5-2018

A Survey of Three-Dimensional Sound and its Applications

Andrew Topping

California State University, Monterey Bay

The past decade of technological advancements has led to the pursuit of continuously more immersive media experiences, as exampled by the growth in interest in virtual reality as a viable medium for content dissemination. With this growth comes a demand for convincing three-dimensional audio capabilities available at the consumer level. Among other reasons, audio professionals and enthusiasts alike may find it useful to have a foundation of knowledge about 3-D audio given the upward projection of this trend (Gordon). This paper details what 3-D audio is, how 3-D audio is accomplished, the paradigm shifts of 3-D audio as both a creative and engineering consideration, and an examination of some of the existing applications of 3-D audio technology.

What is Three-Dimensional Sound?

Broadly, three-dimensional sound refers to audio that enables a source to be positioned at any point in a three-dimensional soundscape (Duda). More specifically, 3-D audio techniques seek to produce sound in such a way that is exactly representative of the sound profile that would ultimately reach the ears of the listener in a real life listening circumstance. The 3-D Audio and Applied Acoustics Lab at Princeton University provides the following description:

> [S]uppose the actual listening experience consisted of you hearing a firecracker explode in the sky; the sound leaving the exploding firecracker would change considerably before it stimulates your eardrums. Now if, somehow, the sound from a pair of loudspeakers located a few feet from you could stimulate your eardrums in exactly the same way as the sound from the firecracker did, you would perceive the explosion as originating from way up in the sky, and not from anywhere near the pair of loudspeakers. The idea, therefore, is

> that the brain would get "tricked" into believing that the sound isn't necessarily originating from the loudspeakers, but rather from any point in 3D space. (Choueiri)

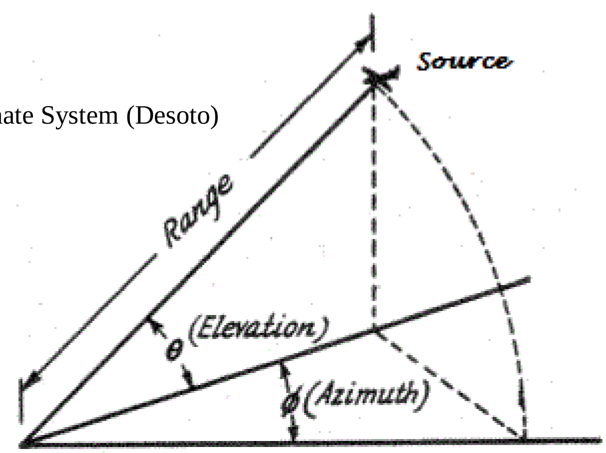
There are several distinctions to point out in such a definition. For one, there exists a salient distinction between a realistic soundscape and a faithful reproduction of a soundscape. As illustration, one can imagine a fantastic environment full of alien, otherworldly sounds, a soundscape which is by no means an encounterable depiction of reality. Regardless, there are intuitive differences between what it would be like to hear soundscape in person versus through a pair of cheap speakers. Additionally, 3-D audio transcends simple localization of sound in this definition. Beyond the basic horizontal positioning capabilities found in the standard stereo paradigm, a 3-D audio system is able to depict a realistic auditory event to an extent that the brain is convinced that a sound originates not from the loudspeakers or headphones, but from an arbitrary point in three-dimensional space.

Sound as a Physical and Perceptive Phenomena

The human auditory system is innately able to spatialize and localize sound, an ability dependent on a set of factors derived from the physics of sound and psychoacoustics (Potisk). Evolutionary forces have designed the ear to react to pressure variations in a particular medium, which are then converted to electrical impulses interpreted by the brain as sound. These variations have measurable amplitude, frequency, wavelength and speed; the values of these features affect a listener's perception of a sound. Further, sound behaves as waves that can be reflected, diffracted, absorbed and constructively or destructively interfere with each other, behavior conditional in understanding the psychoacoustic responses to sound ("Reflection, Refraction and Diffraction").

Psychoacoustics is the combined study of acoustics and auditory physiology that determines the relationship between the characteristics of a sound and the auditory sensation it provokes. The field ultimately provides an explanation for the human capability to localize sound in three dimensions (Lorenzi). The ears, head, torso and shoulders all play a role in the final depiction of a sound that reaches the brain, and as such, a role in sound localization (Duda). Sound localization is generally described in spherical coordinate systems, where the standard coordinates are azimuth, elevation and range. More specifically, the vertical-polar coordinate system is the most popular definition for these coordinates as depicted below (Duda).

Vertical-Polar Coordinate System (Desoto)



Azimuthal cues, or cues that indicate the position of a sound in horizontal space, are made possible because the brain is fed two signals of auditory information through the left and right ears. The Duplex Theory of Sound Localization proposes that discrepancies present in the two signals provide the brain the necessary indications to decipher the direction a sound is coming from (Kendall). These discrepancies include:

- *Interaural Time Difference* (ITD) – the difference in the time it takes for a sound to reach the left and right ears. Sound travels at roughly 343 m/s at sea level static atmospheric conditions ("Speed of Sound"). Even at this speed, there is a perceptible delay in the time it takes for sound to reach one ear to the other (Duda).

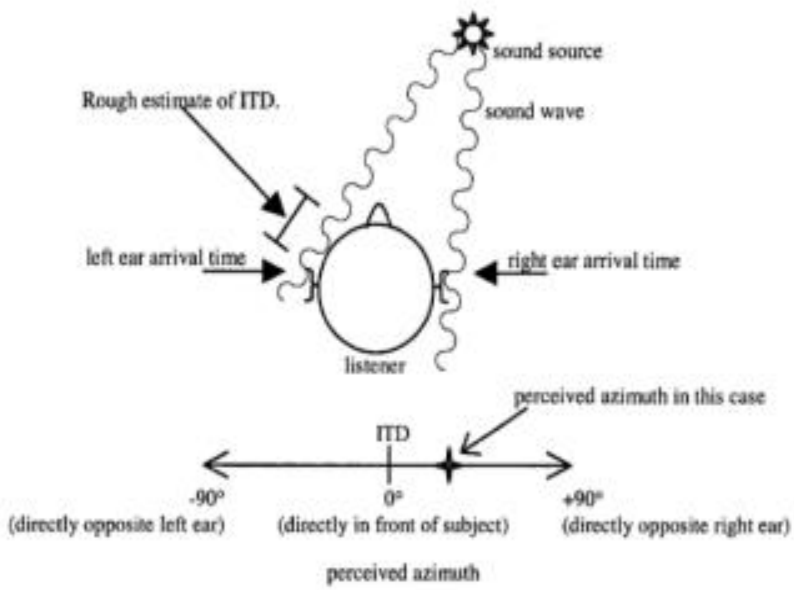
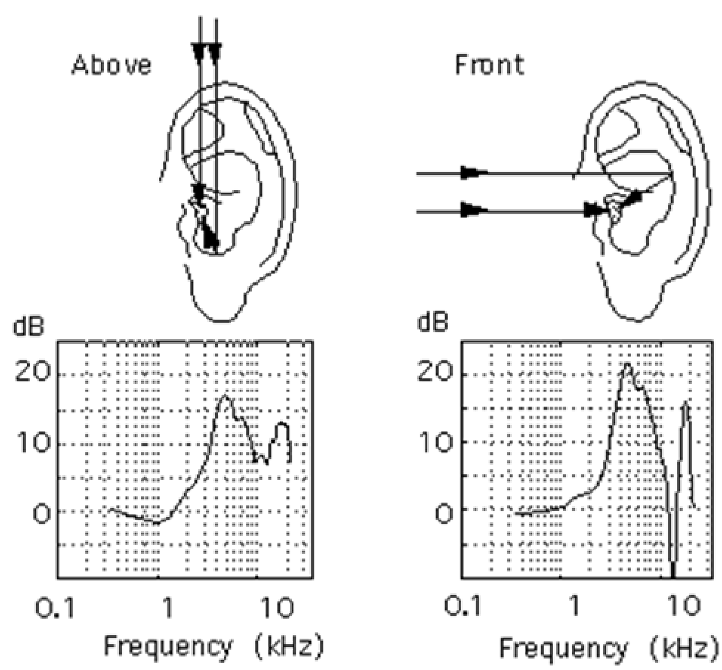


Illustration of ITD (Potisk)

- *Interaural Phase Difference* (IPD) – the difference in the phase of a sound between the left and right ears. Human beings are able to detect phase differences as small as three degrees ("Interaural Phase Difference (IPD)"). IPD is less effective as a spatial cue for frequencies corresponding to wavelengths less than two head diameters, as longer wavelengths will bring a smaller phase change (Elert).
- *Interaural Level Difference* (ILD) – Also referred to as the Interaural Intensity Difference, ILD is the difference in the amplitude of a sound present between the left and right ears. In the stereo paradigm of a left and a right channel, ILD is the basis for panning. Human beings are able to perceive a 1 dB difference between ears ("Interaural

Intensity Difference (IID)"). An important aspect of ILD is head-shadowing, an effect that renders ILD effective only in higher frequencies. Sounds with wavelengths larger than the diameter of the human head will diffract around the head, and the intensity difference will be negligible. However, higher frequencies with shorter wavelengths are not able to diffract around the head. Instead, the head attenuates the amplitude of a sound as it travels between the ears (Elert).

Where the cues for azimuth are created through the discrepancies between the signals fed to the brain each ear, elevation cues are monaural, derived from the spectral filtering provided by the ear's pinna. The pinna consists of several resonant cavities that induce reflections, causing interference patterns that amplify certain frequencies and attenuates others. As shown in the diagram below provided by UC Davis, the specific frequency response is directionally dependent and thus responsible for elevation cues. It is important to note that the pinna does not cause significant spectral filtering until higher in the frequency spectrum due to negligible differences in the wavelengths of the reflected signal and the direct signal (Duda).



Directionally Dependent Spectral Filtering of the Pinna (Duda)

Range cues, the perceptive cues that provide information about how far a source is from the listener, are the least understood localization indicators (Duda). In general, human beings do not accurately perceive the distance of a sound through hearing alone but in conjunction with other sensory input (Potisk). Auditory cues of range include:

- *Loudness* – The inverse square law states that the intensity of a sound decreases inversely with the square of distance. Accordingly, louder sounds are generally perceived as closer to the listener and quieter sounds further away (Duda).
- *Motion Parallax* – a consequence of azimuthal cues. Turning one's head necessarily produces a shift in azimuthal discrepancies. The further a source is from a listener, the smaller such azimuthal shifts will be. Interaural level differences are particularly of note as a sound gets extremely close to the head (Duda).
- *Ratio of Direct to Reverberant Sound* – related to the effects of environmental acoustics. The image of a sound rarely consists solely of the direct signal in isolation. Rather, it is a conglomerate of direct and indirect signals, a composition of the reverberant qualities of the environment, or lack thereof (Savioja). In most environments, the level of reverberated sound does not change much with distance, but the level of direct sound will. As such, the ratio of direct to reverberant sound will be higher the closer a source is to a listener and lower for sources further away (Duda).

Additionally, localization cues are weighted differently from environment to environment due the effects of reverberation and echoes. One such effect is that interaural timing differences are essentially useless for localizing low-frequency information in highly reverberant environments as the wavelength is too long for the auditory system to process any meaningful information from the signal before several environmental reflections make contact with the ear. (Duda). In this situation, the important timing information comes from the Interaural Envelope Difference (IED) as demonstrated by the Fransen Effect (Duda). Richard O. Duda, a professor in the Department of Electrical Engineering and San Jose State, describes it:

> If a sine wave is suddenly turned on and a high-pass-filtered version is sent to Loudspeaker A while a low-pass filtered version is sent to Loudspeaker B, most listeners will localize the sound at Loudspeaker A. This is true even if the frequency of the sine wave is sufficiently low that in steady state most of the energy is coming from Loudspeaker B. Basically, the starting transient provides unambiguous localization information, while the steady-state signal is very difficult to localize, and in this circumstance the auditory system simply ignores the ambiguous information. With some risk of oversimplification, we can generalize and say that in reverberant environments it is the high-frequency energy, not the low-frequency energy, that is important for localization. (Duda)

All cues discussed depend on the idiosyncrasies of anatomy, particularly of the ears. As no two people have heads or ears of exactly the same size or shape, no two people will have the exact same perception of a sound coming from an equivalent position in three dimensional space. The conclusion to be made is that localization of sound is the result of several spatial cues, the extremes of which are personal to the unique morphology of individuals. Any spatial audio system, including traditional stereo, will take advantage of basic cues to provide spatial information. However, systems that create a true impression of three dimensional audio address with greater precision the depths of psychoacoustic phenomena to localize sound, and the most sophisticated account for minute details in anatomical variabilities between individuals for more convincing emulation and powerful immersive capabilities.

3-D Sound Techniques

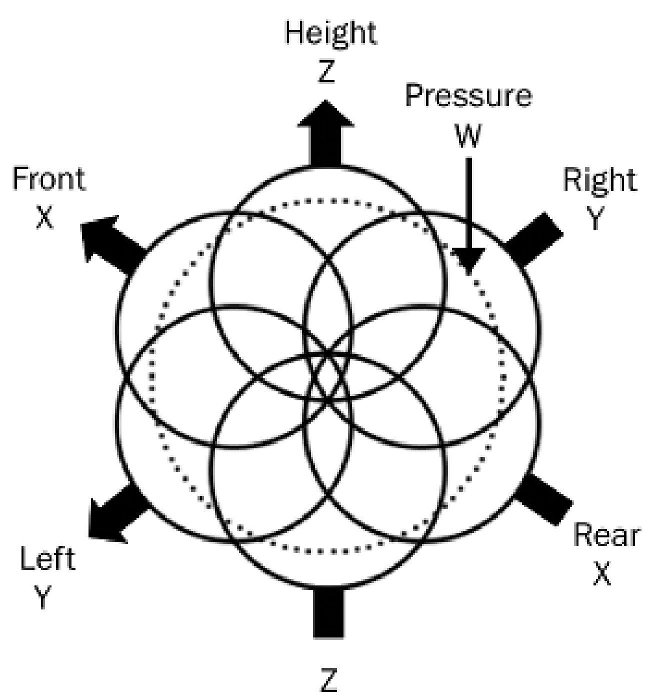
Having established the parameters through which sound is perceived in three dimensions, the following section is devoted to discussing the general defining properties of 3-D audio systems as well as some of the more prominent techniques employed by such systems to satisfy these parameters. Each technique has related advantages and disadvantages that stem from their specific treatment of the barriers gating authentic three dimensional spatialization.

3-D audio systems accomplish their namesake by one of two approaches, both of which share the common objective of producing sound at the eardrums that is ultimately identical to the sound heard in the real auditory conditions of the emulated soundscape. The first approach focuses on recreating the sound field of a given spatial region, thereby producing a *phantom image* of said sound field. Following the propagation of this phantom image, psychoacoustic reactions naturally occur, and the sound field is perceived in three dimensions, removing the necessity to account for anatomical individuality (Malham and Myatt). The second approach encodes the morphological interactions of the body in the output sound and compensates for the interactions subsequent to its propagation. Human localization cues are calculated as characteristics of the audio prior to its output, and therefore individual morphology must be factored in for accurate perception of sound in three dimensions.

Ambisonics

Developed in the early 1970s in large part due to the work of the audio engineer Michael Gerzon (Moreau), ambisonics denotes a method of capturing and reproducing full 360-degree representations of sound fields using specialized microphones made up of a tetrahedral array of four sub-cardioid microphones that capture sound from every direction (Virotek). There are several different ambisonics formats, but the most basic and widely implemented is the first-order B-format, which uses as few as four channels (w, x, y and z) to faithfully reproduce a captured sound field ("Ambisonics Explained: A Guide for Sound Engineers"). In the first-order B-format, the four channels can represented as follows:

- **W** is an omni-directional polar pattern that records sound content in all directions and equal gain
- **X, Y and Z** are figure-8 polar patterns pointing along the axes of the three-dimensional Cartesian coordinate system



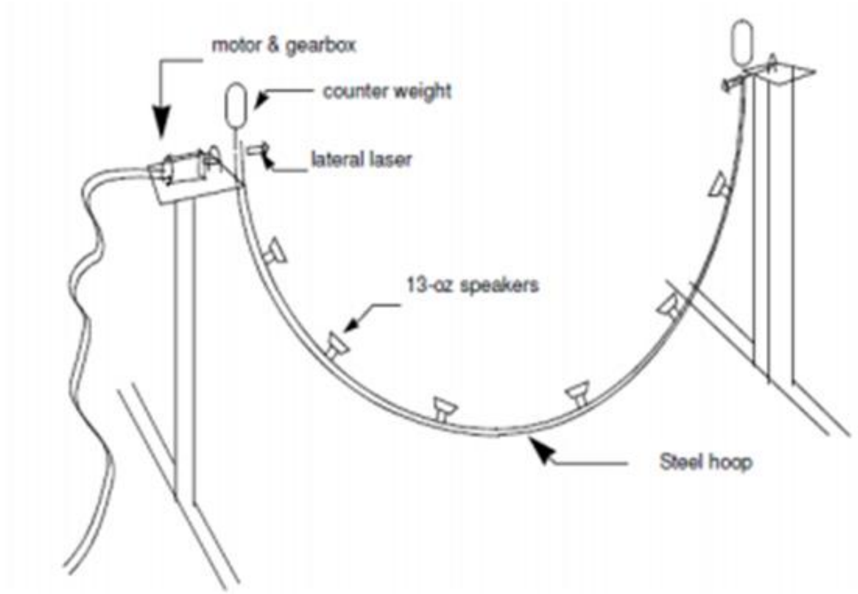
A Depiction of First Order Ambisonics (Yeary)

The most significant advantage of ambisonics is that the recorded information can be decoded to any speaker layout. In other words, ambisonics is not restricted by the limitations of specific playback systems and can impress a sound field in 3-dimensions without a designation of a specific number of playback channels. One other advantage is that ambisonics creates smooth auditory transitions when a sound field rotates. In contrast, sounds can "jump" from speaker to speaker in other formats ("Ambisonics Explained: A Guide for Sound Engineers"). However, ambisonics is not without drawbacks. Ambisonic microphones are prohibitively expensive, as is a multi-speaker spherical array for ambisonics playback. Additionally, ambisonics cannot be accomplished on headphones without additional binaural processing, the type of which is covered in the next section.

Head Related Transfer Functions

Another widely used 3-D audio protocol, *Head Related Transfer Functions* (HRTF), falls into the latter category. A complex function that captures the transformations of sound wave propagation from the source to the ears, the HRTF is concerned with modeling the physiological responses. These transformations account for the diffractive and reflective influences of the head, pinnae, shoulders and torso, and the combination of two unique HRTFs for the left and right ears satisfactorily addresses the aforementioned parameters for localization.

The HRTF is the Fourier transform of the Head Related Impulse Response (HRIR), a recorded impulse response from the source of a sound to the ear. HRIRs are recorded by placing microphones in the ears that record signals excited from all directions in which the HRTF is to be measured (Potisk). To accomplish this, the subjects head is placed in the center of a semicircular assortment of speakers that rotate along the interaural axis (Potisk). These recordings usually take place in an anechoic room to negate any acoustic effects on the impulse response (Potisk). In place of an actual human head, a replica of comparable shape, size and density is often used, which eliminates any potential positioning errors (Potisk). The signals employed vary from sine sweeps to pseudo-random noise generators (Potisk).



A Diagram of An HRTF Recording Rig (Potisk)

The recordings that make up the HRIR are representative of the specific temporal, spectral and volumetric responses of the ear to a source at a given point in space, which can then be extrapolated to the HRTF through Fourier transformation. Given an HRTF for each ear, accurate spatial localization can be synthesized from monaural sources. However, true accuracy would require recordings from an infinite number of points in three-dimensions, and therefore interpolation between HRIRs is needed (Duraiswami).

Other Protocols

It is worth mentioning that other sophisticated consumer-oriented protocols exist, largely as proprietary developments. Such protocols include Dolby Atmos, an expansion of cinematic surround sound that, similar to the aims of ambisonics, treats sounds in a mix as individual entities detached from channel designations ("Freeing Sound From Channels – The Dolby Atmos Concept"). Additionally, Atmos enables sounds to be placed above the listener, introducing elevation effects and with it 360-degree audio depictions (Freeing Sound From Channels – The Dolby Atmos Concept). Atmos is discussed at further length later.

Known Issues in 3-D Audio

Beyond the handful specific to the previously mentioned techniques, there are several additional issues impeding accurate 3-D audio as a result of both biological factors and technical limitations. Some of the most pertinent examples can be illustrated in comparisons between headphones and loudspeakers. Headphones have the advantage of two discrete channels of information sent in isolation to one ear apart from the other. This eliminates any issues associated with cross-talk, or the presence of one signal in both ears (Miller).

On the other hand, the isolation of headphones can be un-immersive, as it prevents external sound interactions. Additionally, headphones are susceptible to notchy frequency responses that resemble the responses of ear pinna. If not compensated for, this severely compromises elevation effects (Duda). To complicate matters further, sounds heard over headphones often seem too close, as the physical source is genuinely close to the ear. Compensating for this effect requires precise headphone positioning (Duda). A binaural room impulse response (BRIR) can recreate the spatial perception of a room, helping to mitigate in-head localization (Potisk). In turn, loudspeakers do produce crosstalk but eschew the other problems of headphones. However, such crosstalk can be minimized to achieve playback analogous to a pair of virtual headphones. A loudspeaker system can incorporate as many channels as possible with each additional channel adding more spatial information and increased localization accuracy, negating the damaging effect of crosstalk. Additionally, 3-D audio can be effectively accomplished on a stereo pair of monitors through crosstalk-cancellation, or transaural stereo (Kaiser). Audio engineer and film producer Robin Miller describes how transaural stereo is accomplished:

> In any XTC method, the signal of one channel is delayed, attenuated, inverted, and mixed with the other speaker's signal so that it arrives at its ear in time for the crosstalk from the first speaker. For one or two listeners on the median line bisecting the speakers, this XTC signal cancels the crosstalk acoustically, leaving each ear to hear only the speaker on the same side, i.e. with no crosstalk. (It is the equivalent of acoustically isolating each speaker to its ear, such as with a barrier.) (Miller)

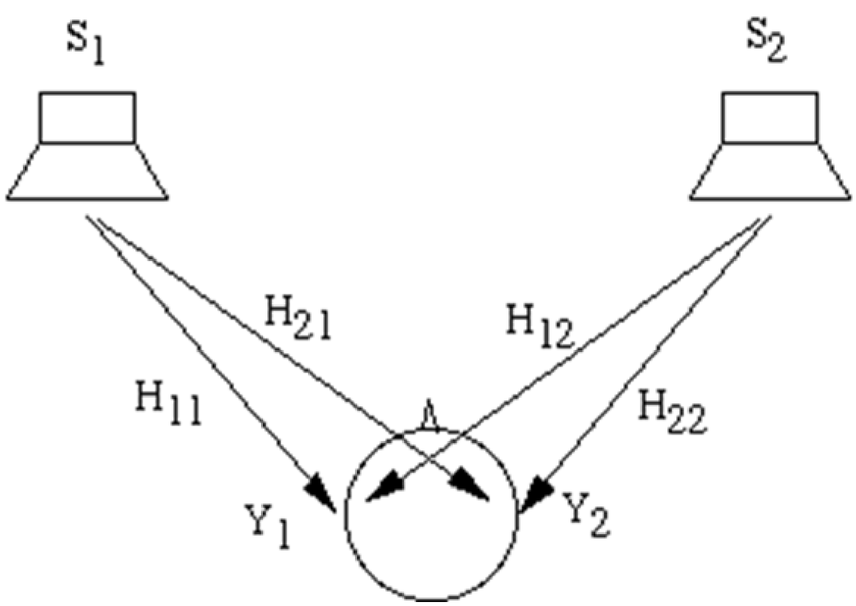
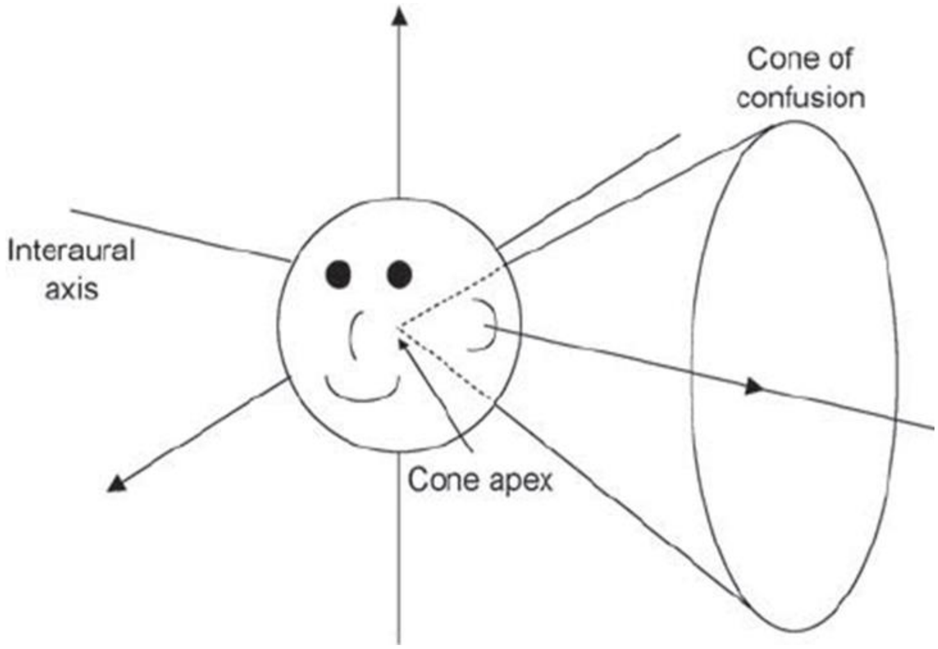


Illustration of Stereo Crosstalk (Duda)

Other examples of the issues facing 3-D audio reproduction are:

- *Anatomical Uniqueness* – As previously discussed, human anatomy varies widely with respect to ear, head, torso and shoulder size and shape, and with it, the psychoacoustical role in sound localization. Although azimuth effects are not as susceptible to accurate physiological modeling, elevation cues are particularly sensitive to individual differences. An HRTF made with an individual's anatomical specification will be most accurate for that person, but less so for others depending on their anatomical similarity to the individual the HRTF was recorded for. This poses scalability issues as an individualized HRTF is inconvenient and time-consuming to measure (Duda). There are two alternative approaches: the use of population standardized HRTFs, which will compromise localization to various degrees for many listeners, or a model HRTF with adaptable parameters (Duda). Rahulram Sridhar and Edgar Choueiri of Princeton University propose a method that synthesizes both approaches in which a listener's head scan points correspond to the best fit from a database of HRTFs. The points are able to be obtained from consumer-grade cameras.
- *Cone of Confusion* – The cone of confusion describes a cone centered along the interaural axis shown in the figure below. Along the surface of the cone, source locations are indistinguishable from each other because they induce the same ITD values (Duda).

The Cone of Confusion, or Cone of Constant Azimuth (Kapralos)



- *Sensatory Interdependence of Localization* – Similar to the issues posed by the cone of confusion, the azimuthal cues for sounds directly ahead or behind a listener will be the same. More specifically, sources located in these positions relative to the listener will not introduce significant ITD and IID cues. In the absence of such cues, the brain will incorporate visual information in interpreting a sounds spatial location. If there is no visual indicator of where a sound is coming from, the auditory system defaults to locating the sound behind the listener (Duda). This naturally makes it difficult to differentiate front-back placement in a purely auditory environment.

Headtracking

The ability of a 3-D audio system to account for head movements is valuable for several reasons. Not only does it introduce an extra immersive element to the experience, but it also represents a solution to many of the aforementioned localization issues. For example, the ambiguity of placement along the cone of confusion disappears as soon as the listener moves his or her head and with it the relative position of the sound source. In a similar vein, head movement will reposition sounds in front of a listener in the opposite direction of sounds placed behind them.

Unfortunately, headtracking is also not without its own set of problems. Systems that incorporate head tracking carry increased concerns for cost, reliability and accuracy (Duda). Two other concerns are: allowable latency, because lag is noticeable down to 50ms of latency (Deber), and the necessity of crossfading between the impulse responses of given spatial locations as the head moves to avoid unwanted clicks and transients between states (Duda).

Creative and Technical Applications of Spatial Audio

There are countless possibilities enabled through sophisticated three-dimensional auditory models, and several have come to fruition. The following section is a survey of the types of applications that apply 3-D audio technology, though by no means an exhaustive list.

Virtual reality makes up a significant portion of content utilizing three-dimensional audio reproduction. 3-D audio is an influential component of a virtual reality experience, as the more convincing the perception of three dimensions audio is, the more immersive the overall experience will be. Such experiences span a wide range of arenas, include gaming, cinema, sporting events and more. A particularly exciting example that places special emphasis on accurate auditory environment emulation is TheWaveVR, a platform enabling users to view, host and socialize in virtual music concerts online (*TheWaveVR*). Performers are able to change venues in real time, thereby changing the audience's acoustical perception of the music with it. Several prominent artists and projects have staged events through TheWaveVR, including OWSLA producer Kill The Noise and Warner Bros. *Ready Player One* ("Past Waves").

Although there is an argument to be made that a majority of the content produced using three-dimensional audio is a result of the surge in popularity of VR following the release of consumer VR headsets, an increasing amount of attention is directed towards the audio concerns of augmented reality, the usefulness of which is widespread across several industries such as medicine, engineering, architecture, leisure, military and business ("What is Virtual Reality?"). Microsoft Soundscape is an app that uses 3-D audio to help visually impaired people navigate cities. Using Soundscape, listeners are provided auditory signals that indicate progress towards a destination or points of interest in 360 degrees around them. In this case, three-dimensional localization of sound is being applied to foster greater independence and mobility to those with limited vision ("Soundscape Features").

Another possible category of 3-D audio applications is tools that solve existing issues in other audio-related fields. An example of this is Nx Virtual Mix Room (Nx VMR), a plug-in modeling virtual studio environments, by Waves Audio. Acoustic treatment is an expensive endeavor not afforded to many mix engineers, who default to headphones over an unpredictable room as a cheaper monitoring medium. Monitoring on headphones is less than ideal for a number of reasons including a faulty image of the stereo field and the lack of acoustical interactions with the environment, leaving the engineer unable to make precise decisions involving the reverberant or spatial identity of a mix. Nx VMR solves these problems by recreating the interaural effects of loudspeaker monitoring and the acoustics of a high-end studio over headphones. Critically, it also support head tracking, emulating the exact reactions of head movements in the sweet spot of a professional control room. Other features of Nx VMR include emulation of popular surround sound formats, allowing users to mix for 7.1, 5.1 or 5.0 surround over stereo headphones, and EQ calibration for specific headphone models, eliciting a flatter response ("Nx – Virtual Mix Room over Headphones").

One last consideration is the application of three-dimensional audio as a creative tool. Visual artists, musicians and content creators of all types have employed 3-D audio to creative ends in ways not possible through traditional stereo, and the modern accessibility of 3-D audio has enabled this content to reach wide audiences. This accessibility explains the cult fandom of ASMR videos, in which creators make certain sounds in efforts to produce ASMR, or an autonomous sensory meridian response (Barratt). The top creators are acutely aware of spatial phenomena and often employ binaural microphones in their videos as an aid in producing this response (Lalwani).

In the music industry, several mainstream ventures into 3-D audio have been put forth. One such endeavor is the special 25th anniversary edition of 1992 album *Automatic for the People* by R.E.M, an album hyped to be the first major commercial release in Dolby Atmos (Machkovech). The original producers of the album, Scott Litt and Clif Norrell, were tapped to do the remix for Atmos, and their thoughts on the process provide insights to the specific nature of mixing in 3-D and its departures from mixing in traditional stereo (Machkovech). A readily apparent advantage of such a configuration like Atmos is that there is more space in the sound field than in left-right stereo allowing for less competition among the elements of a mix for spatial real estate. As described by Norrell: "Instruments weren't fighting as much… Before, I was carving out frequencies. Now, I could do a lot less EQing and let instruments have a more full-frequency sound. We could have a more full-bodied approach… especially in the low end where you have, say a cello and a fuzz bass occupying the same sonic space" (Machkovech). Mixing in three dimensions gives an additional two axes of placement beyond the single horizontal axis of stereo. One could easily imagine a mix that takes extreme advantage of this increase in space through such things like radical automation of the position of an element in the mix through its three-dimensional space, like what can be heard on the Dolby Atmos remix of The Beatles' *Sgt. Pepper's Lonely Hearts Club Band* (Machkovech). However, the production team for *Automatic for the People* adopted a more subdued approach: "let's just make the album sound bigger" (Machkovech). While these insights are specific to Dolby Atmos, there are possible extrapolations to other formats that make mixing in 3-D an attractive platform for a project that values maintaining the integrity of the components in a mix without them unnecessarily competing with one another.

Conclusion

In summary, this paper addresses various facets of three-dimensional audio, covering the physical and biological perception of sound in three dimension, establishing a need for protocols to induce these perceptions and detailing some of the creative and technological applications of these protocols. It is the intention of the author that the information herein acts as a primer to 3-D audio and can serve as the basis for further delving into the topics presented.

Works Cited

"Ambisonics Explained: A Guide for Sound Engineers." *Waves.com*, Waves Audio, Oct. 10, 2017 www.waves.com/ambisonics-explained-guide-for-sound-engineers

Barratt, Emma L., et al. "Autonomous Sensory Meridian Response (ASMR): a Flow-like Mental State." *PeerJ*, PubMed, March 26, 2015. peerj.com/articles/851

Choueiri, Edgar. "Individualization of 3D Sound." *3-D Audio and Applied Acoustics Lab*, Princeton University, www.princeton.edu/3D3A/HRTFModeling.html

Deber, Johnathan. "How Much Faster Is Fast Enough? User Perception of Latency & Latency Improvements in Direct and Indirect Touch." *Understanding & Extending Touch Surfaces*. Tactual Labs. http://www.tactuallabs.com/papers/howMuchFasterIsFastEnoughCHI15.pdf

DeSoto, Clinton B. "Radar Techniques - Primer Principles." *QST*, April 1945

Duda, Richard O. "3-D Audio for HCI." *UC Davis Electrical and Computer Engineering*, University of California, Davis, www.ece.ucdavis.edu/cipic/spatial-sound/tutorial/

Duraiswami, Ramani. "Introduction to HRTFs." *University of Maryland*, legacydirs.umiacs.umd.edu/~ramani/cmsc828d_audio/HRTF_INTRO.pdf

Elert, Glenn. "The Nature of Sound" *The Physics Hypertextbook*, physics.info/sound

"Freeing Sound From Channels - The Dolby Atmos Concept." *Dolby Laboratories, Inc*., www.dolby.com/us/en/technologies/cinema/dolby-atmos.html

Gordon, Kyle. "Virtual Reality (VR) - Statistics & Facts." *Statista*, www.statista.com/topics/2532/virtual-reality-vr/

"Interaural Intensity Difference (IID)." *InSync*, Sweetwater, June 27, 2004 www.sweetwater.com/insync/interaural-intensity-difference-iid-2

"Interaural Phase Differences (IPD)." *InSync*, Sweetwater, June 29, 2005 www.sweetwater.com/insync/interaural-phase-differences-ipd.

Kaiser, Fabio. "Transaural Audio - The Reproduction of Binaural Signals over Loudspeakers." *Toningenieur-Projekt*, IRCAM, March 2011 iem.kug.ac.at/fileadmin/media/iem/projects/2009/kaiser.pdf

Kapralos, Bill, et al. "Virtual Audio Systems." *Presence Teleoperators & Virtual Environments*

Kendall, Gary S. "A 3-D Sound Primer: Directional Hearing and Stereo Reproduction." *Computer Music Journal*, vol. 19, no. 4, 1 Dec. 1995, pp. 23–46. JSTOR, www.jstor.org/stable/3680989

Lalwani, Mona. "Surrounded by Sound: How 3D Audio Hacks Your Brain." The Verge, Vox Media, 12 Feb. 2015, www.theverge.com/2015/2/12/8021733/3d-audio-3dio-binaural-immersive-vr-sound-times-square-new-york.

Lorenzi, Antoine. "Psychoacoustics." *Journey into the World of Hearing*, December 27, 2016 www.cochlea.eu/en/sound/psychoacoustics

Malham, David G., and Anthony Myatt. "3-D Sound Spatialization Using Ambisonic Techniques." *Computer Music Journal*, vol. 19, no. 4, 1995 www.jstor.org/stable/3680991

Miller, Robert E. "Key to Ambiophonic Crosstalk Cancellation." *Filmaker Technology*, www.filmaker.com/papers/RM-sto_ITDspkr%20is%20Key.pdf

Moreau, Sébastien, et al. "3D Sound Field Recording with Higher Order Ambisonics – Objective Measurements and Validation of a 4th Order Spherical Microphone." *120th Convention of the Audio Engineering Society*, 20 May 2006 e“Nx – Virtual Mix Room over Headphones.” *Waves.com*, Waves Audio, www.waves.com/plugins/nx#introducing-nx-virtual-mix-room

“Past Waves” *TheWaveVR,* thewavevr.com/waves

Potisk, Tilen. “Head-Related Transfer Function.” Dissertation, University of Ljubljana, January 9, 2015

“Reflection, Refraction, and Diffraction.” *The Physics Classroom,* www.physicsclassroom.com/class/sound/Lesson-3/Reflection,-Refraction,-and-Diffraction

Savioja, Lauri. “Room Acoustics Modeling.” *Interactive Acoustics,* interactiveacoustics.info

“Soundscape Features.” *Microsoft Store*, Microsoft, www.microsoft.com/en-us/research/product/soundscape/features/

“Speed of Sound.” Edited by Nancy Hall, *Glenn Research Center*, NASA, March 23, 2018 www.grc.nasa.gov/www/k-12/airplane/sound.html.

Sridhar, R., and E. Y. Choueiri. “A Method for Efficiently Calculating Head-Related Transfer Functions Directly from Head Scan Point Clouds” *143rd Convention of the Audio Engineering Society*, October 8, 2017

TheWaveVR, thewavevr.com

Virotek, Paul. “An Introduction to Ambisonics.” *Creative Field Recording*, March 1, 2017 www.creativefieldrecording.com/2017/03/01/explorers-of-ambisonics-introduction

“What Is Virtual Reality?” *Virtual Reality Society,* www.vrs.org.uk/virtual-reality/what-is-virtual-reality.html

Yeary, Jay. "Ambisonics B-Format for Immersive Audio." *TvTechnology*, December 27, 2015 www.tvtechnology.com/expertise/ambisonics-bformat-for-immersive-audio