A REVIEW ON METHODS AND APPLICATIONS OF 3D SPATIAL SOUND SYNTHESIS

JAYANTA KURUMATHUR

R.V. College of Engineering
E-mail: jayanta.k@outlook.com

Abstract— The global advent of Virtual Reality and Augmented Reality has brought on the need for dynamic 3D sound synthesis systems. Spatial Audio Object Coding coupled with Head Related Transfer Functions and Wave Field Synthesis provide a good base for the development of such a robust system.

Index Terms— Audio, Hrtf, Localisation, Spatial.

I. INTRODUCTION

3D sound has been a focus of audio engineering for many years, beginning with the onset of stereo support. Stereo audio signals allowed for a single audio piece to contain information pertinent to both ears. Initially a large part of the technology was used not for spatial differentiation, but for aesthetic improvement, with a certain audio track being louder or sharper in a certain channel.

The next step in the development of 3D audio technology was the concept of binaural audio. Simple binaural audio consisted of a setup that simulates a human head, with a microphone in each ear, with all the audio sources around it in the orientation that is required for the audio. No special algorithms were required as the sounds were physically separated.

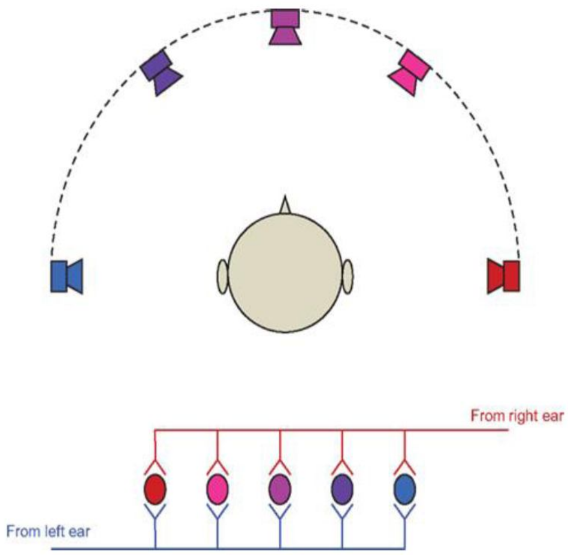

Fig. 1. Binaural Audio

The advent of synthetic sound brought with it a much larger set of challenges. Not only was it required to individually layer sound tracks to create new audio, but a new generation of visual simulation technology was emerging which required its own audio simulations as well.

Driving simulators, aeroplane simulators for pilots, any POC (proof-of-concept) support simulation technology required a robust 3D audio synthesis technology to go with it. Virtual Reality (VR) systems that supported similar features were in need of even more advanced 3D audio synthesis, as the audio must dynamically shift and move as the user interacts with the virtual environment.

The human brain uses certain parameters obtained from the ears to determine where a sound is coming from. These are known as localisation cues.

II. LOCALISATION CUES

A. Interaural time

The time difference between the time of arrival at the left and right ears can be used to calculate the angle of approach.

B. Interarual level

The amplitude difference between the signal at the left and right ears can be used to accurately calculate the distance and the direction.

C. Doppler shift

The Doppler shift between the ears allows for definition of the speed of the source relative to the listener. This is important for keeping the localisation information accurate for a moving source.

D. Low Frequencies

For low frequencies, the wavelength is larger than the diameter of the head, allowing for accurate determination of phase difference between the two ears.

E. Precedence effect

The precedence effect is an effect that explains the effect of reflected sounds on localisation. It has been seen that only the first iteration of any audio wave on the ears is used for localisation.

F. Critical bands

The human ear is sensitive to sounds within what are known as the 24 critical bands of sound. Localisation is only done for sounds within the critical bands.

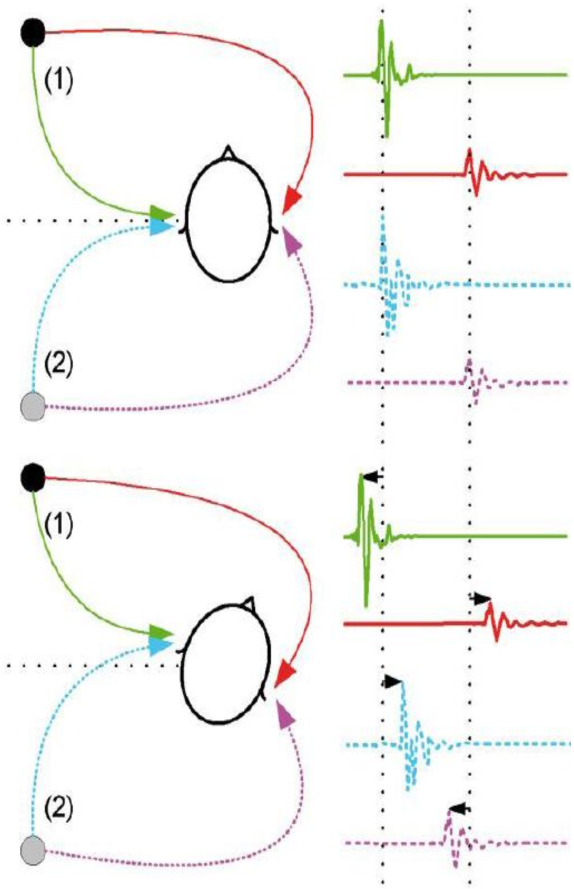

Fig. 2. Interaural differences

G. Cocktail party effect

The importance of the critical bands also allows for the listener to isolate the sounds which are important to them. This allows for better localisation, the ability to ignore the parts of the sound that are not relevant to the localisation cues.

III. HEAD RELATED TRANSFER FUNCTION

3D audio synthesis can be developed using the Head Related Transfer Functions (HRTF). The HRTF shows the properties of a sound source relative to a listener depending on the direction [1].

There are many different parameters that affect the HRTF. Azimuth, elevation and frequency are some important distance-based examples of these. These are modeled into the HRTF using principal component analysis (PCA), which is a linear technique. PCA is a popular technique used in data mining. It takes the most similar vectors of the dataset, and represents the database using these vectors [2].

Even though the easiest method of HRTF creation is experimental measurement, it is not a feasible technique. Since the HRTF must be designed in an echo-free environment and each HRTF is significantly affected by the distance from each audio source, they are very specific for every different scenario. Therefore, databases providing HRTF information are very limited.

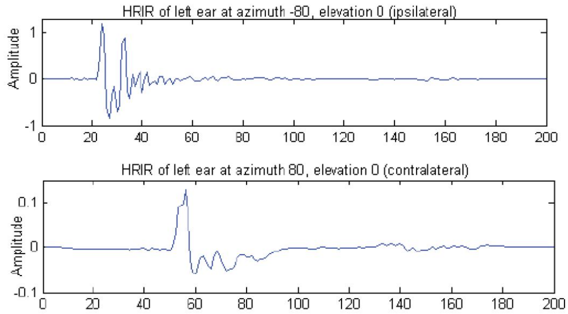

Fig. 3. HRTF of different ears

The automatic generation of a suitable HRTF is very difficult, but may be done by various methods. Sparse representation is an important example of one such technique. The data obtained from the trained HRTFs are layered linearly over the subject's individual characteristic parameters. Only the magnitudes are taken, as opposed to the more intensive techniques.

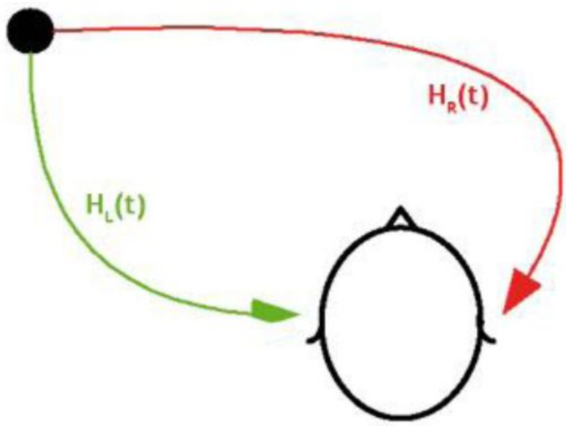

Fig. 4. HRTF for each ear

The characteristic features are known as anthropometric features, and are divided according to their scale. They are the features of the ear, head and full body respectively.

IV. SPATIAL AUDIO OBJECT CODING

SAOC is a technique in which many audio objects are dealt with at the same time, which makes it significantly more effective than traditional audio coders. Each audio object is also given its own individual parameters and structure that persists, allowing the SAOC to control each one according to the user's requirements [3].

This is extremely important for the application of spatial sound as it allows the system to conduct dynamic modification of the HRTF, which is conducive to environments such as Virtual Reality systems or Augmented Reality systems.

The SAOC works by down mixing the audio signal with the spatial features. The bit rate need not be significantly high as it only uses the bandwidth for the down-mix signal. It also allows for less suppression of each channel, due to the fact that each object may be individually suppressed, allowing for less distortion.

V. WAVE FIELD SYNTHESIS

Wave field synthesis is an alternative methodology used for the reconstruction of 3D acoustic signals. It is developed as a direct usage of the Huygens' Principle. The Huygens'

Principle states that and point on a wave front can be represented as a point source which is creating spherical wave fronts [4].

Similarly, in wave field synthesis, each wave field is created by replacing the source point with a continuous array of secondary sources. This allows the system to dynamically generate the spatial sounds equivalent to that produced at the source.

At the receiver, the audio signal will need multichannel system which are capable of using extended displays. The wave field synthesis has the disadvantage of a very big speaker array.

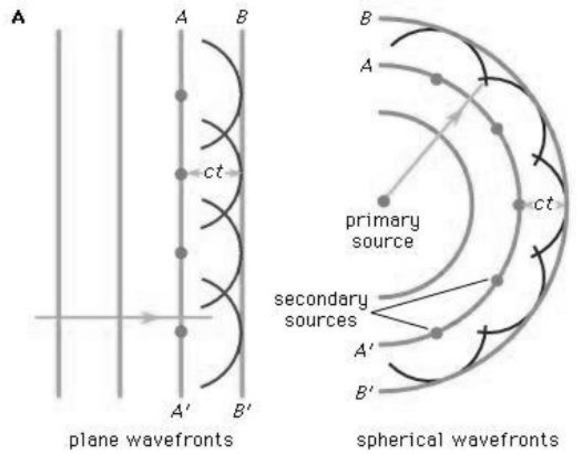

Fig. 5. Huygens principle

A large number of modern simulation technology is described in the form of VR, which has shown the importance of 3D audio. Even though video games have been employing this technology for many years, it has not developed into a state where it can be used in a self sufficient manner for Virtual Reality applications.

Wave field synthesis works best in conjunction with audio object coding, due to the relative ease with which the objects may be manipulated, compared to the multichannel approach.

VI. VIRTUAL REALITY

The recent advancement in the field of Virtual Reality has been key in the push for advancement of 3D audio. Any VR experience must be accompanied with a suitable audio experience to allow for a good system, because the user must be able to hear all the physical objects that they see in the virtual reality environment that they are placed in. For example, if the virtual reality system shows a machine to the left of the user, the audio must support that image and express the sound of the machine from the user's left, otherwise the illusion is broken [5].

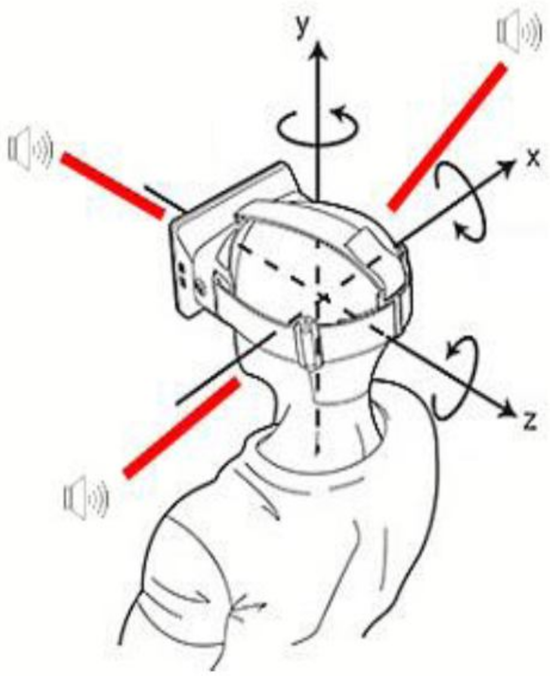

Fig. 6. 3D audio in VR

VII. AUGMENTED REALITY

Similar to VR systems, Augmented Reality (AR) systems require similar audio technology to support all the augmented objects projected into the world, all with their own audio object data.

VIII. PEDESTRIAN LOCALISATION

3D audio systems are also used in short distance wireless communication systems. Each individual with their own transmitter and receiver is treated as an audio object for each of the others. The audio processing systems created an HRTF for each user, causing the sounds to seem like they are approaching from the actual angle where they are. This is extremely useful in military and other high stress scenarios where the location of each user is critical [6].

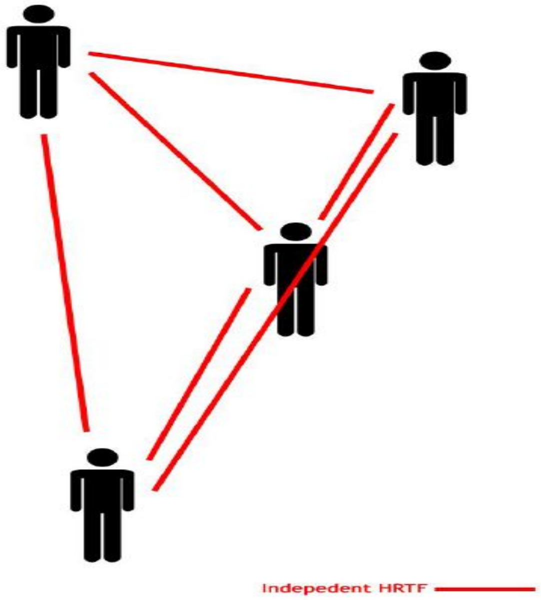

Fig. 7. Pedestrian localisation

CONCLUSION

3D audio has a wide scope for improvement and the applications it supports will cause the field to increase the development. The scalability supported by SAOC and HRTF allow for the wide scale development. The advent of VR and AR will continue to expand, growing alongside the 3D audio system.

REFERENCES

[1] Qinghua Huang, Kai Liu, Yong Fang, "Tensor modeling and interpolation for distance-dependent head-related transfer function", ICSP 2015 © IEEE 2014

[2] Piotr Bilinski, Jens Ahrens, Mark Thomas, Ivan Tashev, John Platt, "HRTF magnitude synthesis via sparse representation of anthropometric features", 2014 IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP) © IEEE 2014

[3] Kwangki Kim, Jeongil Seo, Seungkwon Beack, Kyeongok Kang, and Minsoo Hahn, "Spatial audio object coding with two-step coding structure for interactive audio service", IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 13, NO. 6, DECEMBER 2011 © IEEE 2011

[4] H. Lim, C. Kim, E. Ekmekcioglu, S. Dogan, A.P. Hill, A.M. Kondoz, and X. Shi, "An approach to immersive audio rendering with wave field synthesis for 3d multimedia content", ICIP 2014 © IEEE 2014

[5] Kyungryeol Koo, Kwangki Kim, Jeongil Seo, Kyeongok Kang, Minsoo Hahn, "Variable subband analysis for high quality spatial audio object coding", ICACT 2008

[6] John-Olof Nilsson, Christian Schuldt, and Peter Handel, "Voice radio communication, pedestrian localization, and the tactical use of 3D audio", 2013 International Conference on Indoor Positioning and Indoor Navigation, 28th-31th October 2013

★★★