situations.

(ii) *Detection.* We tested realism of auralized sounds in detection tests, where the listener had to detect whether a sound was auralized or real. We used several methods to test this, including pair-wise comparison of real and auralized sounds ("Which one was auralized?" [10, 31]), presentation of single auralized or real sounds ("Was this sound auralized?", [33]), and sorting of sounds ("Sort these sounds in two groups, one with auralized and one with real sounds", [33]).

(iii) *Similarity.* Auralized and real sound should of course be perceived as highly similar. The degree of perceived similarity may be measured using psychoacoustic methods, for instance similarity ratings of pair of sounds or sorting of sounds in groups of high similarity. Perceptual dimensions underlying such similarity data may be revealed using multidimensional statistical analyses [10, 31, 33]. If these analyses separates between real and auralized sounds, this indicates that the real-versus-auralized aspect of the sounds was perceptually salient, which, in turn, indicates that the auralization was unsuccessful.

(iv) *Attributes*. Specific attributes of sounds may be assessed using psychoacoustic methods. In the project, we used pair-wise comparisons to evaluate perceived annoyance and perceived speed ("Which sound was more annoying?", "In which sound was the vehicle moving faster?" [10, 33]). Noise annoyance is a critical attribute for this project, because decisions on noise mitigation methods will depend on the achieved reduction in perceived annoyance. Perceived speed is important for the realism of the auralized traffic noise. In other applications, other attributes may be more important. For example, in room acoustic applications, speech intelligibility may be the most critical attribute (as we measured in the initial test series, [31])

We also developed methods for perceptual optimization, designed to handle the specific issues investigated, as described in Section 5.4 below.

5.2 Evaluation of auralization approaches

The first auralizations produced in the project were tested in a series of experiments. Two approaches to auralization of noise from a car passage were validated: The Time-Domain (TD) and the Additive Synthesis (AS) approach. For each approach, we also tested auralizations with and without added tonal components (for further description of the auralizations, see Section 3.2).

The experimental sounds were five recordings of a car passages at 27, 45, 65, 84 and 106 km/h and four auralizations of each of these recordings (TD and AS method, with and without added tonal components). The recordings were mono recordings of an Opel Astra at 7.5 m distance from the roadside. Head-related transfer functions (HRTF) were applied to both recorded and auralized mono sounds to obtained binaural sounds for the listening tests (for details, see Section 4.3).

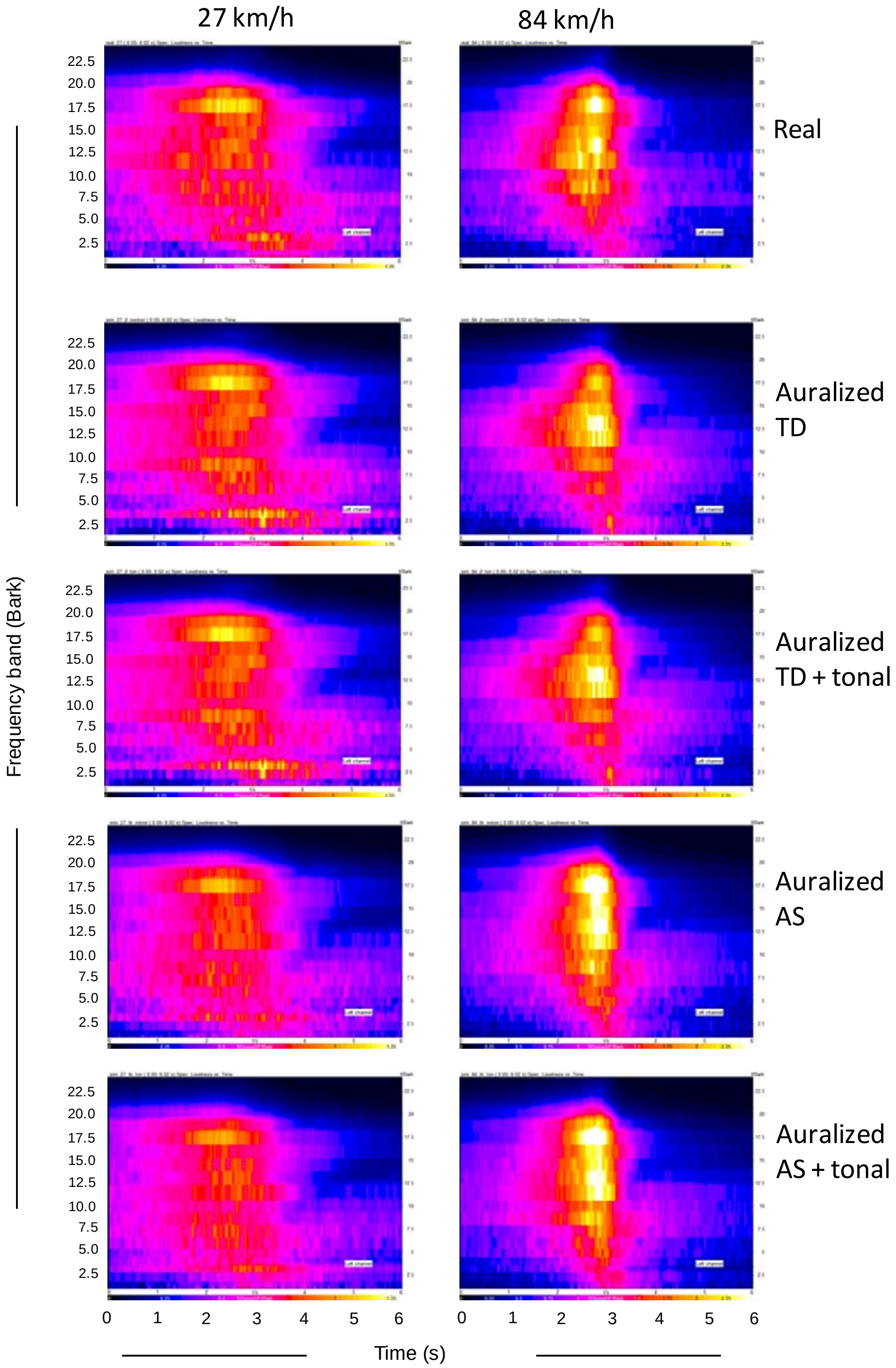

Figure 5.1. Zwicker spectrograms of real and auralized sound from a car passing by at 27, 45, 65, 84 and 106 km/h. Receiver position: 7.5 m from the road side. The spectrograms show frequency region (in Bark) as a function of time (2, 10 and 20 Bark corresponds to 150, 1170, and 5800 Hz). Color indicates degree of loudness (dark to bright colors indicate low to high levels of loudness). For readability, the color scale was normalized for each column of spectrograms. In reality, the 84 km/h passage was louder than the 27 km/h passage.

Figure 5.1 shows spectrograms of real recording and auralization of a car passage at 27 km/h (left) and at 84 km/h. The spectrograms show how specific loudness in different frequency regions evolves over time. Specific loudness, unit sone, refer to calculations according to Zwicker's method [34], and frequency is expressed in critical band rate (unit Bark). The similarity between the top panel (real recording) and the panels below the top panel (auralizations) indicates that all the methods produces auralizations that captured the main aspects of the real recording, such as gross temporal pattern and frequency composition. Therefore, we proceeded to the main part of the validation procedure: the listening experiments.

Three listening experiments were conducted, in which the listeners compared real recordings of the car passage with the auralizations of the same passage. The three experiments were designed to evaluate (1) detection of real and auralized sounds, (2) perceived annoyance and perceived speed of real and auralized sounds, and (3) perceived similarity of real and auralized sounds. Table 5.1 describes the main features of the experiments (for details, see [10]). We did not include a discrimination experiment, because the detection experiment indicated that real and auralized sounds could be discriminated.

In the *first experiment*, pairs of sounds were presented. The pairs consisted of one real recording and one auralization of the same recording, the order of the two sounds varied randomly between presentations. The listener's task was to decide which of the two sounds that was the real recording. If the listener easily could hear which sound was real and which was auralized, the proportion of correct responses would be 1.0. If the listener could not discriminate between the two sounds, the expected proportion of correct responses would be 0.5.

The results showed that the average proportion of correct responses varied between 0.66 and 0.75 for the two methods, TD and AS, with and without tonal components. Thus, performance was better than expected by chance, which shows that listeners could discriminate between a real and an auralized sound. However, the performance was far from perfect, which shows that listeners often mistook auralized for real sounds and vice versa.

Table 5.1. Overview of the listening experiments in the first series of validation experiments.

	1. Perceived realism	2. (a) Perceived annoyance, (b) Perceived velocity	3. Perceived similarity
Experimental sounds	Pair of sounds, each consisting of one real recording of a car pass by and its auralization	Pair of sounds, each consisting of one real recording of a car pass by and its auralization	30 sounds: Real recordings of car passages at five speeds (5), duplicates of the real recordings (5), auralizations with the TD approach, with or without added tonal components (5+5), and auralizations with the AS approach, with or without added tonal components (5+5)
Instruction	”Which of the two sounds was a real recording?”	(a) ”Which of the two sounds was more annoying” (b) ”In which of the two sounds was the car moving faster”	”Sort the sounds into as many or as few groups as you wish in terms of perceived similarity. You are free to rearrange, break, or remake them until you reach an arrangement that is satisfactory to you. The only requirement is that the number of groups is greater than one and smaller than the number of sounds”
Participants	1 man and 11 women (mean age = 28 years)	6 men and 6 women (mean age = 28 years)	2 men and 6 women (mean age = 27 years)

The performance was about equal for the two methods. For the AS-approach, tonal component did not improve the results, whereas the listeners performed slightly better for the TD-approach with tonal components than without tonal components. There was a tendency that the proportion of correct responses decreased with increasing speed, suggesting that auralizations of low speed passages were perceived as less realistic than passages of higher speeds.

In the *second experiment*, pairs of one real and one auralized sound were presented. In one part of the experiment, the listeners assessed which of the two sounds that was more annoying. In the other part, the listeners assessed in which of the sounds the car was perceived as having the higher speed (the order of the two tasks was counterbalance across listeners).

Figure 5.2 shows the proportion of assessments that the auralized sounds was more annoying (left) or moving faster (right) as a function of vehicle speed, separately for each auralization method. The AS-approach (filled symbols) produced auralizations that were assessed as more annoying and perceived as faster than the real recordings (the exception being annoyance of the lowest speed, were proportions were close to 0.5). For the TD-approach, the proportions were closer to 0.5 for most conditions. On average, the auralized sounds were assessed as more annoying in 60 % of the cases for the AS approach and about 50 % of the cases for the TD approach. For perceived speed, the corresponding percentages were 64 and 52 %. There were no systematic effect of tonal components on assessment of annoyance and speed. Overall, the results suggest that the TD-approach produced auralizations that were closer to the real recordings than the AS-approach with respect to perceived annoyance and speed.

In the *third experiment*, the listeners sorted real and auralized sounds using a computer interface developed for free sorting of sounds. The task was to sort the sounds in groups of equal perceived similarity, such that similar sounds were located in the same groups and dissimilar sounds in different groups. As a test of internal consistence, each real sound was included twice.

The sorting data was subjected to Multiple Correspondence Analysis (for sorting data applications of MCA, see [36, 37]). This analysis converts nominal data (group membership) to distances between objects (sounds) in an N-dimensional space. Figure 5.3 shows the two-dimensional MCA solution, with crosses representing real recordings and the other symbols representing auralized sounds. The data form a semi-circular pattern, as often is found in analysis of sorting data. In the present case, it is obvious that the perceived speed was the single dominant perceptual factor behind the sortings, since the sounds are nicely located according to speed along the horseshoe shape.

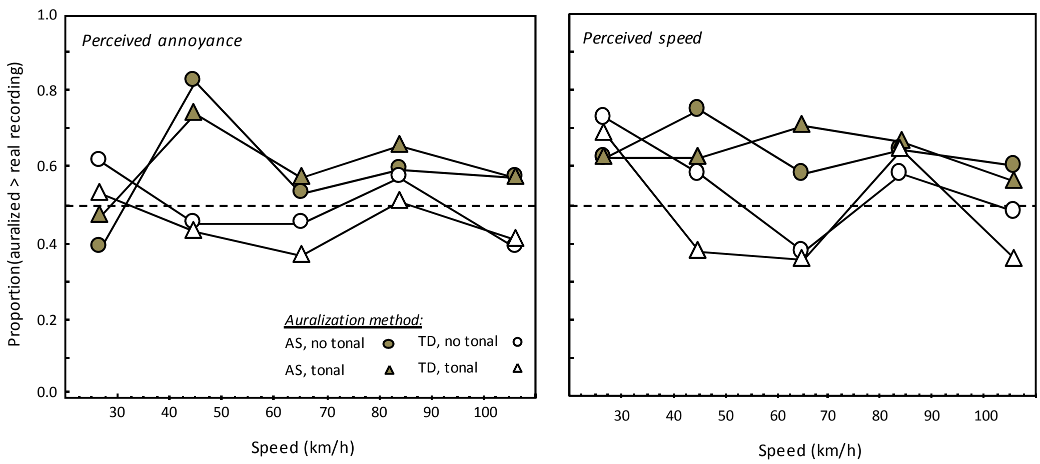

Figure 5.2. Perceived annoyance (left) and perceived speed (right) of real and auralized sound from a car passing by at 27, 45, 65, 84 and 106 km/h. Receiver position: 7.5 m from the road side.

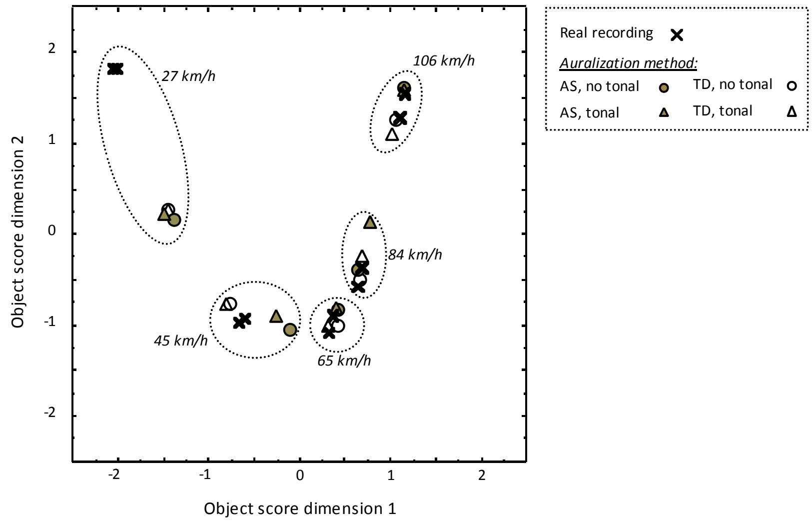

Figure 5.3. Two dimensional solution of a Multiple Correspondence Analysis of similarity sortings of real and auralized sounds from a car passing by at 27, 45, 65, 84 and 106 km/h. Receiver position: 7.5 m from the roadside.

The greatest distance between real and auralized sounds was found for the lowest speed, 27 km/h (crosses located far from symbols representing auralized sounds). For the other speeds, real and auralized sounds were found close to each other. In fact, for speeds at 65 km/h or higher, the distance between auralized and real sounds was not much greater than the distance between the real sound and its replica (the two crosses), indicating a great similarity between auralized and real sounds. An exception was the auralization according to the AS-approach with tonal components (filled triangle) at 84 km/h, which seems to deviate slightly from the other sounds at the same speed.

In summary, the experiments showed good results for both auralization methods. The TD approach was slightly better in terms of perceived annoyance and speed, and auralizations with added tonal components were slightly better than without components. In addition, the AS approach is computationally heavier than the TD approach. Further work in the project was therefore based on the TD approach with added tonal components. The auralizations of the lowest speed, 27 km/h was least successful (see e.g., Fig. 4.3) which suggested that improvements may be archived by improved modeling of tonal components and fine-tuned the balance between propulsion noise and rolling noise.

5.3 Optimization of auralization parameters

The purpose of the experiments described in this section was to find ways to optimize the auralization process. There is no need to auralize aspects of the sound environment that cannot be perceived. Thus, a way to optimizing the auralization process is to determine the degree of detail that is needed, to simplify the problem by short-cuts that will not affect the realism of the auralizations.

Three potential ways of optimizing the auralizations were explored in three experiments. The first experiment assessed speed discrimination of auralized vehicles, to determine if a fixed number of speeds may be used without loss in perceived realisms. The second experiment assessed if realistic source variation might be achieved by varying the ratio if propulsion and rolling noise. This would keep down the number of sources needed to obtain traffic flows. The third experiment assessed whether it was possible to distinguish between traffics flow were all vehicles travel in the same lane and a travel flow where traffic in different directions travel in different lanes. The experimental sounds in three experiments were all created from auralizations of single vehicle passages, using the TD approach with tonal components and the source recordings used in eth first series of experiments. Table 5.1 describes the main features of the experiments (for details, see [32, 38])

Table 5.2. Overview of experiments on perceptual optimization.

	1. Velocity	2. Balance propulsion and rolling noise	3. Distance
Experi-mental sounds	21 auralized car passages (7 m distance), 30 - 70 km/h (steps of 2 km/h)	21 auralized car passages (40 km/h, 6 m distance), with propulsion component varied in 3 dB steps from -18 to +18 dB	Pairs of auralized traffic flows, with average speed 78 km/h, density 6600 vehicles per hour, at either 7 m distance or 25 m distance. One reference sound (R): Traffic flow with vehicles at constant distance (one lane for both directions), and two example sounds: One traffic flow with all vehicles in one lane (as for R) and one with vehicles in two separate lanes.
Instruction	”Rank the sounds with respect to perceived velocity”	”Rank the sounds with respect to perceived realism”	“Which of the example sounds is most similar to the reference sound?
Partici-pants	16 men and 6 women (mean age = 28 years)	16 men and 6 women (mean age = 28 years)	16 men and 6 women (mean age = 28 years)

In the *first experiment*, auralized car passages between 30 and 70 km/h, in steps of 2 km/h were scaled using the Visual Sort and Rate method (VSR, [39]). The method uses a graphic interface, where each sound is represented by an icon. The participant listened to a sound by clicking its icon, and then rated its perceived speed by placing the icon along a vertical line, with endpoints labeled “low speed” (score = 0) and “high speed” (score = 1000). This method allows the listener to compare different sounds as well as rating then according to perceived speed.

Differences in ratings were calculated for pair of sounds of different speeds. The differences in ratings were not statistically significant for pairs of sounds with a difference in speed of 2 km/h (30 & 32, 32 & 34 km/h, etc). The same was true for pair of sounds differing by 4 km/h (30 & 34, 32 & 36 km/h, etc), whereas for 6 km/h difference between sounds, the ratings differed significantly for speed greater than 50 km/h (50 & 56, 52 & 58 km/h, etc). The results of this experiment thus suggest that a difference between vehicle speeds of less than about 5 km/h is difficult to detect. By restricting the auralizations to speeds in 5 km/h intervals, the auralization may be optimized without a loss in perceived realism.

In the *second experiment*, the propulsion noise of thirteen auralized sounds was varied in 3 dB steps, from -18 to +18 dB (the level of the rolling noise was kept constant).The auralized vehicle was driving at 6 m distance to the receiver with a speed of 40 km/h. A low velocity was chosen, because the rolling noise starts to dominate at higher speeds. Perceived realism of the sounds was scaled using the VSR method. The participant listened to a sound by clicking its icon, and then rated its perceived realism by placing the icon along a vertical line, with endpoints labeled “unrealistic” (score = 0) and “realistic” (score = 1000).

Figure 5.4 shows mean ratings as a function of propulsion-rolling noise ratio. The highest degree of realism was found for a ratio of 3 dB (level of propulsion noise 3 dB higher than for rolling nose). High degree of realism was also found for ratios of 0 and -3 dB, which suggests that ratios between -3 and +3 dB may be used for increasing source variability without substantial effects on perceived realism. Note that perceived realism seem to be sensitive to increase in propulsion-rolling noise ratios, as suggested by the sharp decline in realism at ratios greater than +3 dB.

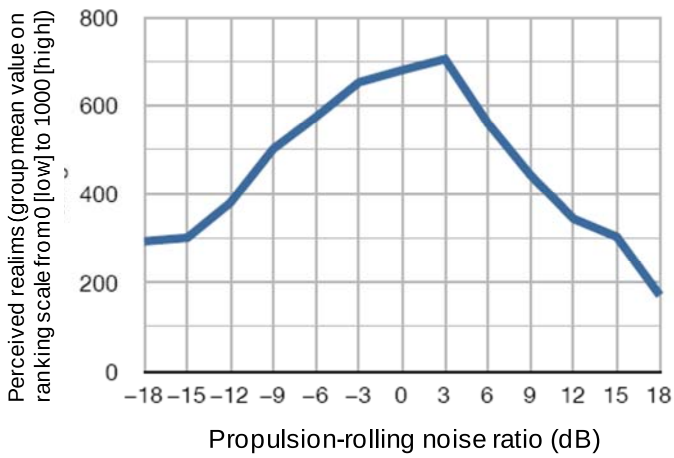

Figure 5.4. Perceived realisms as a function of the propulsion-rolling noise ratio.

In the *third experiment*, auralized traffic flows were used as experimental sounds. The average speed was 78 km/h and the traffic density corresponded to 6600 vehicles per hour. The critical comparison was between auralizations of traffic flows traveling in a single lane (traffic in opposite direction passing through each other) and flows with traffic in two separate lanes. The distance to the source was either 6 m or 23.5 m for the single lane scenario. For the double lane scenario, the corresponding distance was 6 or 23.5 m for the nearest lane and 9 or 26.5 for the second lane.

The method was a two-alternative-forced choice with a reminder (2AFCR). A graphical interface developed in the project was used with three icons representing three 20-s traffic flows. One single lane condition was used as reference (R). Two other conditions, labeled A and B in the interface, were either a single lane or a double lane situations. The listener's task was to decide which of sounds A and B that was most similar to the reference sound R. A 7-13-16 rule was used. This means that a test series was classified as a success if the listeners either answered correctly on the 7 first trials, or answered correctly on 12 out of first 13 trials, or answered correctly on 14 out of the first 16 trials. Otherwise, the test series was classified as a failure. A success means that the listener clearly can hear a difference between sound A and B (performance corresponding to 99 % correct answers, see [40])

The results showed that only four out of 22 listeners passed the test for the closets distance (6 m versus 6&9 m), and only one listener passed the test for the longer distance (23.5 m / 23.5&26.5 m distance).These results suggest that auralizations may be simplified by only simulating a single lane, especially for large distances to the road.

In summary, the experiment showed that optimization may be achieved by limiting the number of modeled speeds, by restricting the traffic flow to a single lane, and by using single source recordings to simulate several different sources through variation of the propulsion to rolling noise ratio.

5.4 Evaluation of final auralization approach

Based on results from the first series of evaluations, and development of tonal component extractions, a final auralization method was developed. This was a refined version of the TD-method with added tonal components (see above Section 3.2, describing the final road auralization method). The purpose of the experiments described in this section was to perceptually validate the auralizations.

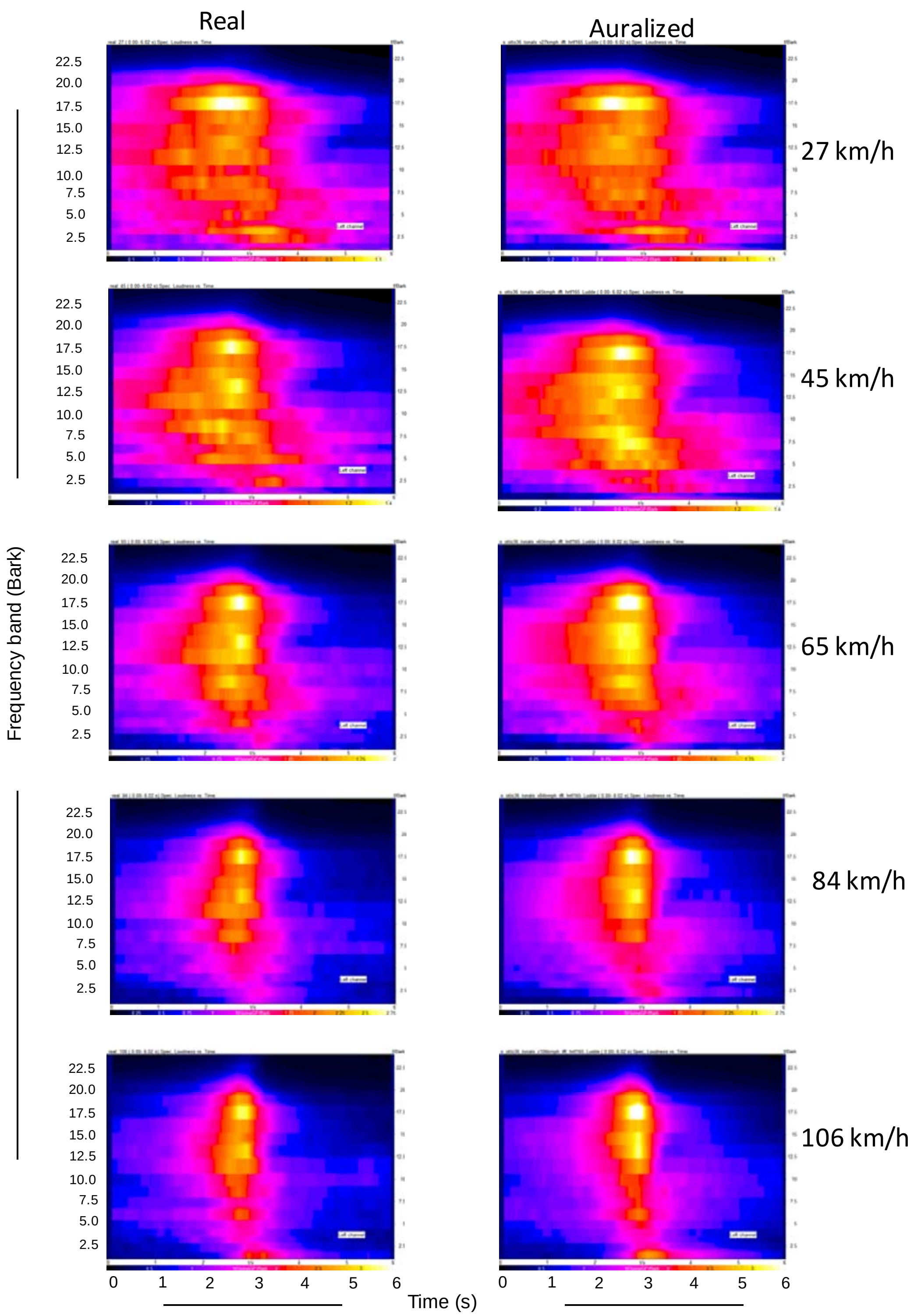

Figure 5.5. Zwicker spectrograms of real and auralized sound from a car passing by at 27, 45, 65, 84 and 106 km/h. Receiver position: 7.5 m from the road side. The spectrograms show frequency region (in Bark) as a function of time (2, 10 and 20 Bark corresponds to 150, 1170, and 5800 Hz). Color indicates degree of loudness (dark to bright colors indicate low to high levels of loudness). For readability, the color scale was normalized for each row of spectrograms. In reality, loudness increased with vehicle speed.

The experimental sounds were the same five recordings as used in the previous experiments, and five auralizations of these using a refined version of the TD-method with added tonal components. Figure 5.5 shows spectrograms for real and auralized sounds (same units as in Figure 5.1). The similarity between spectrograms of real recordings (left panels) and auralizations (right panels) indicates that the auralizations captured the main aspects of the real recording, such as gross temporal pattern and frequency composition.

Four listening experiments were conducted to perceptually evaluate the auralizations. The experiments were designed to evaluate (1) ability to discriminate between real and auralized sounds, (2) detection of real and auralized sounds, (3) perceived similarity of real and auralized sounds, and (4) perceived annoyance and perceived speed of real and auralized sounds. Table 5.3 describes the main features of the experiments (for details, see [33]).

In the *first experiment*, pairs of sounds were presented to the listeners. The sounds consisted of 2-s segments of the car-passage: first, middle or last part. For each combination of segments of a real recording (R) and its auralization (A), four pairs were created: RR, AA, RA, AR. The listeners' task was to decide if the two sounds in each pair were different or identical (same-different method [41]). This method provides a strong test of the quality of auralizations, because the listener only has to be able to hear if two sounds are identical or not.

Table 5.3. Overview of the listening experiments in the second and final series of validation experiments.

	1. Discrimination (same-different)	2. Detection	3. Similarity a) free sorting b) fixed sorting (real vs auralized)	4. Perceived attributes a) Annoyance b) Speed
Experi-mental sounds:	Pair of sounds each consisting of one real recording of a car pass by and its auralization	Single sounds, either real recording or a auralization of a car pass	20 sounds: 5 Real recordings of car passages at five speeds, 5 duplicates of the real recordings, 5 auralizations of the real recordings, 5 duplicates of the auralizations.	Pair of sounds each consisting of one real recording of a car pass by and its auralization
Instruc-tions:	Were the two sounds the same sound or two different sounds?	Was the sound a real recording, yes or no? (Feedback provided: "Right" or "Wrong")	a) Sort the sounds into as many or as few groups as you which in terms of perceived similarity. You are free to rearrange them until you are satisfied with your arrangement. The only requirement is that the number of groups is greater than one. b) The sounds consist of both real and auralized sounds. Sort the sounds into two groups one consisting of sounds that you believe are real recordings and the other consisting of auralizations.	a) Which of the two sounds was more annoying? b) In which of the two sounds was the car moving faster?
Partici-pants:	1 man and 9 women (mean age =27 years)	4 men and 8 women (mean age=27 years)	5 men and 7 women (mean age=28 years)	5 men and 7 women (mean age=26 years)

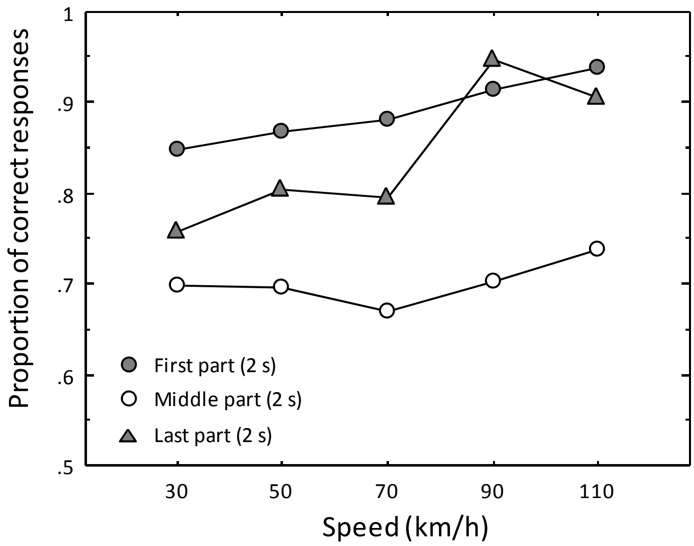

Figure 5.6. Proportion of correct responses in the same-different experiment, in which listeners were asked to decide whether two sounds were identical or not (pair of sounds: real–real, real–auralized, auralized–real, or auralized–auralized). Experimental sounds included real and auralized sound from a car passing by at 27, 45, 65, 84 and 106 km/h Results presented separately for sounds taken from the first, middle or last part of the 6-s vehicle pass by.

Figure 5.6 shows that the proportion of correct responses, p(c) was lower for the middle part of the car passage, that is, during the period when the car was closest to the receiver (middle part), than for the parts when the vehicle was approaching or receding (first and last part). This difference may be explained by temporal variations in the approaching or receding part of the sounds that was difficult auralize perfectly. Overall, the experiment showed that it was possible to discriminate between real and auralizes sounds, especially for the first and last part of the vehicle pass by. However, performance was not perfect despite the use of a very sensitive method (same-different discrimination), which shows that listeners quiet often could not discriminate between real and auralized parts of the vehicle pass by.

In the *second experiment*, single sounds (6-s car passage) were presented and the listener's task was to decide whether it was a real recording or an auralized car passage. Feedback was provided after each trial, which helped the listener to identify aspect of the sounds that differed between real and auralized sounds. The average proportion of correct responses did not differ significantly between the five speeds, ranging from 0.75 to 0.83. Thus, despite a better than chance performance, auralized sounds were often mistaken for real recordings and vice versa. This shows that the auralized sounds did not have a distinct perceptual character that could be used to perfectly discriminate them from real recordings.

In the *third experiment*, the listeners sorted real and auralized sounds using a computer interface developed for sorting of sounds. The experiment included a free sorting task and a fixed sorting task. In the free-sorting task, the listeners were asked to sort the sounds in groups based on perceived similarity, such that similar sounds were located in the same groups and dissimilar sounds in different groups. In the fixed-sorting task, the listeners were asked to sort the sound in two groups, one with auralized sound and one with real sounds. In both tasks, each auralized and real sound was included twice to assess the consistency of the sorting.

In the free sorting part of the experiment, the sounds were sorted in 3 to 10 groups. The median number of groups was 5, which is the same as the number of different speeds. This suggests that main perceptual aspect used for the free sortings was the perceived speed of the sounds. This was confirmed by a Multiple Correspondence Analysis of the sorting data. Figure 5.7 shows the two-dimensional MCA solution. The data form a semi-circular pattern, as often is found in analysis of sorting data including our previous study (cf. Figure 5.3). It is obvious from Figure 5.7 that the perceived speed was the dominant perceptual factor behind the sortings, since the sounds are located according to speed along the semi-circle. The analysis did not form separate clusters for auralized and real sounds, which indicates that this property of the sounds was not perceptually salient (this was also true for a three dimensional MCA analysis).

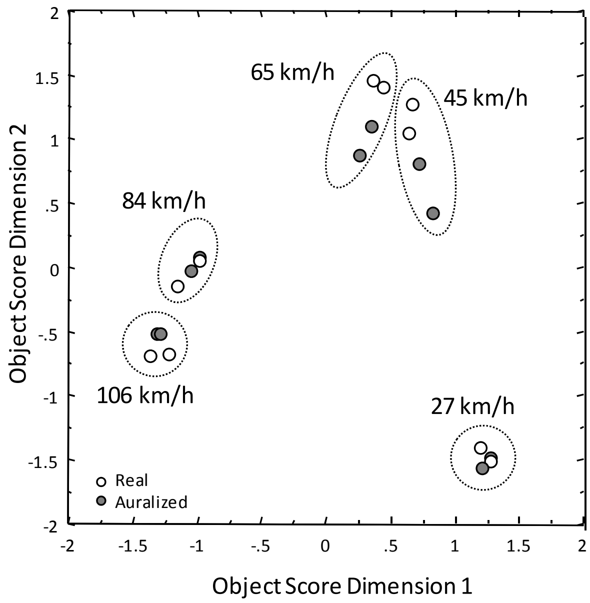

Figure 5.7. Two dimensional solution of a Multiple Correspondence Analysis of similarity sortings of real and auralized sounds from a car passing by at 27, 45, 65, 84 and 106 km/h. Receiver position: 7.5 m from the roadside.

The fixed-sorting data was analyzed by calculating the proportion of real sounds in each of the two groups. For a perfect sorting, these proportions would be 1.0 and 0.0 (one group with only real sounds, and one group with only auralized sounds). If random, both proportions would be close to 0.5. The average proportion was 0.62, with a range from 0.5 to 0.8. Thus, although not random, the performance was far from perfect, again indicating that it was difficult to discriminate between real and auralized sounds.

In the *fourth experiment*, the listeners assessed which of two car passages, a real recording and its auralization, that was more (a) annoying or (b) was perceived as having the higher speed. Figure 5.8 shows the results averaged over the 12 listeners. The left diagram shows the proportion of assessments that the auralized sound was more annoying than the real recording. The right diagram shows the corresponding proportions for perceived speed. For annoyance, the proportions were close to 0.5 (equal annoyance) for the two lowest speeds, and between 0.6 and 0.74 for the three highest speeds. Thus the proportions were not higher than 0.75, the typical criterion for a “just noticeable difference” [42]. For speed, the proportions were close to 0.5 suggesting that real and auralized sounds evoked the same perception of speed. The exception was the lowest speed, where the proportion was higher (0.7), but still below 0.75.

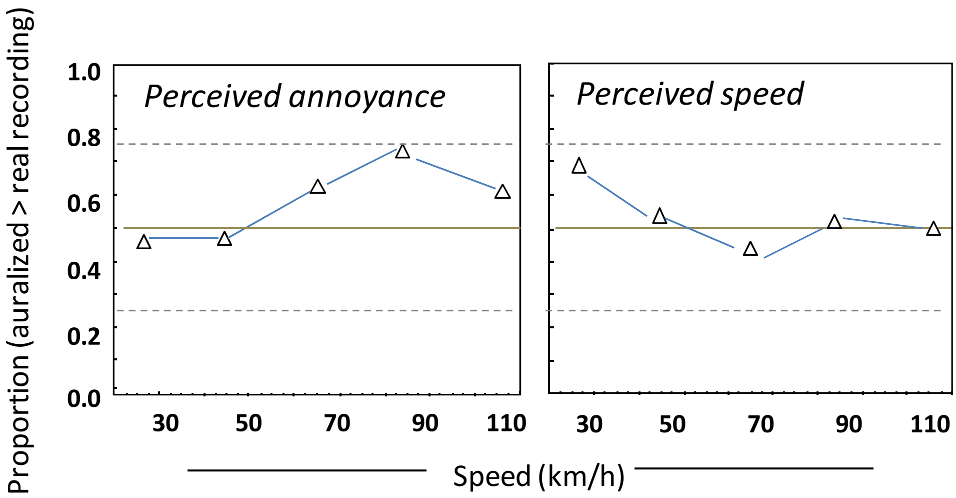

Figure 5.8. Perceived annoyance (left) and perceived speed (right) of real and auralized sound from a car passing by at 27, 45, 65, 84 and 106 km/h. Receiver position: 7.5 m from the roadside.

In summary, Listeners could identify whether a sound was real or auralized more often than expected by chance, but they were not 100 percent accurate, and often close to chance level. The most challenging test, the same-different discriminations, showed that auralizations of the first and last part of the vehicle pass by did not sound identical to the real recording. But also in this task, listeners often performed less than perfect, indicating that real and auralized sounds were highly similar. In addition, auralized and real sounds were found to be about equally annoying and have the same perceived speed, and the sorting experiment suggested that they were perceptually similar in other aspects as well. Overall, the results of this series of experiments provide support for the perceptual validity of the projects approach to auralization of road-traffic noise.

5.5 Railway and barrier auralizations

The railway auralizations were evaluated by informal listening tests and acoustic analyses. The informal listening test revealed that the railway auralizations were perceived as less realistic than the car noise auralizations. The acoustic analyses confirmed this, as spectrograms of calculated loudness (Zwicker's method) were not comparable for real and auralized sounds. As an example, Figure 5.9 shows spectrogram for the Regina train. The overall time-pattern and duration is captured by the auralization. However, there is an obvious discrepancy between real and auralized sound in the middle frequency region (around 10 Bark, which corresponds to about 1.2 kHz) additionally, it is currently unclear how to incorporate the pulse-like sound as the locomotive heading towards and departing from the receiver . The informal listening tests and the acoustic analyses thus suggest that more work would be needed to obtain a fully realistic auralization of railway noise.

Auralizations of the effect of a roadside barrier on perception of car noise were also evaluated by informal listening tests and acoustic analyses. Informal listening tests suggested that the auralization of a car passage fully screened by a barrier had a high degree of realism. This was expected, because the barrier makes the event less distinct and thereby easier to auralize than an unscreened car passage. Auralizations of a car passage partly screened by a barrier were however less realistic, as suggested by informal listening tests and acoustic analyses. Figure 5.10 shows spectrograms of a real and auralized car passage as heard close to the edge of a noise barrier. As can be seen, the overall level of the auralization changes more abruptly than the real recording, as the cars travels form an unscreened to a screened position. This suggests that more work would be needed to obtain a fully realistic auralization of the effect of a finite noise barrier on perception of traffic noise.

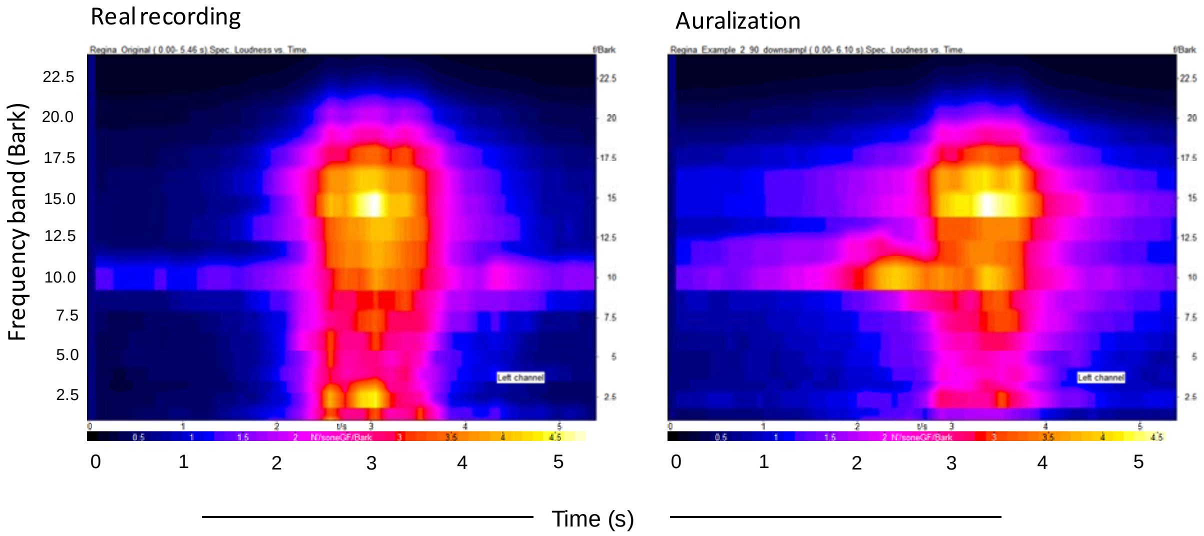

Figure 5.9. Spectrogram of real and auralized sound from a Regina train passing by. The spectrograms show frequency region (in Bark) as a function of time (2, 10 and 20 Bark corresponds to 150, 1170, and 5800 Hz). Color indicates degree of loudness (dark to bright colors indicate low to high levels of loudness).

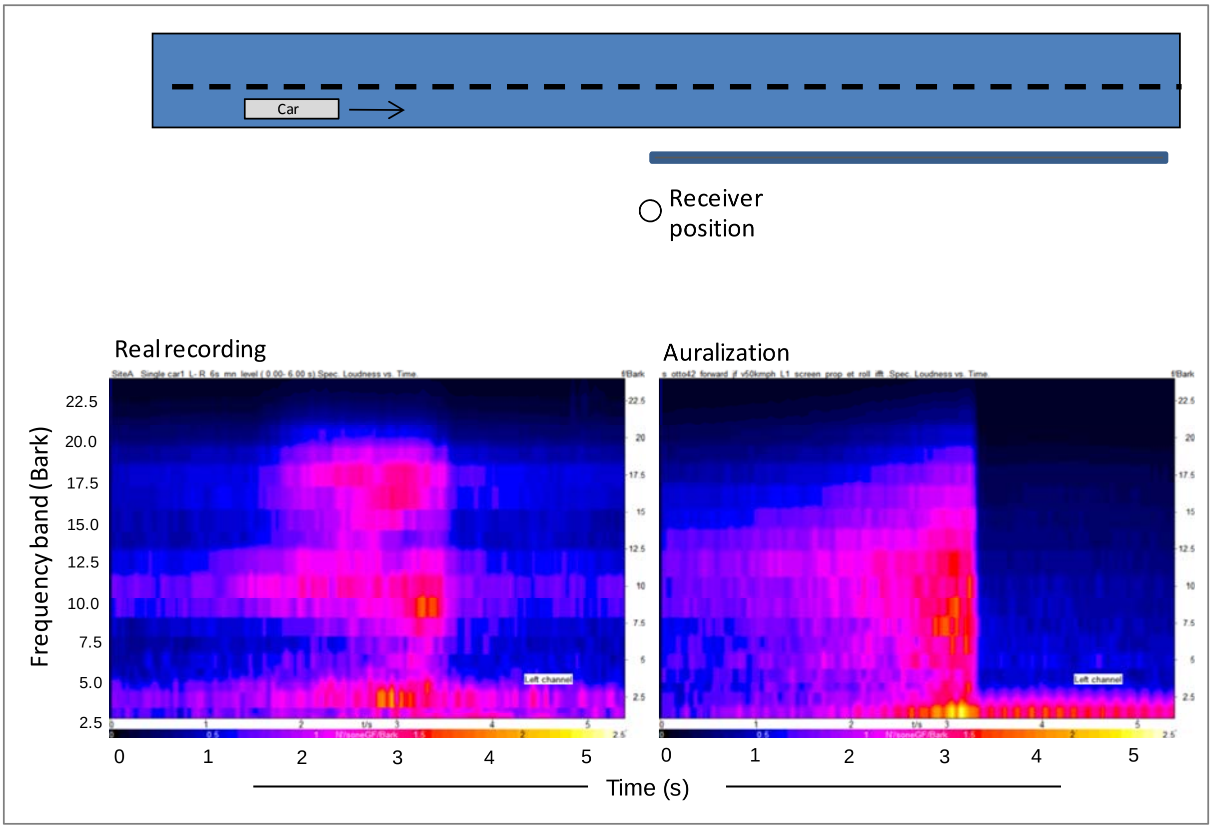

Figure 5.10. Spectrograms of real and auralized sound from a car passing by a barrier edge (see top illustration). Receiver position (R): 7.5 m from roadside. Barrier height: 2.4 m. The spectrograms show frequency region (in Bark) as a function of time (2, 10 and 20 Bark corresponds to 150, 1170, and 5800 Hz). Color indicates degree of loudness (dark to bright colors indicate low to high levels of loudness).

5.6 Summary of results perceptual validation and optimization

1) Methods were developed for perceptual validation and optimization of auralized traffic noise
2) The project's auralizations of single car passages were found to have a high degree of perceptually validity
3) Optimization experiments showed that the project's auralization methodology may be simplified without loss in perceived realism. Specifically, a limited number of different speeds and different distances to the source need to be auralized, and variation in source characteristics may be achieved from a single source by varying the relative level of propulsion and tire noise.
4) Auralizations of railway noise and finite barrier were promising, but informal listening test and acoustic analyses revealed that further improvements are needed before it would be meaningful to proceed to formal listening tests

6. Conclusions and Future Perspectives

In conclusion, the LISTEN-project has provided a demonstration of how auralization may be used for evaluating urban soundscapes, specifically the effect of various environmental factors on the perception on traffic noise.

The LISTEN-Demonstrator allows real-time listening to the effects of road-traffic noise a number of environmental factors, for example distance from listener to road, traffic density, traffic speed, presence of noise barrier between listener and traffic, and effects of different window types for perception of noise indoors.

The project also demonstrated the possibility to auralize more complex situations, such as road-traffic noise behind barriers of finite length and road-traffic behind a building (the shielded side scenario). However, these situations were found to be too complex to achieve real-time auralizations with the performance of today's personal computers. This was also true for railway noise, which was more difficult to auralize than road-traffic noise, mainly due to the larger number of sources that had to be modeled.

The development of the LISTEN-Demonstrator was based on results from the project's research on acoustic modeling, auralization and perceptual evaluation. The main conclusions from this research are:

1. A general methodology was developed for auralization sound from road-traffic and railway vehicles. It included (a) a method for estimating source characteristics of a single vehicle from recordings, (b) models of the effect of distance, ground, air and head-related transfer functions affecting the sounds as it travels from source to the ear canal of the listener, and (c) simulation of traffic flow by combining sounds from several vehicles.

2. A method was developed for auralizing the effect of noise barriers on road-traffic and railway noise. The method included effects of diffraction both at the top of the barrier and at barrier sides. The latter allows for evaluation of the effect near the end of barriers, which may be crucial for planning decisions on the length of barrier.

3. A real-time implementation of the acoustic models was accomplished. This was possible for most of the environmental factors included in the project. The exceptions were the effect of barriers and shielding building with diffraction both at the top and at the side of the structure. The computational demands of these complex situations were found to exceed the performance of today's personal computers. However, it is anticipated that real -time implementation of these complex situations will be possible in the near future as the performance of personal computers is likely to increase with time.

4. A methodology was developed for perceptual validation of auralizations, based on (a) informal listening, (b) acoustic analyses, and (c) formal listening experiments. The methodology was used to test the auralizations of sounds from an automobile at different speeds. The results showed high

agreement between auralized sound and real recordings, supporting the validity of the LISTEN approach to auralization. For auralizations of railway noise and effects of finite barriers, more work remains before the perceptual validity of the auralizations will be as high as for the single road-traffic vehicles.

Further developments of the approach taken in the LISTEN project should include:

(a) Real-time implementation of auralizations of complex situations, with multiple diffraction and reflection paths, was found to be difficult using the present LISTEN-approach. Real-time implementation of such complex situations will require developments that effectively use parallel processing in multi-core computers,

(b) The present approach assumed vehicles of constant speed. A further development would be to also include auralizations of decelerating and accelerating vehicles, for example, as found close to crossing with stop signs. This requires more detailed models of the sources than attempted in the present project.

(c) The present approach to traffic flow simulation assumed that vehicles operate independently of each other. Further developments should include models of traffic flows that may simulate situations where the behavior of individual vehicles depend on the behavior of other vehicles, such as in situations with stop signs or with traffic congestions.

(d) The LISTEN-project did not attempt to add visualizations to its auralizations. However, an obvious development of the LISTEN-work would be to link auralizations to visual images of the environment. This would include visual presentation of the traffic synchronized with auralizations in terms of perceived size (related to distance) of vehicles, their speed and the number of vehicles that passes the visual field of the perceiver.

6.1 SWOT-analysis

A SWOT analysis (Strengths, Weaknesses, Opportunities, Threats) applied to the LISTEN project scientific research as a whole has been undertaken. Although not a quantitative approach, SWOT can aid in assessing the status or potential outcome of a given project as successfully employed in many international and national projects. It must be stated here that all the points listed under SWOT may not be complete, but the information provides an opportunity for an in-depth study.

Table 6.1. SWOT Analysis applied to LISTEN project.

	Strengths	**Weaknesses**
Opportunities	Capability (Referenced personnel – all hold doctorates) Experience (Senior researchers > 10 years each) Knowledge (Forefront of scientific research)	Financial (Limited scope due to budget constraint) Supply Chain (Emerging technologies MATLAB to PD transfer) Continuity (Across all disciplines)
Threats	Technology (HRTFs & PD coding) New markets, vertical horizontal (Commercial opportunities in building and urban space design) Geographical (Nationwide and European opportunities) Business and product development (Scope for development is clear)	Competitors (Commercial software available) Sustaining internal capabilities (Risk in scientific research: new areas. For example train auralization) Loss of key staff (Risk of staff change high: may lead to loss of new thread of research) New technologies (Threats from competing commercial groups)

Strengths related to the LISTEN project are primarily related to the members of the group involved in the project, the practitioners. This is to be expected as the project is based on a specialized application which requires experienced personnel.

External opportunities are strongly linked to the unique features of the more technical parts to the design of the demonstrator. These could have commercial value further down the line. Not only do the software aspects present opportunities in sound design but the fundamental concepts in both physical and psycho-acoustic can add extra layers of understanding to environmental urban design. Opportunities for nationwide and European dimension for the Swedish Transport Agency, WSP acoustics, and Rambøll Denmark and university departments have already been realized.

Weaknesses in the project were limited to budgetary constraints affecting the scope of the project and limiting resources in certain areas and unexpected complexity required in synthesizing railway noise and real-time implementation of the more complex scenarios including finite barriers and shielding buildings. Possibly as a consequence this had a knock-on effect against the complex “supply chain” related to source code and software for which available resources were unable to complete certain tasks. Threads of continuity of railway noise synthesis were not fulfilled since railway noise was not considered realistic for evaluation.

The inevitable threats come from commercial “competitors” as this technology is not new. SNCF (Rail Research) in France are currently working on a railway noise virtual reality software and CadnaA supports traffic noise auralization. However, the level of reliability, traceability and certainty of realistic sounds from vehicle traffic are probably not available in current off-the-shelf software. The purpose of the project was research based and ultimately a demonstrator, not a full working piece of software, was the targeted output. The threat could be end-user acceptance of the original demonstrator software as it may be viewed as not being original.

References

1. European Environment Agency, *Transport at a Crossroads (EAA Report No 3/2009)*. 2009, Copenhagen: European Environment Agency.
2. WHO, *Burden of Disease from Environmental Noise*, ed. L. Fritschi, et al. 2011, Copenhagen: World Health Organization Regional Office for Europe.
3. Kihlman, T., *Sustainable development in an urbanizing world–The noise issue.* Noise News International, 2006. **14**: p. 14-19.
4. Kang, J., *Urban Sound Environment*. 2007, London: Taylor & Francis.
5. Gidlöf-Gunnarsson, A., et al., *Ljudlandskap för bättre hälsa. [Soundscape Support to Health. Final Report edited by A. Gidlöf-Gunnarsson]*. 2008, Gothenburg, Sweden: University of Gothenburg (In Swedish).
6. Nilsson, M.E., M. Andéhn, and P. Lesna, *Evaluating roadside noise barriers using an annoyance-reduction criterion.* Journal of the Acoustical Society of America, 2008. **124**(6): p. 3561-3567.
7. Nota, R., R. Barelds, and D. van Maercke, *Engineering method for road traffic and railway noise after validation and fine-tuning (Technical Report HAR32TR-040922-DGMR20)*. 2005: EC under the Information Society and Technology (IST) Programme.
8. Algazi, V.R., et al., *The CIPIC HRTF Database*, in *2001 IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics*. 2006, Mohonk Mountain House: New Paltz, NY. p. 99-102.
9. Serra, X. and J. Smith, *Spectral modeling synthesis: A sound analysis/synthesis system based on a deterministic plus stochastic decomposition.* Computer Music Journal, 1990. **14**(4): p. 12-24.
10. Forssén, J., et al., *Auralization of traffic noise within the LISTEN project - Preliminary results for passenger car pass-by*, in *Euro Noise 2009*. 2009, EAA: Edinburgh, UK.
11. Kaczmarek, T., *Road-vehicle simulation for psychoacoustic studies*, in *International Congress of Acoustics*. 2007, International Commission for Acoustics: Madrid.
12. Roads, C., *Introduction to granular synthesis.* Computer Music Journal, 1990. **12**(2): p. 11-13.
13. Pendharkar, C., *Auralization of Road Vehicles Using Spectral Modeling Synthesis (M.Sc. Thesis)*. 2011, Gothenburg, Sweden: Chalmers University of Technology.
14. Bongini, E. and E. Bonnet, *Railway noise sources definition within the scope of pass-by sound synthesis*, in *Euro Noise 2009*. 2009, EAA: Edinburgh, UK.
15. Plovsing, B., *Proposal for Nordtest Method: Nord2000 – Prediction of Outdoor Sound Propagation. Nordtest Proposal AV 1106/07*. 2007, Hørsholm, Denmark: Delta.
16. Attenborough, K., *Ground parameter information for propagation modelling.* Journal of the Acoustical Society of America, 1992. **92**: p. 418-427.
17. Ostashev, V.E., et al., *Propagation of sound in a turbulent medium. II. Spherical waves.* . Journal of the Acoustical Society of America, 1997. **102**: p. 2571-2578.
18. Hornikx, M. and J. Forssén, *A scale model study of parallel urban canyons.* Acta Acustica united with Acustica, 2008. **94**(2): p. 265-281.
19. Puckette, M., *Pure Data*, in *Proceedings of the International Computer Music Conference*. 1996, International Computer Music Association: San Francisco. p. 224-227.
20. Begault, D., *3-D Sound for Virtual Reality and Multimedia*. 1994, San Diego, CA: Academic Press.
21. Möller, H., *Fundamentals of binaural technology.* Applied Acoustics, 1992. **36**: p. 171-218.
22. Kajastila, R., et al., *A distributed real-time virtual acoustic rendering system for dynamic geometries*, in *Proceedings of the 122th Audio Engineering Society Convention in Vienna* 2007, Audio Engineering Society: Vienna.

23. Algazi, V.R., et al., *The CIPIC HRTF Database*, in *2001 IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics*. 2001, Mohonk Mountain House: New Paltz, NY. p. 99-102.
24. Blauert, J., *Spatial hearing: The psychophysics of human sound localization*. 1997, Cambridge, MA: MIT Press.
25. Huopaniemi, J. and M. Karjalainen, *Review of digital filter design and implementation methods for 3-D sound,* in *102nd Audio Engineering Society Convention*. 1997, Audio Engineering Society: New York. p. Paper Number 4461.
26. Kulkarni, A., S.K. Isabelle, and H.S. Colburn, *On the minimum-phase approximation of head-related transfer functions,* in *Proc. IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*. 1995, Mohonk Mountain House: New Paltz, NY.
27. Oppenheim, A.V. and R.W. Schaefer, *Digital Signal Processing*. 1975, Englewood Cliffs, NJ: Prentice Hall.
28. Allen, J.E. and L.R. Rabiner, *A unified approach to short- time Fourier analysis and synthesis,* in *Proceedings of the IEEE*. 1977, IEEE: New Paltz, NY. p. 1558-1564.
29. Matsumoto, M. and T. Nishimura, *Mersenne Twister: A 623-dimensionally equidistributed uniform pseudorandom number generator.* ACM Trans. on Modeling and Computer Simulation, 1998. **8**(1): p. 3-30.
30. Vorländer, M., *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*. 2008, Berlin: Springer.
31. Nilsson, M.E., et al., *Perceptual evaluation of a real time auralization tool*, in *Euro Noise 2009*. 2009, EAA: Edinburgh, UK.
32. Lundén, P., et al., *Psychoacoustic evaluation as a tool for optimization in the development of an urban soundscape simulator*, in *Proceedings of the 5th Audio Mostly Conference: A Conference on Interaction with Sound*. 2010, Association for Computing Machinery (ACM): New York, NY.
33. Nilsson, M.E., et al., *Perceptual validation of auralized road traffic noise*, in *Procedings of Inter-Noise 2011*. 2011, INCE: Osaka, Japan.
34. ISO, *Acoustics - Method for Calculating Loudness Level. ISO 532-1975 (E)*. 1975, Geneva, Switzerland: ISO.
35. Zwicker, E. and H. Fastl, *Psychoacoustics: Facts and Models*. 1990, Berlin: Springer Verlag. 354.
36. Abdie, H., et al., *Analyzing assessors and products in sorting tasks: DISTATIS, theory and applications.* Food Quality and Preference, 2007. **18**: p. 627-640.
37. Cadoret, M., S. Le, and J. Pages, *A factorial approach for sorting task data (FAST).* Food Quality and Preference, 2009. **20**: p. 410-417.
38. Gustin, M., *Outdoor Auralization. Evaluation of a Road Simulation Model*, in *Computer Science and Communication*. 2010, KTH Royal Technical University: Stockholm.
39. Granqvist, S., *The Visual Sort and Ratre method for perceptul evaluation in listening testst*. 2003, Stockholm: KTH Royal Institute of Technology (www.speech.kth.se/~svante/Thesis/paperI.pdf).
40. Granqvist, S., *Elektroakustik: Laboration B2, lyssningstest*. 2008, Stockholm: KTH Royal Institute of Technology (www.csc.kth.se/utbildning/kth/kurser/DT2400/ablab.pdf).
41. Macmillan, N.A. and C.D. Creelman, *Detection theory: A user's guide*. 2 ed. 2005, London: Lawrence Erlbaum Associates.
42. Gescheider, G.A., *Psychophysics: The Fundamentals*. 3 ed. 1997, London: Lawrence Erlbaum Associates.