Procedural 3D Audio for AR Applications

Angeliki Skandalou
John Jeremy Ireland

Supervisor:
Michael Rose

Kongens Lyngby 2017

Abstract

This is the first paragraph

THis is the second

THird paragraph of abstract

Four paragraphs is enough I guess

Acknowledgements

We would like to express our gratitude and appreciation to our supervisor for his support and guidance throughout this thesis work. Several discussion sessions and advice helped us take the most out of this project and make this study possible.

We would like to express special thanks also to the rest of our classmates who did their thesis at the same time under the same supervisor and offered us their advice. And last but not least to our family and friends whose support throughout this thesis was invaluable.

List of Figures

List of Tables

Contents

Abbreviations

ASW Apparent Source Width.

AVIL Audio Visual Immersion Lab.

BRIR Binaural Room Impulse Response.

CS Compressive Sensing.

DOA Direction of Arrival.

ERB Equivalent Rectangular Band.

HATS Head And Torso Simulator.

HOA Higher Order Ambisonics.

HRTF Head-Related Transfer Function.

IACC Inter-Aural Cross Coherence.

ILD Inter-Aural Level Difference.

ITD Inter-Aural Time Difference.

STFT Short-Time Fourier Transform.

WFS Wave Field Synthesis.

Nomenclature

$\mathbf{\Omega}_{LS}$ Vector containing directions of Loudspeakers in reproduction.

$\mathbf{\Omega}_L$ Grid of directions used for the CS algorithm.

$\mathbf{\Omega}_s$ Subvector of $\mathbf{\Omega}_L$ containing only the prominent directions after CS processing.

$\breve{\mathbf{H}}$ Combined transfer matrix for mixed-norm problem.

$\breve{\mathbf{p}}$ Combined measurement pressure vector for mixed-norm problem.

$\mathbf{x}$ Combined amplitude.

ℓ_p Norm-p.

$\mathbf{H}$ Transfer Matrix for plane waves impinging on rigid sphere.

$\mathbf{p}$ Measurement vector for the pressure on the spherical array.

$\mathbf{x}$ Amplitude vector for plane waves impinging on the sphere.

$\widetilde{\mathbf{p}}$ Pressure vector reconstructed from prominent plane waves.

B_n^m Ambisonics coefficients.

L Number of plane waves in a discrete grid of directions.

LS Number of Loudspeakers in reproduction.

N Truncation order for the spherical Harmonic Functions.

P_n^m The associated Legendre polynomials of the first kind.

Q Number of sampling points on the spherical microphone array.

R_0 Radius of reproduction area.

Y_n^m Spherical harmonic Functions.

Ω Angular Dependency on both azimuth and inclination angle.

λ Regularization factor for natural field HOA processing.

$\mathbf{B}_N$ Ambisonics coefficients vector truncated at order N.

$\mathbf{S}$ Loudspeaker signals resulting from HOA decoding.

$\mathbf{W}$ Vector containing radial functions W_n.

$\mathbf{Y}_N(\mathbf{\Omega}_L)$ Spherical harmonics vector truncated at order N for all measurement angles in vector $\mathbf{\Omega}_L$.

$\mathbf{p}'$ Residual pressure.

ε Noise parameter for Compressive Sensing Algorithm.

a Radius of microphone array.

“w/ Residual” Exploiting the residual pressure (full implementation of signal path in Figure **??**).

“w/o Residual” Residual pressure is neglected (only upper path in Figure **??**).

Chapter 1

Introduction

Immersion and all these stuff that makes our thing good. Why we are doing it and what do we want to give to the community?

Audio in interactive projects like video games and VR/AR applications, plays a significant role for user immersion and realism. Visual and acoustic experiences are interconnected and lacking one of them spoils the whole experience.

The most difficult task is to produce realistic virtual sounds inside the application, difficult to distinguish them from the real ones. This can be achieved not only by playing back a realistic sound, but also by taking care of the environment effects and the context. For example, striking a nail on a board when it still vibrates from the previous struct, produces a different sound that gets added to the previous one [2].

Why is our method better that others? (eg wavetable)? And why we think this is the future of the audio in video games?

Chapter 2

Theoretical Background

Short overview of the theory parts
This is a way to link to explanations Direction of Arrival (DOA)

THis is a todo: **To do** do smth (1)
THis is smth done:

2.1 State-Of-The-Art

2.2 Modal Analysis

In this thesis we are using solid objects that are struck in different ways to produce sound. These ways could be falling on the floor or colliding with another object. The sounds produced can be impact, rolling or scratching sounds. When an object is struck, the forces applied cause deformations to it, emitting sound waves through the vibration of its outer surfaces [4].

Modal analysis studies the response of models under excitation. It uses the 3D model of an object to calculate its modal modes (vibration modes). There are multiple ways to do this, with the most accurate being FEM (Finite Element Method). The objective of FEM is to calculate the natural frequencies of a structure when it vibrates freely.

2.2.1 Data Extraction

Modal analysis is performed before modal synthesis, to extract the necessary data. Modal synthesis is the sum of damped oscillators each corresponding to a modal frequency, as it will be discussed further below. The data needed for synthesis are shown in the table 2.1.

Since every different point being struck produces different deformations on the object, we need matrices of size N (N being the number of struck points of the object). More specifically,

Symbol	Description	Derivation
A_n	Initial amplitude	Modal analysis
d_n	Damping	Material properties
f_n	Modal frequency	Modal analysis

Table 2.1: Data extracted in modal analysis.

we need a vector **f** of size **N** corresponding to the modal frequencies of every point, a vector **d** of size **N** corresponding to the damping ratios and a matrix **A** of size **NxK**, where K is the number of modal frequencies calculated in one point, which corresponds to the amplitudes of each mode in every point of the object. All the above gives the modal model which can be symbolized as $\boldsymbol{M = \{f,\ d,\ A\}}$ [4].

2.3 Modal Synthesis

In the modal synthesis part, using the data extracted above, we synthesize the struck sound corresponding to the object. There are different ways to synthesize impact sounds, two of them being "Sinusoidal Additive Synthesis" and "Filter-based Modal Synthesis". The former uses exponential damping and the latter band-pass filters where the damping is the Q-factor of the filter.

2.3.1 Sinusoidal Additive Synthesis

At a struck point k when vibrating in mode n, the impulse response of the model is:

$$y_k = \sum_{n=1}^{N} A_{nk}\ e^{-d_n t}\ \cos(2\pi f_n t) \tag{2.1}$$

if $t > 0$ and $y_k = 0$ if $t <= 0$ [4].

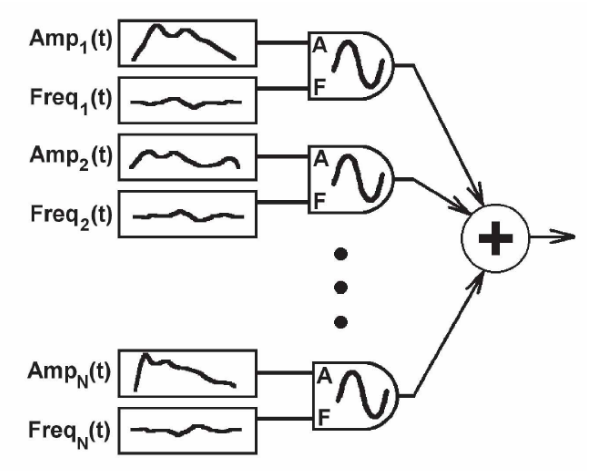

Figure 2.1: Sinusoidal Additive Synthesis Algorithm [2].

2.3.2 Filter-based Modal Synthesis

Band-pass Filters

At this point we will give some basic description of the band-pass filter since it is widely used in this thesis. Band-pass filters (BPFs) take a signal as input and give only a range of it as output, attenuating the rest of the frequencies. This range depends on the central frequency f_c. A filter of this kind is a result of a cascading of a low-pass and a high-pass filter circuit.

The passing range or "band" of frequencies is called **Bandwidth (BW)**. Defining as 0db the resonant peak, we can find the two cut-off frequencies ($f_{c_{\text{LOWER}}}$ and $f_{c_{\text{HIGHER}}}$) at -3dB. The range between them is the bandwidth (equation 2.2). In figure 2.2 we can see the frequency response of a BPF. [1].

$$BW = f_{c_{\text{HIGHER}}} - f_{c_{\text{LOWER}}} \tag{2.2}$$

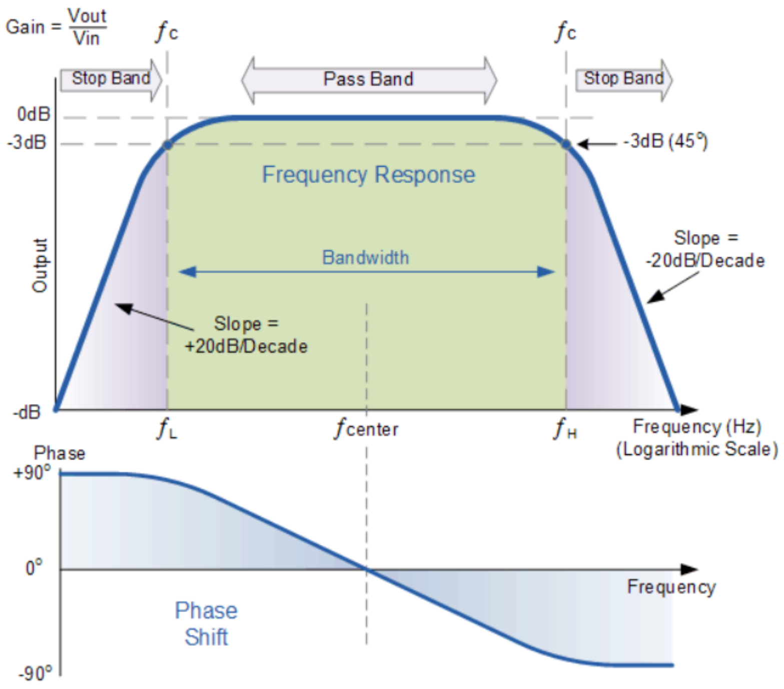

Figure 2.2: Frequency Response of a Band-pass Filter [1].

Synthesis

This method is also additive, since we are adding the outputs of a number of band-pass filters. To synthesize a sound using this method, we use as many filters as the modal frequencies. The filter takes as input an impulse, the center frequency which is the modal frequency and a **Quality factor (Q-factor)** which specifies the bandwidth of the filter. The Q-factor is calculated heuristically, depending on the material of the sound and is inversely proportional to the bandwidth, so the lower the Q-factor, the wider the bandwidth and vice-versa. Hence, more and less frequencies respectively will be included in the audible range. We call the above structure a *resonator*, which also includes a multiplication with the corresponding amplitude, taken from the A matrix.

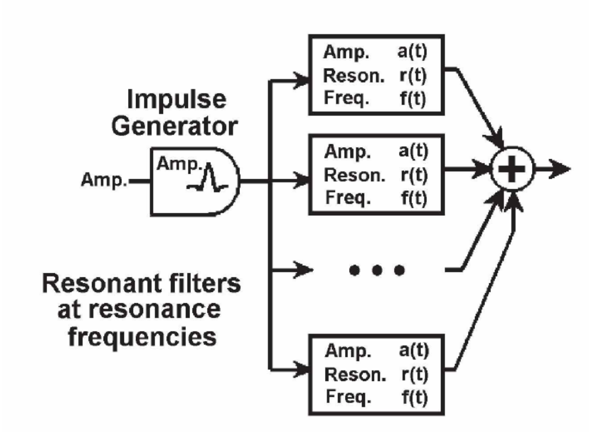

Figure 2.3: Filter-based Modal Synthesis Algorithm [2].

CHAPTER 3

Method

A combination of the methods described in Chapter 2 is proposed in the present study.

3.1 Chuck language

Modal features extraction code

3.2 PureData

Resynthesis patches

3.3 Heavy Compiler

3.4 Unity

3.5 Overview

Chapter 4

Measurements

Here we can describe the audio recordings and put pictures

Chapter 5

Implementation

Here we can put pictures and codes snippets

5.1 Impact Sounds

5.1.1 Sinusoidal Additive Synthesis

5.1.2 Filter-based Modal Synthesis

5.2 Rolling Sounds

5.3 Scratching Sounds

5.4 User Interface

Chapter 6

Results & Discussion

6.1 Which Synthesis Method Is Better?

6.2 Did we manage to achieve what we wanted?

6.3 How can we improve our work?

CHAPTER 7

Conclusion

This is the conclusion
4-5 paragraph approx

Appendix A

Results of tests to users

Appendix B

User Guide to our product

Bibliography

[1] Inc. AspenCore. Passive band pass filter. `http://www.electronics-tutorials.ws/filter/filter_4.html`, Accessed: April 28,2017.

[2] Perry R. Cook. *Real Sound Synthesis for Interactive Applications.* A. K. Peters, Ltd., Natick, MA, USA, 2002.

[3] D Brandon Lloyd, Nikunj Raghuvanshi, and Naga K Govindaraju. Sound synthesis for impact sounds in video games. In *Symposium on Interactive 3D Graphics and Games*, pages PAGE–7. ACM, 2011.

[4] Kees Van Den Doel, Paul G Kry, and Dinesh K Pai. Foleyautomatic: physically-based sound effects for interactive simulation and animation. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 537–544. ACM, 2001.

To do. . .

- ☐ 1 (p. 3): do smth
- ☑ 2 (p. 3): this is done