Perceptual Propagation - Episode VI: Return of the Psychoacoustic Kick-Ass

Chris Malloy and Aidos Abzhanov
University of North Carolina at Chapel Hill

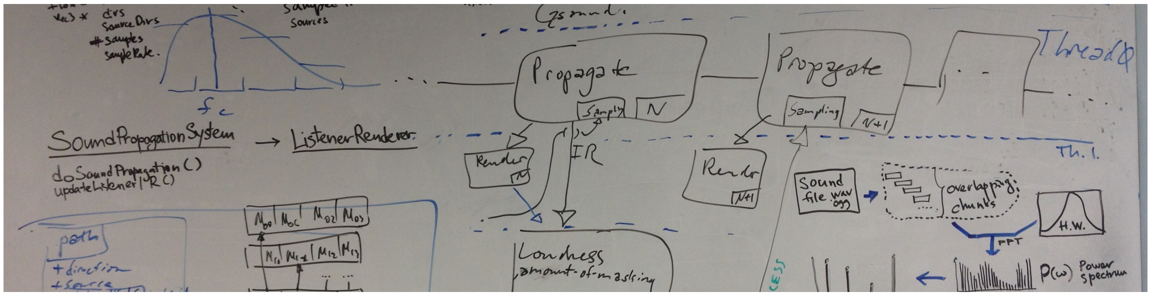

Figure 1: *The System Design.*

Abstract

We introduce a new method for optimizing the performance of path-based sound propagation phase of 3D audio simulation system by leveraging the perceptual feedback from the auralization (and rendering) stage. Our method combines using of psychoacoustics measures to prioritize perceptually important paths and guiding the ray tracing algorithm to better distribute rays to sources ...

CR Categories: I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling—Physically based modeling I.3.5 [Computer Graphics]: Applications—Sound rendering;

Keywords: sound propagation, perceptual importance sampling

1 Introduction

This project aims to produce auditory display pipeline with perceptually optimal sound propagation component (subsystem, state). We force a simulator to focus its computational resources on the perceptually salient regions of an audio (aural/sound) scene, while maintaining the high (same) level of perceptual quality and remaining computationally efficient.

The sound simulation is important area of research and application. Interactive sound simulation is used in many field (games, VR, architectural acoustics). Computation of the sound propagation in complex scenes is either expensive or imprecise. Wave based sound propagation methods can produce accurate results [**?**], but take too much time to be considered for real-time applications. Geometry based methods, on the other hand, provide fast, but approximate solution to the wave equation. The accuracy of geometric (ray-based) approaches can be improved by shooting (generating) more rays (paths), however this directly correlates with the computation time of the simulation. The core idea of our project is to exploit the perceptual limitations of the human auditory system to shoot the rays robustly.

[Small intro on perceptual part ...] It is reasonable to use this psychoacoustic principles to guide the propagation system during ray sampling phase.

We introduce a combination of techniques that improve the running time or/and perceptual quality of the sound propagation system. We first compute perceptual loudness corresponding to the sound signal and the impulse response (IR) of each path from the listener to the sound source. We next use the directional information of the paths to generate spherical distribution of a loudness around the listener, which, combined with HRTF, characterizes how an ear receives a sound from a point in space. We finally integrate the spherical loudness distribution into the next frame's sampling strategy.

Concretely, following are the contributions of our project work:

- a novel approach of using the psychoacoustic principles during the sound propagation stage;
- a new technique for constructing loudness distribution map;
- a partial preprocessing of the perceptual information for maintain real-time level of computations.
- an implementation the proposed methods and integration of them into gSound, an interactive sound propagation and rendering system.

Operating together these contributions constitute a system that can produce a Jedi from any hopeless Padawan.

2 Related Work

Our work is based on ray-tracing methods of sound propagation simulation, and psychoacoustic principles of human auditory system. In this section we give brief overview of related work in geometric sound propagation, perceptual coding, and perceptual audio rendering.

Geometric Sound Propagation [Sound Simulation pipeline overview? short: synthesis, propagation and rendering (diagram).]

There are two main approaches in solving the problem of sound propagation: wave-based [Savioja 2010; Thompson 2006; Gumerov and Duraiswami 2009] and geometric [Funkhouser et al. 2003], with an accuracy and computational complexity being the major trade-offs between them. In wave-based approach the acoustic wave equation being solved numerically, while in geometric approach the problem is reduced to geometric computations assuming the rectilinear propagation of sound waves. Due to the high computational demands of wave-based methods, the field of interactive sound propagation algorithms is primarily dominated by geometric methods, although there are interactive wave-based methods for static scenes [Raghuvanshi et al. 2010; Mehra et al. 2013].

Geometric approach includes algorithms based on image source [Borish 1984], beam tracing [Tsingos et al. 2001], frustum tracing [Chandak et al. 2009], ray tracing [Taylor et al. 2012; Schissler et al. 2014]. Recent advances in ray tracing based approach together with its highly parallel nature makes it stand out among other geometric approachers for interactive sound [auralization (simulation)] applications [Taylor et al. 2010; Schissler and Manocha 2011]. The general idea behind ray tracing based algorithms is to model sound propagation effect by considering different paths between a source and a listener. These propagation paths encode information about the delays and attenuations of sound traveling along the paths, and may consist of any number of reflections and diffractions.

See [Hulusic et al. 2012] for a more comprehensive survey on acoustic rendering and auditory.

Psychoacoustics When sound wave reaches human ear the mechanical energy transforms into neural signal [pulse?], which eventually travels to the brain. This [what?] suggests that taking into account the final signal transformations due to ear and brain may be advantageous for some sound processing applications.

The field of psychoacoustics has made significant progress toward characterizing the time-frequency analysis capabilities of the inner ear [Painter and Spanias 2000]. One of the vivid examples of applying psychoacoustic principles to digital signal processing is perceptual coding. It exploits the human auditory system's inability to hear quantization noise under condition of auditory masking [Pan 1995] to perform perceptually lossless audio signal compression [Ambikairajah et al. 1997]. These were applied to different audio compression formats, including MPEG-1 Audio Layer III (MP3).

The main psychoacoustic principles consist of absolute hearing thresholds, critical band frequency analysis, simultaneous masking, the spread of masking, and temporal masking. Making use of these psychoacoustic notions in the audio simulation system allows ...

Perceptual Audio Rendering Our work is related to the work of [Tsingos et al. 2004], where the psychoacoustic principles are utilized to handle large number of sources and accelerate sound rendering. Similar approach was done in [Moeck et al. 2007] for producing scalable or progressive rendering of complex mixtures of sounds. Our work differs from the previous two in that we intend to leverage psychoacoustic feedback throughout the entire pipeline (including propagation), not only in rendering phase or for clustering.

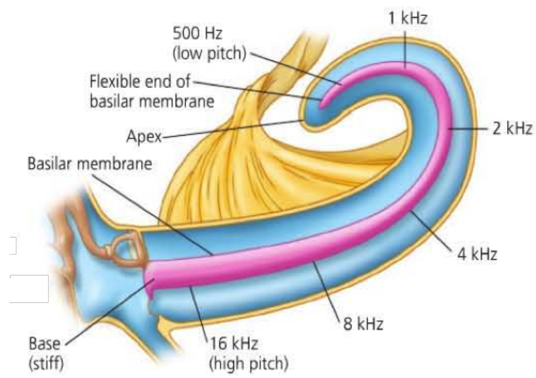

Figure 2: *The basilar membrane*

Spherical Projection Testing. Cite it! Right here, Right now!

3 The Psychoacoustic Model

Psychoacoustics is the science of auditory perception and the human experience of hearing. Limitations induced by the psychological and physiological mechanisms responsible for hearing are of particular interest and have been successfully applied to musical composition, architecture, digital audio coding, and mixing/rendering sound. Individual psychoacoustic effects and illusions are refered to as *psychoacoustic principles* and include:

- The Absolute Threshold of Hearing
- Critical Band Frequency Analysis
- Simultaneous Masking
- Non-Simultaneous Masking
- The Spread of Masking

In this method we seek to demonstrate how these basic psychoacoustic principles can be leveraged in a geometric sound propagation pipeline, with our most novel contribution being the use of masking in the propagation phase of sound simulation.

Auditory Filters and the Bark Scale of Critical Bandwidths The perception of sound frequency is mediated by the **basilar membrane**, a coiled structure of the inner ear that vibrates in response to sound energy. Regions of the basilar membrane have a characteristic frequency they are most sensitive to. This characteristic frequency is highest at the base of the basilar membrane and continuously decreases along it's length. Pure tones result in patterns of excitation on the basilar membrane that may overlap for sounds that are nearby in frequency, and for this reason the basilar membrane is abstracted in literature as a bank of overlapping **auditory filters**. **The Bark frequency scale** is measured in units of the frequency-dependent bandwidths of the auditory filters of the basilar membrane, refered to as the **critical bandwidths**, and is a commonly used frequency scale in psychoacoustics because it is directly proportional to the frequency resolution of human hearing.

4 Overview

The general (high-level) description of the algorithm here.

Spectral Masking Masking thresholds are obtained by performing critical band analysis (with spreading), making a determination of the noise-like or tone-like nature of the signal, applying thresholding rules for the signal quality, then accounting for the absolute hearing threshold [Painter and Spanias 2000]. We assume that the input sound samples are given in advance, and can be pre-processed to extract spectral information about the signal. N samples long Hann window with 50 percent overlap is used to produce the parts of the input sound signal that are processed by psychoacoustic module.

The short time spectrum $X(\omega)$ is computed from the windowed signal chunk $x(i) \;\; i \in [0, N-1]$ using discrete Fourier transform. From this we generate the *power density spectrum* by

$$P(\omega) = Re(X(\omega))^2 + Im(X(\omega))^2, \;\; \omega \in [0, N-1].$$

Signal power for each critical band i is calculated by summing corresponding power density spectra within that frequency band

$$B_i = \sum_{\omega=\omega_{i\_low}}^{\omega_{i\_high}} P(\omega).$$

To account for spreading of masking the critical band signal power is convolved

$$C_i = B_i * SF_i$$

with the *spreading function* SF_i, representing masking across critical bands and given by analytical expression:

$$SF_i = 15.81 + 7.5(\delta i + 0.474) - 17.5\sqrt{1 + (\delta i + 0.474)^2}.$$

Due to the asymmetry of masking there are two masking thresholds: for tone masking noise the threshold is estimated as $14.5 + i$ dB below C_i, while for noise masking tone it is estimated as 5.5 dB below C_i, for each band i.

In order to recognize a tonal or noise-like signal within a certain number of samples, the *spectral flatness measure* (SFM) is estimated

$$SFM_{dB} = 10 \log_{10} \frac{\mu_g}{\mu_a},$$

where μ_g and μ_a are geometric and arithmetic means of the C_i.

The SFM is compared with the SFM of a sinusoidal signal (entirely tonelike signal with SFM = -60 dB) and the *tonality index* is calculated [Johnston 1988] by

$$\alpha = \min(\frac{SFM_{dB}}{-60}, 1).$$

SFM = 0 dB corresponds to a noise-like signal and leads to $\alpha = 0$, whereas an SFM = 75 dB gives a tone-like signal $(\alpha = 1)$.

The tonality index is then used to weight the thresholding rules for each band to form an *offset* between signal level and the masking threshold in critical band i

$$O_i = \alpha(14.5 + i) + (1 - \alpha)5.5 \quad \text{(in dB)}.$$

Finally, a set of *just noticeable difference* JND estimates in the frequency power domain are then formed by subtracting the offsets from the Bark spectral components

$$T_i = 10^{\log_{10}(C_i) - (O_i/10)}.$$

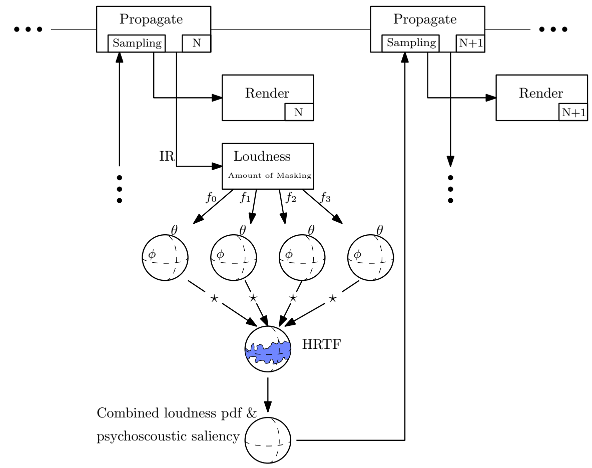

Figure 3: *This is a tiger.*

Coupling with GSound We used gSound, interactive ray-tracing based sound propagation system, to perform the testing of our algorithm, and quickly realized that it is an exceptionally good example of a tool for human torture. Especially, when combined with the MS Visual Studio IDE.

gSound [Schissler and Manocha 2011] contain the following features: backward ray tracing, multi-source clustering, HRTF rendering capabilities.

Here we discuss the design (see Figure 3) for coupling our methodology with gSound.

Results

5 Conclusion and Outlook

Evaluate the existing efficiency of the pipeline based on the distribution of paths in the existing renderer. Conduct user study.

Decouple the propagation to include both backward and forward paths (carrying importance and source information respectively)

Try strategies to connect the paths and guide sampling on both ends to converge more efficiently

Improve sampling schemes based on the state-of-the-art algorithms from graphics field.

Binaural masking.

Incorporate the support of temporal masking (forward, backward)

Go deeper into Rendering part of the system (how can we use psychoacoustic principles there)

Acknowledgements

To Celeste, for the pizza and coffee, and Hasan, for all the bagels.

References

AMBIKAIRAJAH, E., DAVIS, A., AND WONG, W. 1997. Auditory masking and mpeg-1 audio compression. *Electronics & communication engineering journal 9*, 4, 165–175.

BORISH, J. 1984. Extension of the image model to arbitrary polyhedra. *The Journal of the Acoustical Society of America 75*, 6, 1827–1836.

CHANDAK, A., ANTANI, L., TAYLOR, M., AND MANOCHA, D. 2009. Fastv: From-point visibility culling on complex models. In *Computer Graphics Forum*, vol. 28, Wiley Online Library, 1237–1246.

FUNKHOUSER, T., CARLBOM, I., ELKO, G., PINGALI, G., SONDHI, M., AND WEST, J. 1998. A beam tracing approach to acoustic modeling for interactive virtual environments. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, ACM, 21–32.

FUNKHOUSER, T., TSINGOS, N., AND JOT, J.-M. 2003. Survey of methods for modeling sound propagation in interactive virtual environment systems.

GUMEROV, N. A., AND DURAISWAMI, R. 2009. Wideband fast multipole accelerated boundary element methods for the three-dimensional helmholtz equation. *The Journal of the Acoustical Society of America 125*, 4, 2566–2566.

HULUSIC, V., HARVEY, C., DEBATTISTA, K., TSINGOS, N., WALKER, S., HOWARD, D., AND CHALMERS, A. 2012. Acoustic rendering and auditory–visual cross-modal perception and interaction. In *Computer Graphics Forum*, vol. 31, Wiley Online Library, 102–131.

JOHNSTON, J. D. 1988. Estimation of perceptual entropy using noise masking criteria. In *Acoustics, Speech, and Signal Processing, 1988. ICASSP-88., 1988 International Conference on*, IEEE, 2524–2527.

KIRK, A. G., AND O'BRIEN, J. F. 2011. Perceptually based tone mapping for low-light conditions. *ACM Trans. Graph. 30*, 4, 42.

MANOCHA, D., AND LIN, M. C. 2009. Interactive sound rendering. In *Computer-Aided Design and Computer Graphics, 2009. CAD/Graphics' 09. 11th IEEE International Conference on*, IEEE, 19–26.

MEHRA, R., RAGHUVANSHI, N., ANTANI, L., CHANDAK, A., CURTIS, S., AND MANOCHA, D. 2013. Wave-based sound propagation in large open scenes using an equivalent source formulation. *ACM Transactions on Graphics (TOG) 32*, 2, 19.

MOECK, T., BONNEEL, N., TSINGOS, N., DRETTAKIS, G., VIAUD-DELMON, I., AND ALLOZA, D. 2007. Progressive perceptual audio rendering of complex scenes. In *Proceedings of the 2007 symposium on Interactive 3D graphics and games*, ACM, 189–196.

PAINTER, T., AND SPANIAS, A. 2000. Perceptual coding of digital audio. *Proceedings of the IEEE 88*, 4, 451–515.

PAN, D. 1995. A tutorial on mpeg/audio compression. *IEEE multimedia 2*, 2, 60–74.

RAGHUVANSHI, N., SNYDER, J., MEHRA, R., LIN, M., AND GOVINDARAJU, N. 2010. Precomputed wave simulation for real-time sound propagation of dynamic sources in complex scenes. In *ACM Transactions on Graphics (TOG)*, vol. 29, ACM, 68.

SAVIOJA, L. 2010. Real-time 3d finite-difference time-domain simulation of low-and mid-frequency room acoustics. In *13th Int. Conf on Digital Audio Effects*, vol. 1, 75.

SCHISSLER, C., AND MANOCHA, D. 2011. Gsound: Interactive sound propagation for games. In *Audio Engineering Society Conference: 41st International Conference: Audio for Games*, Audio Engineering Society.

SCHISSLER, C., AND MANOCHA, D. 2015. Interactive sound propagation and rendering for large multi-source scenes.

SCHISSLER, C., MEHRA, R., AND MANOCHA, D. 2014. High-order diffraction and diffuse reflections for interactive sound propagation in large environments. *ACM Transactions on Graphics (SIGGRAPH 2014) 33*, 4, 39.

STUART, J. R. 1996. The psychoacoustics of multichannel audio. In *Audio Engineering Society Conference: UK 11th Conference: Audio for New Media (ANM)*, Audio Engineering Society.

TAYLOR, M., CHANDAK, A., MO, Q., LAUTERBACH, C., SCHISSLER, C., AND MANOCHA, D. 2010. i-sound: Interactive gpu-based sound auralization in dynamic scenes. Tech. rep., Tech. Rep. TR10-006.

TAYLOR, M., CHANDAK, A., MO, Q., LAUTERBACH, C., SCHISSLER, C., AND MANOCHA, D. 2012. Guided multi-view ray tracing for fast auralization. *Visualization and Computer Graphics, IEEE Transactions on 18*, 11, 1797–1810.

THOMPSON, L. L. 2006. A review of finite-element methods for time-harmonic acoustics. *J. Acoust. Soc. Am 119*, 3, 1315–1330.

TSINGOS, N., FUNKHOUSER, T., NGAN, A., AND CARLBOM, I. 2001. Modeling acoustics in virtual environments using the uniform theory of diffraction. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, ACM, 545–552.

TSINGOS, N., GALLO, E., AND DRETTAKIS, G. 2004. Perceptual audio rendering of complex virtual environments. In *ACM Transactions on Graphics (TOG)*, vol. 23, ACM, 249–258.

TSINGOS, N. 2005. Scalable perceptual mixing and filtering of audio signals using an augmented spectral representation. In *8th International Conference on Digital Audio Effects (DAFx 2005)*, 6.