Perceptual Spatial Audio recording, Simulation and Rendering

A report by Sai Patil
(19200089)

Contents:

1. Introduction

Physically motivated techniques

Physically motivated techniques aim to reproduce a physically accurate approximation of desired sound fields by employing a very high equipment load and sophisticated, computationally intensive.

Perceptually motivated techniques

Perceptually motivated techniques, however, aim to render only the perceptually relevant aspects of the sound scene by means of modest computational and equipment

2. The Spatial Audio Scene

2.1 Auralization

2.2 Ways to generate an experience of spatial sound scene :
2.2.1) Binaural
2.2.2) Accurate physical approximation

2.3 Recording and reproduction techniques:
2.3.1) Physically motivated methods
2.3.2) Artificial reverberators

2.1 Auralization

-Spatial sound system is virtual
- the required playback signals need to be synthesized rather than recorded.
-To that end, an accurate approximation of the desired sound field would ideally be computed and then recorded using a virtual microphone array, to be played via the corresponding actual loudspeaker array. This process is referred to as auralization.

2.2 Ways to generate an experience of spatial sound scene : Binaural

- convincing experience over two channels by presenting stereophonic audio cues
- work best over headphones
- Limitations : Individualization

2.2 Ways to generate an experience of spatial sound scene : Accurate Physical Approximation

- Examples :wave field synthesis (WFS) and higher-order Ambisonics (HOA)
- WFS uses Huygens principle and Kirchhoff–Helmholtz integral. It can be widely used in film theatres.
- HOA is capable of achieving results comparable to WFS close to the center of the reproduction rig.
- they have high equipment load requirements thus their domain is confined

2.3 Recording and reproduction techniques: Physically motivated methods

- aim to calculate an approximate solution of the wave equation.
- several numerical methods have been developed that achieve a very high level of accuracy.
- prohibitively high computational costs.
- Examples : the finite-difference time domain, finite element method (FEM), and boundary element method (BEM).
- computational cost is still too high for real-time operation at interactive rates and on low-cost devices.

2.3 Recording and reproduction techniques: Artificial Reverberators

- try to render only some higher-level perceptual effects.
- require only a fraction of the computational load associated with physically motivated room simulators and
- typically aim to mimic only certain characteristics of the tail of typical room impulse responses, such as modal density, echo density, and timbral quality.
- Do not explicitly model a given physical space but, rather, are used to obtain a pleasing reverberant effect and have been widely used for artistic purposes in music production.

4. The history of perceptually motivated spatial Audio

4.1 Binaural audio

4.2 Two-channel stereophony

History : Binaural Audio

Assumption : if the signals that would be received at the ears of a listener as a result of an acoustic event are provided to the listener with sufficient accuracy, the person will perceive an auditory event corresponding to the original acoustic event.

- Uses Dummy Head microphones which are made to resemble a human head
- **Requirements of binaural synthesis** are
  1) the availability of a set of HRIR measurements densely sampled on a spherical shell and
  2) the match between these HRIRs and the actual HRIRs of the listener.

- Binaural synthesis also allows interactivity if the position and orientation of the listener's head can be tracked

- In the case of a sound field created by P sources in the far field, the right and left ear signals can be synthesized

- Binaural audio can also be presented via a pair of loudspeakers (consider Crosstalk)

History : Two channel Stereophony

- Requires the minimal number of channels to produce the impression of spatial sound.

- uses two loudspeakers, each at the same distance from the listener, positioned 30° to either side of the front direction, providing a frontal auditory scene within a base angle of 60°.

- The ideal listening position, referred to as the sweet spot, thus forms an equilateral triangle with the loudspeakers.

- Two-channel stereophony creates the illusion of a sound source in a given direction within the base angle by means of the inter- channel time differences (ICTDs) and interchannel level differences (ICLDs) of the two channels over which the source signal is presented.

5. Perceptually motivated multichannel recording and reproduction

5.1 Vector-base amplitude panning(VBAP)

5.2 Spatial encoding methods

5.3 Perceptual sound field reconstruction

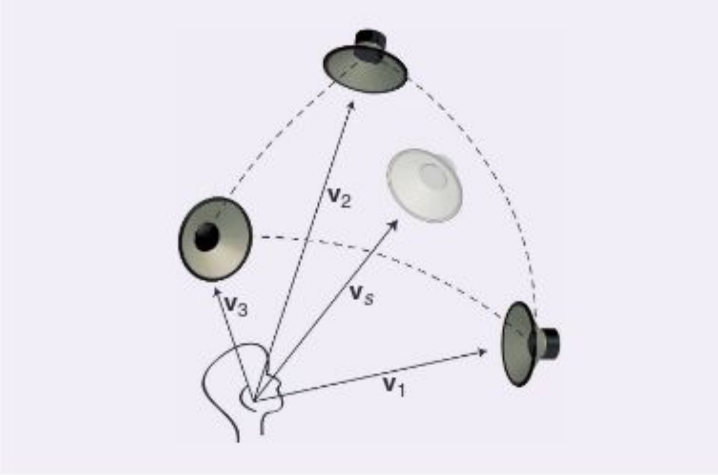
5.3 Enlarging the optimal listening area

5.1 VBAP (Vector-based amplitude panning)

Originally, VBAP was designed for a loudspeaker array with elements placed on a geodesic dome vertices that are situated at the acoustic far field of the listener. Figure shows
a section of such a sphere with three loudspeakers, with a listener positioned at the center of the array.

Assumption : VBAP in three dimensions is that summing localization would occur not only with two sources
but also with three.

This assumption was subjectively tested for different setups and virtual source directions,
and it was shown to result in a good subjective localization accuracy for elevated virtual sources



5.2 Spatial Encoding Methods

A class of multichannel audio methods involves dividing recorded signals into time or time–frequency bins and estimating certain spatial attributes within each bin.

One of these methods is the spatial impulse response rendering (SIRR) method

It is assumed that each time–frequency bin corresponds to a single plane wave and thus that the direction of the acoustic intensity vector also represents the direction of that plane wave.

Perceptual sound field reconstruction

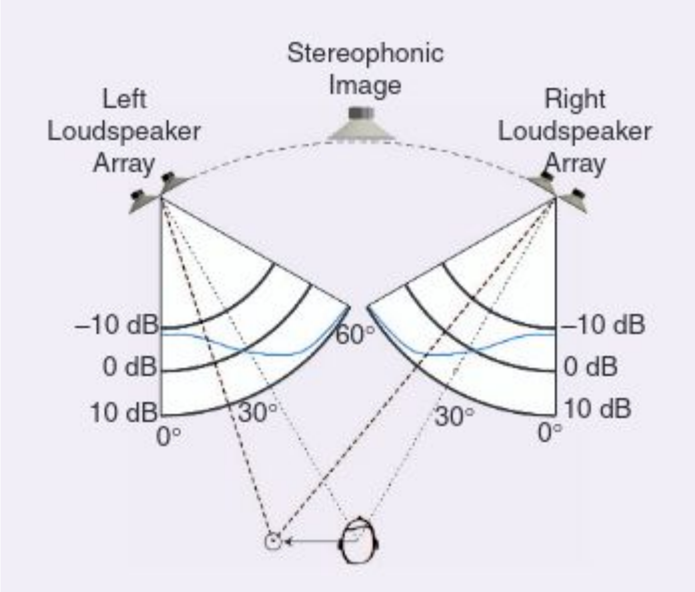
it provides a systematic framework for recording and reproduction of sound scenes.

PSR relies on designing the underlying microphone arrays in a way that captures the required directional cues.

When the recorded signals are played back with no additional processing, the directions of the wave fronts of all sound sources and all reflections are rendered accurately.

5.3 Enlarging the optimal listening area

When a listener moves away from the center of the sweet spot, the auditory event shifts in the direction of the closest loudspeaker.
This is due to the fact that the signal from the closest loudspeaker arrives earlier when compared with what is observed at the sweet spot.
Position-independent stereo [57], [58] aims to alleviate this problem by designing loudspeaker directivity patterns in a manner that compensates for the incongruent time delay via appropriate intensity differences.



The design method consists of two separate **optimization** procedures.

1) The first procedure involves finding a common directivity pattern for the left and right loudspeakers of a stereophonic setup to provide the level differences needed to compensate for the incongruent time differences over a desired listening area. Such directivity patterns can be obtained by beamforming using an array of loudspeakers.

2) The other optimization procedure involves finding the filter coefficients to be used for beamforming

6. Perceptually motivated room auralization

Auralization is the process of making the acoustics of a real or virtual environment such as a room or a concert hall audible

6.1 Simplification of room acoustics models

6.2 Perceptually motivated artificial reverberation

6.3 Audio source culling

6.1 Simplification of room acoustics models

- Comprehensive mathematical model lacks the precedence effect.
- Thus, analogous to models of monaural masking, has made it difficult to predict whether an individual reflection would be audible in the presence of the direct sound and other reflections.
- This is mainly due to the fact that the audibility of a reflection depends on many parameters.
- One of the first models that aimed to parameterize the audibility of reflections, named reflection masked threshold (RMT),

6.2 Perceptually motivated artificial reverberation

Artificial reverberation in a room auralization system is dictated not only by perceptual considerations but also by computational cost, and the holy grail in artificial reverberator design is an algorithm that can achieve good perceptual quality at a reasonable computational cost.

The earliest digital artificial reverberators consisted of comb filters connected in parallel to simulate the frequency modes of a room and all-pass filters to simulate a dense reverberation tail

6.3 Audio source culling

State-of-the-art game engines typically use volumetric culling of sound sources. Each sound source has an associated culling volume (cube, sphere, or cylinder), and when the listener is within this volume, the sound is rendered.

Loudness values are used to calculate a masking threshold from a time–frequency representation of the sound sources and stored for use during

runtime. As a new sound event occurs, the decision to render the new sound source is made at the audio-frame level. Each

frame is compared with the existing mix for evaluating whether the mix can mask it. If it can, the frame is culled.

Conclusion

- While developments in computer hardware could make it possible to overcome issues due to computational limitations, physical limitations such as the size of electroacoustic transducers or data bandwidth will remain.

-Similarly, the energy cost of carrying out simple operations such as multiplication or memory access is likely to diminish but will never vanish,and the power efficiency of mobile devices will also continue to be relevant.

-These issues will make it even more desirable to design simpler audio systems and algorithms. Thus we should use our existing knowledge of auditory perception to make things simple.