PROCEEDINGS of the
23rd International Congress on Acoustics
9 to 13 September 2019 in Aachen, Germany

Rendering of scattering effects from finite objects using neural network-controlled parametric digital filters

Ville PULKKI[1,2], U. Peter SVENSSON[3]

[1]Dept Signal Processsing and Acoustics, Aalto University, Finland, Ville.Pulkki@aalto.fi
[2]Hearing systems, DTU, Denmark
[3]Dept. of Electronic Systems, NTNU, Norway, peter.svensson@ntnu.no

Abstract

We are proposing a technique to render the acoustical effect of scattering from finite objects in virtual reality. The effect is implemented using parametric filtering structures and the parameters for the filters are estimated using artificial neural networks. The networks are trained with modelled or measured data. The input data consists of a set of geometric features describing a large amount of source-object-receiver configurations, and the target data consists of the filter parameters computed using measured or modelled data that implement the spectral effect of scattering of the configurations. In a dynamic test scenario with a 3D plate object that reflects and diffracts sound, the approach is shown to provide a similar spectrogram when compared with a reference case, although some spectral differences are present.

Keywords: Virtual reality, audio rendering, machine learning

1 INTRODUCTION

Scattering occurs when a sound wave reaches a finite object, where the presence of the object causes the sound to be redirected to all directions. The directional radiation of scattering depends heavily on the geometry of the object and on the acoustical characteristics of the surface; typically it follows a complex frequency-dependent pattern that can not be described in a straight-forward manner [1].

Acoustical virtual reality aims to provide a listener with the same perception of sound as would occur in a corresponding real scenario [2]. Typical use cases are in acoustical design, computer gaming, and in telepresence. In most cases the target is to produce a dynamic rendering of the virtual world, where the sources, receivers and objects may be moving. To achieve plausible rendering results for such dynamic conditions, the update rate of spatial audio rendering should be relatively high, of the order of 30 Hz.

The rendering of sound scattered from a finite object is challenging in such virtual realities. If the object has a relatively simple geometry, the scattering phenomenon can be computed accurately, although even in such cases the simulation of scattering requires such high computational resources that is is not practical in real use cases. Furthermore, in some cases the geometry is too complex to be simulated, or the acoustical parameters of the object are not known, which makes computational modelling not possible in practice.

A method [15] is described in this article where the scattering effect is rendered using a parametric filter structure, and proposes an efficient method to estimate filter-parameters directly from the geometry of a source-object-receiver scenario using artificial neural networks. The networks are taught using a large number of simulations and/or real measurements where the feature vectors describing the source-object-receiver configurations are associated to parametric descriptions of scattered sound. When rendering a virtual reality, the scattering can then be estimated efficiently giving the perception of high plausibility to a listener, with low computational requirements. The possibility to use real measurements in teaching of the system makes it possible to estimate scattering with acoustically and geometrically complex objects and even with living animals, which opens a large number of potential applications for the approach.

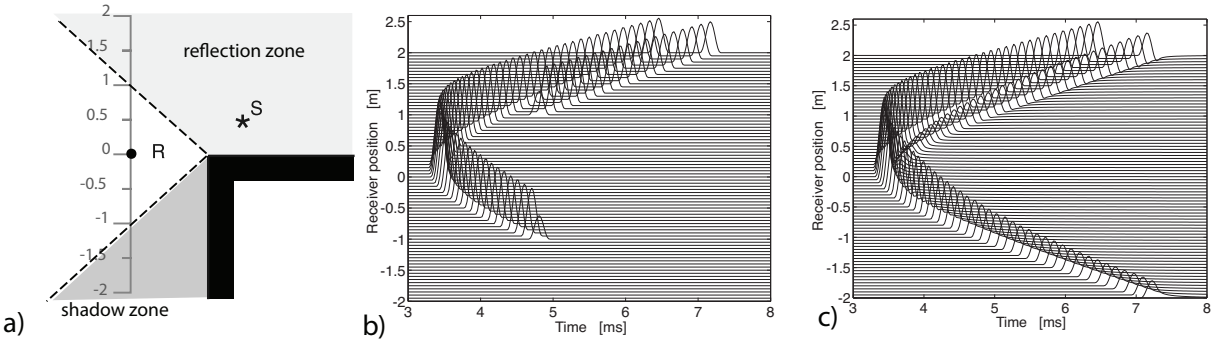

Figure 1. An example of a first-order diffraction wave for an infinite wedge with closed angle of 90°. a) The geometry, with a fixed source, S, and a receiver moving across the zone boundaries. b) The geometrical acoustics (GA) impulse responses, low-pass filtered, for a number of receiver positions, indicating the discontinuities for the direct sound and for the specular reflection. c) After the addition of diffraction waves modelled with the edge source (ES) model, the wavefronts are perfectly smooth with changing receiver position.

2 BACKGROUND

The simulation of sound reaching the avatar, a virtual representative for the user, is typically conducted using geometrical acoustics (GA) modelling as described above. In practice this is done by ray-based rendering of direct and reflected sounds and by adding some diffuse reverberation to the scene [3, 4]. Ray-based acoustics can be implemented with relatively low computational resources. In practice each ray is implemented as a DSP structure, where sound is delayed and attenuated according to its propagation path. Additionally, the acoustical properties of reflecting surfaces along the propagation path, in the form of frequency-dependent absorption, may be implemented with digital filtering. Each ray is then made audible to the listener typically using either head-related transfer function (HRTF) based filtering for headphones or loudspeaker-based techniques [5]. However, the downside of such methods is that surface scattering and diffraction phenomena are are usually neglected.

Edge diffraction effects have been implemented as complements to ray-based models. A simulation method is utilized to estimate a FIR or IIR for each diffracted path [6, 7], and they are treated as secondary sources along edges, which scatter sound to all directions in the virtual room. A number of simplified and computationally less demanding approaches to modelling the scattering from rigid objects have been suggested, which can be implemented in real-time dynamic scenarios [8, 9, 10, 11]; however, they usually suffer from giving uncertain accuracy for lower frequencies and/or scattering objects that are in close proximity to the sources or the receivers in the simulated geometries. This is caused since they typically have to rely on high-frequency asymptotic solutions such as the Kirchhoff approximation, or the Uniform Theory of Diffraction [12].

3 MACHINE-LEARNING BASED RENDERING OF SCATTERING

A method to extend the GA model with object scattering is proposed in [15], as follows. The spectral effect of scattering of sound is rendered using a parametric filter structure, which in practise has low-pass, high-pass, or shelving effects on the output spectrum. The implementation with parametric filters differs from earlier approaches, where generic structures such as warped IIRs were used [6], where the filters were directly designed to fit the frequency response of diffracted sound instead of utilizing parametric filters. The motivation to use parametric filters is that they can be controlled by such meaningful measures as cut-off frequencies in Hz and frequency-specific output amplitude in dB. This is assumed to provide an intuitive and a computationally efficient means to deliver the effect of scattering to the listener.

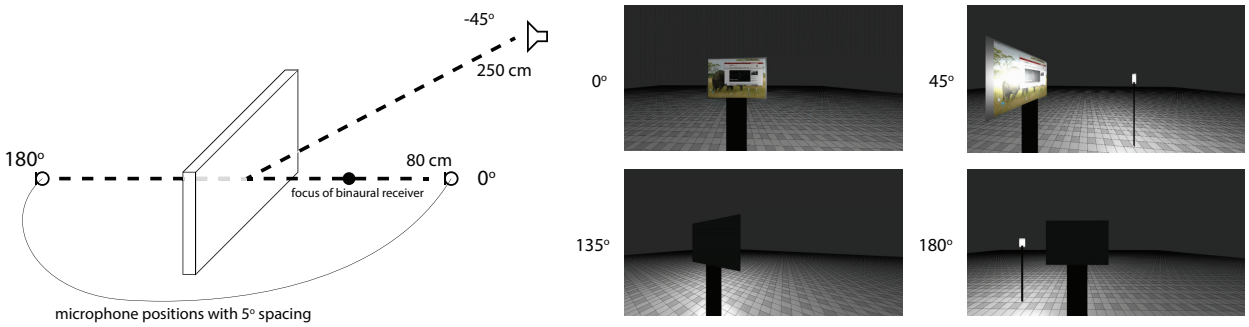

Figure 2. Left: the geometry of the dynamic scene utilized as a test scenario. Both the source and the receiver are located in the horizontal plane, with the source fixed in the direction of -45° and the receiver moving from azimuth angle of 0° to 180°. Right: four snapshots from the corresponding video with annotated viewpoints.

Furthermore, it is assumed that an artificial neural network can learn to associate the geometry of the source-object-receiver setup to filter parameters. For example, in reality, when a receiver approaches an occluding object in shadow, the effect of occlusion is a low-pass effect where the cut-off frequency decreases with decreasing distance between the listener and the object. The machine learning should then associate the angular size of the occluding object from the view point of the listener to the cut-off frequency of corresponding low-pass filter.

In the proposed approach a large set of real measurements or computer simulations of scattering are utilized to teach artificial neural networks. Each network is trained to associate the geometric description of source-object-receiver to a specific filter parameter value set in the teaching stage. The target filter parameters values in teaching are in turn obtained by fitting the parametric filter response to the corresponding measured or simulated data. During the run time of the virtual reality audio engine, the description of the current geometric configuration is then used as an input to the trained networks, which compute the parameters for the filter structure, which then renders the scattering effect.

4 DYNAMIC TEST SCENARIO

A simple test scenario was designed to illustrate the methods and to compare the proposed methods with reference simulations. The scenario is shown in Fig. 2. A 62 cm x 47 cm x 3 cm plate, corresponding to the size of a normal computer screen is located with the center of the front plate in the origin. The source is at 250 cm distance in the horizontal plane, in the direction of -45° azimuth. The receiver, which is either a pressure sensor or a binaural human listener, rotates around the plate from 1° to 181° with 5° steps. The offset of 1° from five-based values was introduced since the physical modelling tool used in the study did not produce a result with the direction of 90° of azimuth. The source was positioned at a relatively long distance to emphasize the effect of the scattered sound component in the presence of direct sound; when the source is further away, the propagation attenuation has a similar range for both direct and scattered sound components. This is opposed to the case where the source is close to the receiver, where the propagation attenuation is much more prominent for scattered sound than for direct sound.

Scattering was simulated with the ES model including 15th-order diffraction for 50 frequency points spaced logarithmically between 50 Hz and 12 kHz for each receiver location. The edge source (ES)-based model to compute diffraction waves presented in [12] and [13] and is implemented in a freely available Matlab toolbox, “EDtoolbox" [14], which was utilized to provide the reference for machine learning. Results computed by this method are denoted the “ES model" from henceforth. The frequency resolution was selected to cover the most important hearing range of humans with slightly better resolution than humans have [5]. The resulting location-frequency spectrogram of only the scattered component is shown in Fig. 3 a). It can be noted that the

scattered sound has a discontinuity at the shadow zone boundaries, at angles of 120° and 150°, respectively, to compensate for the direct sound's discontinuity. The high-pass-nature of specular reflection can be seen from receiver azimuth angles 30° to 60°, and the low-pass-nature of occlusion is visible from receiver angles 120° to 150°. The scattering strength is low towards receivers in directions between 80°-100°. There exists also a comb-filter structure that changes the notch frequencies dynamically as the direction of the receiver changes.

In the dynamic test scenario the sound arriving directly at the receiver is present most of the time, and it is therefore interesting to monitor the interference between the scattered sound and the direct sound, as similar interference occurs also in the ears of the listener. The resulting spectrogram is shown in Fig. 4 a). It can be seen, that the scattered component does have a prominent effect near the regions where the specular reflection occurs and where the occlusion occurs. Consequently, the effect is mild in the region where the level of scattered sound component is low. When the source becomes audible near the azimuth of 150°, there is a prominent change in the level of sound, which seems larger than one would expect in real cases. It is not known if this effect corresponds to reality, or if the ES model exaggerates the change.

For the sake of comparison, the GA model output is shown for only reflected component in Fig. 3 d), and in Fig. 4 d) the reflected component is summed with direct sound. Prominent differences can be seen when compared with the ground truth cases shown in Figs 3 a) and 4 a), respectively. It can be seen that the image source model does not render the high-pass effect of the reflected component, it renders occlusion as silence, and also it does not produce smooth spectral transitions when reflection and occlusion effects occur.

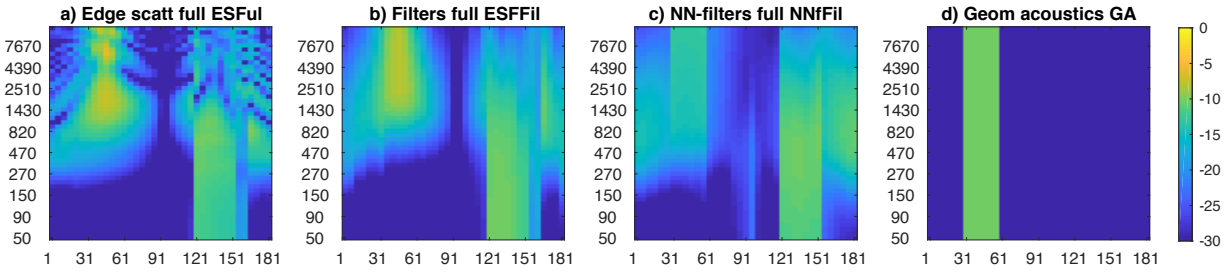

Figure 3. The scattered component arriving from the plate to the receiver in the dynamic scene shown in Fig. 2 rendered with different methods. x-axis – Azimuth [°]; y-axis – Frequency [Hz], color scale shown in right down panel in [dB]. a) 15th-order modelling of diffraction b) parametric IIR structure directly fitted to physically modelled spectrum c) parameters of the structure estimated from geometry using neural networks d) geometrical acoustic model with direct sound and specular reflection

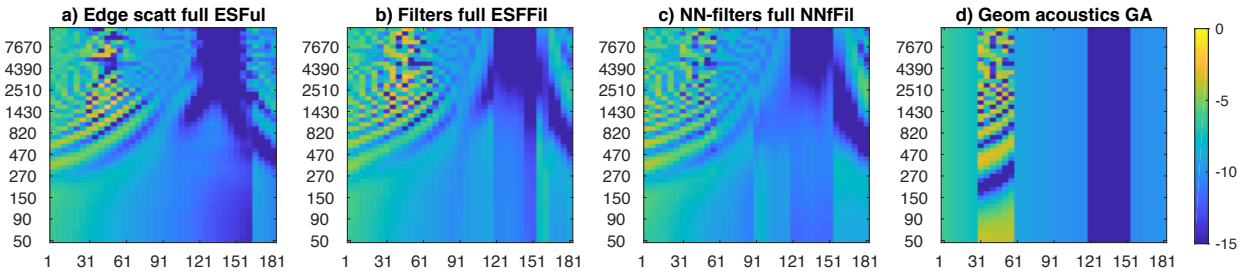

Figure 4. Same as Fig. 3 but with the inclusion of the direct sound contribution.

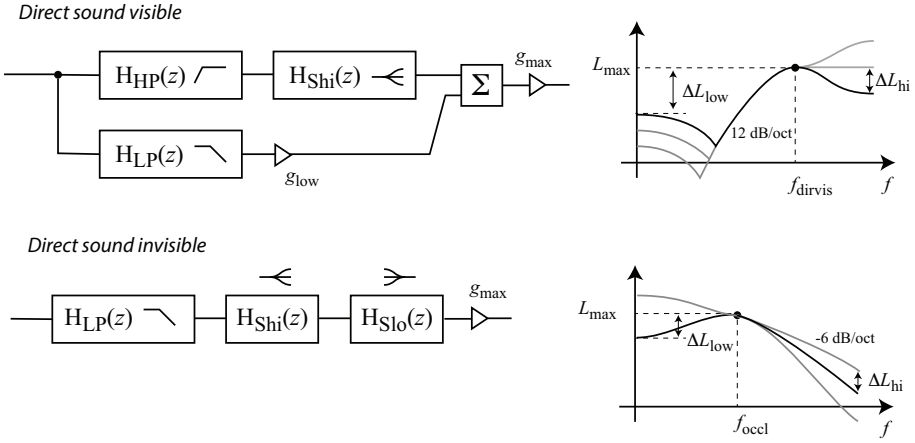

Figure 5. Top: Filter structures used to render scattering from an object. Bottom: prototype responses.

5 REPRESENTATION OF SCATTERED COMPONENTS USING LOW-ORDER FILTER STRUCTURE

Two approaches have been proposed to render scattered sound using filter structures in [15]. The approach discussed here is to implement the total scattered sound component emanating from an object with a single filter, whose output is then spatialized to the direction of the apex point that corresponds to the shortest sound route via the scattering object.

In this section the scattering effect caused by an object is implemented using a single filter, without considering the edges separately as was described in the previous section. The produced spectrum can be assumed to be more complex than the first-order diffraction spectrum of a single edge, as the first- and higher-order diffraction contributions from each edge add up at the listener position with arbitrary phase relationships, causing interference effects.

The main spectral characteristics of scattering should be implementable with the filter selected for the task. To find such a filter structure, a large amount of different symmetric and asymmetric boxes with dimensions from 40 cm to 2 m were simulated with the EDToolbox in a preliminary study and the spectral shapes of scattered components were monitored. Two distinct types of spectra were identified, and filter structures are suggested for each case.

1. If the direct sound is visible, the effect was found to be in most cases a high-pass effect with 12dB/oct stopband response. In some cases also a plateau was found at low frequencies. Above the cut-off frequency of the high-pass filter, the spectral shape typically varied widely. The corresponding filter is shown in the top part of Fig. 5.

2. If the direct sound is not visible, i.e., the source is occluded, a low-pass effect with about -6dB/oct stopband response was seen, although there could be large frequency-dependent variations in the spectrum. The corresponding filter is shown in the bottom part of Fig. 5.

The switching from direct-visible case to occluded case can be made based on the visibility of direct sound. However, in reality the object scattering produces some low-pass effect already before the shadow zone boundary and the estimation of scattering should be very accurate, since the direct sound and scattered contributions are close to each other in the temporal domain. To prevent errors due to such cases, the switching is made already a bit before the shadow zone, where the direct sound is just visible, but the edge diffraction produces

already a low-pass effect. In this case the target filter parameters are computed based on the combined effect of scattered sound and direct sound.

The parameters of the filters are described in the prototype frequency responses shown in Fig. 5. The filter parameters are fitted to the responses with a heuristic approach, where first the high- or low-pass filter is fitted to the spectrum, for direct-sound-visible and -invisible cases, respectively. The fitting is performed by finding the lowest estimation error by sliding the position of the cut-off frequency. After this the parameters of the rest of the filter components are computed in a least-squared-error sense, using the methods described in [16].

The direct-sound-visible filter consists of parallel first-order low-pass and second-order high-pass filters with the same cut-off frequency. The low-pass contribution is attenuated by gain factor g_{low}, and the corresponding level parameter is shown in the figure as $\Delta L_{\mathrm{low}} = 20\log(g_{\mathrm{low}}/g_{\mathrm{max}})$. After fitting the high-pass and and low-pass filters to the response a high-frequency first-order shelf filter is subsequently fitted to provide a better approximation of the spectrum. The occlusion filter is formed in similar fashion, however, the high-pass filter does not exist and both high-frequency and low-frequency first-order shelf filters are utilized to approximate the spectral slopes in the modeled spectrum.

To get an indication of the accuracy obtained with such a filter model, the 15th-order edge source responses computed for the dynamic scene shown in Fig. 3 a) were implemented with the filter structure proposed here and the resulting spectra are shown in Fig. 3 b). It can be seen that the filter-structure captures the high-pass and low-pass-effects relatively well, and also the overall level of scattered sound. However, the comb-filter effects are generally lost.

Fig. 4 b) shows the resulting spectrogram when the direct sound is included. In this case the scattered component is implemented as a sound ray that implements the propagation delay and distance attenuation, where the propagation distance is computed using the shortest route touching at one position on the surface of the object. It can be seen that the interference patterns in the spectrogram are similar to the ground truth case shown in Fig. 4 a). The method provides smooth fade-in and fade-out of the specular reflection, and also low-pass-filtered occlusion effects, although relatively large differences between a) and b) figures are also visible. The introduction of propagation delay to the scattered component produces complex interference patterns in a similar fashion as in the original, although not in identical form. It may be assumed that the drop of plausibility in virtual audio rendering caused by filter-fitting is minimal, which will verified in perceptual testing reported in [15].

6 ESTIMATION OF FILTER PARAMETERS USING NEURAL NETWORKS FROM THE GEOMETRY OF SOURCE-OBJECT-RECEIVER SCENARIO

In the method proposed in this article the filter parameters implementing diffracted or scattered components are estimated using machine learning directly from the geometry of the scenario. A non-linear regression task is perfomed with most of the networks, where the input is geometric descriptors of the source-edge-receiver system, and output is a parameter for the filter structure. For regression the classic feed-forward neural network with Bayesian regularization backpropagation training with one or two hidden layers are used with Matlab command `fitnet`. The selection of linear or logarithmic scale of target parameter was based on the assumption that the error will be distributed evenly on selected scale and in most cases an even distribution is desired. The details of the networks are shown in [15]. The goodness of fit measured with the trained network, using the test data, is also given in the table as correlation value or alternatively as a percentage of correct classification.

A logarithmic scale is used for the target parameters in teaching to ensure even distribution of error. However, in total object scattering net trained to estimate L_{max} linear scale is used, since the best accuracy is desired for high values of L_{max}, because low-level scattered components will be masked by other sounds in many cases. There is no self-evident set of feature values that could be utilized to estimate the filter parameters. The geometry of the scenario is described for the EDToolbox as coordinate values of corners, which were not viewed as a viable approach in this context. A set of features was selected heuristically, which describes the

main angular and absolute dimensions and orientations of the object from the viewpoints of the receiver and the source. Additionally, the geometry of the shortest path of sound traveling from the source via the scattering object to the receiver is described. The features are described in detail in [15].

For training, 78 rectangular plate objects were simulated, where the width and height of the plate varied from 30 cm to 90 cm, and the depth from 2 cm to 4 cm. The responses to 50 receiver positions from 20 source positions were computed, where the sources and receivers were in random directions, with distances varying from 40 cm to 3 m. The geometry of the test was not included in simulation. This resulted in 78000 spectral responses, 67000 of which had source visible, and the source was occluded in the remaining 11000 responses. Separate networks were trained for direct-sound-visible and direct-sound-occluded-cases.

The test scenario described in Sec. 4 was implemented with trained networks and the resulting spectrogram is shown in Fig. 3 c). It can be seen that the networks clearly capture the main features of reflections and occlusion. However, the accuracy is notably lower when compared to the filtering approach. When the direct sound is also taken into account, as shown in Fig. 4 c), the similarity to the ground truth case in Fig. 4 a) is relatively high, although some spectral details are either different or smeared. For example, in the region of occlusion, the interference patterns caused by comb-filtering are different, however, the perceptual prominence of the errors has been tested to be low with human listeners, as described in [15].

7 CONCLUSION

A computationally efficient method to estimate and render sound scattered from finite objects in virtual reality is proposed in this work. The method creates the spectral effect of scattering by filtering the sound arriving at the object with a parametric filter structure consisting of a combination of low-pass, high-pass and shelving filters.

The parameters for the filters are estimated using a machine learning approach. The input data in teaching contains information about the configuration formed by the source, the scattering object, and the receiver. The target data in teaching is obtained by fitting the response of the filter structure into a spectrum obtained from measurement or from acoustical modelling.

A perceptual experiment, where the scattering from a thick-plate object, at a distance of 80 cm from the listener, was rendered for a dynamic scene using different methods. The results of objective and subjective tests fully detailed in [15] show that a relatively simple parametric filter delivered similar plausibility as a detailed acoustical model, and that the filter parameters estimated using a neural network degraded the plausibility only slightly from the detailed acoustical model.

ACKNOWLEDGEMENTS

This research was supported by the Academy of Finland.

References

[1] Frank J Fahy. *Foundations of engineering acoustics*. Elsevier, 2000.

[2] WR Sherman and AB Craig. Understanding virtual reality: interface, application, and design. *The Morgan Kaufmann series in computer graphics and geometric modeling*, 2003.

[3] Lauri Savioja, Jyri Huopaniemi, Tapio Lokki, and Ritta Väänänen. Creating interactive virtual acoustic environments. *Journal of the Audio Engineering Society*, 47(9):675–705, 1999.

[4] Michael Vorländer. *Auralization: fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality*. Springer Science & Business Media, 2007.

[5] Ville Pulkki and Matti Karjalainen. *Communication acoustics: an introduction to speech, audio and psychoacoustics*. John Wiley & Sons, 2015.

[6] Tapio Lokki, Peter Svensson, and Lauri Savioja. An efficient auralization of edge diffraction. In *Audio Engineering Society Conference: 21st International Conference: Architectural Acoustics and Sound Reinforcement*. Audio Engineering Society, 2002.

[7] Ville Pulkki, Tapio Lokki, and Lauri Savioja. Implementation and visualization of edge diffraction with image-source method. In *Audio Engineering Society Convention 112*. Audio Engineering Society, 2002.

[8] Nicolas Tsingos and Jean-Dominique Gascuel. Fast rendering of sound occlusion and diffraction effects for virtual acoustic environments. In *Audio Engineering Society Convention 104*. Audio Engineering Society, 1998.

[9] Nicolas Tsingos, Thomas Funkhouser, Addy Ngan, and Ingrid Carlbom. Modeling acoustics in virtual environments using the uniform theory of diffraction. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '01, pages 545–552, New York, NY, USA, 2001. ACM.

[10] Nicolas Tsingos, Wenyu Jiang, and Ian Williams. Using programmable graphics hardware for acoustics and audio rendering. *Journal of the Audio Engineering Society*, 59(9):628–646, 2011.

[11] Carl Schissler, Ravish Mehra, and Dinesh Manocha. High-order diffraction and diffuse reflections for interactive sound propagation in large environments. *ACM Transactions on Graphics*, 33(4):39, 2014.

[12] U Peter Svensson, Roger I Fred, and John Vanderkooy. An analytic secondary source model of edge diffraction impulse responses. *The Journal of the Acoustical Society of America*, 106(5):2331–2344, 1999.

[13] Andreas Asheim and U Peter Svensson. An integral equation formulation for the diffraction from convex plates and polyhedra. *The Journal of the Acoustical Society of America*, 133(6):3681–3691, 2013.

[14] P. Svensson. Edge diffraction Matlab toolbox (EDtoolbox). https://github.com/upsvensson/Edge-diffraction-Matlab-toolbox, 2000. [Online; accessed 20-April-2018].

[15] Ville Pulkki and U Peter Svensson. Machine-learning-based estimation and rendering of scattering in virtual reality. *The Journal of the Acoustical Society of America*, 145(4):2664–2676, 2019.

[16] Vesa Välimäki and Joshua D Reiss. All about audio equalization: Solutions and frontiers. *Applied Sciences*, 6(5):129, 2016.