Sindh Univ. Res. Jour. (Sci. Ser.) Vol.44 (3) 367-374 (2012)

SINDH UNIVERSITY RESEARCH JOURNAL (SCIENCE SERIES)

On the Requirements of 3D Audio Telephone and Teleconferencing Calls: A Mini Review on Background of 3D Audio – From Human Hearing toTechnology

M. HYDER, G. DAS, I. A. QURESHI*, M. MEMON, A. JALBANI, E. SABA

Information Technology Centre, Sindh Agriculture University Tandojam, Pakistan

Received 2nd April 2012 and Revised 11th June 2012

Abstract: Audio quality of telephones and teleconferencing solutions is much talked issue these days. Despite the significant advances in the technology there is hardly any improvement in the audio call quality. Incorporation of 3D audio is the one way to improve the audio quality of telephone or teleconferencing calls. In order to develop and to design telephone and teleconferencing solutions supporting 3D audio, we have to understand the 3D audio phenomenon and its associated concepts. In this article, we present a mini review on the background of 3D audio. Especially, focusing binaural technology in general, 3D audio localization and localization cues, binaural technology, audio spatialization techniques such as Ambisonics and wave field synthesis. Also, we discuss technology behind 3D audio recording and reproduction by focusing on dummy head recording, B-format, Head Related Transfer Functions (HRTFs) and its characteristics, and amplitude panning techniques for the reproduction of 3D audio over loudspeakers.

Keywords: 3D Audio, Binaural Technology, Telephony, Teleconferencing, HRTFs

1 INTRODUCTION

Telephones have revolutionized the way people communicate personally and/or professionally. According to the International Telecommunication Union (ITUT), at the end of January 2011, the number of telephone subscribers in India only crossed the 800 million mark (ITU-T, 2011). Additionally, there were more than 539 million fixed broadband subscribers globally at the end of second quarter in the year 2011 (ITU-T, 2011). Also, it is evident that there were five billion mobile cellular subscribers globally including 940 million subscription to 3G services at the end of year 2010 (ITU-T, 2010). Despite these significant advances in the number of subscribers for fixed, mobile and broadband services, the audio quality of the calls for the classic, as well as modern, telephony, teleconferencing has hardly improved. It has been observed that audio call quality drops even further when it comes to the teleconferencing situation where more than two persons take part in a conversation. In (Yankelovich, *et al.,* 2004), major audio related problems in teleconferencing scenarios have been reported. The problems include: some people could not be heard, difficulty to identify who was speaking, poor audio quality and too much extraneous noise.

On the other hand, incorporation of 3D audio in telephony and teleconferencing systems is considered advantageous in literature (Kilgore, *et al.,* 2003, Yankelovich, *et al.,* 2006, Ahrens, *et al.,* 2010, Hyder, 2010). Further, 3D audio1 improves overall audio quality of communication and helps to overcome problems Begault, D.R. 1994) such as: poor audio quality, extraneous noise and lake of identifying interlocutors during conversation. In the present article we review background of 3D audio. The main focus is on four specific areas. First, we discuss human hearing organ the ear. Also, we review binaural hearing, sound localization and localization cues (see Sections 2 and 3). Since, humans natural listening ability is inherently 3 dimensional or binaural. We see it as an important task to review binaural hearing and localization cues which help us to locate audio events around us. Second, we discuss technologies behind spatial audio especially we focus on binaural technology and also discuss Ambisonics and Wave field synthesis (WFS) which are sound spatialization techniques (see Sections 4). Binaural technology is basically aimed at reproducing human ability of listening in 3D dimensions in to the computational systems. Third, we discuss 3D audio recording and reproduction systems especially we focus on dummy head and B-format recording (see Section 5). Fourth, for 3D audio reproduction over headphones we discuss Head Related Transfer Functions (HRTFs) by discussing individualized vs non-individualized HRTFs and their characteristics. For 3D audio reproduction over loudspeakers we discuss amplitude panning techniques (SeeSection 6).

In the next section, a brief background of anatomy physiology of human ear is presented.

++ Corresponding author: mansoor.hyder@sau.edu.pk

**Department of Telecommunication Engineering, Mehran University of Engineering and Technology,

2 Hearing: Ear the listening organ

The ear is the series of interlinked structures which provides humans sense of hearing. Hearing involves collection of sound waves by ear as pressure changes in the air and sending these sound waves to brain. A depiction of human ear is provided **(Fig.1).** The outer part of the ear consists of the pinna and concha, which leads to the eardrum via the ear canal. The outer ear acoustically filter incoming sound waves that vibrates the eardrum.

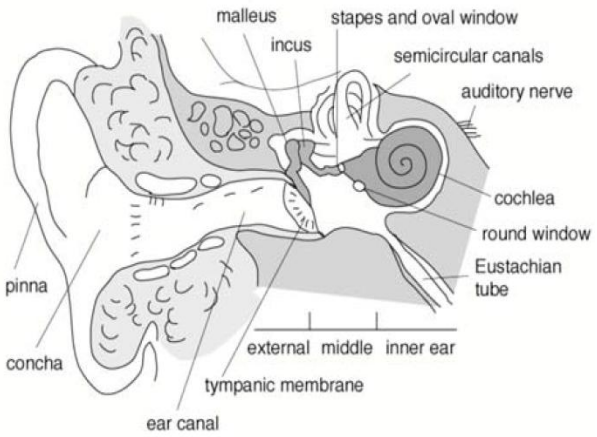


Fig. 1: Interaural time differences (ITDs). (Adopted from (Karjalainen, 2011))

The middle ear is air filled cavity that contains ossicular chain which consist of small bones the malleus, incus and stapes, and also contains eardrum. As reported by (Irwin, 2006), the main function of the middle ear is to pass on the vibrations of sound in air gathered at the tympanic membrane to the fluid of the inner ear at the oval window.

Inner ear contains the Cochlea and semicircular canals. The cochlea is a coiled tube, located in the temporal bone of the skull, which is divided along its length by membranes into three fluid-filled compartments. The vibration of the eardrum causes pressure waves to travel through the fluid of the cochlea, setting up travelling waves in the lower basilar membrane, which is approximately 35 mm long. On the basilar membrane is the organ of Corti, which includes several rows of hairs, which are in contact with tectorial membrane. When the basilar membrance vibrates there is a difference in motion between the basilar and tectorial membrane. This causes that hairs in the hair cell bend. Bending of the hairs causes hair cells to send impulses to the auditory nerve. To simplify, these impulses are understood as sound by the brain. For further details on the anatomy and physiology of human hearing readers are referred to (Irwin, 2006, Pickles, 1988).

3 Binaural Hearing

Binaural hearing is defined as the process required to use two ears to perceive the location of sound sources (Wightman, and Kistler, 1997). The Duplex theory presented by (Strutt, 1907) was the foremost extensive study of the physics of the binaural perception of audio and this theory is still considered as valid. There are two physical cues dominating the perceived location of an incoming sound source (Strutt, 1907) **(Fig. 2 and 3),** where sound arrives slightly earlier in time at the ear which is physically closer to the source and with somewhat greater intensity. Thus, Interaural Time Difference (ITD) is produced because sound takes longer time to reach at the ear which is farther from the source. It also produces, Interaural Intensity Difference (IID) because of the shadowing effect of the head (Cheng, and Wakefield, 2001). Binaural hearing also enables us to selectively attend to individual conversation when there are many people having conversations at the same time which is termed as cocktail party effect (Cherry, 1953, Crispien, and Ehrenberg, 1995, Brungart, *et al.*, 2007). Conclusively, binaural hearing provides directional localization to help precisely focus sounds in horizontal median plane in particular (Vorländer, 2008).

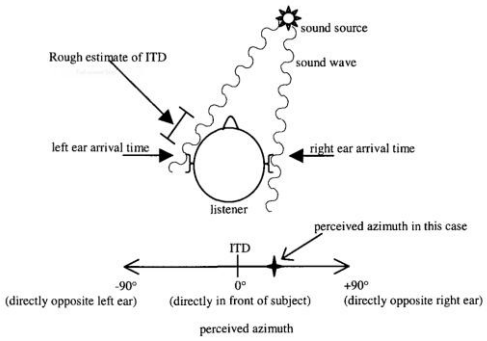


Fig. 2: Interaural time differences (ITDs) (adopted from (Cheng, and Wakefield, 2001))

3.1 Sound Localization

Sound localization is a complex human process. Normally, to calculate the position of sound source humans take advantage of binaural hearing. In (Blauert, 1997) sound localization has been defined as; Localization is the technique by which the location of an auditory event i.e its direction or distance, is related to a specific attribute of a sound event, or of another event that is in some way correlated with the auditory event.

3.2 Sound Localization Cues

Normally, human auditory system utilizes different acoustical cues to achieve static sound localization. According to (Strutt, 1907, Blauert, 1997, Begault, 1994, Tsakostas, *et al.*, 2007) and as explained in (Hirahara, *et al.*, 2011), the first acoustical cue is ITD

which is described as the disparity in arrival times of wavefront of sound wave at the right and left ears and normally calculated from lower frequency components below 1.5 kHz, and the second acoustical cue is Interaural Intensity Difference (IID) which is defined as the difference in amplitude produced between the right and left ears by a sound in the free field and normally calculated from higher frequency components above 1.5 kHz. An illustration of ITD and IID is presented **(Fig, 2 and 3).**

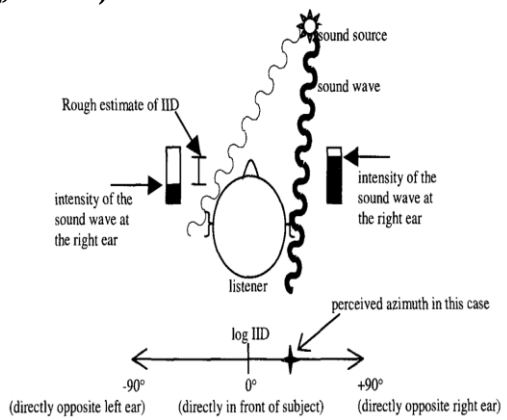


Fig.3: Interaural Intensity differences (IIDs)(adopted from (Cheng, and Wakefield, 2001))

The third acoustical cue is spectral cue which is defined as the spectral notches appearing in the mid frequency rang of around 4 to 14 kHz of the amplitude spectrum of HRTFs. All three above discussed cues are known as static cues (Blauert, 1997).

When either a sound source or the listener moves all these static cues for sound localization change. This change in static localization cues may be called as dynamic cues. With dynamic cues not only moving but static sound can also be localized, since the head movements have tendency to change physically static sound to perceptually dynamic sound (Hirahara, *et al.*, 2011). (**Fig. 4)** For further reading, readers are referred to (Blauert, 1997, Begault, 1994, Hirahara, *et al.*, 2011).

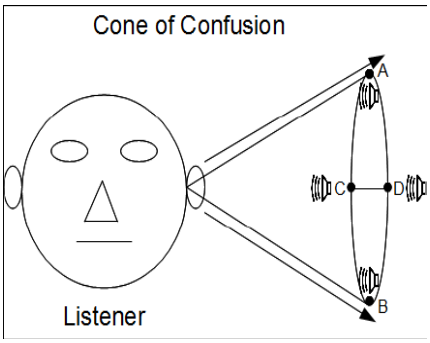


Fig. 4: Cone of Confusion: Points A and B and points C and D are having same ITD and IID and therefore create cone of confusion for the listener according to duplex theory.

4 Binaural Technology and 3D sound

4.1 3D sound

The natural human hearing encounters different sounds everyday from different directions and distances. Natural hearing is defined as the sounds we hear in everyday life with uncovered ears, our head in motion and also in interaction with other sensory inputs. It is also clear that human hearing is inherently three dimensional, we experience not only the horizontal and vertical directions of the sound but also its distance (Begault, 1994, Ericson, and McKinley, 1997), and this is where we encounter the term 3D sound. **(Fig. 5).**

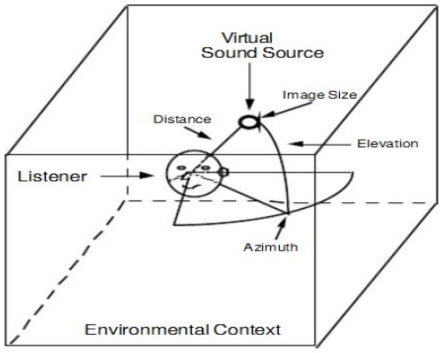


Fig. 5: Three Dimensional Hearing, (Listener's perspective). (Adopted from (Begault, 1994))

In literature, 3D sound's equivalent designations include virtual acoustics, binaural sound and spatial audio. Fundamentally, all of these designations of 3D sound refer to methods where the outer ears (the Pinna) are directly modeled as digital filters (Begault, 1994).

3D sound is defined as the simulation of 3D sound field for any real environment with the use of various techniques. 3D sound system uses processes that either complement or replace spatial attributes that existed originally in association with a given sound source (Begault, 1994). Also, 3D sound refers to a sound which makes a listener differentiate significant spatial cues for a sound source such as distance and direction. Thus in generating 3D sound, the sound could be placed anywhere ,left or right, up or down, near or far at one's disposal in 3 dimensional space (Begault, D. 1994, Kim, *et al.*, 2004, Low, and Babarit, 1998, Lee, *et al.*, 1998).

4.2 Binaural Technology

Binaural technology is aimed at reproducing human ability of binaural hearing in computational systems to generate the same sound pressures at the listeners eardrums as a real sound source would have produced. The principle of binaural technology is to

control the sound field at the listener's ears so that the reproduced sound field coincides with what would be produced when the listener is in the desired real sound field (Kim, *et al.,* 2005). As explained in (Tsakostas, *et al.,* 2007), in binaural technology, the physical cues are incorporated in Head Related Transfer Functions (HRTFs) which represent directional dependent transfer function between the human listener's ear canal and the specific sound source placement (Pulkki, 1997). Therefore, convolving the mono sound source wave with the appropriately selected pair of HRTFs produce the sound waves that corresponds to each of the listener's ears. This process is termed as binaural synthesis. By employing binaural synthesis approach virtual sound sources can be created around the listener with in an open or closed space. This concept is termed as binaural rendering through which 3D virtual sound environments can be created (Kim, *et al.,* 2005, Tsakostas, *et al.,* 2007, Cheng, and Wakefield, 2001). For binaural audio playback either headphones or loudspeakers are utilized.

4.3 Ambisonics

Ambisonics is a technique for sound spatialization. Ambisonics system is a thorough approach of directional sound pickup and reproduction system in studio or live or ideally in both by encoding and decoding a two-dimensional, (planer or horizontal-only), OR three-dimensional (periphonic or full sphere) sound field (Gerzon, 1973, Rumsey, 2001). Ambisonics is logical extension of the work of (Blumlein, A., 1931) which is known as blumlein technique defined as: when using the two figure-of-eight capsules positioned perpendicular to each other any other response of figure of eight could be created (Gerzon, 1970, Gerzon, 1970b). For further details on Ambisonics theoretical and technical basis readers are referred to (Gerzon, 1973, Gerzon, 1974, Gerzon, 1975, Malham, Myatt, 1995).

4.4 Wave-Field-Synthesis

Wave-field synthesis (WFS) concept was introduced in (Berkhout, 1988). With WFS a sound field can be generated with natural temporal and spatial properties within a volume or area bounded by arrays of loudspeakers (De Vries, Boone, 1999). WFS is a 3D audio rendering technique which is based on the concept of Huygens (Huygens, 1690). WFS generates artificial wave fronts synthesized by a large number of individually driven speakers, where all points on the wave front serve as individual point sources of spherical secondary wave fronts. In acoustics this principal can be realized by using a large number of small and closely spaced loudspeakers (loudspeaker strips) (Brandenburg, *et al.,* 2004). Corresponding driving signals are then fed to the each loudspeaker in the array (Brandenburg, *et al.,* 2004, De Vries, 1996). For further details, readers are referred to (Berkhout, 1988, Boone *et al.,*1995, Berkhout, *et al.,*1993, Brandenburg, *et al.,*2004).

5 3D Audio Recording

5.1 Dummy Head

A purpose behind dummy head creation was representation of an average head representing an average listener for the evaluation of hearing aids (Burkhard , and Sachs, 1975, Shaw, 1982). But today, besides other usages, dummy heads are used for recording and reproduction of 3D audio for real and virtual environments.

The dummy head recording is a technique to achieve binaural recording. In dummy head recording a replica of human head is used which is called an artificial head or a dummy head. As explained in (Pekonen, 2008), in dummy head recording two microphones are positioned inside the ear canals of a dummy head. Binaural recording with dummy head has an advantage of spatially-dependent filtering caused by the dummy head since sound waves reaching the dummy head undergo the same transmission on their way to the ear canal as if they were reaching a real listener (Moller, 1992). In the dummy head recording technique the binaural microphones achieve signals that listener ears would have achieved if positioned at the same place as the microphones (Sherman, 1953, Kürer, *et al.,*1969). With recorded sound binaural effects can be reproduced with headphones playback.

5.2 B-format

B-format is a technique of Ambisonics which was developed on the basis of the work of (Gerzon, 1973, Gerzon, 1974). The B-format is a recording standard that utilizes four channels W,X,Y and Z (Gerzon, 1973) and uses a sound field microphone. The W channel represents acoustic pressure at a point in space while other channels represent components of the pressure gradient at left-right (X), front-back (Y) and up-down (Z) signals stem from figure of-eight microphones. Where, the W channel is feed from an omnidirectional microphone (Vorländer, 2008, Barrett, and Berge, 2010). Any source material such as synthesized sound or mono recordings can be positioned or moved within B-format sound field. For further reading, a recent work of (Vilkamo, 2008) and a work of (Malham, and Myatt, 1995) are referred.

6. 3D Audio Reproduction

3D audio reproduction may be defined as recreating the acoustic signals at the ears of the listener in such a way that these signals are equal to a recorded or synthesized audio scene (Blauert, 1997). For 3D audio reproduction with HRTFs headphones are

typically used. However, amplitude panning method has also been described in the following sections to cover 3D audio reproduction using loudspeaker based solution.

6.1 Head Related Transfer Functions (HRTFs)

Knowledge regarding human perception of sound, particularly about human ability to locate a spatial sound is very important to obtain for generating a realistic auditory environment (Zhang, *et al.*,2010). Listeners can differentiate separate sound sources by the sound location which may be determined by the intensity difference and time or phase difference between sound signals at the ears (Middlebrooks, and Green, 1991). Furthermore, Jens Blauert in his book (Blauert, 1997) pointed out that, the sound signals in the ear canals i.e the ear input signals are the most important signals to the subject for spatial hearing and that the quality of spatial hearing could be easily influenced by the alteration of signals at ears.

6.1.1 Individualized vs Generic HRTFs

HRTFs can be either individualized or generic. Since, there are anthropometric shape and size differences among the subjects (Wenzel, *et al.*,1993). Due to these differences, literature study suggests to measure individual HRTFs experimentally for each subject (Begault, *et al.*,2001). Individualized HRTFs are based on measurements considering each persons unique physical head properties. Measurement of HRTFs require special equipments, facilities and also expertise which are very difficult to achieve (Hu, *et al.*,2008).

Also, literature study (Genuit, 1984, Shaw, 1997) suggests that exact HRTFs are complicated to achieve rather their general behavior can be estimated from simple geometric models of torso, head and pinna. For understanding and estimating HRTFs as suggested by (Genuit, 1984) and reported by (Algazi, *et al.*, 2001), a set of 27 anthropometric measurements 17 for the head and torso **(Fig. 6)** are shown. Further, like finger prints, human pinna are not identical and vary quite widely in shape and size and consequently HRTFs also varies which makes it difficult to generalize the spectral characteristics across large numbers of individuals.

On the other hand, generic HRTFs are mathematical combination of multiple, individualized HRTFs. For speech content, it does not matter whether individualized or generic HRTF are used, also, if an HRTF is omitted, the externalization would be weak; the sound would come from inside-the head (Begault, *et al.*, 2001).

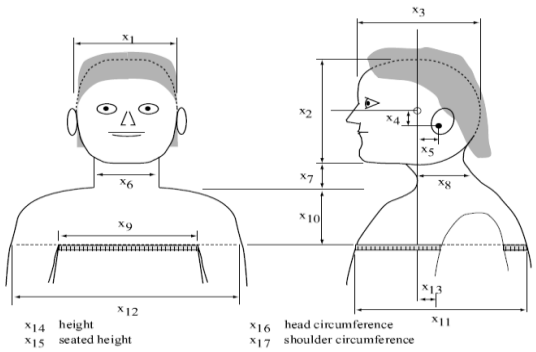


Fig. 6: Head, Torso and Pinna Measurements (Adopted from (Algazi, *et al.*, 2001))

As compare to individualized HRTFs, non individualized HRTFs in literature (Fisher, *et al.*, 1968, Weinrich, 1982, Bronkhorst, 1995) Møller, *et al.*, 1996) have been cited as degrading localization accuracy, decreasing externalization and increasing reversal errors. However, mentioned research reports are based on full-spectrum noise stimuli. Additionally, in (Bronkhorst, 1995) found no significant effect of using individualized HRTFs on reversals, but on the other hand, research results reported by (Wenzel, *et al.*, 1993) indicated that individualized HRTFs mitigated reversal confusions. According to (Møller, *et al.*, 1996) non individualized HRTFs resulted in an increased reversals, however, they reported no effect on externalization in their experimental results, since their experiments were based on speech stimuli.

6.2 Amplitude Panning

Amplitude panning method is typically applied for multi-loudspeaker setups. Under this arrangement the signal is applied to two or more loudspeakers at an equal distance from the listener with appropriate non-zero amplitude (Blumlein, 1931). In this method, listener perceives an impression of a virtual source at a location which actually does not coincide with any of the physical sound sources. In normal condition virtual sound source should be reproduce the way it is targeted, but in practice there are some imperfections as reported in (Pulkki, 2001).

However, through the Vector Base Amplitude Panning (VBAP) method multiple moving or stationary sounds can be positioned in any direction in the sound field spanned by the loudspeakers. VBAP is introduced by (Pulkki, *et al.*, 1996) and is utilized to position virtual sources in arbitrary 2-D or 3-D loudspeaker setups where the same sound signal is applied to a number of loudspeakers with appropriate non-zero amplitudes. VBAP can be generalized as a triplet-wise

panning method for 3-D loudspeaker setups (Pulkki, 1997). VBAP has certain advantages over other virtual source positioning methods. For example earlier methods either used all loudspeakers to produce virtual sources, which results in some artifacts, or they used loudspeaker triplets with a non-generalizable 2-D user interface (Pulkki, and Karjalainen, 2001).

7. CONCLUSION

In this article a mini review on the background of 3D audio is presented in general. Firstly, an introduction of the advantageous use of 3D audio for telephony and teleconferencing solutions has been highlighted. Secondly, this article covers a background of the terms and technologies used particularly in the area of human hearing, binaural hearing, and sound localization phenomena, binaural technology and 3D sound including Ambisonics, wave-field-Synthesis concepts have been covered. Also, 3D sound recording and reproduction have been presented. A detailed introduction and background of HRTFs have also been presented covering individualized and non-individualized HRTFs. Additionally; amplitude panning methods have been described in this article. It is important to mention here that the current article is the first in the series of two articles to cover foundational knowledge which is important to obtain in order to incorporate 3D audio into telephony and teleconferencing systems.

8 ACKNOWLEDGEMENT

Authors of this article are very much thankful to authorities of Sindh Agriculture University (SAU) Tandojam, Mehran University of Engineering and Technology (MUET) Jamshoro and Higher Education Commission Pakistan (HEC) for their help and support to achieve this research work.

REFERENCES:

Ahrens, J., M. Geier, A. Raake, and C. Schlegel, (2010) Listening and conversational quality of spatial audio conferencing, in Audio Engineering Society Conference: 40th International Conference: Spatial Audio: Sense the Sound of Space.

Algazi, V., R. Duda, D. Thompson, and C. Avendano, (2001) The cipic HRTF database, in Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop, 99-102.

Barrett, N. and S. Berge, (2010) A new method for b-format to binaural transcoding, in 40th AES International conference. Tokyo, Japan, 8-10, 2010.

Begault, D. R. (1994) 3-D sound for virtual reality and multimedia. San Diego, CA, USA: Academic Press Professional, Inc.

Begault, D. R., E. M. Wenzel, and M. R. Anderson, (2001) Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source, JAES, vol. **(49):** 904-916.

Berkhout, A., (1988) A holographic approach to acoustic control, Journal of the Audio Engineering Society, vol. **(36):** no. 12, 977-995.

Berkhout, A., D. De Vries, and P. Vogel, (1993) Acoustic control by wave field synthesis, Journal of Acoustical Society of America, vol. **(93):** 2764-2764.

Blauert, J. (1997) Spatial Hearing: The Psychophysics of Human Sound Localization. MIT Press, 1997.

Blumlein, A., (1931) British patent specification 394325, Reprinted in John Eargle (ed.) Stereophonic Techniques, 32-40.

Boone, M., E. Verheijen, and P. Van Tol, (1995) Spatial sound field reproduction by wave-field synthesis, Journal of the Audio Engineering Society, vol. **(43):** no. 12, 1003-1012.

Brandenburg, K., S. Brix, and T. Sporer, (2004) Wave field synthesis: From research to applications, in Proceedings of 12th European Signal Processing Conference (EUSIPCO), Vienna, Austria.

Bronkhorst, A., (1995) Localization of real and virtual sound sources, The Journal of the Acoustical Society of America, vol. **(98):** 2542Pp.

Brungart, D.B., Simpson,C. Bundesen, S. Kyllingsbaek, A. Burton, and A. Megreya, (2007) Cocktail party listening in a dynamic multi-talker environment, Perception and Psychophysics, vol. **(69):** no. 1, 79Pp.

Burkhard , M. and R. Sachs, (1975) Anthropometric manikin for acoustic research, The Journal of the Acoustical Society of America, vol. **(58):** 214Pp.

Cheng, C. and G. Wakefield, (2001) Introduction to Head-Related Transfer Functions (HRTFs): Representations of HRTFs in Time, Frequency, and Space, J Audio Eng Soc, vol. **(49):** no. 4, 231Pp.

Cherry, E. (1953) Some experiments on the recognition of speech, with one and with two ears,_ Journal of the acoustical society of America, vol. **(25):** no. 5, 975-979.

Crispien, K. and T. Ehrenberg, (1995) Evaluation of the "Cocktail Party Effect" for Multiple Speech Stimuli

within a Spatial Auditory Display, Journal of the Audio Engineering Society, vol. **(43):** no. 11, 932-941.

De Vries, D., (1996) Sound reinforcement by wave field synthesis: Adaptation of the synthesis operator to the loudspeaker directivity characteristics, J. Audio Eng. Soc, vol. **(44):** no. 12, 1120-1131.

De Vries, D., M. Boone, (1999) Wave field synthesis and analysis using array technology, in Applications of Signal Processing to Audio and Acoustics, 1999 IEEE Workshop on, 15-18.

Ericson, M. and R. McKinley, (1997) Binaural and Spatial Hearing in Real and Virtual Environments, ch:. The intelligibility of multiple talkers separated spatially in noise, 701-724. Erlbaum, Mahwah, NJ, r. h. Gilkey and T.R. Anderson edition, 1997.

Fisher, H. and S. Freedman, (1968) The role of the pinna in auditory localization. Journal of Auditory Research **(4):** 32-40.

Gerzon, M., (1970b) Ultra-directional microphones: Applications of blumlein difference technique: Part 2, Studio Sound, vol. **(12):** 501-504.

Gerzon, M., (1973) Periphony: With-height sound reproduction, J. Audio Eng. Soc, vol. **(21):** no. 1, 2-10.

Gerzon, M., (1974) Gerzon, Surround-sound psychoacoustics, Wireless World, vol. **(80):** no. 1468, 483-486.

Gerzon, M., (1975) Ambisonics. Part Two: Studio Techniques, Studio Sound, vol. **(17):** no. 8, 24-26.

Hirahara, T., Y. Sawada, and D. Morikawa, (2011) Impact of dynamic binaural signals on three dimensional sound reproduction, Proc. INTERNOISE, 2011.

Hu, H., L. Chen, and Z. Y. Wu, (2008) The estimation of personalized HRTFs in individual VAS, in IEEE Fourth International Conference on Natural Computation, 203-207.

Hyder, M., M. Haun, , O. Weidmann, and C. Hoene, (2010) Assessing virtual teleconferencing rooms, in 129th Audio Engineering Society Convention, San Francisco CA, USA.

ITU-T, (2011) International telecommunication union, itu-t. http://www.itu.int/ITU-D/ict/newslog/ Category View, category, FixedNov

ITU-T, (2010) International telecommunication union, itu-t webpage. http://www.itu.int.

Kilgore, R., M. Chignell, and P. Smith, (2003) Spatialized Audioconferencing: what are the bene_ts?, in Proceedings of the 2003 conference of the Centre for Advanced Studies on Collaborative research, 144Pp, IBM Press.

Kim, H. J., D.G. Jee, M.H. Park, B.S. Yoon, and S.I. Choi, (2004) The Real time Implementation of 3D Sound System Using DSP, in IEEE 60th Vehicular Tech. Conference (VTC2004), vol. **(7):** 4798-4800.

Kim, J., S. Kim, Y. Kim, J. Lee, and S.I. Park, (2005) New HRTFs (Head Related Transfer Functions) for 3D Audio Applications, 118th AES Convention.

Kürer, R. G. Plenge, and H. Wilkens, (1969) Correct spatial sound perception rendered by a special 2-channel recording method, in Proceedings of the 37th AES Convention, New York, USA.

Lee, H., H. Kim, M. Lee, P. Kim, S. Suh, and K. Kim, (1998) Development of 3D sound generation system for multimedia application, in Proceedings 3rd IEEE conference on Computer Human Interaction, 120-123.

Linkwitz, lab., (2011) Blumlein Microphone Setup. URL: http://www.linkwitzlab.com/Recording/record-play- map.htm

Low, C. and L. Babarit, (1998) Distributed 3D Audio Rendering, Computer Networks and ISDN Systems, vol. **(30):** 407-415.

Malham, D., A. Myatt, (1995) 3D sound spatialization using ambisonic techniques, Computer Music Journal, vol. **(19):** no. 4, 58-70.

Middlebrooks, J. and D. Green, (1991) Sound localization by human listeners, Annual Review of Psychology, vol. **(42):** no. 1, 135-159.

Middlebrooks, J., (1999) Individual differences in external-ear transfer functions reduced by scaling in frequency, The Journal of the Acoustical Society of America, vol. **(106):** 1480-1494.

Moller, H., (1992) Fundamentals of binaural technology, Applied Acoustics, vol. **(36):** 171-218.

Møller, H., M. Sørensen, C. Jensen, and D. Hammershøi, (1996) Binaural technique: do we need individual recordings?, Journal of the Audio Engineering Society, vol. **(44):** no. 6, 451-469.

Park, M., S. C. Choi, S. Kim, and K. Bae, (2005) Improvement of front-back sound localization characteristics in headphone-based 3D sound generation, The 7th International Conference on Advanced Communication Technology (ICACT 2005), vol. **(1):** 273-276.

Pekonen, J., (2008) Microphone techniques for spatial sound, in Proceedings of the 2008 Acoustics Seminar on Spatial Sound Modeling, (Espoo, Finland), TKK Helsinki University of Technology, Department of Signal Processing and Acoustics.

Pickles, J.O, (1988) An Introduction to the Physiology of Hearing, Academic Press London, Vol. **(3):**

Pulkki, V., J. Huopaniemi, and T. Huotilainen, (1996) DSP tool for 8-channel audio mixing, in Proc. Nordic Acoustical Meeting, vol. **(96):** 307-314.

Pulkki, V. (1997) Virtual Sound Source Positioning using vector based Amplitude Panning, Journal of the Audio Engineering Society, vol. **(45):** no. 6, 456- 466.

Pulkki, V., and M. Karjalainen, (2001) Localization of amplitude-panned virtual sources I: Stereophonic panning, Journal of the Audio Engineering Society, vol. **(49):** no. 9, 739-752.

Shaw, E. (1982) External ear response and sound localization, Localization of sound: Theory and applications, 30-41, 1982.

Sherman, H., (1953) Binaural sound reproduction at home, Journal of the Audio Engineering Society, vol. **(1):** no. 1, 142-145.

Studiocare,lab., (2011) Dummy head, http://www.studiocare.com/store/index.php?main_page =product_info&products_id=63816 2011.

Strutt, J., (1907) On our perception of sound direction, Philosophical Magazine, vol. **(13):** 214-232.

Tsakostas, C., A. Floros, and Y. Deliyiannis, (2007) Binaural rendering for enhanced 3D audio perception, in Proceedings of the AudioMostly 2007, 2nd Conference on Interaction with Sound, 27-29.

Vilkamo, J., (2008) Spatial sound reproduction with frequency band processing of b-format audio signals. PhD thesis, Helsinki University of Technology.

Vorländer, M. (2008) Auralization: fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality. Springer Verlag (2008).

Webb, C.J. and Bilbao, S. (2011) Virtual Room Acoustics: A Comparison of techniques for computing 3D-FDTD schemes using CUDA, 130th Audio Engineering Society Convention. London, UK

Weinrich, S., (1982) The problem of front-back localization in binaural hearing., Scandinavian audiology. Supplementum, vol. **(15):** 135-138.

Wenzel, E, M. Arruda, D. Kistler, and F. Wightman, (1993) Localization using non-individualized head-related transfer functions, Journal-acoustical Society of America, vol. **(94):** 111-114.

Wightman, F. and D. Kistler, (1997) Monaural sound localization revisited, The Journal of the Acoustical Society of America, vol. **(101):** 1050-56.

Yankelovich, N., W. Walker, P. Roberts, M. Wessler, J. Kaplan, and J. Provino, (2004) Meeting central: making distributed meetings more e_ective, in Proceedings of the 2004 ACM conference on Computer supported coop- erative work, ACM, 419-428

Yankelovich, N., J. Kaplan, J. Provino, M.Wessler, and J. M. DiMicco, (2006) Improving audio conferencing: are two ears better than one?, in Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work (CSCW '06), ACM (New York, NY, USA), 333-342

Zhang, W., Abhayapala, T.D. Kennedy, R.A. and Duraiswami, R (2010), Insights into head-related transfer function: Spatial dimensionality and continuous representation, J. Acoustical. Society of America, Vol. **(127):** 4-8.