DEVELOPMENT OF AN OUTDOOR AURALISATION PROTOTYPE WITH 3D SOUND REPRODUCTION

Erlend Magnus Viggen

Acoustics Research Centre, SINTEF ICT
Trondheim, Norway
`erlendmagnus.viggen@sintef.no`

Audun Solvang

Acoustics Research Centre, SINTEF ICT
Trondheim, Norway
`audun.solvang@sintef.no`

Jakob Vennerød

Acoustics Research Centre, SINTEF ICT
Trondheim, Norway
`jakob.vennerod@sintef.no`

Herold Olsen

Acoustics Research Centre, SINTEF ICT
Trondheim, Norway
`herold.olsen@sintef.no`

ABSTRACT

Auralisation of outdoor sound has a strong potential for demonstrating the impact of different community noise scenarios. We describe here the development of an auralisation tool for outdoor noise such as traffic or industry. The tool calculates the sound propagation from source to listener using the Nord2000 model, and represents the sound field at the listener's position using spherical harmonics. Because of this spherical harmonics approach, the sound may be reproduced in various formats, such as headphones, stereo, or surround. Dynamic reproduction in headphones according to the listener's head orientation is also possible through the use of head tracking.

1. INTRODUCTION

Noise from transportation and other activities impact many people over large areas. The sources in question (e.g. airplanes and motor vehicles) are often quite noisy, frequent, and cover large distances. For this reason, authorities throughout the world require computations of community noise in order to determine how many people and which areas are impacted. The computations use methods such as [1, 2] and result in *noise maps*, where noise levels are given as equivalent sound pressure levels in dB. A widely used such measure is the day-evening-night equivalent sound level L_{den}, where penalties are added to evening and night-time noise to reflect its additional impact.

However, from these numbers it is very difficult to get an intuitive understanding of the experienced impact of a particular noise situation, and next to impossible to get such an understanding for non-acousticians, such as the people making decisions that cause community noise, and the people in the affected communities. For instance, the short-term but strong noise from overflights may result in the same equivalent levels as the long-term but weaker noise from traffic, even though the two situations are radically different.

In order to aid the understanding of a given noise situation, it would be valuable to have a tool for *auralising* the situation. The term auralisation refers to an auditory representation of something; like visualisation, but with hearing instead of vision [3]. Additionally, such a tool could be used in order to auralise the effect of various noise mitigation scenarios, such as the placement and height of noise screens and the use of noise-reducing road surfaces. Letting

noise affected communities listen to various scenarios in this way and make decisions among several options may also help reduce their noise annoyance [4].

The topic of this paper is the development of such an auralisation tool for providing a realistic representation of an outdoor noise situation such as traffic or industry. Our prototype tool is based on near-field recordings on a car, the Nord2000 model for outdoor sound propagation [1, 5, 6], and use of spherical harmonics [7] for spatial audio reproduction.

1.1. Outdoor auralisation

The topic of outdoor auralisation has been less explored than that of indoor auralisation, and these two topics pose different challenges. Outdoors, the distances between source and receiver are typically much longer, so that fully stochastic ray tracing is no longer a viable option due to the large number of rays required to cover faraway regions with sufficient density. As diffraction over and around screens and buildings must be well-handled, diffraction is a non-negligible effect. The number of relevant reflections is typically far lower outdoors than indoors, since the sound is not trapped in an enclosed geometry as it is in a closed room. On the other hand, due to complex ground effects, ground reflections in long-range sound propagation cannot be treated as simply as reflections typically are treated in indoor ray tracing.

Research on outdoor auralisation in the context of videogames and virtual worlds has been done by the GAMMA group at the University of North Carolina at Chapel Hill. In such cases, the geometry can often be decomposed into an open space domain and small domains around spatially separated objects of possibly complex shape. For example, Mehra et al. handled scattering and diffraction by various objects through an equivalent source method [8], while Yeh et al. coupled a ray tracing approach for the open space with a wave-based simulation method around the objects [9].

Some related work was done as part of the Swedish project *Listen – Auralization of urban soundscapes*, e.g. [10, 11, 12, 13]. The auralisation used for this project used a mix of the Nord2000 and Harmonoise [2] propagation models to construct 1/3 octave band filter levels to be applied to a source signal.
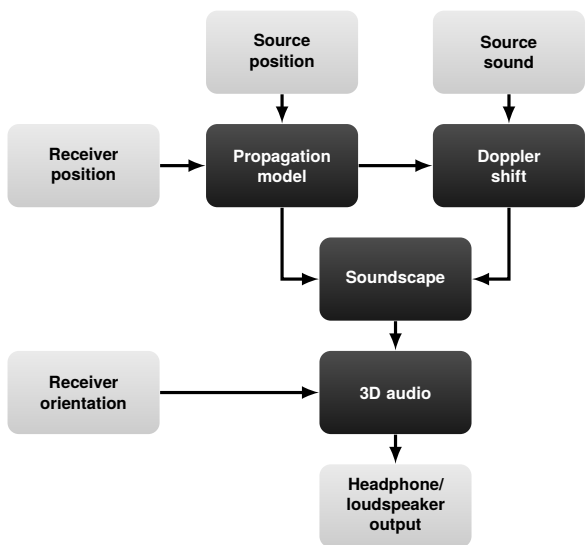
Figure 1: Simple schematic overview of the system, with inputs and outputs in light grey and core components in dark grey

1.2. The Nord2000 model

The Nord2000 model [1, 5, 6] is one of the world's most advanced comprehensive models for computing outdoor sound propagation. It compiles knowledge on many different aspects of outdoor sound, including detailed treatment of ground and structure reflections using the concept of Fresnel zones, the treatment of sound propagation over obstacles or valleys, atmospheric effects, and so forth.

In this model, the possible sound paths from source to receiver are first identified. Then, the sound propagation is subsequently calculated along each path, taking into account the terrain types and building heights along the vertical cross-section of the path. These sound paths may be reflected off buildings and screens, or diffracted around corners. At the receiver point, the various effects are summed to produce 1/3 octave band sound pressure levels. In cases where there are multiple paths from source to receiver, these may be summed fully or partially incoherently.

1.3. Our prototype

We aim to develop an augmented reality tool for an on-site listener to reproduce the noise from nearby virtual noise sources. The listener's head position and orientation may be followed using a positioning system and a head tracking system, and sound may be reproduced using a virtual auditory display (VAD). With a VAD, we combine headphones and head tracking to provide a realistic virtual localisation of the auralised sounds. For instance, if a new road is planned next to a residential area, the tool will allow a listener to walk around the area listening to a natural auralisation of the noise from the future road. The listener may also compare various noise-reducing scenarios.

The usefulness of on-site auralisation is supported by the literature, which indicates that a realistic visual representation of the auralised system is important [14, 15], and that listener mobility [16] and head movement [17, 18, 19] is important for the correct perception of the acoustic situation. With such an augmented reality tool, we would give the user the true visual view of the acoustic space while allowing them up to six degrees

Figure 2: Example of microphone positions for direct vehicle sound measurements

of freedom: three in their head position, and three in their head orientation. It is worth mentioning that the related field of virtual reality has matured rapidly in the last few years; a similar virtual reality tool where both the visual representation of the scene and the auralised sound is updated with the user's head orientation would also be a possibility.

Currently, we are developing an off-line prototype of the auralisation model for traffic noise, where the receiver's position and orientation is specified manually. In this model, possible sound paths from source to receiver are determined and Nord2000 is used separately for each path to determine the sound propagation. Together with noise recordings, these sound paths can be used to determine a spherical harmonics representation of the sound at the listener point. From these spherical harmonics, it is possible to reproduce sound in a number of formats: headphones, stereo, surround, or virtual auditory display. We will specifically come back to the topic of stereo and VAD reproduction in section 2.6.

A simple overview of the system can be seen in Fig. 1. The various components of the system will be covered in the following section, followed by a description of the full prototype system and important lessons learned during development.

2. COMPONENTS OF AURALISATION PROTOTYPE

2.1. Source materials

The source materials were generated from recordings on the front and back of a Ford Mondeo, shown in Fig. 2. As a starting point, represented the vehicle as an omnidirectional source point, simply adding the microphone signals. This simple approach could be improved on by separating the various vehicle sound source types (e.g. engine and wheels) with separate recordings and source
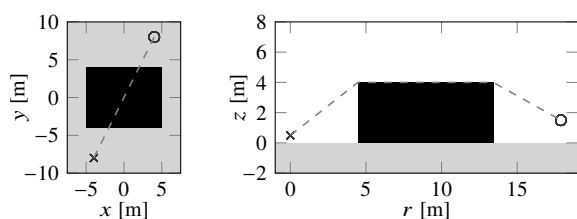
Figure 3: Vertically diffracted sound path (dashed) between source (cross) and receiver (circle), shown in a top-down view (left) and in the vertical cross-section along the path (right). The source and receiver are separated by a building (black).

directivities for each type. While this would require quite some additional work on establishing the different sources and their directivities, it would not be much more computationally heavy if the different source types could still be considered co-located point sources; the propagation model results could be re-used for each source type, each with an additional source directivity filter. However, as the research and implementation work of a more complex source model would be considerable, and as it is not certain that such a level of detail could be perceived in the aforementioned heavy traffic scenario, we retained the simple approach.

Recordings were made at 30, 50 and 80 km/h and the source materials for the velocity closest to the target vehicle velocity for the auralisation were chosen. The recordings are of finite length and in order to create soundscapes of arbitrary duration the recordings were prepared for repeated playback without audible splices. This was achieved by an overlap fade in/out with a fading function that keeps the intensity to unity for uncorrelated signals.

We assume that audible copies of the same sound degrade the fidelity due to the impression of multiple instances of the same car being present. In addition one will also have comb filter effects when adding time shifted correlated signals together. We therefore create a number of excerpts from the recordings. The number is dependent on the traffic density, the velocity of the cars and the length of the road. The number is chosen so that the expected time interval between a repeated sound signal is at least the length of the sound signal itself. We also include a restriction of minimum five excerpts to avoid the risk at low traffic densities of having very few excerpts which could be recognised in the auralisation.

2.2. Propagation model

At the base of the propagation model is a model of the geometry of the area to be simulated, including simple buildings and sound screens with constant height and constant horizontal cross-section, and a map of ground areas of various ground types. Different ground types are characterised by different flow resistivities which lead to different reflection properties. While the prototype currently only supports flat terrain with simple buildings and screens, the Nord2000 model can also handle more general geometries. In this model geometry, a single listener point is also placed at a given horizontal position and height.

We model a road with traffic as a line with evenly spaced source points placed along it. These static source points represent the different positions of a moving vehicle. The traffic is identified by a given vehicle speed and traffic density in passings per hour.

From each source point, possible sound paths to the listener

are found. Straight-line sound paths are simple to find, and from the vertical terrain cross-section along these paths, the Nord2000 model can be used to determine the transfer function as 1/3 octave band levels from the source point to the receiver. In particular, the model uses the ground types and the terrain height profile (including man-made structures) in the vertical cross-section. An example of a vertical terrain cross-section is shown in Fig. 3.

Similarly, the Nord2000 model is used to determine the transfer function along reflected sound paths. The reflected paths are determined using a beam-tracing approach, with similar approaches described in the literature [20]. This beam-tracing is done using a 2D representation of surrounding buildings and noise screens where vertical walls represent objects for specular sound reflection in a horizontal plane.

Any object facing a noise source covers a limited horizontal sector seen from the source. This sector represents a sound propagation beam that will be reflected back from the object, corresponding to the mirror image of the source point. The reflected beam may hit other objects inside its angle, which are handled in a recursive manner. If the object covers a part of the beam, the beam is split into more narrow sub-beams separating between reflected and non-reflected sub-sectors. This builds a beam-tree representing all possible reflection paths down to a wanted maximum order of reflections. The final sound propagation lines (rays) from source points to receiver areas are easy to calculate accurately from this beam-tree. In our case, this technique is very efficient compared to traditional stochastic ray-tracing. One weakness of beam-tracing is that diffuse reflections cannot be determined. However, this does not matter in our case as such diffuse reflections are not accounted for in the Nord2000 model.

Beam-tracing is also useful for assessing reflections for sound propagation from discrete point sources to spatial receiver areas. Assuming reciprocity, it is equally efficient to assess sound propagation from source areas to discrete receiver points. The beam-tracing is simply done in reverse, since the resulting ray path is the same in both directions.

For each source point, each computed straight-line and reflected sound path from source to listener is stored, including its determined source-listener sound propagation time, its source-receiver transfer function as 1/3 octave band levels, and its azimuthal and polar angles from which the sound paths impinge on the listener.

As the Nord2000 model's processing of the sound paths from a source point to the listener is somewhat computationally expensive, a considerable amount of computational time may be saved for road sources by only directly computing a small and evenly spaced selection of the source points. Sound paths are organised by type, and the sound paths for the non-computed source points may be determined by interpolation of the sound paths of the same type from the neighbouring directly computed source points. As examples of sound path types, two sound paths that have been reflected by the same walls in the same order are of the same type, and two sound paths that have been diffracted around the same corner are of the same type.

2.3. Soundscape model

The propagation model serves as the basis of a time-varying multi-path framework for processing the source material. In the next step, we determine the sound that impinges on the listener from each of the possible sound paths. We call this step the soundscape model.

The multipath transfer functions are estimated by a cubic

spline interpolation of the 1/3 octave band gain. We employ DFT-based overlap-add processing where transfer functions are applied corresponding to the vehicle's instantaneous position along its trajectory. The transfer functions are zero-phase, so the source material is pre-processed in order to create the necessary time delays and Doppler shifts.

In this pre-processing, each recording excerpt $p(t)$ is resampled through linear interpolation as

$$p_i(t) = p\left(t - \frac{r_i(t)}{c} + K\right), \tag{1}$$

where $p_i(t)$ is the received sound for sound path type i, $r_i(t)$ is the interpolated instantaneous sound path distance from source to receiver, c is the speed of sound, and K is an arbitrary time shift. As $p(t)$ is only defined for $t > 0$, K is chosen to avoid to avoid negative arguments,

$$K = -\min_{i,t}\left(t - \frac{r_i(t)}{c}\right). \tag{2}$$

Using this approach, each sound path type—the direct sound, each possible reflection type, and each possible diffraction type—gets the correct Doppler shift and relative time delay.

For each overlap-add window the multipath sound signals are encoded in spherical harmonics utilising the angle of incidence of the different sound paths and the relation presented below in (6). Representing the sound field in spherical harmonics allows decoding to arbitrary reproduction formats.

2.4. Spherical harmonics encoding

After the attenuation and angle of each incoming beam has been determined, the audio signals are combined into a spherical harmonics (SH) representation, frequently called Higher Order Ambisonics. The theory of sound field decomposition into a SH format has been extensively covered in the literature [21, 22, 23], so only a short description is included here.

Any sound field $p(r, \theta, \phi, \omega)$ can be decomposed with a modal representation such that

$$p(r, \theta, \phi, \omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} A_n^m(\omega) j_n(kr) Y_n^m(\theta, \phi) \tag{3}$$

where $A_n^m(\omega)$ are the frequency-domain spherical harmonic signals to be determined, $j_n(kr)$ are spherical Bessel functions of the first kind, and $Y_n^m(\phi, \theta)$ are the spherical harmonics [23]

$$Y_n^m(\theta, \phi) = \sqrt{\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!}} P_n^m(\cos\theta)\, e^{im\phi} \tag{4}$$

where $P_n^m(\cos\theta)$ are the associated Legendre functions.

In practice, the series in (3) must be truncated to a finite order N for implementation. The above equation is valid for for interior problems where sources are placed outside the region of interest. Further, a virtual source (e.g. the direct sound from a vehicle or a surface reflection) can be encoded as a plane wave coming from a direction (θ_s, ϕ_s) with spherical harmonics:

$$A_{n,s}^m(\omega) = x_s(\omega) 4\pi i^n Y_n^m(\theta_s, \phi_s)^* \tag{5}$$

where $x_s(\omega)$ is the source signal, including frequency-dependent distance attenuation. The total sound field which is a sum of all virtual sources can be described as:

$$A_n^m(\omega) = 4\pi i^n \sum_s x_s(\omega) Y_m^n(\theta_s, \phi_s)^* \tag{6}$$

The sound field is now represented in a convenient format which can be stored and later reproduced with any 3D audio reproduction system, such as a loudspeaker or headphone Higher Order Ambisonics system, or mixed down to stereo or surround sound systems.

2.5. Traffic modelling

The different single pass-by signals can be used to create an artificial traffic sound signal. The traffic density is taken from the propagation model as passings per hour, and recomputed to a per-second pass-by rate λ.

Traffic pass-bys are modelled in a simple fashion as a Poisson process, so that the probability distribution of the waiting time Δt_{pass} between adjacent pass-bys follows an exponential distribution with probability density

$$P(\Delta t_{\text{pass}}) = \lambda e^{-\lambda \Delta t_{\text{pass}}}. \tag{7}$$

The waiting time is drawn using an inverse transform sampling approach from the inverse cumulative distribution function of the exponential distribution,

$$-\frac{\ln(1-p)}{\lambda} \qquad \text{for } 0 \le p \le 1. \tag{8}$$

When drawing random and uniformly distributed numbers p, this function produces waiting times Δt_{pass} that are exponentially distributed with an average waiting time of $1/\lambda$, as they should be for a Poisson process.

Having drawn a set of waiting times $\Delta t_{\text{pass},n}$ such that the sum of these corresponds to the desired length of the traffic signal (e.g. a minute), the traffic signal can be assembled by summing the signals of individual pass-bys, delayed using the drawn waiting times.

2.6. Audio rendering

2.6.1. Stereo loudspeaker reproduction

The signal from a virtual microphone pointing in a direction given as (θ_m, ϕ_m) can be synthesised using the spherical harmonics representation of the soundscape model described in Sections 2.3 and 2.4 by the following relation

$$p(\theta_m, \phi_m, \omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} b_n A_n^m(\omega) Y_n^m(\theta_m, \phi_m), \tag{9}$$

where b_n dictates the directivity pattern.

The target is x-y stereo cardioid microphone configuration with opening angle of $180°$ that results in the following order-dependent directivity function:

$$b_n = \begin{cases} 1 & \text{for } n = 0, \\ \frac{1}{3} & \text{for } n = 1, \\ 0 & \text{otherwise,} \end{cases} \tag{10}$$

and the left and right loudspeaker signal:

$$p(\theta_l, \phi_l \pm 90°, \omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} b_n A_n^m(\omega) Y_n^m(\theta_l, \phi_l \pm 90°), \tag{11}$$

where (θ_l, ϕ_l) is the listener orientation.

2.6.2. Virtual auditory display

With virtual auditory display, we mean a headphone reproduction system with head-tracking that gives a realistic virtual localisation of the auralised sounds. Such a system is based on modelling the time-varying transfer function from the sound source to the ear drum in real time.

Here, we have implemented a VAD system based on the *mode-matching* approach [24] with virtual loudspeakers. This is done by considering at least $(N+1)^2$ such loudspeakers, evenly distributed on a spherical grid, radiating plane waves towards the origin where listener is placed. If each virtual loudspeaker $l = 1, 2, \ldots, L$ emits a signal $S_l(\omega)$, the sum of plane waves in the SH domain [23] can be expressed as

$$p(r, \theta, \phi, \omega) =$$
$$4\pi \sum_{l=1}^{L} S_l(\omega) \sum_{n=0}^{N} i^n j_n(kr) \sum_{m=-n}^{n} Y_n^m(\theta, \phi) Y_n^m(\theta_l, \phi_l)^* \quad (12)$$

which can again be equated with Equation 3, truncated to a finite order N, yielding [25]

$$\sum_{n=0}^{N} \sum_{m=-n}^{n} A_n^m(\omega) = \sum_{l=1}^{L} S_l(\omega) \sum_{n=0}^{N} \sum_{m=-n}^{n} Y_n^m(\theta_l, \phi_l)^* \quad (13)$$

since the spherical harmonics are orthogonal. This can again be expressed in matrix form

$$\mathbf{A} = \mathbf{SY} \quad (14)$$

where we must find the inverse of $\mathbf{Y}$ to extract the virtual loudspeaker signals. $\mathbf{Y}$ is here a matrix where each row contains complex conjugate spherical harmonics for each virtual loudspeaker angle. Often, the number of virtual loudspeakers will exceed the number of spherical harmonics, $(N+1)^2$, which causes $\mathbf{Y}$ to be non-square and the Moore-Penrose pseudoinverse must be used:

$$\mathbf{S} = \mathbf{DA}, \quad \mathbf{D} = (\mathbf{Y}^T \mathbf{Y})^{-1} \mathbf{Y}^T \quad (15)$$

Note that the number of virtual loudspeakers must be greater than $(N+1)^2$ for $\mathbf{Y}$ to be (pseudo-)invertible.

When the virtual loudspeaker signals have been calculated by multiplying the SH signals with the pseudoinverse of $\mathbf{Y}$, it is straightforward to assign a Head-Related Transfer Function (HRTF) $H(\theta, \phi, \omega)$ to each virtual loudspeaker in order to obtain a binaural format. Mathematically, this is equivalent to converting the HRTF set into a SH-domain representation for each ear,

$$H(\theta, \phi, \omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} H_n^m(\omega) Y_n^m(\theta, \phi), \quad (16)$$

where, in practice, the series must be truncated to a finite order N. The SH representation of the sound field A_n^m must then be filtered with $H_{n,L}^m$ and $H_{n,R}^m$ for the left and right ear, respectively.

In binaural headphone reproduction, head-tracking is essential to obtain a realistic 3D sound experience [19]. Angular input from a head-tracker can be used to rotate the sound field in the opposite direction of the head rotation, stabilising the virtual sound field with respect to the real surroundings. With spherical harmonics, sound field rotation is achievable with Wigner-D weighting [26]. For a given Euler rotation with angles (α, β, γ), a $(N+1)^2 \times (N+1)^2$
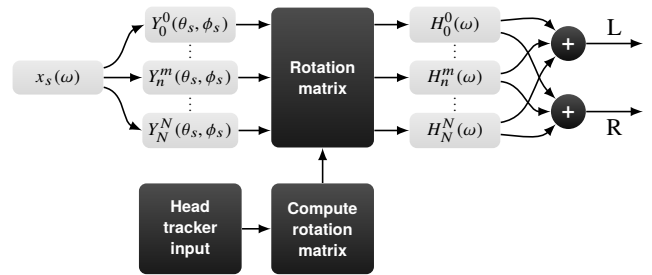

Figure 4: Block diagram of the binaural 3D audio reproduction system. Each SH-domain HRTF has two outputs $H_{n,L}^m$ and $H_{n,R}^m$.

rotation matrix $\mathbf{R}$ is defined. The calculation of this matrix can be calculated with real [27] or complex [28] spherical harmonics, but the details of the calculation are out of the scope of this paper.

Fig. 4 shows how the implementation is done. The block diagram shows only one source $x_s(\omega)$ from one direction (θ_s, ϕ_s), however, in practice all the direct sound and reflection sources must be mixed together after multiplying with the spherical harmonics coefficients. Subsequently, the signals are mixed with the rotation matrix, computed with input from the head-tracker. Finally, each resulting spherical harmonics channel is filtered with the corresponding SH-domain HRTF, and summed for the left and right ear. To reproduce sound for the VAD, we used a SH order $N = 4$ unlike the $N = 1$ order used for stereo reproduction.

2.6.3. High-frequency phase correction

The main advantage with the SH-domain representation is the ability to easily interpolate between measured HRTF angles, scalability and a flexible way of storing or transmitting the 3D sound field data. The main limitation of this approach is the high-frequency spatial aliasing given by the spherical harmonics series truncation. The rule-of-thumb is that near perfect reproduction only occurs for wave numbers $k < N/r$, where N is the truncation order and r is the reproduction radius in the loudspeaker array [29]. With HRTF reproduction, the reproduction radius is the distance from the ears to the centre of the virtual loudspeaker array (normally the centre of the head). This implies that one has to choose a relatively large N to achieve accurate high-frequency reproduction, which results in more computational load.

One possible solution for the high-frequency problem is to reduce the reproduction radius r at high frequencies. Since this radius is determined by the HRTF set, the high-frequency phase response of the HRTFs must be modified to accomplish this. By adjusting the phase such that the effective reproduction radius corresponds to $r_{\text{modified}} = N/k$, the high-frequency reproduction accuracy will improve. As a side effect, the interaural time difference (ITD) at high frequencies will no longer be correct. However, this is negligible since the binaural human auditory system does not rely on ITD at high frequencies, but rather on interaural level difference (ILD). The phase correction approach will improve the ILD reproduction accuracy. Without phase correction, both ITD and ILD will be incorrectly reproduced at high frequencies. This technique is described further and demonstrated in [25].
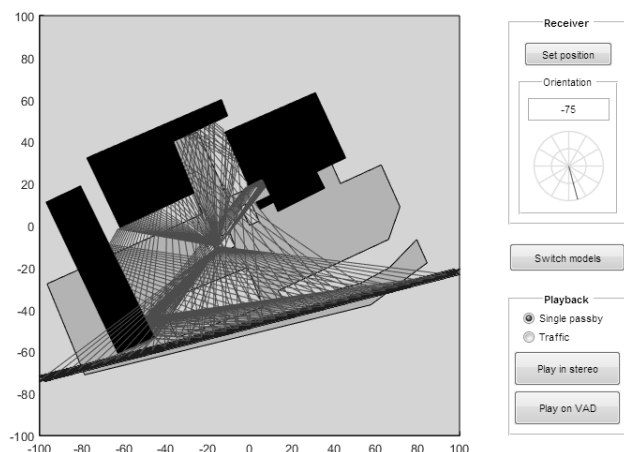
Figure 5: Overview of the prototype GUI for a $200\,\text{m} \times 200\,\text{m}$ model of NTNU's P-25 parking lot with three buildings (black) and two ground types (lighter grey is grass, darker grey is asphalt). Sound paths are drawn from each computed source point (along the bottom black line) to a listener close to the geometry's centre.

3. TOTAL PROTOTYPE SYSTEM

Our prototype is developed as a single-threaded object-oriented MATLAB program. The prototype is presented to the user using a graphical user interface as shown in Fig. 5. The dominant aspect of the interface is a top-down view of the model's geometry, including source positions, listener position, structures, and the different ground areas, thus giving the listener a visual representation of what they are listening to. Additionally, the sound paths from the computed source points to the receiver may be plotted as they are being computed, making the underlying sound propagation computation more transparent to the listener. The listener may then play single pass-by or traffic sounds on stereo loudspeakers (Section 2.6.1) or a VAD (Section 2.6.2).

The user may choose where the virtual listener is positioned in the model geometry. As can be seen from Fig. 1, this requires recomputation of the entire propagation model, and since the other system components depend on this, they must also be recomputed. Such a full recomputation is time-consuming; on the order of half a minute in our implementation.

For stereo reproduction, the orientation of the virtual listener can also be altered. As Fig. 1 shows, this only requires recomputing the rendering from spherical harmonics to stereo, and thus changing the orientation of the listener is a quick process. Using the VAD, the listener's orientation can be changed in real time.

Additionally, the prototype can switch between two variations of the same geometry, in order to facilitate comparison between e.g. two alternative noise-reduction scenarios. If sound signals have already been computed for both geometries, this switch is almost instantaneous.

4. DISCUSSION OF INITIAL RESULTS

As the project is still in development we have not yet started to perform systematic listening tests. However, we have gained some useful insights during the development and internal testing of the prototype.

4.1. Performance

After performing some simple optimisations of our prototype, a full computation like the one shown in Fig. 5 takes on the order of half a minute to complete. This is dependent, however, on the complexity of the model: The required computation time increases with the number of reflections and the number of source points to be computed.

Our prototype has not been explicitly parallelised as MATLAB only supports this through the purchase of an additional toolbox, though we still benefit from MATLAB's parallelisation of some internal functions. However, the computation has a strong potential for parallelisation. In the propagation model, the computation for the various source points are fully independent of each other, and additionally the Nord2000 computation for each identified sound path is independent of the others. This is also the case for the soundscape model; the preprocessing of the direct sound and various reflected sound is independent as well as the overlap-add processing for each time window and sound path. However, the overlap-add processing is dependent on the preprocessing which is dependent on the propagation model.

The computation time is split evenly between the propagation model and the soundscape model. In the propagation model, most of the time is used for preparing and carrying out the Nord2000 calculations for the determined sound paths. Determining the paths themselves and extrapolating the calculated paths takes comparatively very little time. In the soundscape model, most of the time is spent on resampling the different recording excerpts to get the correct propagation delay and Doppler shift for each of the different sound path types.

Currently, our prototype would be too slow to track a listener moving at walking speed in real-time, especially with the $N = 4$ SH order used for VAD reproduction. However, this may be possible with a further optimised and fully parallelised implementation, with an increased amount of computational power. Alternatively, it may be possible to precompute some information for various listener positions in order to avoid having to perform recomputations on-the-fly.

4.2. Fidelity

Generating the source sound material recording the sound of a vehicle using microphones mounted on the vehicle itself is a straightforward and promising alternative to methods described in the literature [10], which are based on pass-by recordings combined with the use of simulated reverse propagation or additive synthesis in order to recreate an artificial source sound. However, our current sample set is limited to only one car. For such a tool to be generally useful for traffic auralisation, additional light and heavy vehicles must be recorded.

In the propagation model, a number of source points were not directly computed; their sound paths were instead interpolated from directly computed source points in their vicinity. While this does not have significant audible effects in an unblocked stretch of road, the interpolation is audible around structure edges with no horizontal diffraction in the model. In a case where all source points are directly computed, the sound from a single moving source would very suddenly become much louder when a virtual vehicle passes a corner so that its straight-line sound path to the listener is no longer blocked. With interpolation, there is instead a smooth artificial transition between the blocked and unblocked sound, essentially, a false diffraction effect.

Including horizontal diffraction in the model is thus important in order to ensure that virtual single vehicles moving behind or from behind a structure sound realistic. However, with no horizontal diffraction the effect is less pronounced in traffic simulations, due to the masking effect of multiple vehicles. Additionally, the false diffraction caused by the interpolation as described above also helps mask the lack of horizontal diffraction.

To begin with, we assumed that a virtual x-y stereo cardioid microphone configuration with opening angle of 90° would be suitable for decoding the spherical harmonics representation to a stereo loudspeaker reproduction. However, informal listening tests showed that the soundscape fidelity was noticeable degraded when the listener was positioned close to the road and facing away. We decided to use an opening angle of 180° instead, i.e. a back-to-back virtual microphone configuration, where the left and right loudspeaker signals correspond to cardioid microphones pointing 90° and −90° relative to the listener orientation, respectively.

Use of the spherical harmonics decomposition means that the soundscape can be auralised using virtual auditory display. In our implementation, this is done by reproducing the sound field with SH-based HRTFs through headphones, requiring head rotation compensation with a head-tracker. This can in theory facilitate perfect sound field reproduction, given a high enough SH truncation order, personalised HRTFs and headphone compensation, as well as a sufficiently low head-tracker latency.

In practice, the spherical harmonics must be truncated to a low order to reduce computational complexity, affecting high frequency reproduction accuracy. In our system we have limited the order to $N = 4$, giving accurate reproduction up to about 2.5 kHz. This can be somewhat compensated for by HRTF phase correction. Personalised HRTFs and headphone compensation is still an open research issue which we seek to investigate further in the future. The end-to-end head-tracker latency has been measured to around 100 ms in the MATLAB version, but should be decreased to < 60 ms in a final implementation [30].

5. CONCLUSION

In this prototype, we have adapted the Nord2000 model for outdoor noise propagation such that we calculate each sound path from the source point to the listener separately from the others instead of combining them. As a result, we find separately the direct, reflected, and diffracted sound paths that impinge on the listener from the sound source, with their 1/3 octave band transfer functions and their incoming azimuthal and polar angles. From this information and appropriate sound source material, we may represent the sound field at the listener point using spherical harmonics. The use of spherical harmonics allows for a wide range of alternatives for 3D sound reproduction: Headphones, stereo, surround, or virtual auditory display.

Conventional stereo reproduction was implemented with two virtual cardioid microphones in a back-to-back configuration, as this gives equal sound levels when the listener is facing towards or away from the road. Virtual auditory display was implemented with spherical harmonics-based HRTFs, and a head-tracker to compensate for head rotation. This gives the listener the ability to determine the direction of sound sources, and a more spatial experience of the soundscape.

As the system described in this paper is a prototype, there are many opportunities for further work. The prototype should be systematically validated, including listening tests and comparing the computed sound levels throughout the geometry with those found by noise mapping tools. Furthermore, source sound materials for a larger number of vehicles should be recorded, including various light and heavy vehicles and perhaps also trains. Synthesising the sound of road vehicles [13] is worth looking into. Realising the full augmented reality system as described in Section 1.3 is also a long-term aim in the continuation of this project.

6. ACKNOWLEDGEMENTS

7. REFERENCES

[1] Birger Plovsing, "Nord2000. Comprehensive Outdoor Sound Propagation Model. Part 1: Propagation in an Atmosphere without Significant Refraction," Tech. Rep. AV 1849/00, DELTA, 2006.

[2] Erik Salomons, Dirk van Maercke, Jerome Defrance, and Foort de Roo, "The Harmonoise sound propagation model," *Acta Acustica United with Acustica*, vol. 97, no. 1, pp. 62–74, 2011.

[3] Mendel Kleiner, Bengt-Inge Dalenbäck, and U. Peter Svensson, "Auralization — an overview," *Journal of the Audio Engineering Society*, vol. 41, no. 11, pp. 861–875, 1993.

[4] Eveline Maris, *The social side of noise annoyance*, Ph.D. thesis, Leiden University, Leiden, 2008.

[5] Birger Plovsing, "Nord2000. Comprehensive Outdoor Sound Propagation Model. Part 2: Propagation in an Atmosphere with Refraction," Tech. Rep. AV 1851/00, DELTA, 2006.

[6] Birger Plovsing, "Proposal for Nordtest Method: Nord2000 — Prediction of Outdoor Sound Propagation," Tech. Rep. AV 1106/07, DELTA, 2010.

[7] Markus Noisternig, Alois Sontacchi, Thomas Musil, and Robert Holdrich, "A 3d Ambisonic Based Binaural Sound Reproduction System," in *Proceedings of the 24th AES International Conference*, Banff, 2003.

[8] Ravish Mehra, Nikunj Raghuvanshi, Lakulish Antani, Anish Chandak, Sean Curtis, and Dinesh Manocha, "Wave-based sound propagation in large open scenes using an equivalent source formulation," *ACM Transactions on Graphics*, vol. 32, no. 2, pp. 19:1–19:13, 2013.

[9] Hengchin Yeh, Ravish Mehra, Zhimin Ren, Lakulish Antani, Dinesh Manocha, and Ming Lin, "Wave-ray coupling for interactive sound propagation in large complex scenes," *ACM Transactions on Graphics*, vol. 32, no. 6, pp. 165:1–165:11, 2013.

[10] Jens Forssén, Tomasz Kaczmarek, Jesper Alvarsson, Peter Lund, and Mats E. Nilsson, "Auralization of traffic noise within the LISTEN project — preliminary results for passenger car pass-by," in *Euronoise 2009*, Edinburgh, Scotland, Oct. 2009.

[11] Peter Lundén, Marja Gustin, Mats E. Nilsson, Jens Forssén, and Björn Hellström, "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, p. 6.

[12] Marja Gustin, *Outdoor Auralization*, Master's thesis, KTH, Stockholm, 2010.

[13] Chinmay Pendharkar, *Auralization of road vehicles using spectral modeling synthesis*, Master's thesis, Chalmers University of Technology, Gothenburg, 2012.

[14] Dick Botteldooren, Bert De Coensel, Timothy Van Renterghem, Luc Dekoninck, and Dominique Gillis, "The urban soundscape: a different perspective," in *Sustainable mobility in Flanders: The livable city*, chapter 8, pp. 177–204. Ghent University, 2008.

[15] Yuliya Smyrnova and Jian Kang, "Determination of perceptual auditory attributes for the auralization of urban soundscapes," *Noise Control Engineering Journal*, vol. 58, no. 5, pp. 508–523, 2010.

[16] Paul Richmond, Yuliya Smyrnova, Steve C. Maddock, and Jian Kang, "Audio-Visual Animation of Urban Space," in *Theory and Practice of Computer Graphics*, 2010, pp. 183–190.

[17] J. Blauert, *Spatial Hearing - Revised Edition: The Psychophysics of Human Sound Localization*, The MIT Press, Cambridge, Massachusetts London England, 1997.

[18] Stephen Perrett and William Noble, "The effect of head rotations on vertical plane sound localization," *The Journal of the Acoustical Society of America*, vol. 102, no. 4, pp. 2325–2332, 1997.

[19] D. R. Begault, E. M. Wenzel, and M. R. Anderson, "Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source," *Journal of the Audio Engineering Society*, vol. 49, no. 10, pp. 904–916, 2001.

[20] Thomas Funkhouser, Ingrid Carlbom, Gary Elko, Gopal Pingali, Mohan Sondhi, and Jim West, "A beam tracing approach to acoustic modeling for interactive virtual environments," in *Proceedings of the 25th annual conference on computer graphics and interactive techniques*. 1998, pp. 21–32, ACM.

[21] J. Daniel, S Moreau, and R. Nicol, "Further investigations of high-order Ambisonics and wavefield synthesis for holophonic sound imaging," in *AES 114th Convention*. 2003, Audio Engineering Society.

[22] B. Rafaely, "Plane-wave decomposition of the sound field on a sphere by spherical convolution," *The Journal of the Acoustical Society of America*, vol. 116, no. 4, pp. 2149–2157, 2004.

[23] E. G. Williams, *Fourier Acoustics: Sound radiation and nearfield acoustical holography*, Academic Press, 1999.

[24] M. A. Poletti, "Three-dimensional surround sound systems based on spherical harmonics," *Journal of the Audio Engineering Society*, vol. 53, no. 11, pp. 1004–1025, 2005.

[25] Jakob Vennerød, *Binaural reproduction of higher order Ambisonics: A real-time implementation and perceptual improvements*, Master's thesis, Norwegian University of Science and Technology (NTNU), Trondheim, 2014.

[26] B. Rafaely and M. Kleider, "Spherical microphone array beam steering using Wigner-D weighting," *IEEE Signal Processing Letters*, vol. 15, pp. 417–420, 2008.

[27] M. A. Blanco, M. Florez, and M. Bermejo, "Evaluation of the rotation matrices in the basis of real spherical harmonics," *Journal of Molecular Structure: THEOCHEM*, vol. 419, no. 1, pp. 19–27, 1997.

[28] C. H. Choi, J. Ivanic, M. S. Gordon, and K. Ruedenberg, "Rapid and stable determination of rotation matrices between spherical harmonics by direct recursion," *The Journal of Chemical Physics*, vol. 111, no. 19, pp. 8825, 1999.

[29] Darren B. Ward and T. D. Abhayapala, "Reproduction of a plane-wave sound field using an array of loudspeakers," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 6, pp. 697–707, 2001.

[30] Douglas Brungart, Alex J. Kordik, and Brian D. Simpson, "Effects of headtracker latency in virtual audio displays," *Journal of the Audio Engineering Society*, vol. 54, no. 1/2, pp. 32–44, 2006.