

Development of Hardware and Software for a Game-like
Wireless Spatial Sound Distribution System

by

Chinmay Dharmadhikari

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved April 2016 by the
Graduate Supervisory Committee:

Andreas Spanias, Chair
Pavan Turaga
Todd Ingalls

ARIZONA STATE UNIVERSITY

May 2016

ABSTRACT

Several music players have evolved in multi-dimensional and surround sound systems. The audio players are implemented as software applications for different audio hardware systems. Digital formats and wireless networks allow for audio content to be readily accessible on smart networked devices. Therefore, different audio output platforms ranging from multispeaker high end surround systems to single unit Bluetooth speakers have been developed. A large body of research has been carried out in audio processing, beamforming, sound fields etc. and new formats are developed to create realistic audio experiences.

An emerging trend is seen towards high definition AV systems, virtual reality gears as well as gaming applications with multidimensional audio. Next generation media technology is concentrating around Virtual reality experience and devices. It has applications not only in gaming but all other fields including medical, entertainment, engineering, and education. All such systems also require realistic audio corresponding with the visuals.

In the project presented in this thesis, a new portable audio hardware system is designed and developed along with a dedicated mobile android application to render immersive surround sound experiences with real-time audio effects. The tablet and mobile phone allow the user to control or “play” with sound directionality and implement various audio effects including sound rotation, spatialization, and other immersive experiences. The thesis describes the hardware and software design, provides the theory of the sound effects, and presents demonstrations of the sound application that was created.

Dedicated to my Mother and Father

ACKNOWLEDGMENTS

Firstly, I would like to express my deepest gratitude and thank my advisor, Dr. Andreas Spanias, who not only taught and motivated me to pursue research, but also helped me achieve certain level of confidence and maturity. I would also like to thank Dr. Pavan Turaga and Prof. Todd Ingalls for taking out time helping throughout my thesis and agreeing to be a part of my thesis defense committee. Without their valuable time, support and guidance, I could not have finished this work.

My Master's term at Arizona State University was productive as well as enjoyable. Thanks to SenSIP LAB and School of Arts, Media and Engineering for providing resources and allowing use of their facilities. Many thanks to their members who made this journey easier and wonderful. Special thanks to Dr. Andreas Spanias for financially assisting me during my Master's study. In addition, I am grateful to have had the opportunity to work with the following people at ASU and I would like to thank them for supporting me in various ways. Thank you Prof. Loren Olson, Assegid Kidane, Peter Weisman and Aaron. I would also like to thank the graduate university staff Lynn, Jenna, Cynthia, Toni, Esther, Darleen and Heather for their timely and kind assistance.

A large portion of this journey was shared with my fellow lab-mates Jongmin, Sai, Alan, Michael, Aaron, Shwetang, Rushil, Vinay, Prasanna, Sophia, Jie and Henry who have been a great source of inspiration. I thank them for motivating me and for offering to help against all odds with every little concern I had.

Finally I would like to reach my arm out to all my dear friends and my family without whom it would have been impossible to accomplish my goals. To start with, I would like

to thank my family for their unending support no matter what I choose to do. Next, I would like to thank Smita Bhawalkar, Jayant Deshpande, Devendra Laulkar and Ajay Gawali for great support and for always being there. A few other people that I would like to thank deeply for making their presence felt in my life despite being far away are Anirudh, Priya and Pranav. Last but certainly not the least, I would like to thank Mrinmaya, Satish, Ameya, Ganesh, Rajesh, Akshay, Aditya Mule, Haripriya and Aditya for being very supportive and making me feel at home in Tempe.

TABLE OF CONTENTS

	Page
LIST OF TABLES	viii
LIST OF FIGURES	ix
CHAPTER	
1. INTRODUCTION	1
1.1 Emerging Technologies	1
1.2 Other Work on Similar Lines	3
1.3 Problem Statement	5
1.4 Proposed Solution	7
1.5 Applications	8
2. LITERATURE REVIEW	9
2.1 Audio Perception	9
2.2 Sound Field Reproduction	11
2.2.1 Binaural Audio	11
2.2.2 Stereophony	14
2.2.3 VBAP	17
2.2.4 Ambisonics	19
2.2.5 Wave Field Synthesis	21
2.3 Localization of Sound in Rooms	22
2.4 Virtual Audio Rendering	23
2.5 Array Signal Processing and Beamforming	25

CHAPTER	Page
2.5.1 Microphone Array	25
2.5.2 Acoustic Beamforming for Loudspeakers	28
3. DESIGN AND CONCEPT.....	29
3.1 Acoustic Design and Sound Reproduction Method	29
3.2 Virtual Source Motion Algorithm	31
3.3 Audio Effects Algorithms	32
4. SOFTWARE	34
4.1 Android	34
4.2 Application Design and Architecture	35
4.3 Audio Player	38
4.4 Bluetooth	38
4.5 Audio Effects and Virtual Source Movement.....	39
5. HARDWARE	41
5.1 Components	41
5.2 Hardware Design and Implementation.....	42
5.3 Practical Issues and Solutions	43
5.4 Interface and Communication	45
5.5 Speaker Arrangement	46
6. OBSERVATION AND ANALYSIS	47
6.1 Experimental Study.....	47
6.2 Applications in Education and Outreach.....	48

CHAPTER	Page
6.3 Importance, Uniqueness and Comparison with Other Work	49
7. CONCLUSION AND FUTURE WORK	50
7.1 Summary	50
7.2 Future Prospects	51
REFERENCES	53

LIST OF TABLES

Table	Page
6.1 Sound Effects for 3D Audio	47

LIST OF FIGURES

Figure	Page
1.1 Virtual Reality Head Gears	1
1.2 Evolution of Audio Systems	5
1.3 Recommended Speaker Arrangement for 5.1 Audio Systems	6
2.1 ILD over 200 – 6000 Hz as a Function of Incident Angle	9
2.2 ITD for 0 ($\phi = 0$) to 690 Microseconds ($\phi = 90$)	10
2.3 Binaural Audio	12
2.4 Virtual Audio Rendering for Headphones	14
2.5 Audio Panning	15
2.6 VBAP 3 Dimensional	18
2.7 Wave Field Synthesis	22
2.8 Microphone Array	26
3.1 Hexahedral Speaker Box	29
3.2 Virtual Point Sources	30
3.3 Distance and Angle Parameters	31
4.1 Android System Architecture	34
4.2 Functional Diagram for Android Application	36
4.3 Android Application Playback Screen	37
4.4 Fragments for Virtual Source Motion	39
5.1 Hardware System Design	43
5.2 Electronic Circuit for Bluetooth Interface and Gain Control	44
5.3 Function in Arduino IDE Code for Rotational and Spatial Control	46

Figure	Page
6.1 Audio System Working and Application	49

Chapter 1

INTRODUCTION

1.1 Emerging Technologies

Virtual reality can be described as a simulated environment which creates an illusion that we are present somewhere we are not. This can be achieved by providing realistic inputs of the virtual environment to our senses. Visual and aural senses play the most prominent role in the awareness of the surrounding for human being. In recent years, prospects of virtual reality have not just become possible but it has become the most anticipated technology of this generation because of advancement in computing power, communication technology and the increasing miniaturization in electronics. Research in signal and image processing is enabling the development of many virtual reality (VR) applications in gaming and entertainment. Virtual reality systems such as Oculus [2] are being developed and assessed for emerging applications in entertainment, gaming, medicine and health [12, 26, 29] as well as data visualization [18, 25]. Such wide ranges of applications and clear possibility of creating an accurate virtual environment have paced the research in technology and tools providing virtual experience.



Figure 1.1 Virtual Reality Head Gears.

A VR system cannot provide an accurate impression of virtual space unless its visual inputs are in perfect synchronization with the audio. A small discrepancy in visuals of the audio source and the corresponding sound pressure at listener's ears can reduce the effectiveness of the experience by large amount. Therefore VR systems typically require a sound system with immersive properties to create realistic sounds associated with visuals. Such sound systems create virtual sound sources and audio scenes to give impression of the realistic audio environment for the listener. Unique hearing system in human beings make it possible to naturally sense different cues of sound source such as direction, distance, loudness of sound sources and idea of surrounding space. Signal processing techniques such as virtual source rendering, HRTF [53, 57], beam formation [9, 21] are used along with the concepts of psycho-acoustics and sound fields to create such immersive experiences depending on the sound distribution methods.

The research [23, 38, 50, 54, 58] for immersive 3D audio has been conducted in various industry and university research laboratories over past few decades. As a result, many different multidimensional audio systems [3, 16, 17] and new multichannel formats [23] have been developed. Recently, new technologies such as RealSpace 3D Audio [39] and DTS: Headphone-X [19] have been developed to produce real 3D audio experiences through headphones. In case of audio systems for large spaces, multiple speaker technologies such as stereophony, Ambisonics and wave field synthesis have been developed and explored in university and research laboratories. Few multi-channel audio systems such as 5.1, 7.1 audio systems are already commercially available though they have their limitations and lack accurate 3D sound experience. But recently Dolby Atmos

system [17] has been developed as an upgrade to home theatre systems to provide most realistic 3D audio experience.

1.2 Other Work on Similar Lines

After widespread introduction of home theatre systems in 1990 and their continuous ongoing improvement, now 3D audio technology have led in a new era of immersive audio once again. Variety of new audio products are being developed and introduced in the market to cope up with requirement of 3D immersive environments and technological advancement in DSP and sound production techniques. To account for physical restrictions in the placement of loudspeakers, the accommodation irregularly placed non-standardized layouts is made possible. This is often accompanied by automatic calibration techniques based on acoustic measurement of the loudspeaker positions. To further simplify the installation and reduction of cabling costs, wireless loudspeaker setups have been introduced [42].

In the home AV systems, the trend has increased to use sound bars using array processing algorithms along with closely placed smaller loudspeakers that allow adequate spatial effects similar to multi-speaker systems. With advanced signal processing algorithms, sound bars are capable of reproducing the virtual audio sources where speakers do not exist. While method such as wave-field synthesis has made its way into very few products but it requires large number of speakers.

Headphones deliver realistic sound to the listener through signal processing such that the acoustic signals that the eardrum would have received in the natural listening scenario can be recreated by measuring the appropriate impulse responses (HRIR's), and

running the convolution in real-time. With head-tracking and individual measurements, the effect can be very convincing with the possibility of presenting users with more natural spatial content than the traditional home theater. Also adding a virtual speaker is only in DSP algorithms and computing power with no need of a physical loudspeaker and amplifier.

Due to the increasing number of different formats [23, 52] and sound reproduction systems for spatial effects and 3D audio, ranging from headphones to 22.2 speakers, The MPEG committee has established a new standard for 3D audio coding [23] to ensure compatibility between formats and systems, and consistency in quality of the spatial audio. Unlike conventional channel based audio contents, approaches such as object based audio [50] and Higher-Order Ambisonics [24] are being developed to deliver content without being constrained to a standardized loudspeaker layout. In the first approach, the individual audio objects are transmitted separately with metadata describing their spatial properties. On the consumer side the audio objects are panned according to the consumer's loudspeaker layout with capability of adjusting the audio mix in real time. In latter approach, Higher-Order Ambisonics (HOA) is a scene based audio technique independent from the reproduction layout and describes the sound field based on spherical harmonics. For the audio reproduction, the HOA data are rendered according to the desired loudspeaker layout and can be created from single- channel audio tracks within a digital audio workstation as well as microphone-array recordings.

1.3 Problem Statement

Audio systems have been greatly evolved over last few years from simple stereo cassette players to multichannel immersive audio systems and 3D audio formats. Nowadays consumer market is comprised of variety of headphones, wireless speakers, sound bars and home theatre systems [8, 42]. Continuous development and improvement in data and media content storage devices has been changing the face of audio devices. In spite of variety of audio products, a gap has been generated between cheaper portable audio appliances and multichannel surround sound systems. Traditional dedicated audio appliances have to be upgraded to multiple source platform based sound distribution systems with sources such as TV, mobile devices, tablets, laptops and gaming device, capable of delivering rich sound. The most common consumer loudspeaker layouts for spatial audio are horizontal only but the next generation of loudspeaker setups incorporates elevated loudspeakers to create immersive audio experiences.



Figure 1.2: Evolution of Audio Systems.

Such multi-channel audio systems are expensive as well as complicated to set up. The total cost of such systems that include speakers, AV system, installation labor and supply materials pile up to a large amount not less than average 1200\$. While installing, the speakers must be positioned properly for different home settings and acoustic dimensions in order to achieve desired surround sound effects. With different home settings and directions, it becomes difficult to arrange system according to layout. For example, according to recommendations in standards ITU-R BS.775-1 and SMPTE [3], a 5.1 loud speaker system should have arrangement as shown in Figure 1.3 [3] and if the setup on the circumference of a circle is not possible, loudspeakers inside the circle should be delayed accordingly. Even if properly installed, the system performance is effective only at the *sweet spot* where results are efficient. Such systems can be installed in hall and living rooms only and cannot be displaced to a different place or small areas. While these setups are horizontal only; the next generation of loudspeaker setups should incorporate elevated loudspeakers to create immersive audio experiences.

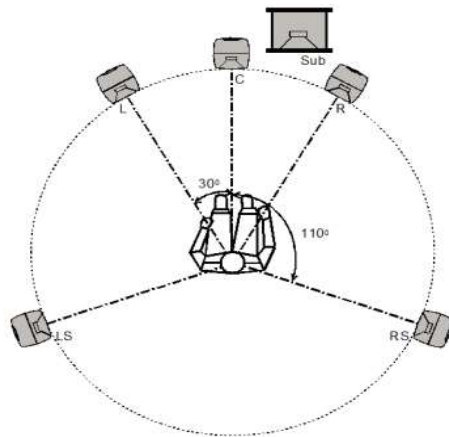


Figure 1.3: Recommended Speaker Arrangement for 5.1 Audio Systems.

Digitization of media have revolutionized audio playback and influenced all sound systems. Digital media content is now easily accessible through high speed internet in multiple data formats. With easily available music and video streaming services and the success of smartphones and tablets as media devices, a shift has occurred in the way most people access the media content. More content is viewed and listened to over headphones, which has led to both a massive increase in headphone sales and new categories of increasingly small portable speakers. But headphones, though portable, do not give the same sound experience as traditional speakers and are limited to personal use. If there are more people in a room, each person will require individual headphone. To have immersive sound experience in space around, a headphone requires head tracking sensors as well as position sensors to accurately change the audio corresponding the listener's movements. It leads to higher cost and a lot of research is being done in this area. Single-unit Bluetooth loudspeakers have become quite popular but are not able to provide a high quality immersive audio experience [43].

1.4 Proposed Solution

This motivated us to develop a cheaper, portable and wireless audio system capable of providing immersive-like audio experience using sound movement. We have designed a portable audio hardware system bundled with an innovative android app to control sound distribution [15]. The system is capable of delivering an immersive surround sound experience and various real time audio effects. This system is a single unit that can be hanged to the ceiling at the center of the room. The hardware consists of an active speaker

enclosure containing five speakers with electronic circuit controlling speaker output and providing Bluetooth connectivity.

The source of sound is a customized music player android application with basic functionality of audio playback from storage and online streams. An additional interface is provided in the app that enables users to manipulate sound directionality and audio movement. This prototype incorporates design principles of amplitude panning [34], localization of sound in rooms [36] and overhead sound objects [16] to create innovative surround sound effects for music. Wireless connectivity makes the system user friendly. This prototype can be developed further in to a commercial music system that provides a simple and yet rich audio experience with game-like features and capabilities.

1.5 Applications

The system is an audio entertainment unit with an anticipated application as affordable immersive-like sound system. Due to its capability of 3D audio, the system can be used in many applications of simulators and games. Variable and easily controllable sound directivity can be helpful where announcement speakers are used in malls or public places. The system can be used as a substitute to portable Bluetooth speakers in certain environments. Another emerging application is the reproduction of spatially distributed environmental noise (engines, streets, car interior, etc.) in laboratory settings.

Chapter 2

LITERATURE REVIEW

2.1 Audio Perception

Virtual audio scenes can be created using different sound field generation methods based on fundamental properties of the perception of sound. Unique human ear structure allows us to localize sounds and comprehend the spatial information of the surrounding. With deeper understanding of human audio perception, different cues of localization and its psycho acoustic principles, it has become possible to reproduce more realistic virtual audio scenes. Some of the important factors affecting sound perception are listed in this section [31, 50, 58].

- **Inter-aural Level Difference (ILD)** – Depending on the sound source direction, intensity of the sound may differ at each ear due to distinct location and direction of ears. This localization cue (ILD) is more effective for high frequency sounds due to diffraction where low frequency sound with long wavelength bends around the head and no shadow is cast by head to other ear as shown in Figure 2,1 [31]. But for high frequency with shorter wavelength, negligible diffraction takes place.

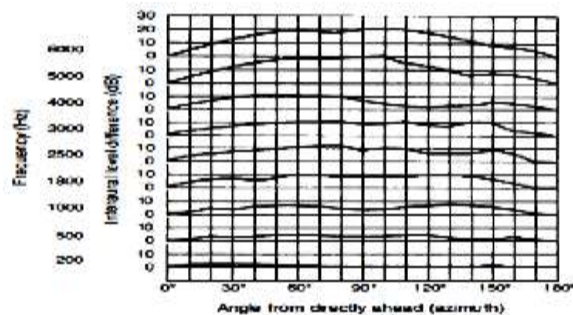


Figure 2.1: ILD over 200 – 6000 Hz as a function of incident angle.

For distant sound sources, ILD negligible below 500 Hz. For sources very close, ILD can occur at low frequencies.

- **Inter-aural Time Difference (ITD)** – Due to distance between ears, sound waves from the same source cannot reach each ear at the same time unless the source is equidistant from both ears. This is a very important cue in localization of nearby sound sources and is more effective for lower frequency sounds as shown in Figure 2.2 [31].

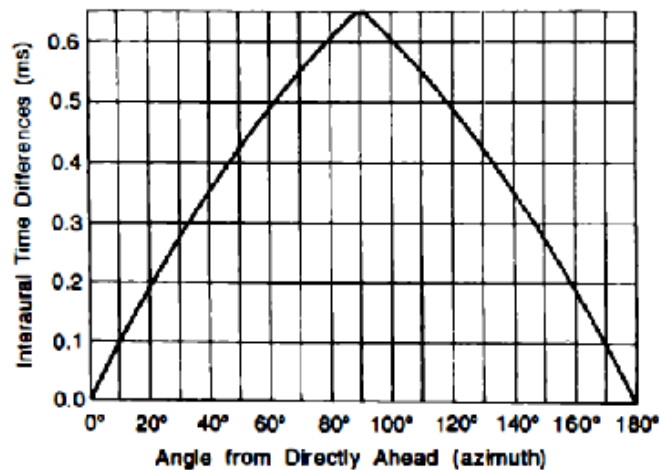


Figure 2.2: ITD 0 for ($\phi = 0$) to 690 microseconds ($\phi = 90$).

For pure tones, ITD is phase difference. For low frequency tones, IPD provides accurate localization of sound. For higher frequencies (1500 Hz), localization becomes highly ambiguous.

- **Perception of Distance** - Overall intensity of sound is a very obvious factor in determining distance from the sound source, though it is not the only one. As distance increases, sound spectrum also changes as higher frequency sounds are

absorbed more quickly by air over long distances. Other parameters are enlarged ILDs for sounds close to the head, and the ratio of direct to reverberant sound [45].

- **Mono-aural Localization** – The peculiar shape of the ear provides natural direction-dependent filtering of sound entering the ear drums. This filtering provides cues for localization in both horizontal and vertical directions. (role of HRTF) It is also important for creating the percept of a sound outside the head rather than inside. Reverberation also contributes to this percept.
- **Reverberation** – This is an important cue that provides an impression of nearby surrounding and ambiance.

All these factors provides cues for localization. By altering these parameters of the source sound, we can create specific audio at the listener’s ears to recreate a virtual sound source. Different methods such as stereo/multichannel systems, orthogonal basis function, object based virtual scene rendering are being used to create virtual source spatial sound effects. In our systems we use multichannel system approach with audio panning principles.

2.2 Sound Field Reproduction

Different sound reproduction techniques [34] have been developed over the years for entertainment and research purposes. These can be roughly classified into binaural techniques, stereophony, Ambisonics and Wave field synthesis.

2.2.1 Binaural Audio –

The principle behind the Binaural audio is that human auditory system perceives audio events as two input signals that are the sound pressure signals at our two

eardrums. In this way human auditory system can perceive spatial audio by localizing and segregating sound sources. Based on this principle, if a sound reproduction device generates the same sound pressure at a listener's eardrum as that would have produced by a real sound source situated in the surrounding space, the listener should not be able to differentiate between the virtual audio from the device and the sound from real source. This technique produces two-channelled signals for each ear to create desired sound pressure based on psychoacoustics of the spatial sound and are used in headphones as shown in Figure 2.3 [55].

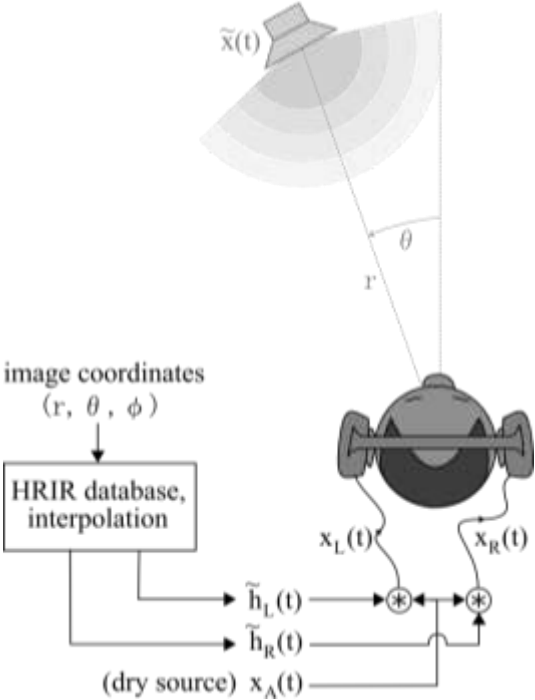


Figure 2.3: Binaural Audio.

Binaural audio concept, its application and of details in the sound transmission have been found in the literature for over more than 80 years. Different recording techniques such as dummy heads and microphone arrays were developed and

improved to create realistic audio content to be played via headphones. But with stellar growth in gaming and virtual media technology, researchers were challenged to work on real-time spatial audio effects in virtual world and improved audio recording techniques. With advancement in psychoacoustic research and more understanding human hearing system, multiple audio perception cues have been explored and transformed into algorithms and digital filters. Basic inter-aural localization cues such as ITD and ILD have now become complex HRTFs that consider a number for factors such as mono-aural cues, reverberations and head movements that influence human hearing to a large extent.

With speedy development of virtual reality gears and their ability to provide realistic experiences, research in binaural audio have thriven enormously. Technologies such as RealSpace 3D audio by VisiSonics, Headphone: X by DTS and many other, have been developed in order to support VR gears like oculus for gaming and 3D experiences.

When both sound source and listener are fixed, the acoustical transmission from a point source to the two ears can be regarded as a linear-time-invariable (LTI) process. Head-related transfer functions (HRTFs) are defined as the acoustical transfer function of this LTI system [53].

$$H_L(r, \theta, \varphi, f, \alpha) = \frac{P_L(r, \theta, \varphi, f, \alpha)}{P_0(r, f)}, \quad H_R(r, \theta, \varphi, f, \alpha) = \frac{P_R(r, \theta, \varphi, f, \alpha)}{P_0(r, f)} \quad (1)$$

where P_L and P_R represent sound pressures at left and right ears, respectively; P_0 represents the free-field sound pressure at head center with the head absent. Generally, HRTFs vary as functions of frequency f and source position (r, θ, φ)

(distance and direction) as well as individual α . For $r > 1.0 - 1.2$ m, HRTFs are approximately independent of source distance and called far-field HRTFs. For $r < 1.0$ m, however, HRTFs are relevant to source distance and called near-field HRTFs.

A complete virtual auditory event is composed of above mentioned free-field virtual source synthesis and other important factors such as virtual auditory environment and dynamic acoustic information of orientation and position of listener's head as shown in Figure 2.4 [57].

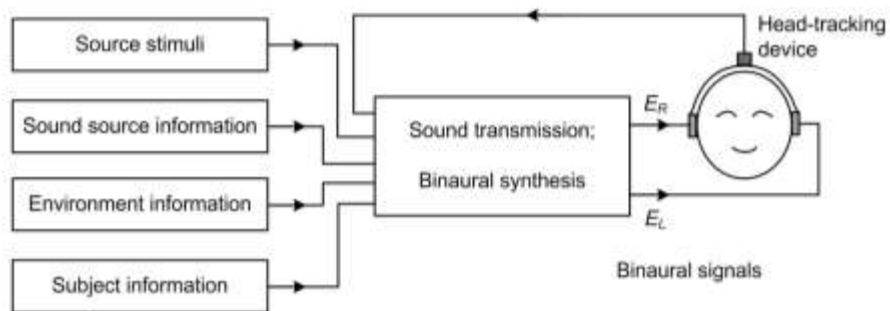


Figure 2.4: Virtual Audio Rendering for Headphones.

Great developments have been achieved in the field of HRTFs [51] and VADs, but many issues need further research. With many solutions being developed for the issues like HRTF non-individualization, spatial interpolation of HRTF and accommodation effects due to head movement [20]. VADs have currently been applied to various fields in scientific research, engineering, entertainment and consumer electronic products.

2.2.2 Stereophony –

Stereophony as the name suggests (Stereo – solid/3D) is an audio reproduction technique developed to create pervasive sound effects and spatial audio. This

technique uses two or more loudspeakers to deliver desired sound pressure at listener's ears. Stereophony is based on psychoacoustic [31] as well as sound field principles. [34] It enables creation of a virtual source in the space between other actual sources by the superposition principle and the physical description of the sound fields created through different sources. Due to superposition of waves, a single virtual sound source is perceived at a different location from the actual source.

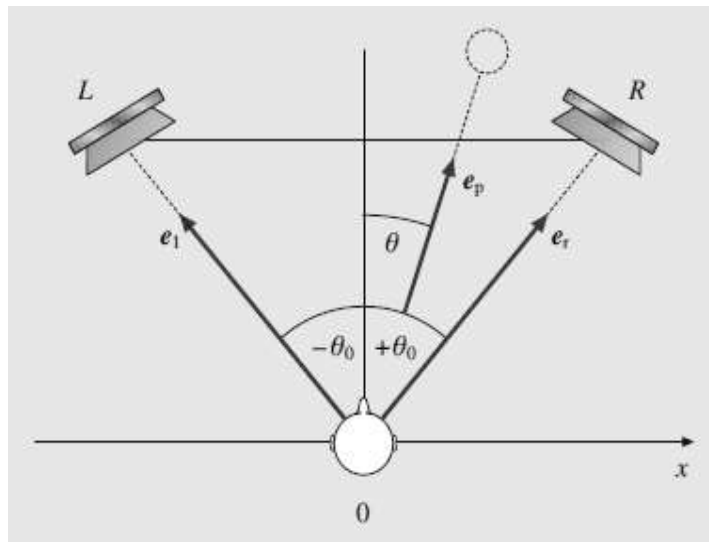


Figure 2.5: Audio Panning.

Figure 2.5 [34] shows a listener in the center of a coordinate system and two loudspeakers in the direction of the angles θ_0 and $-\theta_0$ to the right and left. To reproduce the image of a sound source at some angle θ with $|\theta| < \theta_0$, the same driving signal is fed to both loudspeakers, but with different weighting factors g_r and g_l . These are selected such that the superposition of the sound fields of both loudspeakers makes the listener perceive a single sound source at the desired angle θ . This perception is called a *phantom source* or *virtual source*. This effect is called

amplitude panning and the functional dependency of the weighting factors $g_r(\theta)$ and $g_l(\theta)$ on θ is called a *panning law* such as sine law and tangent law. Sine law and Tangent law can be given as [34]

$$\sin \theta = \frac{g_r - g_l}{g_r + g_l} \sin \theta_0 \quad (2)$$

$$\frac{g_r - g_l}{g_r + g_l} = \frac{1 - g_l/g_r}{1 + g_l/g_r} = \frac{\tan \theta}{\tan \theta_0} \quad (3)$$

Where g_r and g_l are individual weighting factors for the right and left loudspeakers respectively. In this technique, only two sound sources around the virtual source to be created take part in the reproduction of the sound.

The phantom source creation now involves time delay as well in addition to level differences and includes development of multichannel systems such as 5.1 audio system. When compared to a real sound source, the perception is plausible, but differences occur, such as: an increased impression of width, a degraded localization accuracy, and coloration. All panning approaches can be implemented as simple scalar weights and/or delays. Basically, panning approaches work with few loudspeakers, as long as the aperture angle between the loudspeakers is less than 90. In addition, if the listening position is equally distant to the loudspeakers, the auditory event will be largely aligned with the direction of the virtual source. The size of the optimal listening area is often called sweet spot or area, and it enlarges for panning with more loudspeakers. Outside the sweet spot, the auditory scene most often collapses toward the closest active loudspeakers. Basically, the only type of virtual source that can be reproduced by panning approaches is a point

source at the distance of the loudspeaker array. Nevertheless, it is possible to create distance impression by perceptual cues that are possible to reproduce. Such suitable cues are, for instance, a decrease in level and direct-to-reverberation ratio [45, 50].

2.2.3 VBAP –

Vector-based amplitude panning is a multichannel audio reproduction method first introduced by Pulkki [33] and an extension of stereophony. The audio panning is not only applied to two loud speakers but to two or three adjacent speakers. Most commercially available home theatre systems are based on this principle. Vector-based amplitude panning extends the tangent panning law for two loudspeakers to panning between adjacent speakers of a one- or two-dimensional loudspeaker array. In a horizontal plane around the listener, a virtual sound source at a certain position is created by applying the tangent panning law between the closest pair of loudspeaker called two-dimensional VBAP. The position of the virtual source moves without being restricted to certain loudspeaker positions, i. e., $0 \leq \theta < 2\pi$. Then only those two loudspeakers are active which enclose the direction θ of the virtual source. For $\theta = \theta_n$ only one loudspeaker is active ($g_n = 1$). In detail, the weighting factors g for two dimensional vector-based amplitude panning with N loudspeakers are given by [34]

$$g_v(\theta) = \begin{cases} \frac{\sin(\theta_{n+1} - \theta)}{\sin(\theta_{n+1} - \theta_n)} & v = n \\ \frac{\sin(\theta - \theta_n)}{\sin(\theta_{n+1} - \theta_n)} & v = n + 1 \\ 0 & \text{Otherwise} \end{cases} \quad (4)$$

Here, n denotes the current position of the virtual source such that $\theta_n \leq \theta \leq \theta_{n+1}$. Two-dimensional vector-based amplitude panning is the same as stereo panning,

only that the position of the pair of active loudspeakers moves with the sound source.

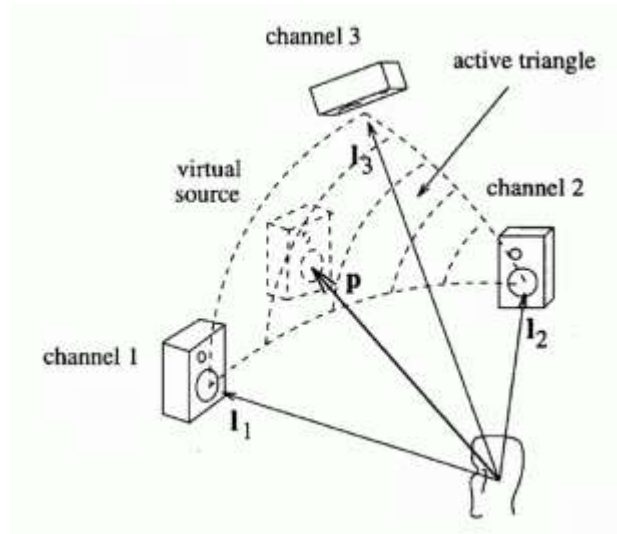


Figure 2.6: VBAP 3 Dimensional.

This principle was also extended to project sound sources onto a three dimensional sphere and assumes that the listener is located in the center of the equidistant speaker setup and it was proposed to triangulate the sphere around the listener and to put one loudspeaker at each vertex as shown in Figure 2.6 [33]. The virtual source is created by amplitude panning between the three loudspeakers of the corresponding triangle. The three weighting factors are again determined from a projection of the unit vector e_p in the direction of the virtual source to the unit vectors e_1 , e_2 , e_3 in the direction of the three loudspeakers [34].

Research experiments conducted in the past suggest that the panning laws for two- and three-dimensional vector-based amplitude panning do indeed associate well with the human perception of the virtual source. However, the localization of a virtual source depends on its targeted position relative to the adjacent speakers. The

localization is most precise if the virtual source direction coincides with the position of a loudspeaker. In this case the loudspeaker is perceived as a real source. For virtual source position in between adjacent loudspeakers, a certain spread of the localization occurs. Research presets methods to achieve a uniform spreading of amplitude panning for virtual source creation.

2.2.4 Ambisonics –

Ambisonics is a global panning approach for audio reproduction where multiple loudspeakers are used to create a virtual source on the basis of superposition principle. Unlike VBAP, a virtual source is not only panned between two or three loudspeakers but all the loudspeakers in the arrangement by continuously formulating a sound field from an encoded signal with virtual, spherical information in a finite order angular transform domain. Ambisonics can be applied for flat 2-dimensional system where spatial reproduction is done for planar loudspeaker arrays as well as 3 dimensional with additional difference channel for height and depth [13].

Ambisonics is therefore comprised of encoding spatial information for the audio to be played and decoding it later according to specific speaker setup to create spatial surround sound. Most widely used basic encoding for Ambisonics is done in spherical harmonics format called the B-format encoding that represents sounds situated in the horizontal plane with 4 signals W , X , Y and Z where W stands for the sound pressure, X for the front-back sound pressure gradient, Y for left-right and Z for up-down. They are given as [24]

$$W = \frac{1}{k} \sum_{i=1}^k S_i \left[\frac{1}{\sqrt{2}} \right] \quad (5)$$

$$X = \frac{1}{k} \sum_{i=1}^k S_i [\cos \phi_i \cos \theta_i] \quad (6)$$

$$Y = \frac{1}{k} \sum_{i=1}^k S_i [\sin \phi_i \cos \theta_i] \quad (7)$$

$$Z = \frac{1}{k} \sum_{i=1}^k S_i [\sin \theta_i] \quad (8)$$

Where S_i are our mono audio signals we want to encode at the according positions ϕ_i (Horizontal angle phi/ Azimuth), and θ_i (vertical angle theta for elevation). These signals can be obtained captured by the means of omnidirectional ‘Sound-Field’ microphone which allows first order Ambisonic recordings of real sound fields which provides positional information of the sound sources or can be synthesized by signal processing for existing audio files and desired spatial information with the basic assumption that all sound waves arriving are plane waves. Hence the position of the virtual sound sources only depend upon azimuth and elevation with an assumed constant distance [13].

Transmission channels for Ambisonics contain a speaker-independent representation of a sound field and are completely independent from the loudspeaker layout. An Ambisonic decoder is always designed for specific loudspeaker system layout where number of loudspeakers k used has to be always greater than or equal to Ambisonic channels N [34, 50].

Lot of research has been accomplished over the years describing spherical Ambisonic panning that include higher order spherical harmonics [14, 24] and

various decoder techniques to improve the angular discrimination and spatial resolution of the virtual signal which reduces sweet spot limitation.

Ambisonics did not become so popular in consumer audio systems and no native Ambisonic recordings were commercially available. Hence many techniques have been developed to make content produced in ambisonics available to consumers in stereo or discrete multichannel formats. But in recent years a great interest in ambisonics have grown among researchers in order to create 3D audio experiences using near field coding higher order ambisonics with distance information by modelling sound fields as spherical waves rather than plane waves.

2.2.5 Wave Field Synthesis -

Wave field synthesis is a spatial audio reproduction technique that uses loudspeaker arrays to physically reconstruct sound fields which construct desired audio wave fronts originating from a virtual source [34]. It is based on the Huygens-Fresnel principle of physical description of the propagation of sound waves [4, 56] which states any wave front can be considered to be superposition of multiple spherical waves. This method eliminates the sweet spot limitation of other systems hence localization of sources is independent of listener's position.

Theoretical basis for this technique is given by the mathematical principle called Kirchhoff- Helmholtz Integral (KHI) [34] which states that if sound pressure and directional pressure gradient (acoustic velocity) at any point on the surface of a source free volume is known, then sound pressure at any pint within this volume can be completely determined. Practically, a computer synthesis independently

drives a large number of separately controlled loudspeakers arranged in an array around the listener as shown in Figure 2.7 [10].

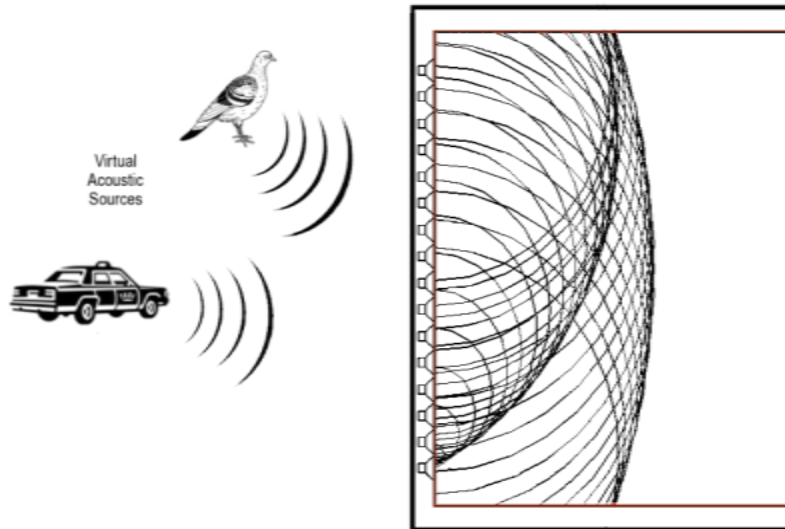


Figure 2.7: Wave Field Synthesis.

Though WFS is capable of reproducing true holophonic audio, it has many disadvantages such as limitation of planar sound, playback room acoustics, aliasing effect, truncation effect and high setup costs. Research has been carried out to eliminate these limitations and bring such systems out of research laboratories for commercial use. Such systems have been installed in few public places and theatres and development of home audio using WFS is still an ongoing process [4].

2.3 Localization of Sound in Rooms

For any audio system, apart from quality of speakers, signal quality and reproduction method, room acoustics plays an important role in the sound perceived by the listener. Hence all commercial audio system developers and engineers study effects of sound reflections and audio source localization in rooms [36] representing general

household environments. A sound heard by a listener in an enclosed room is a combination of direct sound from the audio source to the ears and reflected indirect sound from walls, ceiling, floor and other appliances or furniture. While reflections can add a spaciousness to the sound that provides better experience, they can also sometimes distort the sound due to coloration effects. According to the experimental outcomes about the localization of sound in rooms, localization accuracy drops with an increase in reverberation such as in case of larger reverberant rooms.

The developed audio system has an important function for sound reflection from the walls. As the sound are directed towards the walls to achieve orthogonal reflections of the walls, sound reflections of other walls doesn't affect the resulting audio to a large extent [44]. In such case, reflected sound can be modelled as virtual mirror source across the walls. According to research conducted to study the effects of orchestration of wall reflections, household appliances and furniture doesn't affect the perceived audio significantly. Also localization cues depend more upon the geometrical details of source positions, room surfaces and the listener's position than it does upon the total-to-direct sound power ratio. Hence, if the proposed audio system is properly installed with each speaker facing each wall to produce orthogonal reflections, deterioration in localization of audio signals can be minimized.

2.4 Virtual Audio Rendering and 3D Audio

Virtual sound rendering with 3D audio systems is one of the top research field for past few years in audio processing resulting in some groundbreaking innovations and

product developments. Few known recording and reproduction techniques for 3D audio are briefly discussed here [50].

- **Recording:** Two widely known practical recording methods used for the creation of virtual sound scenes. In first approach, audio signal of different objects are recorded separately and a complete virtual audio scene is composed from the recordings by using spatialization audio processing techniques based on virtual source and its positional information. For later approach, special microphone arrays are used to record multidirectional sound with to capture complete sound scene. In many practical cases, both approaches are combined to produce desired 3D audio.
- **Audio Reproduction:** Apart from binaural audio techniques developed for realistic audio for headphones and VR gears, three major audio reproduction methods over loudspeakers are used. Most widely used method is multichannel surround sound systems where loudspeaker are arranged in a specific layout. This channel based approach have been improved over last few years but faces some limitations. Another approach uses orthogonal basis functions to represent virtual sound fields and each reproduction unit contains decoder to create desired virtual sound event based on a specific speaker layout. Most recent approach used is known as object based virtual audio. In this method, sound source signals are combined with metadata for spatial information in an audio format. The audio signals are then rendered for reproduction via loudspeakers. Efforts are being applied towards the inclusion for various newly developed audio reproduction methods into commercial systems capable of virtual reality applications.

2.5 Array Signal Processing and Beamforming

An acoustic array system is an assembly of acoustic transducers either receiving or transmitting acoustic signals and delivering desired information or audio outputs with the help of signal processing algorithms. Array techniques have been used for decades for beamforming and direction of arrival (DOA) estimation with numerous applications apart from acoustics in radar, sonar, wireless communication, Smart Antennas, medical diagnosis, radio astronomy etc. [9] Current applications of array processing includes speech enhancement, Acoustic Beamforming for Hearing Aid Applications, Noise source identification, Sound field visualization for research in sound field reproduction techniques, Under-determined Blind Source Separation Using Digital 3D/4D Ultrasound Imaging Array, Synthetic aperture radar, advanced underwater mapping and Chemical sensor arrays [9,21]. Different array patterns such as linear, rectangular, circular and even 3 dimensional arrays have been designed and researched for various applications, but two main types in acoustic applications are microphone arrays and loudspeaker arrays.

2.5.1 Microphone Array

The array systems can be categorized in two types depending on the acoustical propagation model in case of microphone arrays, as far field arrays and near field arrays. In case of far field arrays, a simple source is located at a large distance from the array such that wave fronts arriving at the array are planar. The acoustical propagation model from the source to the microphones is normally a SIMO system with only one focus. Whereas nearfield array systems include a distributed source and the sound waves arriving at the

array follow a complex convolution and interference patterns and hence can be categorized as a MIMO system involving multiple focal points [9].

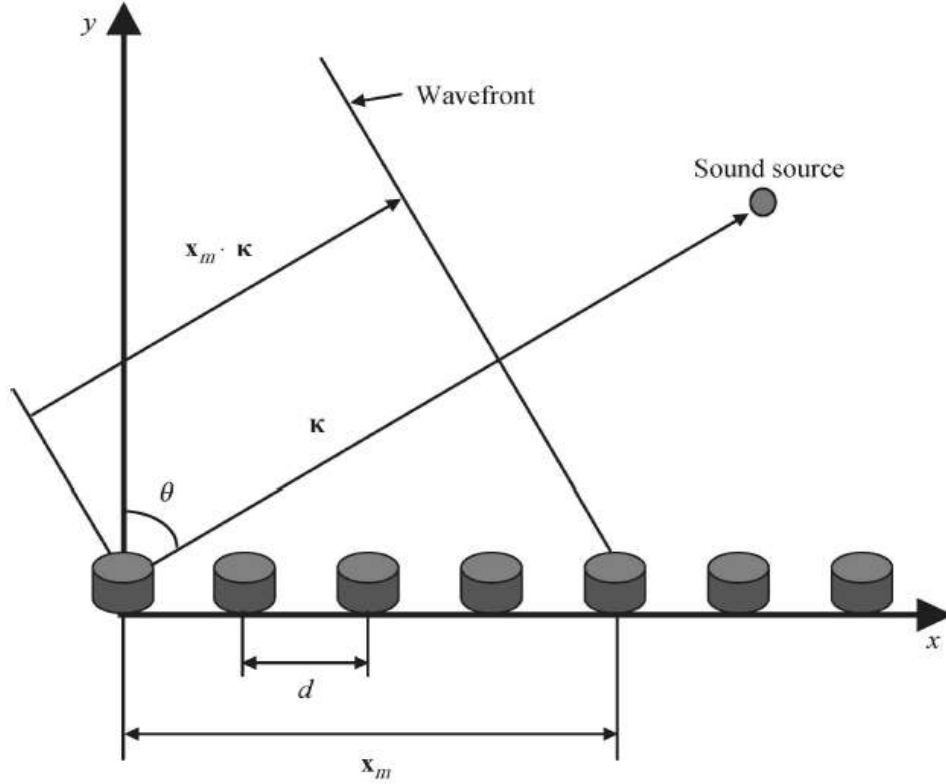


Figure 2.8: Microphone Array.

For basic mathematical representation of linear uniform array system as shown in Figure 2.8 [9], a narrowband source signal $r(t)$ can be represented as:

$$r(t) = s(t)e^{-j\omega t} \quad (9)$$

Where $s(t)$ is the baseband signal and ω is the center frequency of the narrowband signal.

With far field assumption, the sound field wave is considered as planar as shown in Figure 2.8 and sound pressure at position x can be expressed as [9]

$$x(t) = s(t)e^{-j(\omega t - \mathbf{k} \cdot \mathbf{x})} = s(t)e^{-j(\omega t + \frac{\omega}{c} \mathbf{K} \cdot \mathbf{x})} \quad (10)$$

Where $\mathbf{k} = \left(-\frac{\omega}{c}\right) \mathbf{K}$, is a wave vector with \mathbf{K} as a unit vector $(\sin \theta, \cos \theta)$ pointing from array position to the source, \mathbf{x} as position vector of a field point and c as the speed of sound. For a uniform linear array of M microphones with signals $x_1(t), \dots, x_M(t)$ at positions x_1, \dots, x_M ; the data vector $\mathbf{x}(t)$ is given as [9]

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_M(t) \end{bmatrix} = \begin{bmatrix} e^{-j\left(\frac{\omega}{c}\mathbf{K}\cdot\mathbf{x}_1\right)} \\ \vdots \\ e^{-j\left(\frac{\omega}{c}\mathbf{K}\cdot\mathbf{x}_M\right)} \end{bmatrix} s(t)e^{-j\omega t} + \begin{bmatrix} n_1(t) \\ \vdots \\ n_M(t) \end{bmatrix} = \mathbf{e}(\mathbf{K})r(t) + \mathbf{n}(t) \quad (10)$$

Where vector $\mathbf{e}(\mathbf{K})$ containing the spatial information is called steering vector or array manifold. The dot product $(\mathbf{K}\cdot\mathbf{x}_m)$ of unit vector and position vector is given as

$$\mathbf{K}\cdot\mathbf{x}_m = (m - 1) d \sin \theta, \quad m = 1, 2, \dots, M \quad (11)$$

Where θ is angle of the source with respect to y axis of the array reference and d is $\mathbf{n}(t)$ is the vector of uncorrelated noise added to each microphone sensor. For D sources using the superposition principle [9]

$$\mathbf{X}(t) = \sum_{i=1}^D \mathbf{e}(\mathbf{K}_i)r_i(t) + \mathbf{n}(t) = [\mathbf{e}(\mathbf{K}_1) \quad \dots \quad \mathbf{e}(\mathbf{K}_D)] \begin{bmatrix} r_1(t) \\ \vdots \\ r_D(t) \end{bmatrix} + \mathbf{n}(t) = \mathbf{E}\mathbf{s}(t) + \mathbf{n}(t) \quad (12)$$

Where $\mathbf{s}(t) = [r_1(t) \quad \dots \quad r_D(t)]^T$ is source signal vector and \mathbf{E} is called DOA matrix. This mathematical expression contains the signal information in both time and space domain leading to multidimensional signal processing and provides information regarding source position. With further mathematical expansion beamwidth of an array can be given as [9]

$$BW = \frac{2\lambda}{M d \cos \theta_0} \quad (13)$$

Where θ_0 is steering angle, λ is the wavelength and $M*d$ is the aperture size.

2.5.2 Acoustic Beamforming for Loud Speakers

Array processing techniques are also applied to loud speaker arrays to create directional sound fields. Microphone array techniques have been widely used over many years but research in speaker arrays started in recent decades considering dual properties with sensor arrays due to acoustic reciprocity. Hence the design principles of microphone arrays are also applicable to loudspeaker arrays with the roles of source and receiver exchanged. [9] Different techniques and algorithms such as Delay and sum method, Acoustic brightness control, Acoustic contrast control, Pressure matching, Energy cancellation and sound field synthesis have been developed and enhanced over the years to create accurate and focused beam of sound in the desired direction. Current research in speaker array includes point focusing, 3D arrays and holography for the applications in virtual audio rendering and commercial products.

Chapter 3

DESIGN AND CONCEPT

3.1 Acoustic Design and Sound Reproduction Method

The audio system developed here works on the basic principle of superposition and sound localization in enclosed spaces. Unlike the speaker arrangement in multispeaker audio systems, the speakers are arranged in a single rectangular box. The box acts as an omnidirectional audio point source at its center. The four speakers in four directions orthogonal to each other can produce sound at any angle between 0 and 360 degrees with amplitude panning laws. Figure 3.1 depicts the rectangular box of speakers made for the system prototype.



Figure 3.1: Hexahedral Speaker Box.

Sounds produced by four speakers facing four cardinal directions are aimed towards the walls of the room such that reflected sound will be directed towards the center. Every wall reflection acts as virtual mirror source and creates spatial artifacts where sound is perceived to be emitted from the surrounding space as shown in Figure 3.2. It has been

verified that household room configurations with different reflecting properties and obstructions do not change the localization of sounds to a large extent [44]. A fifth speaker facing the ground provides vertical sound directivity and helps to create close distance sound source impressions. This hardware provides portability and requires no specific setup except hanging it in the middle of the room ceiling.

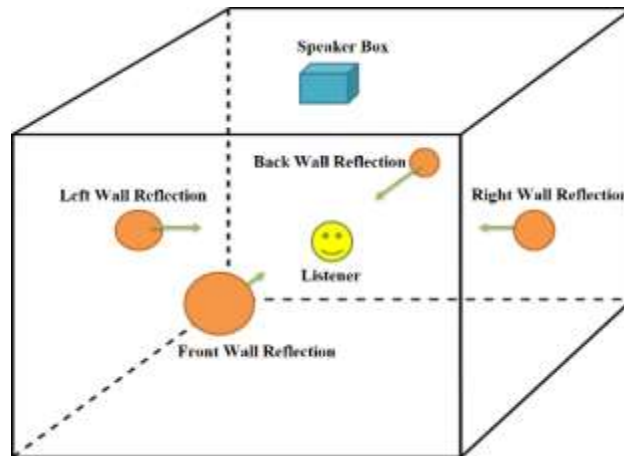


Figure 3.2: Virtual Point Sources.

The system functions ideally in a room where all reflection will be perpendicular to the wall surfaces. As most of the rooms are rectangular cuboid shaped, optimal results can be achieved by facing the speaker surfaces exactly parallel to the walls to ensure perpendicular reflections from the center of the wall. Thus we can assume 4 virtual point sources behind the centers of the walls similar to the 4 speakers in typical surround sound systems.

As the audio sources are not beam sources, the span of the reflected sound will be wider than that of a speaker located at the same location. Instead it will be similar to a sound coming from larger distance creating spherical sound fields with larger radius. When

sound waves are incident on a hard surface, the high pressure part of a sound wave reflects as a high pressure without phase change upon reflection and hence do not cause destructive interference. For nearby virtual source experience, direct sound fields from the vertical channel will play a major role [27].

3.2 Virtual Source Motion Algorithm

Android app provides different tabs to play with sound directivity. A screenshot of such a tab provided for virtual source movement functionality is shown in Figure 3.3. UI for all such tabs have a canvas to control/play with sound directions where center of the canvas is assumed to be the position of the user and the pointer represents the position of the virtual source. When a user touches the canvas, relative distance (rd) and angle (t) of the pointer location touched by user is calculated from the center of the canvas. These values are received as inputs to the Arduino board from the android application. Using these values as inputs, amplitude panning algorithm in the Arduino microcontroller drives speaker gains to give the desired virtual sound source movement.

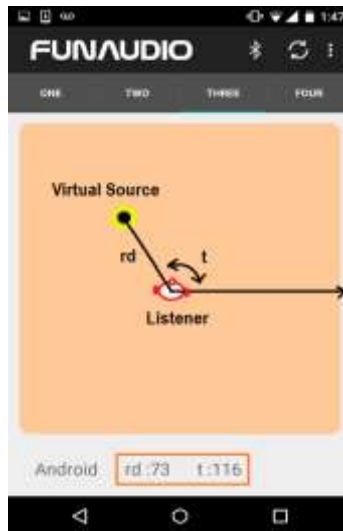


Figure 3.3: Distance and Angle Parameters.

The amplitude panning algorithm is based on the sine law and the equation governing gain of the four speakers in four directions is

$$Volume_{1-4} = 255 * (distance * \frac{\cos(\theta * 2\pi)}{2} + 0.5) \quad (14)$$

where distance (rd) is parameter between 0 and 1 that is used to measure the distance of the virtual source from the user. $\theta(t)$ is the angle of the virtual source from reference in radian. Volume is the gain of the four speakers sampled between 0 and 255 steps. Gain of the fifth speaker is inversely proportional to the distance. For nearby virtual source experience, direct sound fields from the vertical channel will play a major role and controls height of the virtual source. All the computations for obtaining gain for individual speaker are done in Arduino micro-controller.

3.3 Audio Effects Algorithms

Simple digital audio effects such as echo and reverberation are implemented in the android application with variable delay that can be controlled by user. These basic audio effects not only used to offer a sense of surrounding but also in many film, music and game audio for creating impact and

- **Echo** – An echo effect is a simple addition of a copy of the original audio signal normally attenuated and always delayed by a fixed amount in time. A simple FIR filter imitates an echo effect [46, 49] and the function is implemented in the android app before the audio is played using audioTracks library.

$$outframe(n) = frame(n) + attenuation * frame(n - delay) \quad (15)$$

As the samples are received in a byte buffer from audio codec, samples are stored in another new circular buffer which depends upon the delay setting. This delayed

signal is then mixed in with the original signal at a somewhat reduced gain with a feedback for decaying repeats.

- **Reverberation** – It is the result of the many reflections of a sound that occur in a room or enclosed surrounding. Reverberation plays a major role in cues for human audio perception in localization and surrounding environment and therefore is widely used in audio reproduction systems such as binaural audio to accurately recreate virtual environments. A simple reverb audio effect can be implemented using an IIR filter with fixed delay as [46, 49]

$$\text{outframe}(n) = \text{frame}(n) + \text{attenuation} * \text{outframe}(n - \text{smallDelay}) \quad (16)$$

Where this IIR (infinite impulse response) difference equation roughly imitates the reverberating nature of a room. Reverberation is an audio effect that has been studied over the years and still a great amount of research is being conducted to device complex algorithms and new techniques to create artificial reverberation using digital filters for creation of virtual environment experiences.

Chapter 4

SOFTWARE

4.1 Android

Android is open mobile device integrated platform which is independent of the device itself, the middleware and some of the main Application., Android provides open architecture and excellent environment for development that makes full use of handheld devices to provide excellent mobile applications for various purposes of entertainment, healthcare, business, social media and almost everything related to daily lifestyle. Currently there are more than 2 million android applications available on the digital distribution platform called Google Play.



Figure 4.1: Android System Architecture.

Android system architecture is comprised of five layers namely application framework, Binder IPC proxies, Android system services, Hardware abstraction layer (HAL) and Linux kernel as shown in Figure 4.1 [6]. Application framework is mostly utilized by application developers using the APIs available for application development. Binder Inter-Process Communication (IPC) layer allows high level framework APIs to interact with Android's system services by calling into the Android system services code and allowing the application framework to communicate across boundaries without notifying the developer. System services are required for communication of application framework APIs functionalities to access the underlying hardware. Services are distributed into modular components with focused functionality but are grouped into two parts as system and media. The system services include things such as the Window or Notification Manager and the media services include all the services involved in playing and recording media. Hardware abstraction layer (HAL) is a standard interface that allows the Android system to connect with the device driver layer while being unaware of the lower-level implementations of the drivers and hardware. HAL implementations are typically built into shared library modules (.so files). Linux Kernel used in Android is a specialized version with a few special additions that are important for a mobile embedded platform like Android.

4.2 Application Design and Architecture

One of the software components of the system is an Android audio player application with audio effects and virtual source motion features. This application is based on front-back end architecture where the front-end is the player interface along with the

access to content and Bluetooth connection. The back-end is the implementation of playback which is on separate thread created using asyncTask. The communications between front and back end are executed via intent [22].

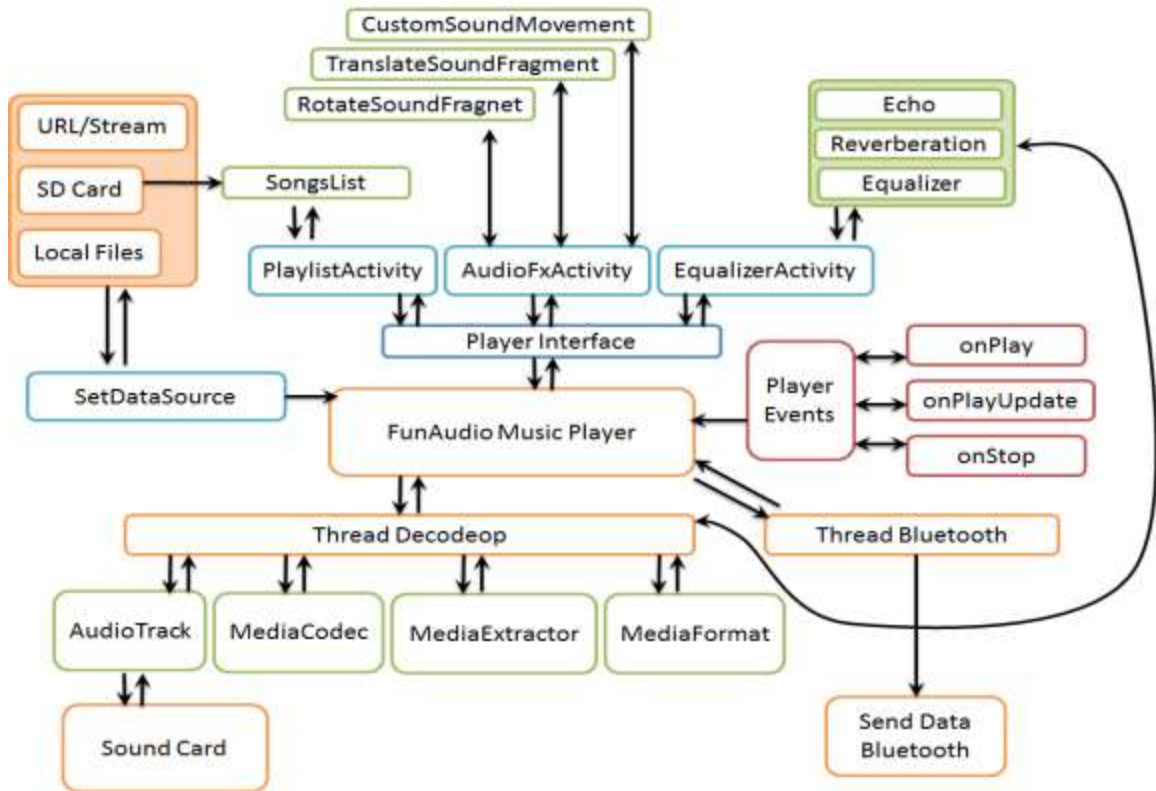


Figure 4.2: Functional Diagram for Android Application.

Figure 4.2 depicts the functional diagram of the application. The application is comprised of three activities [22] namely mainActivity, playListActivity and equalizerActivity. The mainActivity is the main UI class with audio playback functionality as shown in Figure 4.3. The mainActivity provides basic playback functionalities such as play, pause, stop, next previous, repeat etc. and a slide bar to show progress or to set playback start point. It provides buttons for Bluetooth connection, access to equalizerActivity and playListActivity which lets user to select audio content to be played.

It also provides two slide bars for audio effects echo and reverberations which allows user to change delay and apply the effects in real time. Apart from main UI, MainActivity has four additional fragments tabs for individual speaker gain control and virtual source motion which we will discuss separately in next sub section.

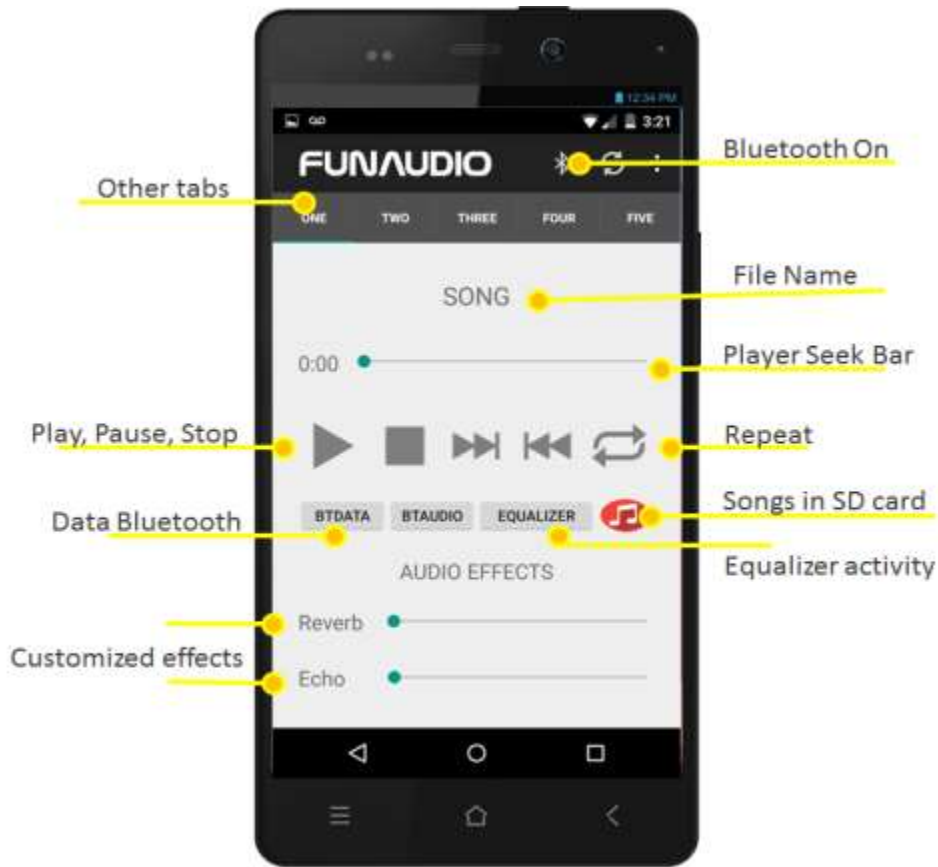


Figure 4.3: Android Application Playback Screen.

A SongsManager class is created to input the details of all audio contents from the device storage in an arrayList. The playListActivity shows the names of all audio files in a ListView format and allows user to select audio files for playback. The equalizerActivity provides vertical slide bars to vary gain for different frequency bands for equalizer control. This equalizer is built using built in AudioFx library provided by Android API.

4.3 Audio Player

This application uses the `audioTracks` library [22] for audio playback and to implement effects such as echo and reverberation. A player class is created for playback [5] that accepts the audio data source, basic commands such as play, pause, stop and repeat and provides events to inform progress and update UI. It can play audio content from device storage or online audio streams. The player class uses the `mediaCodecs`, `mediaExtractor` and `mediaFormat` classes to extract and decode any audio format into raw pulse-code-modulation (PCM) data and make it available for playing in stream mode using `audioTracks` [5]. This decoded audio data from by `MediaCodec` is stored in a buffer which is extracted as a byte array. Different customized filter functions are applied to the byte array depending on the used input for audio effects such as echo and reverberation. Decoding and playback is done asynchronously on a separate thread using `asyncTask` class.

4.4 Bluetooth

The system uses the data Bluetooth module to communicate with the audio hardware. The application enables the user to connect to the Bluetooth device by creating a Bluetooth adapter and sends control data in a separate thread using Bluetooth socket when source motion or gain control tabs are used. For the virtual source motion tabs, it calculates relative distance and angle of the virtual source from its coordinates on screen and sends it to data Bluetooth as control data. `MainActivity` UI provides a button to connect to data Bluetooth module and the connection can be confirmed by led light turning green on the data Bluetooth module. The app also lets connection of mobile device with the audio Bluetooth module which allows wireless audio playback.

4.5 Audio Effects and Virtual Source Movement

The software provides a canvas through different fragments as shown in Figure 4.4 to move virtual source position on screen for the effects such as spatial motion, rotational motion and 3D sound effects. The movement of the cursor on the screen gets translated into space via Arduino controlled speaker system.

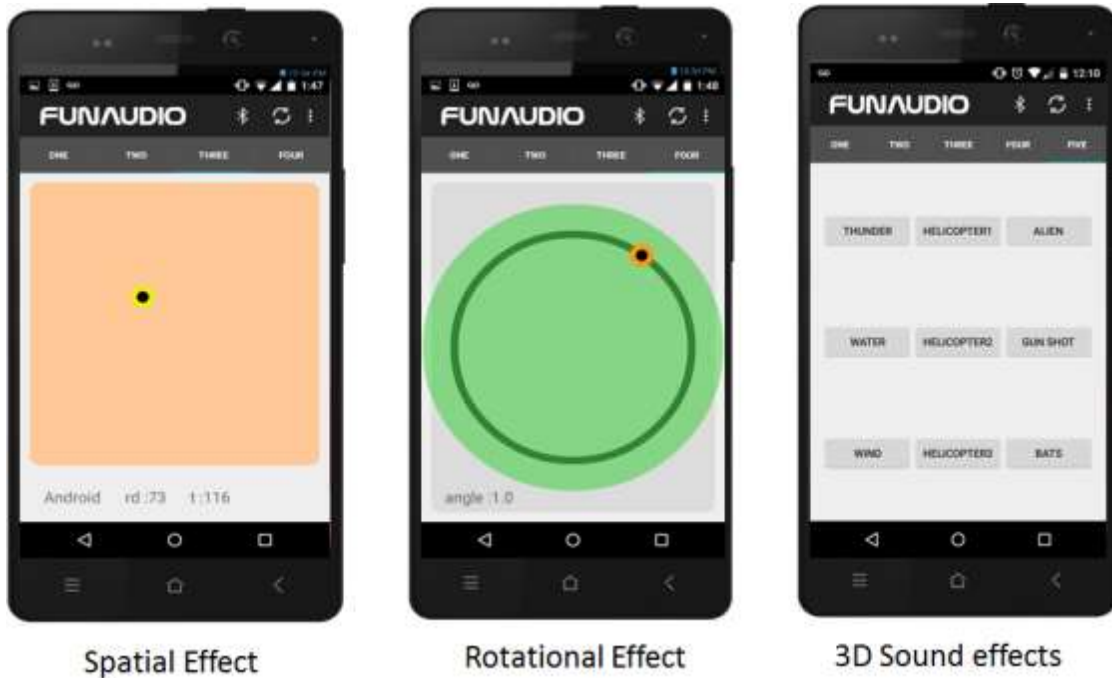


Figure 4.4: Fragments for Virtual Source Motion

The cursor shown at touch coordinates represents the position of the virtual source while listener's position is assumed to be at center. When users touches the screen, as the source distance (rd : 0 to 1) and angles (t : 0 to 360) are calculated and sent to the hardware. For rotational effects r is constant as 1. The applications also has other tab to control volume of each speaker individually.

The software also contains a tab with different sound effects such as vehicles and natural sounds in order to demonstrate 3D sound motion effect. These effects have pre-

determined source motion data to demonstrate sound animation more effectively for the specific sounds.

Chapter 5

HARDWARE

5.1 Components

- **Arduino UNO** - The Arduino Uno is a microcontroller board based on the ATmega328P. It has 14 digital input/output pins, 6 analog inputs, a 16 MHz quartz crystal, a USB connection, a power jack, an ICSP header and a reset button. The board can be configured using the Arduino software (IDE) and microcontroller can be programmed to achieve desired task. The board has a number of services for communicating with a computer, another Uno board, or other microcontrollers using serial communication protocols such as I2C, SPI and UART. The board is a major part of the system hardware in this prototype which receives positional data from mobile device via Bluetooth module and implements audio panning algorithms that provides sound directivity control [7].
- **MCP4100** - The MCP41100 is a single-channel, 8-bit digital potentiometer features 100k Ω end-to-end resistance value. The wiper position varies linearly with 256 taps for each potentiometer and is controlled via the SPI interface. In the project, 5 such digital potentiometers are used to control gain for each speaker wirelessly and are controlled by Arduino board based on the inputs received from mobile device [30].
- **Bluesmirf RN-41Modem** - The Bluesmirf uses the RN41 which is a small form factor, low power, simple to integrate Class 1 Bluetooth radio module. It work as a

serial (RX/TX) pipe and it is used to pass the positional data as a serial stream at 9600bps wirelessly from mobile device to Arduino [40].

- **RN-52** - The RN52 Bluetooth audio module is composed of a Class 2 Bluetooth radio with an embedded DSP processor that provides a fully integrated solution for high-quality wireless stereo audio delivery in a small form factor. The module provides a UART interface, user programmable I/O pins, stereo speaker outputs, microphone inputs, a USB port etc. and it can be programmed and controlled with a simple ASCII command language. It is used for wireless audio playback for this project [41].
- **Amplifiers** - The hardware includes custom made D class 5 channel audio amplifier to boost the audio output from the speakers. The system also uses differential amplifier to obtain amplified output signal from differential audio signals obtained from audio Bluetooth module.

5.2 Hardware Design and Implementation

The audio hardware is built around an Arduino Uno board [1, 7] that controls the speaker system. Two separate Bluetooth modules are used for separate functionalities of the data transfer and wireless stereo audio playback. The board receives data from the android application through a Bluetooth module using UART. Input data is a String made of a tag followed by values that are used to control gain based on the tag functionalities. Based on the data inputs and implementation of panning laws, the Arduino controls 5 digital potentiometers (mcp41100) using serial peripheral interface (SPI) which in turn control the gain of the individual speaker channels. The audio signal received from the

audio Bluetooth is amplified by a differential amplifier being used as input for the digital potentiometer. All channels are connected to their speakers through an audio amplifier that boosts the overall volume of the sound. The system overview and schematic is shown in figure 5.1 and 5.2 respectively. An Rn-52 Bluetooth audio module is used for audio data transfer from Android device to the hardware.

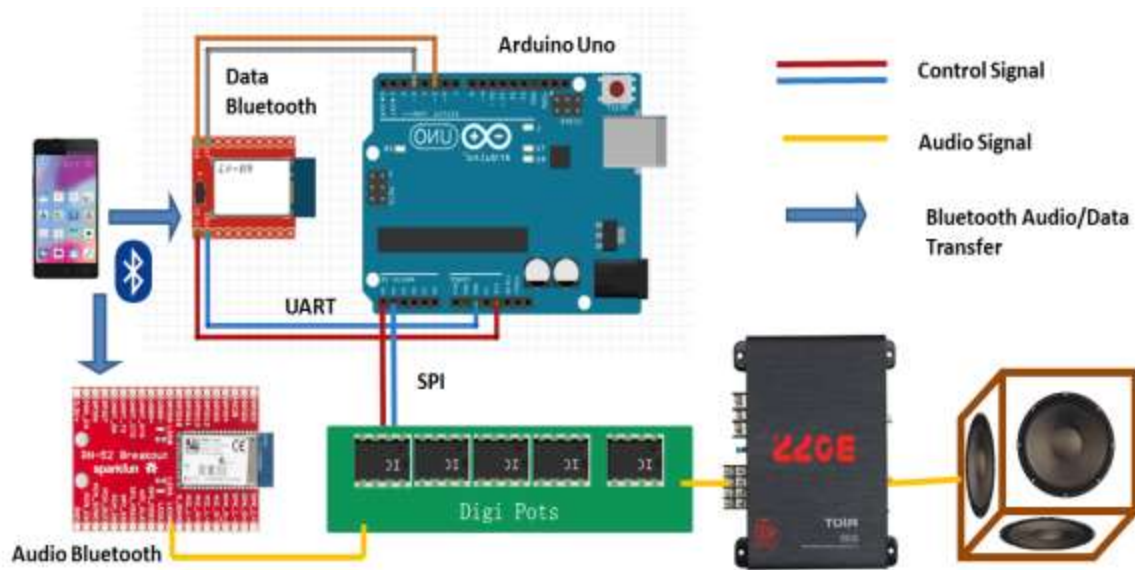


Figure 5.1: Hardware System Design.

5.3 Practical Issues and Solutions

Bluetooth modules add an unacceptable level of high-frequency RF noise to the circuit. The noise is more prominent during data transfer through Bluetooth modules. Audio Bluetooth module adds continuous high pitch noise which deteriorates audio signal as high as it completely eclipse original audio at low volumes. The noise is removed by implementing a separate power source circuit for each of the Bluetooth modules to stop reverse noise signal feed in power circuit.

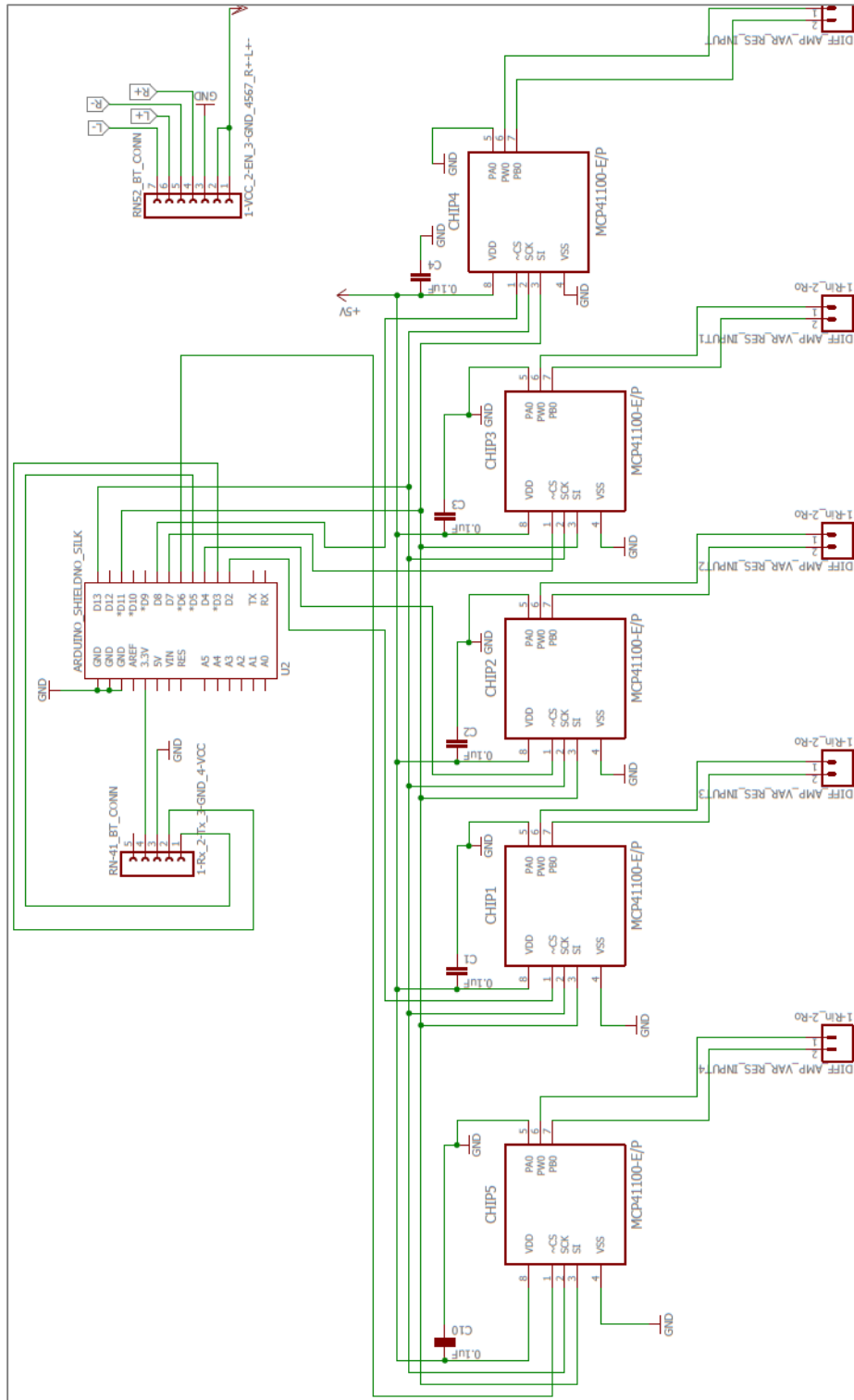


Figure 5.2: Electronic Circuit for Bluetooth Interface and Gain Control

Another major problem arrived with data Bluetooth module getting stuck while receiving data from mobile device at high speed. A touch screen has a reporting rate of 60 Hz for most of the devices in a constant screen contact mode. A function to send Bluetooth data is called every 16.66 milliseconds and data of at least 20 bytes is sent. Hence 160 bits of data is sent at the frequency of 60 Hz which can be calculated to 9600 bps. Any data more than 20 bytes causes rate of data sent to be more than Bluetooth device baud rate of 9600 bps. To resolve this problem, module was set to the baud rate of 19200 bps and the rate of data sent from mobile device was reduced to half by calling the function only once in two times the touch was recorded. This may have caused the slower response of the speaker gain control and sound directivity but the change was observed in the lab and the effect on the response was unnoticeable.

5.4 Interface and Communication

The Arduino Software (IDE) also provides a serial monitor which allows simple textual data to be sent to and from the board hence enables us to observe data transfer between Arduino and other devices. The flashing Rx/Tx LEDs on the board indicates the data is being transmitted via the USB-to-serial chip and USB connection to the computer.

Based on the data input tags received from mobile device through Bluetooth module, Arduino code implements different functions to implement various directivity controls such as rotational effect, spatial effect and also individual speaker gain control. The spatial and rotational effects are controlled with a single function where distance parameter is considered constant as unity where as it is variable data received from used in

case of spatial effect. The function implements equations using audio panning law to calculate gains for each speaker as shown in figure 5.3.

```
void rotateSound(int dimmerR, float dimmerP){

    //Serial.println("I'm in Rotary"+ valorRecebidol); // DELETE LATER
    if(dimmerP > 0){
        dist_weig = dimmerP;
    }else{
        dist_weig = 1;
    }

    rotator_ang = dimmerR;
    const float pi = 3.14;
    vol1 = (int) (dist_weig*255*(cos(rotator_ang*pi/180)/ 2 + .5));
    vol2 = (int) (dist_weig*255*(sin(rotator_ang*pi/180)/ 2 + .5));
    vol3 = (int) (dist_weig*255*(-1*cos(rotator_ang*pi/180) / 2 + .5));
    vol4 = (int) (dist_weig*255*(-1*sin(rotator_ang*pi/180) / 2 + .5));

    digitalPotWrite1(vol2);
    digitalPotWrite2(vol1);
    digitalPotWrite3(vol3);
    digitalPotWrite4(vol4);
}
```

Figure 5.3: Function in Arduino IDE Code for Rotational and Spatial Control

5.5 Speaker Arrangement

The speaker arrangement is the most important aspect of the acoustic design. 4 channels are used for the 4 speakers in 4 directions in the same plane whereas fifth channel is the vertical overhead speaker channel. The speakers are arranged in a rectangular box in order to achieve required acoustic design and portability. The speaker box is made up of thick paper board with top of the box kept open. The space inside the box between speakers is filled with shock absorbent insulating foam. The top face has space to mount electronics hardware.

Chapter 6

OBSERVATION AND ANALYSIS

6.1 Experimental Study

The evaluation of the system was carried out from the results of a user-based study. System features such as surround sound and user controlled sound directivity effects were assessed for effectiveness by a group of students and faculty. These features were found to be very innovative and effective in providing a simple but unique immersive experience. The translation of movement from the cursor on the device canvas to sound movement is interactive and real-time. Sound effects in the app with different predetermined sound animations in 3D space were demonstrated. These effects demonstrated sound directivity and audio transition of sound from one point of the room to another with a game-like interface. In essence the application allowed the user to play a game with sound directive and audio effects. The array of sound effects available with our app and software are listed in Table 1 with their names and path of their motion.

Table 1

Sound Effects for 3D Audio.

Sound	Movement
Helicopter	Passing Overhead
Thunder	Overhead in random directions
Gun shot	Rotation in circle
Alien	Rotating sound clockwise/anticlockwise

6.2 Applications in Education and Outreach

The system was presented to the ASU Digital Culture class in a course called Signal processing for digital culture. Digital Culture [28, 46] is an interdisciplinary undergraduate elective course designed to teach Digital Signal Processing basics [46, 47] and applications in gaming, sound and media performances [11, 28, 35, 46, 47, 48]. The course covers basic theory of DSP such as time and frequency domain analysis, sampling, digital FIR and IIR filters and the FFT [47].

The prototype developed in this project was presented to the students to demonstrate real-time audio effects, sound directivity control methods and sound animation. Simple coding and implementations of theoretical equations of filters and effects such as reverberations and echo were demonstrated. This exposed students to the aspects of developing combined hardware-software projects for arts. Furthermore, it helped students understand how apps can be developed to deliver unique arts and media experiences [32, 37].

The authors used two class sessions and tasked students with using the app evaluating various aspects of the software and assessing immersive sound experience. Students also were tasked with making suggestions for augmenting app functionality. The exposure of the arts students to this application was important in that it promotes the app in multidisciplinary non-engineering environments. Arts students were able to provide a different perspective on experiential media that was unique. An evaluation instrument was developed and disseminated to the students. Interviews were conducted following the demonstration of the system. Students reported that the system and its applications were

intriguing and they appreciated specifically the virtual source movement. All the aspects of the app were assessed and interviews provided ideas to the developers for new functionality [15].

6.3 Importance, Uniqueness and Comparison with Other Work

This system copes with the most of the issues discussed above. The system is a compact single unit and portable. It can be connected to any audio source capable of Bluetooth connection. The system can be used similar to a spotlight with direction control. One can control the audio source delivering in one direction to avoid great sound disturbance to others. The system has the advantage of vertical channel over other major audio systems which has horizontal plane functionality and gives capability of 3D audio. The game-like interface is very uniquely interactive to change sound directivity and spatial movement of virtual sound source. The system can be used in any enclosed space with provision of hanging it from ceiling and wall reflections provide immersive sound effect as depicted in Figure 6.1. The android application provides unique sound player having custom audio effects such as echo and reverberation with variable delay parameters for sounds played from any source. The system can be much cheaper than multichannel audio systems.



Figure 6.1: Audio System Working and Application.

Chapter 7

CONCLUSION AND FUTURE WORK

7.1 Summary

This work presents the details of the design, working principle and development of hardware and software components of a novel spatial sound audio system. It discusses emerging trends and current techniques for audio reproduction systems. The speaker box arrangement and Android application controlling the system provide wireless connectivity, portability and interactive audio effects. This novel approach allows demonstration of 3D virtual sound motion and the ability to play with sound directivity using smart devices. When used in a classroom setting the system was found to effectively demonstrate the sound motion effects and sparked student interest in audio research.

In the android software, a lower level API for audio playback is used and therefore android libraries for obtaining format, audio codecs and content extraction in streaming mode are implemented in a separate thread. This enables addition of customized audio effects such as echo and reverberation and scope for adding digital filters and other audio effects. The application also provides interactive interfaces for orchestrating sound movements with a game-like experience to play with sound directivity in specific trajectories as well as spatial movements in free form. Such interactive features provide a platform and scope of implementing more creative interactive ways to play with sound directivity and interesting real life sound effects.

The electronic hardware used in the system is simple yet effective for the application and uses Bluetooth for wireless communication with Android mobile device. Unique system design saves user from clumsy speaker setup and wiring. The system developed was robust, noiseless and can be made compact by integrating complete circuit on a single PCB and by using small high quality small size speakers.

The project progressed through many challenges in hardware, software as well as in the performance evaluation and most of them were dealt with while some are projected for the future development of the system. Multiple Bluetooth devices connected in audio circuit caused high frequency RF noise through supply network which was eliminated by separating power supply circuits of audio and Bluetooth. Sensitivity of the sound movement with respect to local pointer movement on mobile screen required a great amount of tuning and delays were eliminated by adjusting baud rate of the Bluetooth device in accordance with the average frequency of the touch sensors of the mobile device and the data transmitted per cycle. Another major challenge is the performance variation of the system with large variation in wall dimensions along horizontal plane. This problem planned for future work and can be solved by capturing impulse response of the surrounding and using audio processing algorithms to deliver uniform experience in all environments.

7.2 Future Work

Further development in the prototype can be done by measuring the room frequency response on the mobile device and using signal processing algorithms for the system to adapt to any surrounding to produce precise audio effects for different surroundings.

System can be improved to play new 3D multichannel sound formats for movies and games. The array processing techniques can be used for learning, experimenting as well as creating additional immersive effects. The extensions of the array processing techniques can influence the research in other areas such as antenna arrays. The android application can be uploaded on app store and interface can be made more flexible for users to add new customized virtual object sound effects for education as well as entertainment purpose.

REFERENCES

- [1] A. Alteneiji. (Fall 2013) Multi-output Receiver System, Senior Project, ASU.
- [2] Abrash M. (March 2016). Welcome to the Virtual age, Oculus Blog. Retrieved from <https://www.oculus.com/en-us/blog/welcome-to-the-virtual-age/>
- [3] AESTD1001.0.01-05 (2001) Multichannel surround sound systems and operations. Audio Engineering Society, Technical Council, New York, NY.
- [4] Ahrens, J., Rabenstein, R., & Spors, S. (2014). Sound field synthesis for audio presentation. *Acoust. Today* 10(2), 15-25. Retrieved from <http://acousticstoday.org/sound-field-synthesis-for-audio-presentation-jens-ahrens-rudolf-rabenstein-and-sacha-spors/>
- [5] Android open source Audio player using mediaCodec API. Retrieved from <https://github.com/radhoo/android-openmxplayer>.
- [6] Android System Architecture, Android Interfaces and Architecture. Retrieved from <https://source.android.com/devices>.
- [7] Arduino Uno Board. Retrieved from www.arduino.cc/en/main/arduinoBoardUno.
- [8] Atkins, J., & Giacobello, D. (2014). Trends and Perspectives for Signal Processing in Consumer Audio. arXiv preprint arXiv:1405.4843.
- [9] Bai, M. R., Ih, J. G., & Benesty, J. (2013). *Acoustic Array Systems: Theory, Implementation, and Application*. John Wiley & Sons.
- [10] Bakht S. and Shear G. (2007) Welcome: Technical Background. Retrieved from <http://www.mat.ucsb.edu/sbakht/welcome/background2.html>.
- [11] Banavar, M. K., Rajan, D., Strom, A., Spanias, P., Zhang, X. S., Braun, H., & Spanias, A. (2014, October). Embedding Android signal processing apps in a high school math class—An RET project. In *Frontiers in Education Conference (FIE)*, 2014 IEEE (pp. 1-4). IEEE.
- [12] Blaha, J., & Gupta, M. (2014, March). Diplopia: A virtual reality game designed to help amblyopics. In *Virtual Reality (VR)*, 2014 IEEE (pp. 163-164). IEEE.
- [13] Brice R., (October 2008), *Ambisonics*, White paper, Pspatial audio Retrieved from <http://www.pspatialaudio.com/Ambisonics%202014%20article.pdf>

- [14] Cabrera A. (Fall 2015) Ambisonics, 240D Spatial Sound. Retrieved from <http://mat.ucsb.edu/240/D/notes/Ambisonics.html>.
- [15] Dharmadhikari, C., Spanias, A., & Ingalls, T. (February 2016) Development of Hardware and Software for a Game-Like Wireless Sound Distribution System. MIC, IASTED, Innsbruck, Austria.
- [16] Dolby Atmos home theatre installation guideline. Retrieved from <http://www.dolby.com/us/en/technologies/dolby-atmos/dolby-atmos-home-theater-installation-guidelines.pdf>.
- [17] Dolby Digital Atmos. Retrieved from <http://www.dolby.com/us/en/brands/dolby-atmos.html>
- [18] Donalek, C., Djorgovski, S. G., Cioc, A., Wang, A., Zhang, J., Lawler, E., & Davidoff, S. (2014, October). Immersive and collaborative data visualization using virtual reality platforms. In Big Data (Big Data), 2014 IEEE International Conference on (pp. 609-614). IEEE.
- [19] DTS Headphone:X. Retrieved from <http://listen.dts.com/pages/headphone-x>.
- [20] Fels, J. (2013). Trends in Binaural Technology. AIA-DAGA, Merino, Italy.
- [21] Foutz, J., Spanias, A., & Banavar, M. K. (2008). Narrowband direction of arrival estimation for antenna arrays. *Synthesis Lectures on Antennas*, 3(1), 1-76.
- [22] Google Android developers. Retrieved from <http://developer.android.com/index.html>.
- [23] Herre, J., Hilpert, J., Kuntz, A., & Plogsties, J. (2015). MPEG-H 3D Audio—The New Standard for Coding of Immersive Spatial Audio. *Selected Topics in Signal Processing*, IEEE Journal of, 9(5), 770-779.
- [24] Hollerweger, F. (2008) An Introduction to Higher-order Ambisonic. Retrieved from <http://flo.mur.at/writings>
- [25] Kirner, C., & Kirner, T. G. (2006, October). A data visualization virtual environment supported by augmented reality. In *Systems, Man and Cybernetics, 2006. SMC'06. IEEE International Conference on (Vol. 1, pp. 97-102)*. IEEE.
- [26] Kleven, N. F., Prasolova-Forland, E., Fominykh, M., Hansen, A., Rasmussen, G., Sagberg, L. M., & Lindseth, F. (2014, December). Training nurses and educating the public using a virtual operating room with Oculus Rift. In *Virtual Systems & Multimedia (VSMM), 2014 International Conference on (pp. 206-213)*. IEEE.

- [27] Leishman, T. W., Rollins, S., & Smith, H. M. (2006). An experimental evaluation of regular polyhedron loudspeakers as omnidirectional sources of sound. *The Journal of the Acoustical Society of America*, 120(3), 1411-1422.
- [28] Liu, J., Hu, S., Thiagarajan, J. J., Zhang, X., Ranganath, S., Banavar, M. K., & Spanias, A. (2012, March). Interactive DSP laboratories on mobile phones and tablets. In *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on* (pp. 2761-2764). IEEE.
- [29] Mahalil, I., Rusli, M. E., Yusof, A. M., Mohd Yusoff, M. Z., Zainudin, R., & Redza, A. (2014, November). Study of immersion effectiveness in VR-based stress therapy. In *Information Technology and Multimedia (ICIMU), 2014 International Conference on* (pp. 380-384). IEEE.
- [30] MCP41100. Retrieved from <http://www.microchip.com/wwwproducts/en/MCP41100>
- [31] Moore, B. C. J. (2003). *An introduction to the psychology of hearing*. Academic Press.
- [32] Painter, T., & Spanias, A. (2000). Perceptual coding of digital audio. *Proceedings of the IEEE*, 88(4), 451-515.
- [33] Pulkki, V. (1997). Virtual sound source positioning using vector base amplitude panning. *Journal of the Audio Engineering Society*, 45(6), 456-466.
- [34] Rabenstein, R., & Spors, S. (2008). Sound field reproduction. In *Springer Handbook of Speech Processing* (pp. 1095-1114). Springer Berlin Heidelberg.
- [35] Rajan, D., Spanias, A., Ranganath, S., Banavar, M., & Spanias, P. (2013, October). Health monitoring laboratories by interfacing physiological sensors to mobile android devices. In *Frontiers in Education Conference, 2013 IEEE* (pp. 1049-1055). IEEE.
- [36] Rakerd, B., & Hartmann, W. M. (1985). Localization of sound in rooms, II: The effects of a single reflecting surface. *The Journal of the Acoustical Society of America*, 78(2), 524-533.
- [37] Ramamurthy, K. N., & Spanias, A. S. (2010). MATLAB® Software for the Code Excited Linear Prediction Algorithm: The Federal Standard-1016. *Synthesis Lectures on Algorithms and Software in Engineering*, 2(1), 1-109.
- [38] Ranjan, R., & Gan, W. S. (2015, April). A hybrid speaker array-headphone system for immersive 3D audio reproduction. In *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on* (pp. 1836-1840). IEEE.
- [39] Real Space 3D Audio. Retrieved from <http://realspace3daudio.com/>

- [40] RN-41 Bluetooth module. Retrieved from <https://learn.sparkfun.com/tutorials/using-the-bluesmirf>
- [41] RN-52 Audio Bluetooth Module. Retrieved from <https://learn.sparkfun.com/tutorials/rn-52-bluetooth-hookup-guide>
- [42] Rumsey, F., Schulein, B., Herre, J., & Kelly, M. (2015). Technology Trends in Audio Engineering A report by the AES Technical Council.
- [43] Savioja, L., Ando, A., Duraiswami, R., Habets, E. A., & Spors, S. (2015). Introduction to the Issue on Spatial Audio. *Selected Topics in Signal Processing, IEEE Journal of*, 9(5), 767-769.
- [44] Sharma, G. K., Zotter, F., & Frank, M. (2014). Orchestrating wall reflections in space by icosahedral loudspeaker: findings from first artistic research exploration, ICMC-SCM Athens.
- [45] Shinn-Cunningham, B. G. (2000, December). Distance cues for virtual auditory space. In *Proceedings of the IEEE-PCM* (Vol. 2000, pp. 227-230).
- [46] Spanias, A. (2015, August). An introductory signal processing course offered across the curriculum. In *Signal Processing and Signal Processing Education Workshop (SP/SPE), 2015 IEEE* (pp. 55-58). IEEE.
- [47] Spanias, A. S. (2007). *Digital signal processing: An interactive approach*. J-DSP Editor. 2nd Edition, 403 pages, Textbook, ISBN 978-1-4675-9892-7, Lulu Press On-demand Publishers Morrisville, NC.
- [48] Spanias, A., & Painter, T. (May 1996). An Educational Software Tool for the Study of Speech Coding Algorithms in a DSP Class, *Special Issue on DSP Education, IEEE Trans. on Education*, pp. 143-152.
- [49] Spanias, A., Painter, T., & Atti, V. (2006). *Audio signal processing and coding*. John Wiley & Sons.
- [50] Spors, S., Wierstorf, H., Raake, A., Melchior, F., Frank, M., & Zotter, F. (2013). Spatial sound with loudspeakers and its perception: A review of the current state. *Proceedings of the IEEE*, 101(9), 1920-1938.
- [51] Tang, Y., Fang, Y., & Huang, Q. (2011, May). Audio personalization using head related transfer function in 3DTV. In *3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), 2011* (pp. 1-4). IEEE.

- [52] Thiagarajan, J. J., & Spanias, A. (2011). Analysis of the MPEG-1 Layer III (MP3) algorithm using MATLAB. *Synthesis Lectures on Algorithms and Software in Engineering*, 3(3), 1-129.
- [53] Torres, J. C. B., Petraglia, M. R., & Tenenbaum, R. A. (2003, June). HRTF modeling for efficient auralization. In *Industrial Electronics, 2003. ISIE'03. 2003 IEEE International Symposium on* (Vol. 2, pp. 919-923). IEEE.
- [54] Ward, D. B., & Elko, G. W. (2000). A new robust system for 3D audio using loudspeakers. In *Acoustics, Speech, and Signal Processing, 2000. ICASSP'00. Proceedings. 2000 IEEE International Conference on* (Vol. 2, pp. II781-II784). IEEE.
- [55] Wikimedia Commons, the free media repository. Retrieved from https://commons.wikimedia.org/wiki/File:Hrir_binaural_synthesis.png#filelinks.
- [56] Winter, F., & Spors, S. (2015). Physical Properties of Local Wave Field Synthesis using Circular Loudspeaker Arrays. In: *Proc. of the 10th European Congress and Exposition on Noise Control Engineering*.
- [57] Zhong, X. L., & Xie, B. S. (2014). Head-Related Transfer Functions and Virtual Auditory Display. *Soundscape Semiotics - Localization and Categorization*, Dr. Hervé Glotin (Ed.), 2014, InTech, DOI: 10.5772/56907.
- [58] Zotkin, D. N., Duraiswami, R., & Davis, L. S. (2002, May). Creation of virtual auditory spaces. In *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on* (Vol. 2, pp. II-2113). IEEE.