Interactive Room Acoustics Synthesis for XR

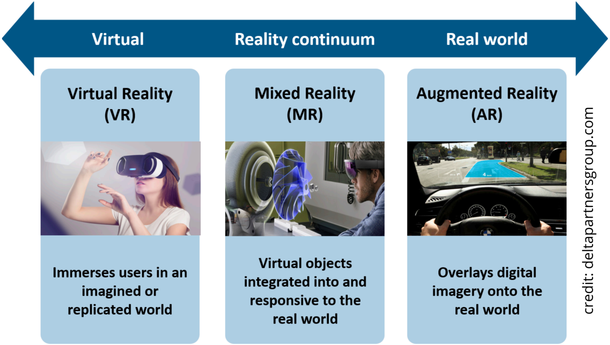
Enzo De Sena, Institute of Sound Recording, University of Surrey (UK)
Hüseyin Hacıhabiboğlu, Graduate School of Informatics, METU (Turkey)
Zoran Cvetković, Department of Engineering, King's College London (UK)

AES International Conference on Audio for Virtual and Augmented Reality 2020

Acknowledgments

Julius O. Smith III (Stanford)
Toon van Waterschoot (KU Leuven)
Patrick Naylor (Imperial College London)
Niccolo' Antonello (IDIAP)

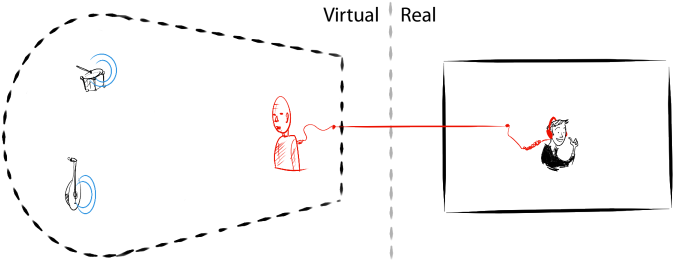
XR – Virtual, Augmented, Mixed Reality



credit: deltapartnersgroup.com

- Applications: entertainment, education, communications...

Audio rendering for VR/MR/AR



- Step 1: room acoustic synthesis
- Step 2: auralization, i.e. making the sound field audible
- Rendered room acoustics in
  - VR: room acoustics of virtual space
  - MR/AR: room acoustics of real space

Why render room acoustics

- Rendering high-quality room acoustics is essential for
  - auditory immersion
  - sense of realism and presence
  - spaciousness
  - envelopment experience
  - effective externalization
  - convincing illusion of sound source distance
  - rendering apparent source width

Aims and scope of the Tutorial

- Overview of fundamental principles and state-of-the-art methodologies in audio for XR:
  - Synthesis of room acoustics
  - Integration with headphones-based auralization

Outline of the Tutorial

- Room Acoustics Synthesis
  - Fundamentals of Room Acoustics
  - Perception of Room Acoustics
  - Room Acoustic Models
- Binaural Rendering for XR
  - Binaural Rendering
  - Integration with Synthesized Room Acoustics in XR
- Examples of open-source VR/AR Audio Rendering Software
  - Audio360
  - Resonance Audio
  - Steam Audio
- Conclusions

Outline of the Section

- Room Acoustics Synthesis
  - Fundamentals of Room Acoustics
  - Perception of Room Acoustics
  - Room Acoustic Models
- Binaural Rendering for XR
  - Binaural Rendering
  - Integration with Synthesized Room Acoustics in XR
- Examples of open-source VR/AR Audio Rendering Software
  - Audio360
  - Resonance Audio
  - Steam Audio
- Conclusions

Wave equation

- Sound propagation governed by the PDE[2]:

$$\triangle p - \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} = s$$

p pressure, s source distribution, c speed of sound

- Initial and boundary conditions are needed to find a solution
- Example boundary condition:

$$\frac{\partial p}{\partial t} = -cZ_w \nabla p \cdot \mathbf{n}$$

$\mathbf{n}$ orthogonal to the boundary, Z_w wall impedance

- Equation admits closed form solution only in few cases

Modal description of reverberation

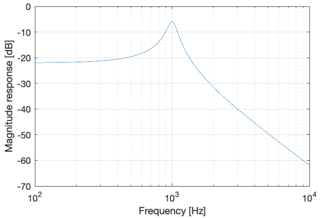
- Monochromatic sound source, Helmholtz equation[2]:

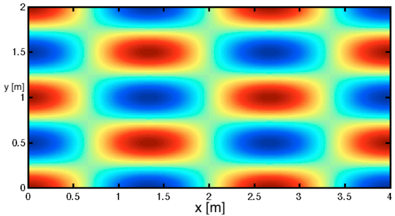
$$\triangle\hat{p} + \left(\frac{\omega}{c}\right)^2 \hat{p} = \hat{s}$$

- Solutions for point sources:

$$\hat{p} = A \sum_{m=0}^{\infty} \frac{\psi_m(\mathbf{x}')\psi_m(\mathbf{x})}{\omega^2 - \omega_m^2 - 2j\xi_m\omega_m}$$

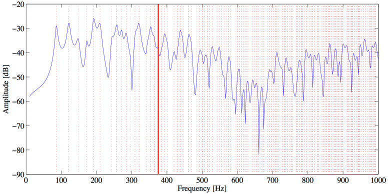
each element of the summation is called "mode", where ω_m modal frequencies, $\psi_m(\mathbf{x})$ eigenfunctions, ξ_m imaginary part of eigenvalue, $\mathbf{x}$ observation position, $\mathbf{x}'$ source position





Modal description of reverberation (cont'd)

- Density of modes increases as f^2 (Kuttruff, 2000)[2]
  - Sparse distribution of modes at low frequencies
  - Overlapping modes at high frequency into a random frequency response



- Transition – Schroeder frequency [Schroeder, 1962][3]:

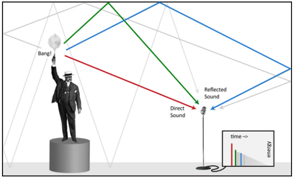
$$F_c = 2000\sqrt{\frac{T_{60}}{V}}$$

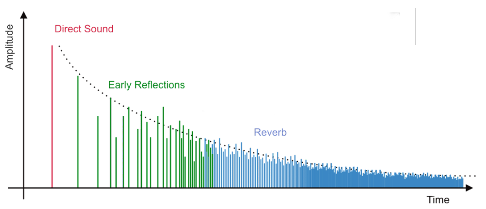
Examples

- Bathroom $V = 10\,\text{m}^3$, $T_{60} = 0.35$ s $\Rightarrow F_c = 374$ Hz
- Concert hall $V = 2700\,\text{m}^3$, $T_{60} = 2$ s $\Rightarrow F_c = 54$ Hz

Acoustic Impulse Response (AIR)

- Acoustic impulse response is (time-domain) solution of wave equation for impulsive sound source
- Response to any input is obtained via convolution with AIR





(Rational Acoustics)

Components of AIR $h(t)$ in rooms

- Direct line-of-sight (LOS)
- Early reflections: sparse low order reflections
- Late reverb: dense, higher order reflections, from all directions

High level characteristics of AIR

Mixing time

- Transition between early reflections and late reverb

Reverberation time (T60)

- Time taken for cumulative sound energy to decay 60 dB below its initial level

Direct-to-reverberant Energy Ratio (DRR)

- The ratio of the energies of early reflections and late reverb

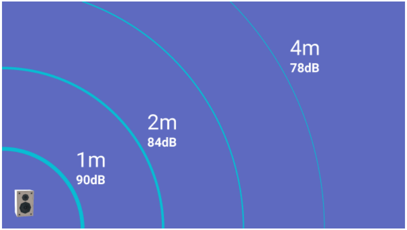
Critical distance (Reverberation radius)

- The distance at which energies of the direct sound and reverberation are equal
- Affects speech intelligibility

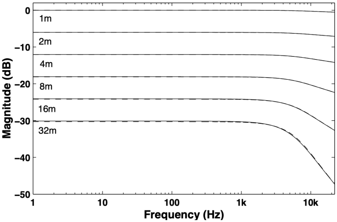
Measured AIR samples

Link here to: https://youtu.be/hWaDaB2B9i8

Propagation



(Google Resonance)



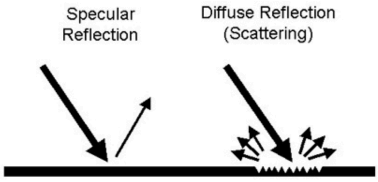
Distance attenuation

Spherical spreading; inverse square law; 6 dB drop for every doubling of distance

Air absorption

Heat losses due to friction and relaxation processes; frequency dependent

Reflections



(Sonography folder)

- Sound wave hitting a boundary will generate a specular reflection and, depending on the material, a more or less energetic diffuse reflection
- In the process, energy will be lost, with losses dependent on material and on frequency

Source directivity

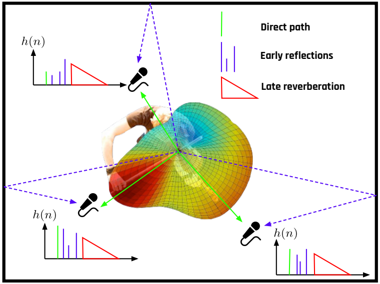
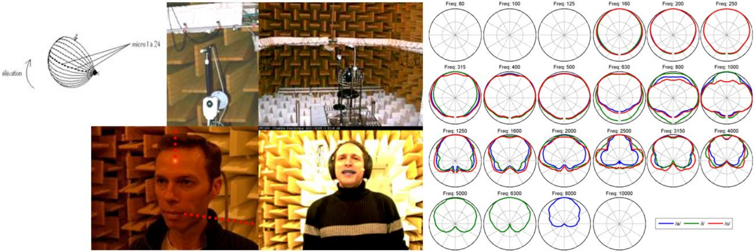


Figure adapted from Behler, 2006 [4]

- Real sound sources are never omnidirectional
- Depending on the position of observation the direct path level and the ER pattern will change
- Critical distance: $r_c \approx 0.1(gV/\pi T)^{\frac{1}{2}}$ (g is the directivity gain)

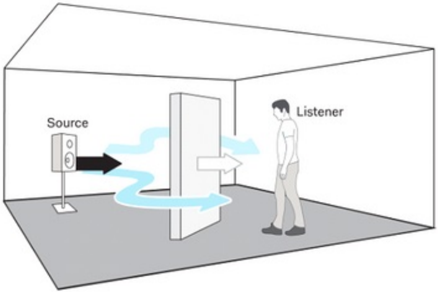
Measuring source directivity



(Katz et al., 2006)[5]

- Typically requires an anechoic chamber (or that the direct path response does not overlap with the first early reflection)
- Smaller static sources (e.g. loudspeakers) easier to measure
- CLF [6] or SOFA [7] can be used to store and deliver directivity data

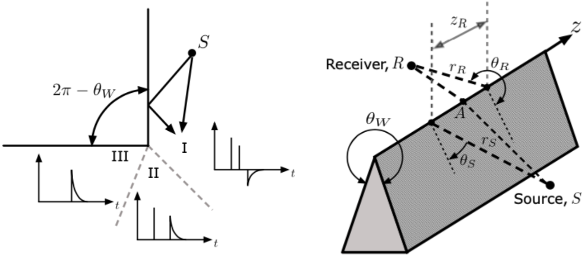
Occlusion



(quora.com)

- Occlusions occur when the line of sight is blocked
- Sound waves can still travel around the edges: diffraction
- Waves with larger wavelengths result in more energy diffracted around the edge
- Occlusions can be caused acoustically transparent materials

Edge Diffraction

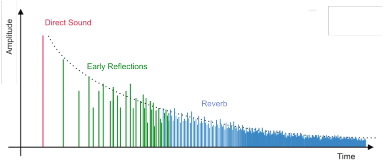


Left figure adapted from (Torres et al., 2001)[8]

- When a source is obstructed, diffraction component is the earliest arriving wave
- Biot-Tolstoy-Medwin (BTM) model[9] provides a closed form solution for infinite wedges

Outline of the Section

Perception of Sound in Rooms

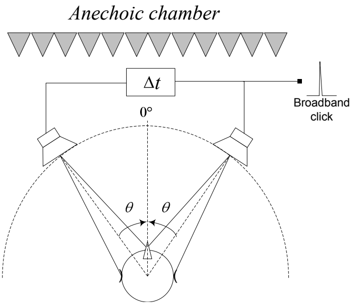


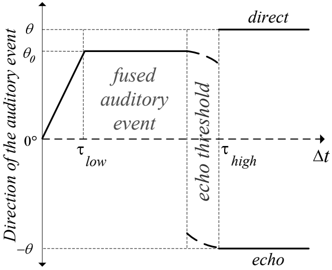
(Rational Acoustics)

1. Early reflections: affect spaciousness, envelopment, and apparent source width.
2. Late reverberation: precise structure not important, but
   2.1 $T_{60}(\omega)$: affects impression of size
   2.2 Echo buildup density: affects the perceived texture of reverberation
   2.3 Mode density: if insufficient can yield metallic sound
   2.4 direct-to-reverberant ratio...

- Governed by complex and not fully understood perceptual phenomena[10, 11]

Perception of Early Reflections

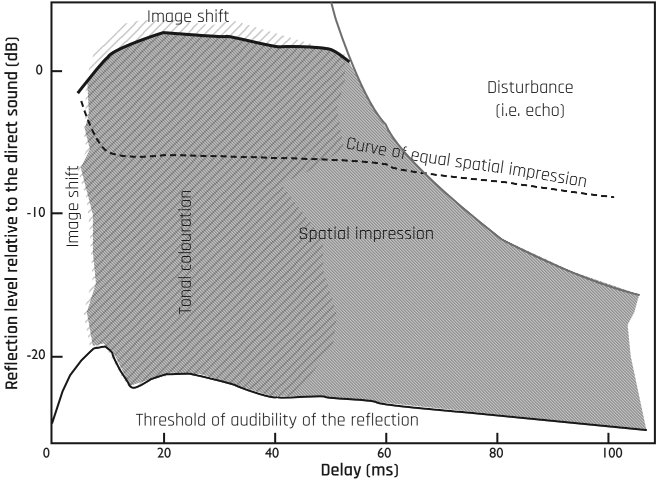




Precedence Effect

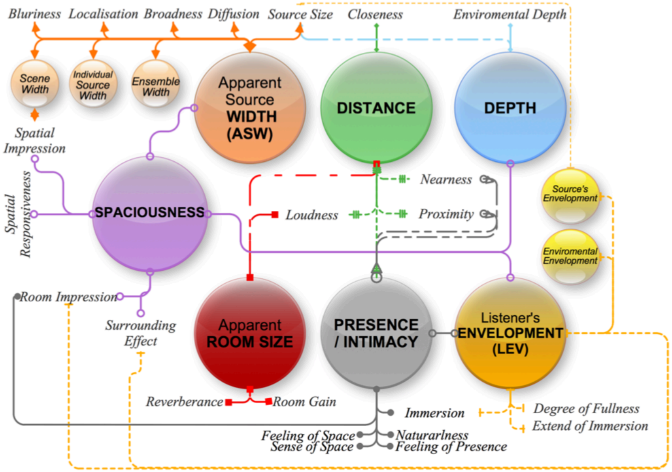
The first arriving sound wave dominates directional localization and most of the localization information conveyed in the reflections is suppressed

Perception of Early Reflections



Perception of a single reflection incident from $40°$ azimuth with respect to the frontal direct sound (Adapted from Barron, 1971[12])

Perception of Reverberation



Perceptual attributes of reverberation (Kaplanis et. al. 2014)[10]

Summary of fundamentals of room acoustics and perception of room acoustics

- AIR in a room: LOS, early reflection and late reverb
- Wave equation gives physical model for propagation
- Wave equation requires initial and boundary conditions to find solution, and solution hard to find in closed form
- Solution for point-like sound source yields modal description of reverberation
- Modes well separated at low frequencies

- Room perception governed by complex phenomena
- Accurate rendering of early reflections is important
- We are not sensitive to precise structure of late reverb

Outline of the Section

Room Acoustics Synthesis

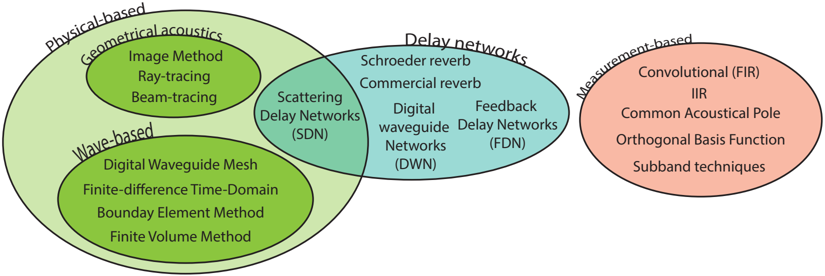
- Fundamentals of Room Acoustics
- Perception of Room Acoustics
- Room Acoustic Models

Binaural Rendering for XR

Examples of open-source VR/AR Audio Rendering Software

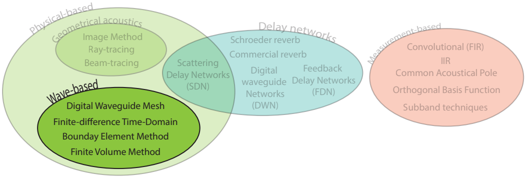
Conclusions

Overview



- Excellent overview paper of past 50+ years of artificial reverberation by Valimaki et al.[13, 14]

Overview



Wave-based models

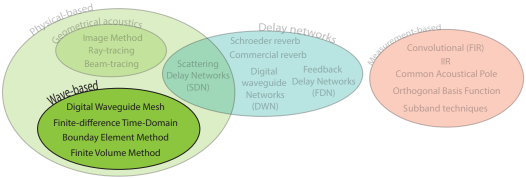
Discretize wave equation in time/frequency and space/boundary/volume

- E.g. FDTD[15] approx. derivatives with finite differences:

$$\frac{\partial^2 p}{\partial t^2} \approx \frac{p_{l,m,i}^{n+1} - 2p_{l,m,i}^{n} + p_{l,m,i}^{n-1}}{T^2} \qquad \frac{\partial^2 p}{\partial x^2} \approx \frac{p_{l-1,m,i}^{n} - 2p_{l,m,i}^{n} + p_{l+1,m,i}^{n}}{X^2}$$

- Convert wave equation into set of linear equations

Overview



Wave-based models

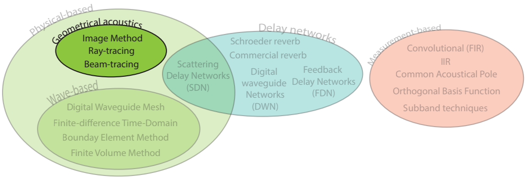
Discretize wave equation in time/frequency and space/boundary/volume

- High physical accuracy...
- ...but extremely high computational complexity

Overview

Please replace with:
https://www.youtube.com/watch?v=PoWpCC5KUmo

Overview

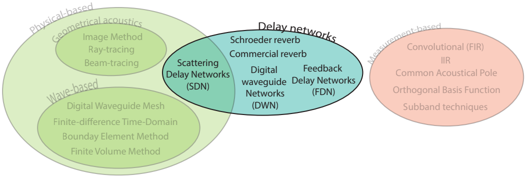


Geometrical acoustics models

Models approximating sound propagation using rays

- Lower computational complexity...
- ...but lower accuracy

Overview

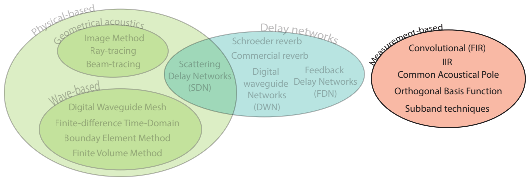


Delay Networks

Methods that do not physically model sound propagation in rooms, but aim to create pleasing reverberant sound

- Very low computational complexity (historically first type of artificial reverberators)...
- ..but no physical accuracy and no explicit physical modelling

Overview

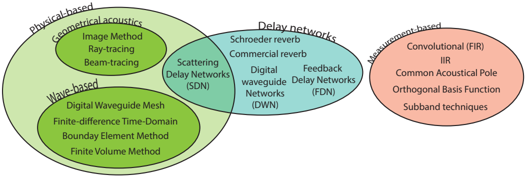


Measurement-based methods

Use measurements in real space to form parametric representation of room acoustics

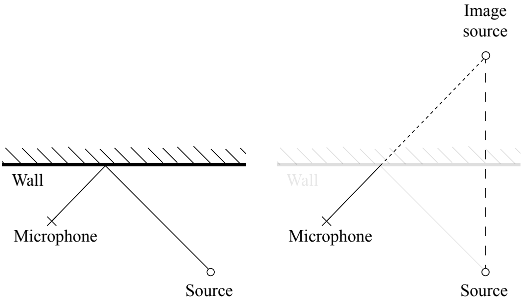
- E.g. convolutional (finite impulse response filter) model:
  - Need to have access to physical space with desired characteristics
  - Very high complexity (e.g. if $F_s = 50$ kHz, $T60 = 2$ s, 3 sound sources and 2 output channels $\Rightarrow 60$ billion FLOPS
  - FFT convolution is faster, if throughput delay is tolerable (and there are low-latency algorithms)

Overview



- Main requirements for XR:
  - Low computational complexity
  - As accurate as possible, either physically or perceptually
- Most suitable room acoustic models are:
  - Geometrical acoustics methods (image method, ray tracing, beam tracing)
  - Delay networks methods (FDN, DWN, SDN)

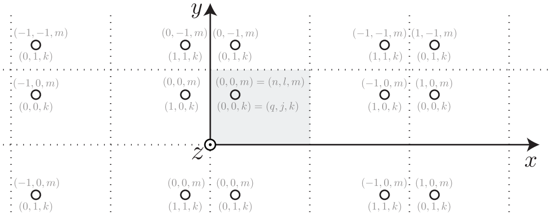
Image method (IM) for single reflector



- Wave propagation in half space is equivalent for:
  1. source and wall
  2. source and image source (no wall)
- Exact for rigid wall ($\nabla p \cdot \mathbf{n} = 0$)
- Approximation for non-rigid wall

Image method (IM) for rectangular room[21]

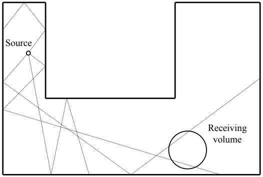
- With multiple reflectors: remove wall, mirror source and opposite wall

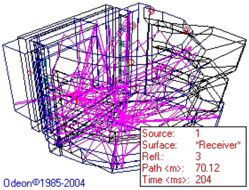


- Spatial periodicity of image sources can be exploited for fast rendering in multiple positions[19]
- Non-rectangular rooms also possible, but need expensive computations of image source visibility[20]

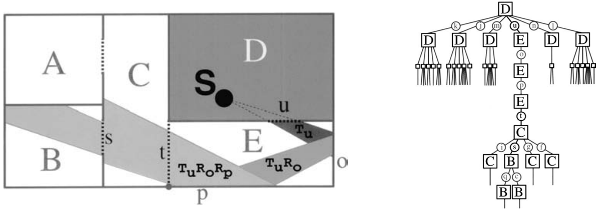
Ray-tracing[22]

- Source emits rays in given number of directions
- Specular reflections; diffraction and scattering also possible
- Build RIR by recording time and amplitude at receiver volume
- Choice of receiver size and number of rays is critical
- Rays can be weighted/filtered according to source directivity and wall absorption





Beam-tracing[23]



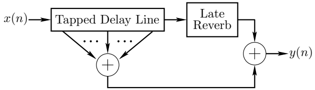
- Precalculate beam tree by intersecting environment polygons (e.g. door u in figure) with the beam
- Precalculated beam tree depends only on source position
- For given observation point, lookup in tree and calculate paths

Comparison Geometric Acoustics Models

- Beam-tracing:
  - fast rendering for moving observer
  - requires recalculate tree if source moves (though recent advancement reduce complexity[24]
- Ray tracing:
  - Complexity can be controlled by number of rays
  - Can model edge diffraction, scattering
  - No guarantee of low-order reflections
- Image Method:
  - Guaranteed all reflections up to certain order present
  - Preferred model to calculate early reflections
  - High computational complexity for long RIRs
- All above output an AIR, so still need to run convolution
- If physical accuracy not needed, perceptual methods provide better option

Perception-based models

- Overview paper by Hacihabiboglu et al.[25]
- Often separate modules for early and late reverb

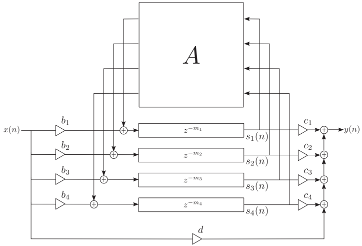


(J. O. Smith, https://ccrma.stanford.edu/ jos)

Desired qualities for late reverb:

- Smooth decay: high reflection density
- Smooth frequency response: high mode density
- Moorer's ideal reverb: exponentially decaying white noise

Feedback delay network



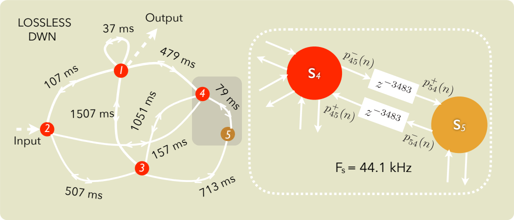
(Schlecht and Habets, 2017)[26]

- Generalization of Schroeder reverberator (Stautner and Puckette, 1982) [27]
- Design: start with lossless prototype ($T_{60} = \infty$) to obtain noise-like reverb and add losses to obtain desired reverberation time in each band

Advancements in FDNs

- Jot and Chaigne (1991) [28]:
  - Practical procedure to design delays and FDN matrix to obtain desired echo density and frequency-dependent reverberation time
- Rocchesso and Smith (2002) [29]:
  - Equivalence with DWN
  - Circulant feedback matrix with increased efficiency
- Schlecht and Habets (2015, 2017) [30, 31, 26]:
  - Time-varying FDNs: reduce artifacts and obtain more lively reverberation tail
  - Unilosslessness: new definition of lossless FDN matrix
  - Closed-form and approximated formulas for echo density
  - Procedure to design delays for desired mixing time

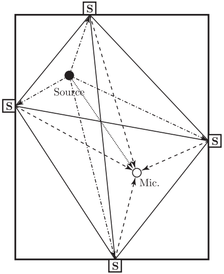
Digital waveguide networks (DWN)



- Network of bi-directional delay lines connected at scattering junctions (Smith, 1985) [32]
- Can be interpreted as network of acoustic tubes
- Question: How to set parameters (delay line lengths, network connections, scattering matrix..)?

Scattering delay network (SDN)

- Design DWN based on characteristics of a physical room



- Position nodes at first-order reflection points
- Fully connected DWN network
- Mono-directional lines for source-junction and junction-mic

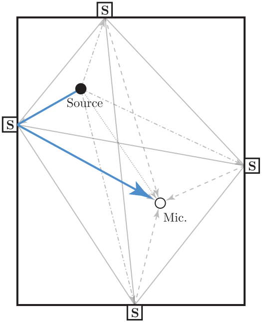
Two interpretations:

- Physical network of acoustic tubes
- Approximation of image method/ray tracing

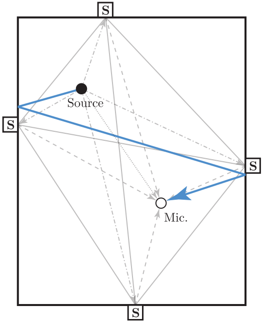
(De Sena, Hacihabiboglu, Cvetkovic, AES 2011)[33] (IEEE/ACM TASLP 2015)[34] (USPTO)[35]

SDN: approximation of image method

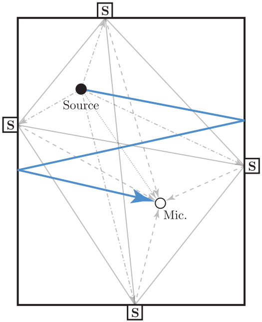
- Correct rendering of LOS and first-order reflections in time, amplitude and direction
- Approximation of second and higher-order reflections, less important perceptually



I-order reflection



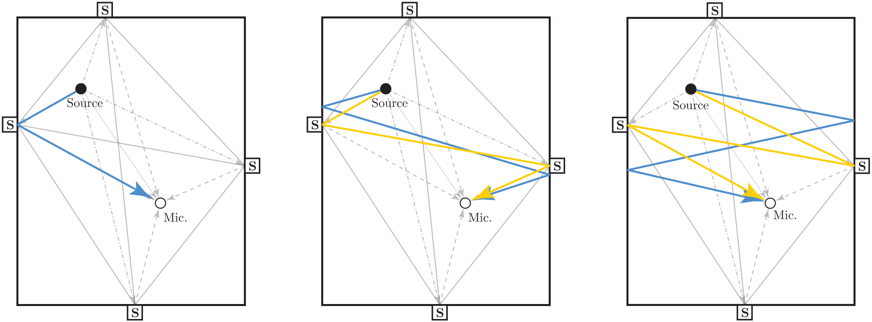
II-order reflection



Another II-order reflection

SDN: approximation of image method

- Correct rendering of LOS and first-order reflections in time, amplitude and direction
- Approximation of second and higher-order reflections, less important perceptually



I-order reflection | II-order reflection | Another II-order reflection

SDN: physical network of acoustic tubes

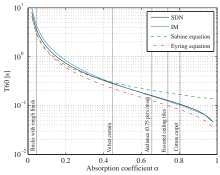
- Can be shown that SDN is a physically accurate model of a network of acoustic tubes

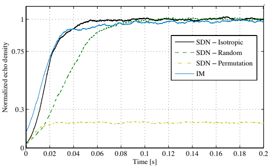
room with source/observer

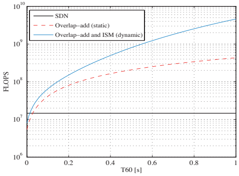
network of acoustic tubes

- Actual room and network of acoustic tubes share a number of perceptually important features (e.g. T60, echo buildup etc)

SDN performance







- Higher perceived naturalness than FDN and ray tracing[36]
- While orders of magnitude faster than (fft) convolution alone
- All parameters of model derived from physical properties

Advantages with respect to delay networks:

- No need for hands-on parameters tuning
- Physical interpretation ⇒ spatialisation possible
- More elegant solution than separate early/late modules

Comparison of SDN-IM-FDTD

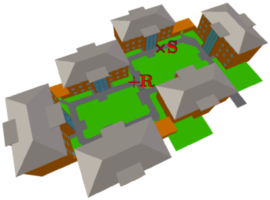
Please replace with:
https://youtu.be/1hdhhrM4juQ

SDN demo (with stereo loudspeakers)

Please replace with:
https://www.youtube.com/watch?v=AbLCJz64oLc

Recent advancements in SDN

- Stevens et al. (2017) [37]:
  - Extension to exact second-order reflections
  - Implementation of direction-dependent scattering (e.g. modelling of trees)
  - Modelling of outdoor scenes (sky absorbing nodes)
- Schlecht and Habets (2017) [26]:
  - Showed scattering matrix is unilossless



(Stevens et al., 2017)[26]

Summary

- Wide variety of room acoustic models and simulators
- Wave-based models: most accurate available but computationally expensive
- Geometric-models: ray-like assumption, lower complexity but also lower accuracy
- Perception-based models: very fast, attempt to reconstruct only perceptually relevant features of reverberation

Outline of the Section

Outline of the Section

Room Acoustics Synthesis

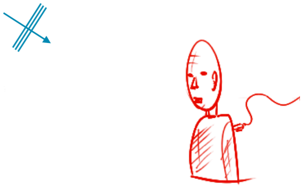
Binaural Rendering for XR

- Binaural Rendering
- Integration with Synthesized Room Acoustics in XR

Examples of open-source VR/AR Audio Rendering Software

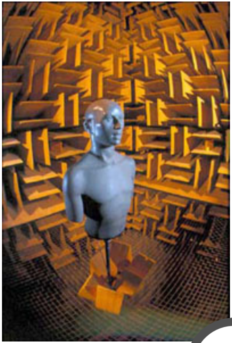
Conclusions

Binaural rendering of a single source

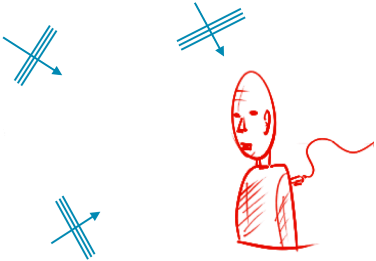


- To render binaural audio, we need to model head diffraction, shadowing etc
- In wave-based models, can be done as part of the simulation, but computationally expensive process

- In other model classes: measured responses from dummy heads
- Typically assumed far-field sound sources (incoming plane waves)
- Results are direction-dependent filters called head-related transfer function (HRTF)
- MIT KEMAR dataset[38], CIPIC[39] and many others.



Binaural rendering of synthesized sound fields



- In sound synthesis we will have multiple incoming plane waves, which can be:
  - individual sound sources
  - room reflections (typically only early reflections are rendered binaurally)
- We know their directions: convolve with HRTFs and sum up
- Head tracking with HRTF update – latency < 85 ms

Ambisonics transcoding

- All sources and reflections are rendered via a virtual multichannel system
  - The number of channels and HRTF filtering operations remains the same regardless of the number of sources and reflections
  - Each virtual loudspeaker is rendered via a pair of HRTFs
  - Sources and reflections are rotated in the direction opposite of head rotations $\rightarrow$ there is no need to update HRTFs
  - N-th order Ambisonics $\rightarrow$ $(N+1)^2$ channels

(Resonance Audio, resonance-audio.github.io/resonance-audio)]

HRTF filter design and individualization

- HRTF filter design and interpolation
  - Filter design via spectral smoothing[40, 41]
  - Interpolation via manifold learning[42, 43]
  - Interpolation via SHD[44]
- HRTF individualization
  - CV based (Genelec™AuralID™, **https://auralid.genelec.com**)
  - Sparse representations[45]
  - PCA-based[46]
  - Deep-learning based [47, 48]

Outline of the Section

Room Acoustics Synthesis

Binaural Rendering for XR

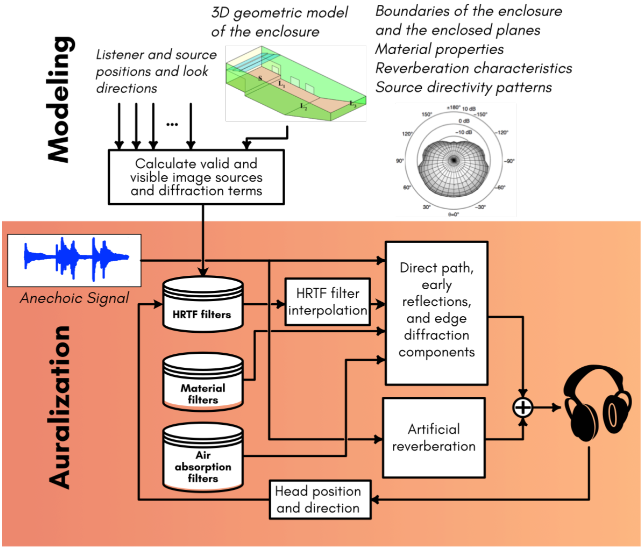
Binaural Rendering

Integration with Synthesized Room Acoustics in XR

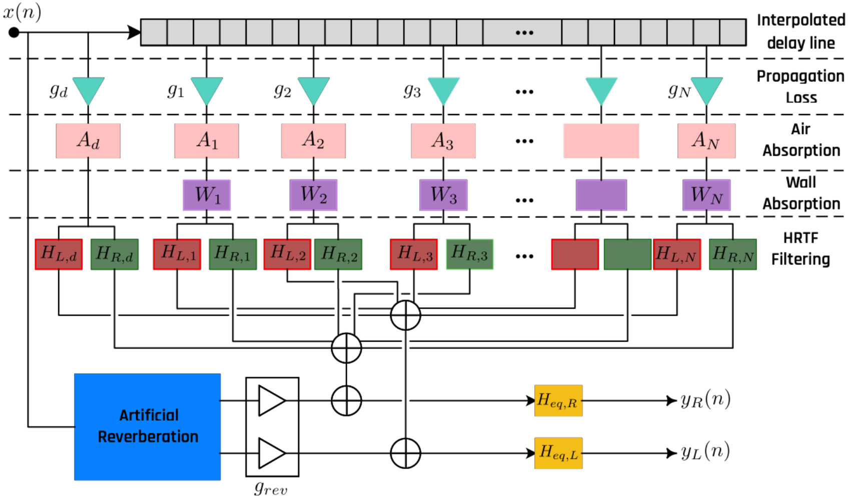
Examples of open-source VR/AR Audio Rendering Software

Conclusions

Overview of an Auralization Model



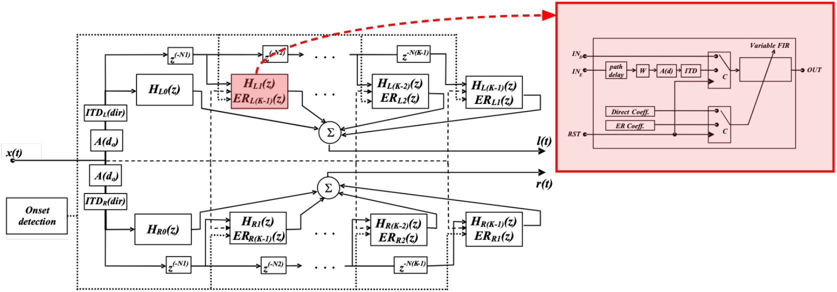
Direct auralization approach



Shortcomings of direct auralization

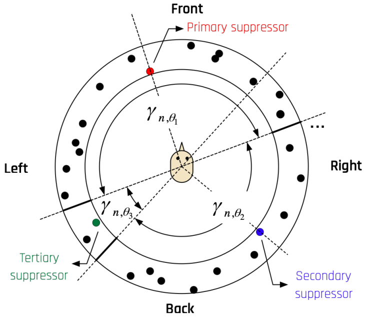
- Calculation of a high number of image sources at interactive rates $\Rightarrow$ Increased computational cost
- The interpolated delay line should be as large as the latest arriving ER to be simulated $\Rightarrow$ Increased memory footprint
- The number of reflections to be simulated determines the number of filters to be used $\Rightarrow$ Increased computational cost
- Interpolation of all filters used in response to moving sources or listener $\Rightarrow$ Increased computational cost
- Quality of artificial reverberation determines the overall experience $\Rightarrow$ Perceptual quality depends on design choices
- OVERALL: Not a scalable approach!

Perception-based methods
Fixed-cost Auralization of early reflections[49]



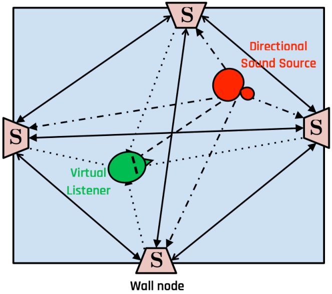
- Dynamic model relying on onset detection
- Uses a single, cascade FIR structure (per ear) to auralize the LOS and the early reflections
- Onsets can be calculated offline

Perception-based approaches
Reflection culling[50]



- Early arriving sounds will suppress late arriving sounds (Precedence effect)
- Cluster and select suppressor early reflections to be auralized
- Represent each cluster by the suppressor secondary source

SDN auralization



- Use pair of HRTF filters for each node-head connection
- Head tracking can be (almost trivially) integrated
- Simulation of source directivity involves weighting the output to each delay line
- SDN is naturally bundling reflections (no need for culling)
- Fixed number of filters per wall (as opposed to direct auralization)

SDN binaural demo

Please replace with:
https://www.youtube.com/watch?v=PmWTXWDQu5U

Outline of the Section

Room Acoustics Synthesis

Binaural Rendering for XR

Examples of open-source VR/AR Audio Rendering Software

- Audio360
- Resonance Audio
- Steam Audio

Conclusions

Audio360, Facebook

[Facebook]

- Multi-platform software for mobile and desktop devices:
  - Windows, macOS, Android, iOS
- Light weight SDK

Audio360

- Room simulation
  - source directivity, spread
  - distance attenuation curve options, min/max distance
  - only early reflections are rendered assuming cuboid spaces
    - room dimensions
    - attenuation of reflections, i.e. wall absorption
    - reflection order
    - additional control of the level of early reflections
    - high frequency room absorption
  - possibility for combining with any late reverb plugin
  - focus effects
- Auralization
  - binaural rendering
  - up to 3rd order Ambisonics transcoding

Resonance Audio, Google

Resonance Audio
by Google

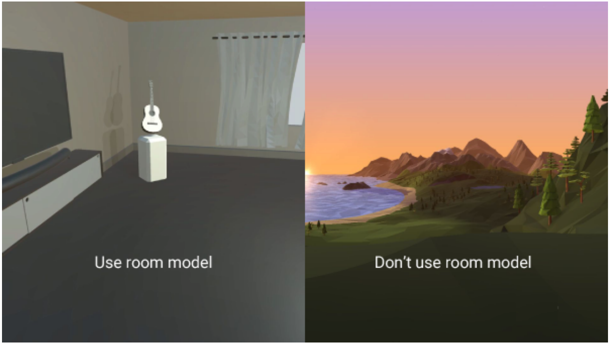
[Resonance Audio, resonance-audio.github.io/resonance-audio/]

- Multi-platform software for mobile and desktop devices:
  - Unity
  - Unreal
  - Wwise
  - DAW
  - Android
  - iOS
- Geared towards limited resources of mobile devices

Resonance Audio

- Room acoustics modelling
  - source directivity controlled by two parameters:
    - alpha – shape (cardioid, circular, figure eight)
    - sharpness – width
  - distance attenuation of direct sound
  - early reflections rendered accurately
  - late reverberation rendered by "reverb engine"
  - occlusions and diffractions – smoothly-changing low-pass filter
- Auralization
  - binaural rendering
  - up to 3rd order Ambisonics transcoding

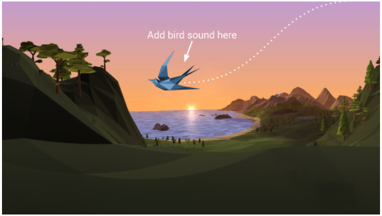
Resonance Audio – Environments



[Resonance Audio, resonance-audio.github.io/resonance-audio/]

- Audio Rooms – early reflections and reverb
  - input parameters
    - room dimensions, cuboid rooms
    - surface materials from a large bank
    - brightness – high/low frequency balance to emulate empty/full rooms
  - late reverb changes in real time with room dimensions/materials

Resonance Audio – Acoustic Sources



[Resonance Audio, resonance-audio.github.io/resonance-audio/]

- Point source
  - point directivity, distance attenuation, dynamic movement
  - monophonic dry sounds
- Ambisonic sound fields
  - react to head rotations only
  - used for distant ambiance/atmosphere sounds

Steam Audio, Valve

[credit wccftech]

- Multi-platform software for mobile and desktop devices:
  - PC, macOS, SteamOS Linux, Android
  - Unity, Unreal Engine 4, FMOD Studio, C API, Wwise (soon)

Steam Audio

- Custom HRTFs
- Gradation of rendering options
  - Occlusions – partial, frequency (in)dependent transmission,
  - Acoustic materials – low/mid/high frequency, scattering
  - Dynamic geometry – two ray tracing options
- High quality revereberation – high quality ray tracing
- Multi-core CPU and GPU acceleration

Steam Audio

- High quality reverberation custom presets
  - length of AIR
  - reflection order
  - number of rays in calculating AIR
  - number of secondary rays for diffuse reflections
  - maximal number of sources
  - CPU time (%) allocated for room simulation

- Optimized ray-tracing options:
  - Intel® Embree – CPU optimized
  - AMD Radeon Rays – GPU optimized

- AMD TrueAudio Next
  SDK for accelerated GPU and multi-core audio

Outline of the Section

Room Acoustics Synthesis

Binaural Rendering for XR

Examples of open-source VR/AR Audio Rendering Software

Conclusions

Conclusions

- Physical room acoustic rendering provides high accuracy but computationally very intensive, even in course approximations
- Leveraging perceptual phenomena enables to reduce complexity
- Hybrid (early reflections+artificial reverb) methods most suitable for XR
- Recent developments in interactive room auralization (e.g. SDN, source culling) promise increased auditory user experience in XR without increasing complexity

Current research topics (I)

- Acoustics of maze-like structures[51, 52]
- Room geometry estimation[53]
- Computationally effective simulation of edge diffraction
- Sound source[54] and diffraction[55] culling
- GPU-based processing[56]
- Applications of deep learning in room acoustics[57]

Current research topics (II)

- Object-based audio (MPEG-H 3D Audio)
  - Position dependent
  - Fully compatible with 3DOF VR
  - Parameterized representation of audio objects and reverberation
  - Transcoding tools for channel-based (e.g. binaural), scene-based (i.e. Ambisonics), and OBA representations
- Upcoming MPEG-I standard
  - Based partly on MPEG-H
  - 3DOF, 3DOF+ and 6DOF modes
  - Work in (somewhat slow) progress
  - Audio to accompany fully immersive 6DOF video

Thank you!

Questions?
Comments?

References I

[1] Jasmina Catic, Sébastien Santurette, and Torsten Dau. "The role of reverberation-related binaural cues in the externalization of speech". In: J. Acoust. Soc. Amer. 138.2 (2015), pp. 1154–1167.

[2] H. Kuttruff. Room Acoustics. 4th. SPON Press, 2000.

[3] M. Schroeder and H. Kuttruff. "On frequency response curves in rooms. Comparison of experimental, theoretical, and Monte Carlo results for the average frequency spacing between maxima". In: J. Acoust. Soc. Amer. 34.1 (1962), pp. 76–80.

[4] G Behler. "How to compare concert halls by listening to music,"" in: Proceedings of the 4th ASA/ASJ Joint Meeting. 2006.

References II

[5] Brian F. G. Katz, Fabien Prezat, and Christophe d'Alessandro. "Human voice phoneme directivity pattern measurements". In: The Journal of the Acoustical Society of America 120.5 (2006), pp. 3359–3359. eprint: https://doi.org/10.1121/1.4781486. URL: https://doi.org/10.1121/1.4781486.

[6] CLF Group et al. "A Common Loudspeaker File Format". In: Syn-Aud-Con Newsletter v32 4 (2004).

References III

[7] Piotr Majdak, Yukio Iwaya, Thibaut Carpentier, Rozenn Nicol, Matthieu Parmentier, Agnieszka Roginska, Yôiti Suzuki, Kankji Watanabe, Hagen Wierstorf, Harald Ziegelwanger, et al. "Spatially oriented format for acoustics: A data exchange format representing head-related transfer functions". In: Audio Engineering Society Convention 134. Audio Engineering Society. 2013.

[8] Rendell R Torres, U Peter Svensson, and Mendel Kleiner. "Computation of edge diffraction for more accurate room acoustics auralization". In: The Journal of the Acoustical Society of America 109.2 (2001), pp. 600–610.

References IV

[9] Maurice Anthony Biot and Ivan Tolstoy. "Formulation of wave propagation in infinite media by normal coordinates with an application to diffraction". In: The Journal of the Acoustical Society of America 29.3 (1957), pp. 381–391.

[10] N. Kaplanis, S. Bech, S. H. Jensen, and T. van Waterschoot. "Perception of Reverberation in Small Rooms: A Literature Study". In: Proc. of the 55th Audio Eng. Soc. Int. Conf.: Spatial Audio. Audio Engineering Society. 2014.

[11] L. L. Beranek. Acoustics. Acoustical Society of America, 1993.

[12] Michael Barron. "The subjective effects of first reflections in concert halls—the need for lateral reflections". In: Journal of sound and vibration 15.4 (1971), pp. 475–494.

References V

[13] V. Välimäki, J. D. Parker, L. Savioja, J. O. Smith, and J. S. Abel. "Fifty years of artificial reverberation". In: IEEE Trans. Audio, Speech, Language Proc. 20.5 (2012), pp. 1421–1448.

[14] V. Välimäki, J. Parker, L. Savioja, J. O. Smith, and J. Abel. "More than 50 years of artificial reverberation". In: Proc. of the 60th Audio Eng. Soc. Int. Conf.: Dereverberation and Reverberation of Audio, Music, and Speech. Audio Engineering Society. 2016.

[15] D. Botteldooren. "Finite-difference time-domain simulation of low-frequency room acoustic problems". In: J. Acoust. Soc. Amer. 98.6 (1995), pp. 3302–3308.

References VI

[16] H. Brandenstein and R. Unbehauen. "Least-squares approximation of FIR by IIR digital filters". In: IEEE Trans. Sig. Proc. 46.1 (1998), pp. 21–30.

[17] Y. Haneda, Y. Kaneda, and N. Kitawaki. "Common-acoustical-pole and residue model and its application to spatial interpolation and extrapolation of a room transfer function". In: IEEE Transactions on Speech and Audio Processing 7.6 (1999), pp. 709–717.

[18] G. Vairetti, E. De Sena, M. Catrysse, S. H. Jensen, M. Moonen, and T. van Waterschoot. "A scalable algorithm for physically motivated and sparse approximation of room impulse responses with orthonormal basis functions". In: IEEE/ACM Trans. Audio, Speech, Language Process. 25.7 (2017), pp. 1547–1561.

References VII

[19] J. Martinez. "Low complexity computer simulation of multichannel room impulse responses". PhD thesis. TU Delft, 2013.

[20] Jeffrey Borish. "Extension of the image model to arbitrary polyhedra". In: J. Acoust. Soc. Amer. 75.6 (1984), pp. 1827–1836.

[21] J. B. Allen and D. A. Berkley. "Image method for efficiently simulating small-room acoustics". In: J. Acoust. Soc. Amer. 65.4 (1979), pp. 943–950.

[22] A. Krokstad, S. Strøm, and S. Sørsdal. "Calculating the acoustical room response by the use of a ray tracing technique". In: J. of Sound and Vibration 8.1 (1968), pp. 118–125.

References VIII

[23] T. Funkhouser, N. Tsingos, I. Calborn, G. W. Elko, M. Sondhi, J. E. West, G. Pingali, P. Min, and A. Ngan. "A beam tracing method for interactive architectural acoustics". In: J. Acoust. Soc. Amer. 115.2 (Feb. 2004), pp. 739–756.

[24] Dejan Marković, Fabio Antonacci, Augusto Sarti, and Stefano Tubaro. "3D beam tracing based on visibility lookup for interactive acoustic modeling". In: IEEE transactions on visualization and computer graphics 22.10 (2016), pp. 2262–2274.

[25] H. Hacıhabiboglu, E. De Sena, Z. Cvetkovic, J. Johnston, and J.O. Smith. "Perceptual Spatial Audio Recording, Simulation, and Rendering: An overview of spatial-audio techniques based on psychoacoustics". In: IEEE Signal Process. Mag. 34.3 (2017), pp. 36–54.

References IX

[26] S. J. Schlecht and E. Habets. "On lossless feedback delay networks". In: IEEE Transactions on Signal Processing 65.6 (2017), pp. 1554–1564.

[27] J. Stautner and M. Puckette. "Designing multichannel reverberators". In: Computer Music J. 6.1 (1982), pp. 52–65.

[28] J-M. Jot and A Chaigne. Digital delay networks for designing artificial reverberators. presented at the 104th Audio Eng. Soc. Conv., Preprint #3030, Paris, France. Feb. 1991.

[29] D. Rocchesso and Julius O. Smith. "Circulant and elliptic feedback delay networks for artificial reverberation". In: IEEE Trans. on Speech and Audio Process. 5.1 (Jan. 1997), pp. 51–63.

References X

[30] S. J. Schlecht and E. Habets. "Feedback delay networks: Echo density and mixing time". In: IEEE/ACM Trans. Audio, Speech, Language Process. 25.2 (2017), pp. 374–383.

[31] S. J. Schlecht and E. Habets. "Time-varying feedback matrices in feedback delay networks and their application in artificial reverberation". In: J. Acoust. Soc. Amer. 138.3 (2015), pp. 1389–1398.

[32] J. O. Smith. "A new approach to digital reverberation using closed waveguide networks". In: Proc. 11th Int. Comput. Music Conf. Burnaby, Canada, 1985, pp. 47–53.

[33] E. De Sena, H. Hacıhabiboğlu, and Z. Cvetković. "Scattering delay network: an interactive reverberator for computer games". In: Proc. 41st Audio Eng. Soc. Int. Conf.: Audio for Games. London, UK, Feb. 2011.

References XI

[34] E. De Sena, H. Hacıhabiboğlu, Z. Cvetković, and J. O. Smith. "Efficient Synthesis of Room Acoustics via Scattering Delay Networks". In: IEEE/ACM Trans. Audio, Speech, Language Proc. 23.9 (Sept. 2015), pp. 1478–1492.

[35] Enzo De Sena, Zoran Cvetkovic, and Huseyin Hacihabiboglu. Electronic device with digital reverberator and method. US Patent 8,908,875. Dec. 2014.

[36] Stojan Djordjevic, Hüseyin Hacihabiboglu, Zoran Cvetkovic, and Enzo De Sena. "Evaluation of the Perceived Naturalness of Artificial Reverberation Algorithms". In: Audio Engineering Society Convention 148. Audio Engineering Society. 2020.

References XII

[37] F. Stevens, D. T. Murphy, L. Savioja, and V. Valimaki. "Modeling Sparsely Reflecting Outdoor Acoustic Scenes using the Waveguide Web". In: IEEE/ACM Trans. Audio, Speech, Language Process. (2017).

[38] William G Gardner and Keith D Martin. "HRTF measurements of a KEMAR". In: The Journal of the Acoustical Society of America 97.6 (1995), pp. 3907–3908.

[39] V Ralph Algazi, Richard O Duda, Dennis M Thompson, and Carlos Avendano. "The CIPIC HRTF database". In: IEEE Workshop Appl. of Signal Proc. to Audio and Acoustics. New Paltz, NY, USA, 2001, pp. 99–102.

References XIII

[40] H. Hacihabiboglu, B. Gunel, and F. Murtagh. "Wavelet-based spectral smoothing for head-related transfer function filter design". In: Proc. of the 22^{nd} AES Conference on Virtual, Synthetic and Entertainment Audio. Espoo, Finland, June 2002, pp. 131–136.

[41] Michal Pec and Pawel Strumillo. "Low-order modelling of head related transfer functions based on spectral smoothing and principal component analysis". In: Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments 2009. Vol. 7502. International Society for Optics and Photonics. 2009, 75021J.

References XIV

[42] Felipe Grijalva, Luiz Martini, Dinei Florencio, and Siome Goldenstein. "A manifold learning approach for personalizing HRTFs from anthropometric features". In: IEEE/ACM Transactions on Audio, Speech, and Language Processing 24.3 (2016), pp. 559–570.

[43] Felipe Grijalva, Luiz Cesar Martini, Dinei Florencio, and Siome Goldenstein. "Interpolation of head-related transfer functions using manifold learning". In: IEEE Signal Processing Letters 24.2 (2017), pp. 221–225.

[44] Matthieu Aussal, François Alouges, and Brian Katz. "HRTF interpolation and ITD personalization for binaural synthesis using spherical harmonics". In: Audio Engineering Society Conference: UK 25th Conference: Spatial Audio in Today's 3D World. Audio Engineering Society. 2012.

References XV

[45] Piotr Tadeusz Bilinski, Jens Ahrens, Mark RP Thomas, Ivan J Tashev, John C Platt, and David E Johnston. HRTF personalization based on anthropometric features. US Patent 10,284,992. May 2019.

[46] Mengfan Zhang, Zhongshu Ge, Tiejun Liu, Xihong Wu, and Tianshu Qu. "Modeling of individual hrtfs based on spatial principal component analysis". In: IEEE/ACM Transactions on Audio, Speech, and Language Processing 28 (2020), pp. 785–797.

[47] Tzu-Yu Chen, Tzu-Hsuan Kuo, and Tai-Shih Chi. "Autoencoding HRTFs for DNN based HRTF personalization using anthropometric features". In: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE. 2019, pp. 271–275.

References XVI

[48] Riccardo Miccini and Simone Spagnol. "HRTF individualization using deep learning". In: 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). IEEE. 2020, pp. 390–395.

[49] Huseyin Hacihabiboglu. "A fixed-cost variable-length auralization filter model utilizing the precedence effect". In: 2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (IEEE Cat. No. 03TH8684). IEEE. 2003, pp. 1–4.

[50] H. Hacihabiboglu and F. Murtagh. "Perceptual simplification for model-based binaural room auralisation". In: Appl. Acoust. 69.8 (Aug. 2008), pp. 715–727.

References XVII

[51] Sami Oksanen, Julian Parker, Archontis Politis, and Vesa Välimäki. “A directional diffuse reverberation model for excavated tunnels in rock”. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE. 2013, pp. 644–648.

[52] Regina E Collecchia, Miriam A Kolar, and Jonathan S Abel. “A computational acoustic model of the coupled interior architecture of ancient Chavı□n”. In: Audio Engineering Society Convention 133. Audio Engineering Society. 2012.

[53] Hansung Kim, Luca Hernaggi, Philip JB Jackson, and Adrian Hilton. “Immersive spatial audio reproduction for vr/ar using room acoustic modelling from 360 images”. In: 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). IEEE. 2019, pp. 120–126.

References XVIII

[54] Nicolas Tsingos, Emmanuel Gallo, and George Drettakis. "Perceptual audio rendering of complex virtual environments". In: ACM Transactions on Graphics (TOG) 23.3 (2004), pp. 249–258.

[55] Paul T Calamia, Benjamin E Markham, and U Peter Svensson. "Diffraction culling for virtual-acoustic simulations". In: Acta Acustica United with Acustica 94.6 (2008), pp. 907–920.

[56] Travis Skare and Jonathan Abel. "Gpu-accelerated modal processors and digital waveguides". In: Linux Audio Conference. 2019.

References XIX

[57] Zhenyu Tang, Nicholas J Bryan, Dingzeyu Li, Timothy R Langlois, and Dinesh Manocha. "Scene-Aware Audio Rendering via Deep Acoustic Analysis". In: IEEE Transactions on Visualization and Computer Graphics 26.5 (2020), pp. 1991–2001.