

3D Audio Rendering and Evaluation Guidelines

version 1.0

**Prepared by the
3D Working Group of the
Interactive Audio Special Interest Group**

June 9, 1998

revision 1.0

Published By:
MIDI Manufacturers Association
Los Angeles CA

Interactive Audio Special Interest Group
3D Audio Rendering and Evaluation Guidelines — Level One

Acknowledgments:

Content for the primer and evaluation guidelines chapters was contributed by Aureal Semiconductor, Harman International, QSound, Rockwell Semiconductor and Spatializer.

Disclaimer:

The MMA, IASIG, and all members shall not be held liable to any person or entity for any reason related to the adoption or implementation of, nor adherence to the recommendations in, nor any other use of this document nor any accompanying software.

Copyright © 1997-1998 MIDI Manufacturers Association Incorporated

Portions of the Primer are Copyright © 1997 Aureal Semiconductor, used with permission.

ALL RIGHTS RESERVED. NO PART OF THIS DOCUMENT MAY BE REPRODUCED IN ANY FORM OR BY ANY MEANS, ELECTRONIC OR MECHANICAL, INCLUDING INFORMATION STORAGE AND RETRIEVAL SYSTEMS, WITHOUT PERMISSION IN WRITING FROM THE MIDI MANUFACTURERS ASSOCIATION INCORPORATED.

PRINTED 1998

MMA

POB 3173

La Habra CA 90632-3173

Table of Contents

INTRODUCTION	1
ABSTRACT	1
3D AUDIO PRIMER	2
INTRODUCTION TO 3D AUDIO	2
REAL 3D AUDIO VS. NOT-SO-REAL 3D AUDIO	2
TABLE 1: SUMMARY COMPARISON OF AUDIO PLAYBACK METHODS	6
THE BASICS OF ACOUSTICS	7
THE BASICS OF HUMAN HEARING	8
3D AUDIO REPRODUCTION	10
HEAD MOVEMENT	12
AUDIO-VISUAL SYNERGY	12
SUMMARY	12
GLOSSARY	13
3D AUDIO EVALUATION GUIDELINES	16
GENERAL CONSIDERATIONS	16
CONSIDERATIONS FOR CONTROLLED EXPERIMENTATION	18
SPECIFIC EVALUATION CRITERIA	19
TOTAL EXPERIENCE:	22
SYSTEM ISSUES	23
IASIG INTERACTIVE 3D AUDIO (LEVEL 1) REQUIREMENTS	25
FUTURE DIRECTIONS (LEVEL TWO)	27

Table Of Figures

Figure 1 — Stereo Expansion	3
Figure 2 — Surround Sound vs. Virtual Surround	4
Figure 3 — System Model	5
Figure 4 — Typical sound field with a source, environment, and listener	7
Figure 5 — Illustration of IID	8
Figure 6 — Illustration of ITD	8
Figure 7 — IID - ITD cone	8
Figure 8 — Spectrum differences between original sound source and pinna reception	9
Figure 9 — Pinnae frequency modulation at varying elevations	9
Figure 10 — Source attenuation and absorption due to range (listener-source distance)	10
Figure 11 — Direct path, first and second order reflections in a typical room	10
Figure 12 — Speaker output and microphone input are combined to compute impulse response	11
Figure 13 — Impulse response synthetically applied to sound source to create illusion of a speaker ...	11

Introduction

This document represents the work of a group of 3D hardware and software vendors and application developers meeting as the 3D Audio Working Group (3DWG) of the Interactive Audio Special Interest Group (IASIG) of the MIDI Manufacturers Association. The goal of the group is to influence and improve the development of platforms for interactive multimedia, in the area of 3D audio playback.

This document defines the minimal expectations of the group with regards to 3D audio functionality, and differentiates that functionality from other approaches which are commonly called 3D audio as well. The functionality expected by the 3DWG is based on what is possible with technology today at acceptable price points. Future guidelines – specifically a “Level 2” document – will define additional functions and parameters that developers of 3D content, hardware and software believe is most appropriate for accurate 3D object and environment modeling as technology improves and 3D audio becomes more pervasive.

These guidelines and recommendations are not intended to be OS or API specific. However, portions of this document necessarily refer to specific platforms and APIs (particularly Microsoft’s) since those are typically where development is occurring today. Ultimately, these guidelines should help move the entire PC industry in a common direction, towards greater quality and superior performance, and more realistic audio in multimedia applications.

Abstract

The new generation of 3D audio localization technology can provide startling positional accuracy. It is now possible for listeners to perceive sounds emanating from above, below and from behind them, even when rendered using two speakers. However, this technology is being applied with great variation in features and performance in typical consumer PC products today — all under the same banner of “3D audio”. The result is it is difficult for consumers, let alone developers, to know for sure what to expect from PC audio systems and PC entertainment titles sporting “3D Audio”.

This document attempts to create the groundwork for correcting this situation, by:

- identifying and categorizing the different flavors of “3D audio” on PCs.
- describing and providing tools for the process of properly evaluating 3D Audio technology.
- defining a preferred implementation called “IASIG Interactive 3D Audio”, which includes real-time positioning of multiple sounds.

This document is expected to influence how 3D technology is applied to PC systems (and interactive entertainment products in general) so that there is less variation among products, and especially more consistent use of terminology, resulting in consistent performance and less confusion. Besides representing industry consensus on the topic to the development community, this document also assists product reviewers (test labs, magazines, etc.) to understand and evaluate the differences between various approaches to 3D imaging, and to report their finding to consumers.

The IASIG 3D Audio Rendering and Evaluation Guidelines consist of a general specification for 3D audio functionality (Level 1), along with information for understanding and evaluating performance of individual 3D audio products (e.g. sound cards).

3D Audio Primer

This primer presents the general concepts and performance of existing three-dimensional audio technology, in categories defined by the IASIG. The “IASIG Interactive 3D Audio” technology label is introduced with the purpose of identifying a single category of technologies that offers “acceptable” 3D audio performance. This section also explains how we hear sound and provides the foundation from which to evaluate 3D audio performance. Lastly, a glossary of terms is included for reference and is also intended to standardize the usage of terminology in the industry.

Introduction to 3D Audio

At its simplest level, the term “3D” (3-dimensional) audio means sound which comes from all around the listener. This is, in fact, exactly how sound occurs and is perceived in the real world. When reproducing sounds with an ordinary stereo system, however, all sense of 3-dimensions is lost. Instead, stereo only produces sounds which can be recognized as coming from one speaker or the other, or perhaps somewhere in between, but certainly not from any other location. The result is that stereo recordings, though pleasurable, fall far short of reproducing a “realistic” listening environment.

One obvious approach to improving upon stereo is to use more speakers. Additional speakers behind the listener create a “surround” effect which is much more exciting and realistic than stereo, and is very popular in movie theaters (and home theaters) today. However, the addition of more speakers still does not produce a totally realistic listening environment, because the additional speakers do nothing to improve the perception of distance or height which is part of our real-world experience.

Psycho-acoustic researchers have been studying this problem for decades, focusing on understanding the functioning of the human hearing system which makes sound localization possible in the real world: the principles of *binaural* human hearing. Binaural means that we hear using two ears. From the two signals that our ears perceive, we can extract enough information to tell where a sound is located in the three dimensional space around us.

Since we can hear three-dimensionally in the real world using just two ears, it is logical to expect that a realistic 3D audio effect can be produced using just two speakers or a set of headphones. On this basic assumption, many 3D audio products have been built. This primer focuses on explaining the technology and science behind the different kinds of 3D audio technology, as well as how we hear, and what one can (and cannot) expect from the different technologies.

Real 3D Audio vs. Not-So-Real 3D Audio

Even though there are many different technologies which are called “3D audio”, they do not all produce the same effect. Each technology is discussed below with an explanation of the applications they are geared towards, and why they are *not* considered to be part of the IASIG preferred category of technologies, called “Interactive 3D Audio”.

The IASIG Interactive 3D audio category combines the ability to position sounds anywhere in 3-dimensional space, and interactively reposition those sounds “on the fly”. These two components are especially significant to the multimedia entertainment market which the IASIG was formed to address and these combined effects offer a new kind of audio listening experience.

Stereo Expansion

Stereo expansion technologies process an existing stereo (two channel) soundtrack to add spaciousness and to make sound appear to originate from outside the left/right speaker locations. This category of technologies uses psychoacoustic cues to create the effect, and is particularly useful for improving stereo performance where speakers are placed very closely together, as is common with many integrated multimedia desktop systems.

Stereo Expansion technologies do not produce sounds in 3-dimensions, and are not applied to individual sounds, but rather to the final mixed output. They differ substantially from IASIG Interactive 3D Audio because they only offer passive spreading of an existing soundtrack, and not interactive 3D positioning of individual sounds.

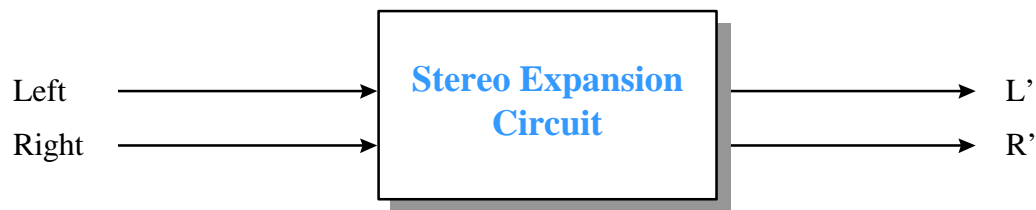


Figure 1 — Stereo Expansion

Multi-Speaker Environments (Surround Sound)

This category includes technologies and products that create a larger-than-stereo sound stage by playing back multi-channel Dolby® or MPEG surround sound soundtracks on multi-speaker setups. Surround sound is based on using audio compression technology (for example Dolby Pro Logic® or Digital AC-3®) to encode and deliver a multi-channel soundtrack, and audio decompression technology to decode the soundtrack for delivery on a surround sound 5-speaker setup.

Surround sound soundtracks are applied most often to movies. They are non-interactive, and therefore not specifically well-suited for inclusion in interactive software titles (games, web sites, and so on), except for background music and non-interactive scenes. Current solutions do not offer any means for interactive positioning of sounds, so surround sound systems are not considered in the IASIG Interactive 3D Audio category.

Quad Output

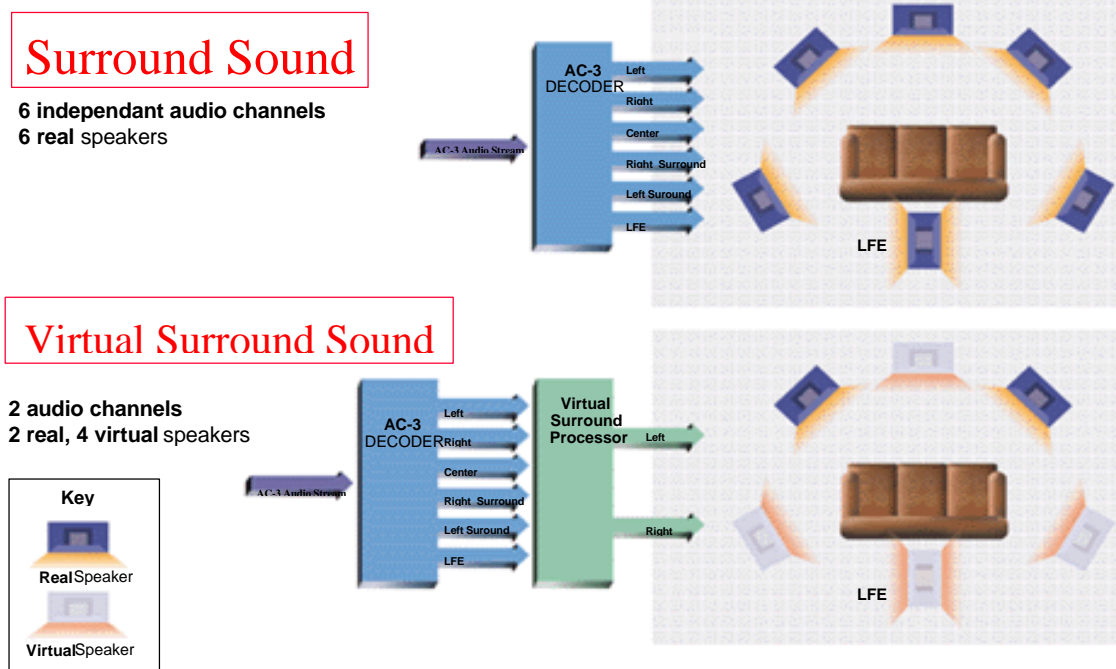
Some sound card vendors now provide two (or more) stereo outputs on their products and provide drivers which allow developers to address those outputs as four discrete channels (“Quad”) as opposed to (or in addition to) using them solely for playback of encoded surround sound tracks. This should not be confused with earlier implementations of “Quad” technology which attempted to matrix the four channels of sound into a stereo stream — in this version Quad is only an output format, not an input (source) format, so it has none of the problems of the earlier implementations.

Without any standard method for implementing quad, there are very few titles which will take advantage of this configuration, but interactive control of sound placement is possible with these products. Still, it is not yet clear which, if any, of these solutions will also produce distance and height cues, so for now these systems in general are not considered part of the IASIG Interactive 3D Audio category.

Virtual Surround Sound

Virtual Surround Sound systems use binaural technology to create the illusion of five speakers emanating from a regular set of stereo speakers. Each of the 5 speaker signals is processed independently to add the proper binaural cues to position that signal where a speaker might actually be (had there been one!), therefore enabling a surround sound listening experience without the need for five physical speakers.

This category of technologies provides significant benefit in situations where full home theater set-ups are not practical, such as on computer systems, or in small homes. However, since the source material is encoded, only the speaker position, not the position of any individual sound, can be changed. Since current solutions do not provide any means for interactive positioning of specific sounds, virtual surround sound technologies are not included in the IASIG Interactive 3D Audio category.



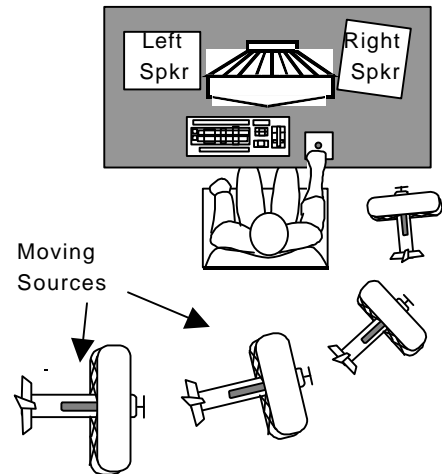
The upper section shows a typical multi-speaker surround sound configuration, where the 5.1 channels of audio are fed to corresponding speakers. The lower portion depicts a typical virtual surround system, where decoded AC-3 source material is down-mixed for accurate 2-speaker playback.

Figure 2 — Surround Sound vs. Virtual Surround

IASIG Interactive 3D Audio

IASIG Interactive 3D Audio allows for on-the-fly positioning of sounds anywhere in the three-dimensional space surrounding a listener. Support for such technologies can be incorporated into software titles such as video games to create a natural, immersive, and interactive audio environment that closely approximates a real-life listening experience.

These technologies replicate the 3D audio cues that the ears hear in the real world, as explained in the following two sections, “The Basics of Acoustics” and “The Basics of Human Hearing”. IASIG Interactive 3D Audio is achievable on all audio playback environments: headphones, stereo speakers and multi-speaker (surround or quad) arrays.



System Model

The IASIG Interactive 3D Audio System consists of three layers: the application, the application programming interface (API) and the audio renderer (see figure 3). The software application could be a game, or maybe a music playback/composition program, which takes audio files and assigns to each some default positional character. The application also accepts data from the user via the joystick, mouse, keyboard or other input device which provides the interactive element, modifying the final positional information.

The syntax and structure of the 3D audio events may be platform-dependent, and thus an IASIG Interactive 3D Audio compatible API is required to do the translation between the application and the renderer. The renderer can be either hardware or software, and must be able to interpret the received events and successfully produce a believable 3D image for a minimum of 8 simultaneous objects. The complete list of instructions and rendering features for an IASIG Interactive 3D Audio System is described in the third section of this document.

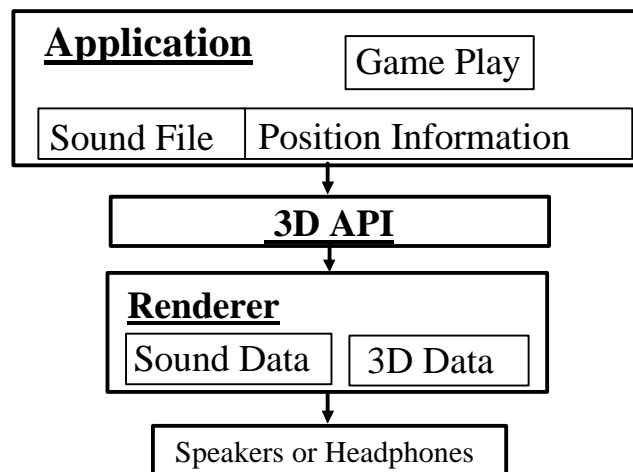


Figure 3 — System Model

**Table 1: Summary Comparison of
Audio Playback Methods**

<i>Type of processing</i>	<i>headphone compatible</i>	<i>stereo speaker compatible</i>	<i>speaker array (quad, home theater)</i>	<i>dimensionality</i>	<i>interactive user controls</i>	<i>perceptual performance</i>
Mono	yes	yes	N/A	0-D	no (on/off)	single point source from speaker location
Stereo	yes	yes	N/A	1-D (left/right)	left/right panning	sounds placed on line between speakers
“3D” Stereo Expansion	some	most	N/A	1-D (added spaciousness)	no (on/off)	sounds fill area around speakers
Multi-Speaker Surround Sound	no	no	yes	2-D (left/right, front/back)	no (soundtracks are pre-encoded)	sounds placed on a circle formed by real speakers
Virtual Surround Sound	some	most	N/A	2-D (left/right, front/back)	no (soundtracks are pre-encoded)	sounds placed on circle formed by virtual speakers
IASIG Interactive 3D Audio	most	most	some	3-D (left/right, front/back, up/down)	full 3D placement using XYZ coordinates	sounds placed at any distance and position from listener

The Basics of Acoustics

Human beings extract a lot of information about their environment using their ears. In order to understand what information can be retrieved from sound, and how exactly it is done, we need to look at how sounds are perceived in the real world. To do so, it is useful to break the acoustics of a real world environment into three components: the sound source, the acoustic environment, and the listener:

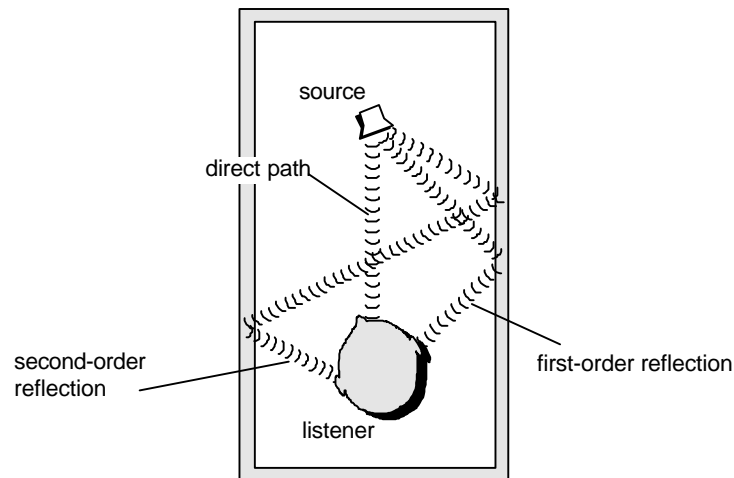


Figure 4 — Typical sound field with a source, environment, and listener

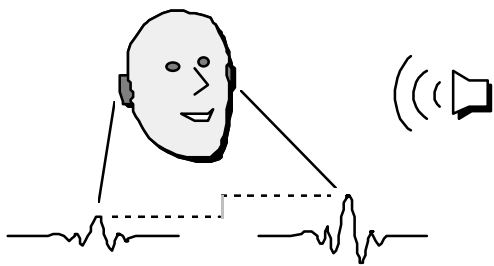
- **The sound source:** this is an object in the world that emits sound waves. Examples are anything that makes sound - cars, humans, birds, closing doors, and so on. Sound waves get created through a variety of mechanical processes. Once created, the waves usually get radiated in a certain direction. For example, a mouth radiates more sound energy in the direction that the face is pointing than to side of the face.
- **The acoustic environment:** once a sound wave has been emitted, it travels through an environment where several things can happen to it: it gets absorbed by the air (the high frequency waves more so than the low ones. The absorption amount depends on factors like wind and air humidity); it can directly travel to a listener (direct path), bounce off of an object once before it reaches the listener (first order reflected path), bounce twice (second order reflected path), and so on; each time a sound reflects off an object, the material that the object is made of has an effect on how much each frequency component of the sound wave gets absorbed, and how much gets reflected back into the environment; sounds can also pass through objects such as water, or walls; finally, environment geometry like corners, edges, and small openings have complex effects on the physics of sound waves (refraction, scattering).
- **The listener:** this is a sound receiving object, typically a “pair of ears”. The listener uses acoustic cues to interpret the sound waves that arrive at the ears, and to extract information about the sound sources and the environment.

The Basics of Human Hearing

As explained above, people can be considered sound receiving objects in an environment. We have an auditory sensing system consisting of two ears and a brain. Additionally, very low frequency sounds can be sensed through the human body. The brain uses a number of cues that are embedded in the two sound signals it receives from the two ears to learn about the sounds and their environment. Most people are unaware that the effects described in the following sections greatly impact our continuous perception of reality, every day of our lives. On the other hand, there are certain people, for example non-sighted people, that are very much aware of these effects, because they heavily rely on their ears for querying and navigating their surroundings.

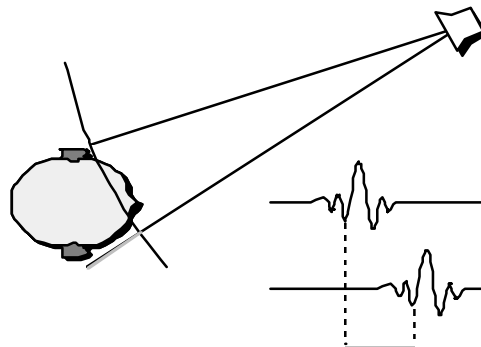
Primary Localization Cues - IID and ITD

The two primary localization cues are called interaural intensity difference (IID) and interaural time difference (ITD). IID refers to the fact that a sound is louder at the ear that it is closer to, because the sound's intensity at that ear will be higher than the intensity at the other ear, which is not only further away, but usually receives a signal that has been shadowed by the listener's head (see fig. 2). ITD means that a sound will arrive earlier at one ear than the other (unless it is located at exactly the same distance from each ear - for example directly in front). If it arrives at the left ear first, the brain knows that the sound is somewhere to the left (see fig. 3).



Interaural Intensity Difference (IID)

Figure 5 — Illustration of IID



Interaural Time Difference (ITD)

Figure 6 — Illustration of ITD

The combination of these two cues allows the brain to narrow the position of an individual sound source to somewhere on a cone centered on the line drawn between the listeners ears (see fig.4).

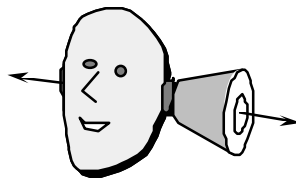


Figure 7 — IID - ITD cone

The Outer Ear Structure - Pinna

Before a sound wave gets to the ear drum, it passes through the outer ear structure, called the pinna. The pinna accentuates or suppresses mid- and high-frequency energy (see fig. 5) of a sound wave to various degrees, depending on the angle at which the sound wave hits the pinna (see fig. 6). This means that the two pinnae act as variable filters that effect every sound that passes through them. The brain knows how to figure out the exact location of a sound in space by receiving a signal that has been filtered in a way that is unique to the sound source's position relative to the listener.

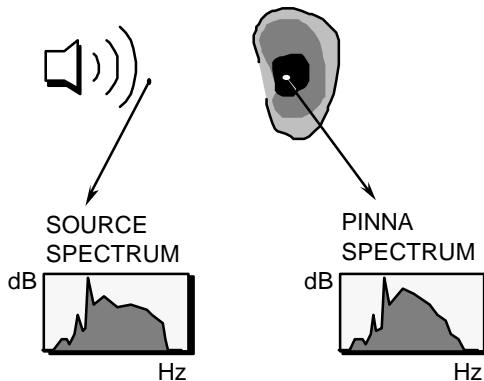


Figure 8 — Spectrum differences between original sound source and pinna reception

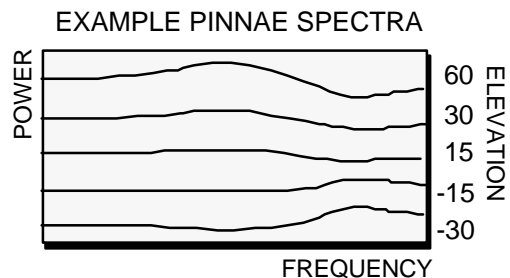


Figure 9 — Pinnae frequency modulation at varying elevations

The pinnae are the key to accurately localizing sounds in space. However, since the outer ear and its folds are on the scale of a few centimeters, only sound waves with wavelengths in the centimeter range or smaller can be affected by the pinna. In addition, the two ears are about 15 centimeters apart, so even IID and ITD cues are greatly reduced for wave lengths bigger than that. For example, a 3.3 kHz sound signal oscillates 3300 times per second, while sound travels at about 330 meters per second. The wave length is therefore about $330/3300 = 0.1$ meters, or 10 centimeters. This means that a sound at 3300 Hz lies in the area where primary cues are still noticeable, but pinna cues start to be diminished. In general, the higher the frequency of a sound, the shorter its wave length, and the better it can be localized. This phenomena can be verified by placing two speakers, a sub-woofer and a high-frequency tweeter, in a room and playing music through them. With closed eyes you will be able to immediately tell where the tweeter is located, the sub-woofer however will sound like it is “coming from everywhere”.

Propagation Effects, Range Cues, and Reflections

Many things happen to a sound as it travels through an environment before it is received by a listener. All of these effects provide us with clues to what we are hearing and what kind of environment we are in:

- a somewhat muffled, quiet sound is likely off in the distance (see fig. 7).

- if it is heavily muffled, we might be in an enclosed space, listening through glass, or other wall materials.
- the effect of sound reflections in an environment is very important, because we are able to hear the difference in time of arrival and location between the direct path signal, first-order, and n^{th} order reflections (see fig. 8). The reflections give us a way to further pin-point a sound source's location, as well as the size, shape and type of room or environment that we are in (people with very "good ears" are able to exactly locate a wall, or tell the difference between a open or closed door, simply by listening to reflections). While humans are capable of individually perceiving first order reflections, second and higher order reflections usually combine to form what are called late field reflections, or reverb.

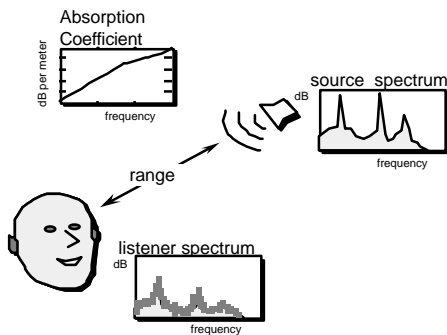


Figure 10 — Source attenuation and absorption due to range (listener-source distance)

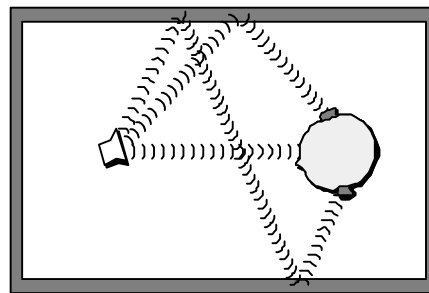


Figure 11 — Direct path, first and second order reflections in a typical room

3D Audio Reproduction

A 3D audio system aims to digitally reproduce a realistic sound field. To achieve the desired effect a system needs to be able to re-create portions or all of the listening cues discussed in the previous chapter: IID, ITD, outer ear effects, and so on. A typical first step to building such a system is to capture the listening cues by analyzing what happens to a single sound as it arrives at a listener from different angles. Once captured, the cues are synthesized in a computer simulation for verification.

What is an HRTF?

Details of different 3D audio technologies cannot be described in this document because they are proprietary and technologies vary in their implementations. However, the majority of technologies (except for ones that are strictly limited to playback on multi-speaker arrays) are at some level based on the concept of HRTFs, or Head-Related Transfer Functions. An HRTF can be thought of as set of audio filters (for each ear) that contain all the listening cues that are applied to a sound as it travels from the sound's origin (its source, or position in space), through the environment, and arrives at the listener's ear drums. The filters change depending on the direction from which the sound arrives at the listener. The level of HRTF complexity necessary to create the illusion of 3D realistic hearing is subject to considerable discussion and varies greatly across technologies.

HRTF analysis

The most common method of measuring the HRTF of an individual is to place tiny probe microphones inside a listener's left and right ear canals, place a speaker at a known location relative to the listener, play a known signal through that speaker, and record the microphone signals. By comparing the resulting impulse response with the original signal, a single filter in the HRTF set has been found (see fig. 9). After moving the speaker to a new location, the process is repeated until an entire, spherical map of filter sets has been devised.

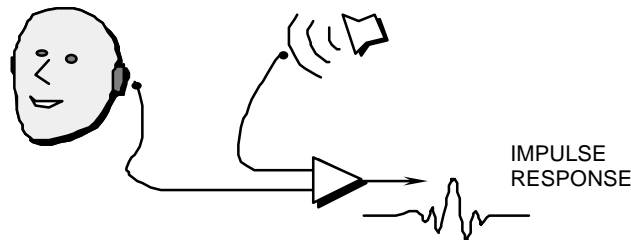


Figure 12 — Speaker output and microphone input are combined to compute impulse response

Every individual has a unique set of HRTFs, also called an ear print. However, HRTFs are interchangeable, and the HRTF of a person that can localize well in the real world will let most people localize well in a simulated world. While generic, interchangeable HRTFs are suitable for general applications such as video conferencing or games, individualized HRTFs are useful for performance critical applications of binaural audio, such as jet fighter cockpit threat warning systems, or air traffic control systems.

HRTF synthesis

Once an HRTF has been devised, real-time DSP (digital signal processing) software and algorithms are designed. This software has to be able to pick out the critical (psycho-acoustically relevant) features of a filter and apply them in real-time to an incoming audio signal to “spatialize” it. The system works correctly if a listener cannot tell the difference between listening to a sound over the speaker setup from the analysis process above (the speaker is in a specific position), and the same sound played back by a computer and filtered by the HRTF impulse response corresponding to the original speaker location (see fig. 10).

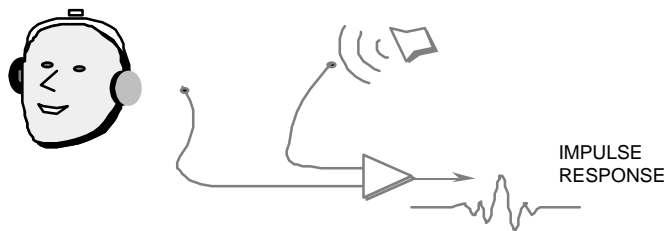


Figure 13 — Impulse response synthetically applied to sound source to create illusion of a speaker

Head Movement

Audio cues change dramatically when a listener tilts or rotates his or her head. For example, quickly turning the head 90 degrees to look to the side is the equivalent of a sound traveling from the listener's side to the front in a split second. We often use head motion to track sounds or to search for them. The ears alert the brain about an event outside of the area that the eyes are currently focused on, and we automatically turn to redirect our attention. Additionally, we use head motion to resolve ambiguities: a faint, low sound could be either in front or back of us, so we quickly and sub-consciously turn our head a small fraction to the left, and we know if the sound is now off to the right, it is in the front, otherwise it is in the back. One of the reasons why interactive audio is more realistic than pre-recorded audio (soundtracks) is the fact that the listeners head motion can be properly simulated in an interactive system (using inputs from a joystick, mouse, or head-tracking system).

Audio-Visual Synergy

The eyes and ears often perceive an event at the same time. Seeing a door close, and hearing a shutting sound, are interpreted as one event if they happen synchronously. If we see a door shut without a sound, or we see a door shut in front of us, and hear a shutting sound to the left, we get alarmed and confused. In another scenario, we might hear a voice in front of us, and see a hallway with a corner; the combination of audio and visual cues allows us to figure out that a person might be standing around the corner. Together, synchronized 3D audio and 3D visual cues provide a very strong immersion experience. Both 3D audio and 3D graphics systems can be greatly enhanced by such synchronization.

Summary

Applications of 3D audio range from 3D Web sites or video games, where monsters can roar from behind and helicopters can circle overhead, to phone- or video-conferences with multiple participants placed in 3D audio space, and mission critical applications such as air traffic controllers receiving communications that are 3D localized to the radar position of a plane, or jet fighter pilots receiving threat warnings that are synchronized to the 3D position of an incoming missile.

Interactive 3D audio systems, and the software titles that are designed to take advantage of them, are being introduced to the PC market place in 1997. Early indications show that affordable, yet high-quality, interactive 3D audio is a very impressive technology that will be embraced rapidly by both software developer and end-user communities, and that shows a long lasting path of growth and technology improvements ahead of it.

Glossary

- **3D sound/3D audio:** refers to the fact that sounds in the real world are three-dimensional. Human beings have the ability to perceive sound spatially, meaning that they can figure out where a sound is coming from and from how far away. Additionally, we can hear where sounds are in relation to their surroundings and in relation to each other. There are three main pieces of information that are essential for the human brain to perform these functions:
 - **ITD**, or Interaural Time Difference, means that unless a sound is located at exactly the same distance from each ear (e.g. directly in front), it will arrive earlier at one ear than the other. If it arrives at the right ear first, the brain knows that the sound is somewhere to the right.
 - **IID**, or Interaural Intensity Difference, is similar to ITD. If a sound is closer to one ear, the sound's intensity at that ear will be higher than the intensity at the other ear, which is not only further away, but usually receives a signal that has been shadowed by the listener's head.Finally, the trickiest part is the fact that a sound bounces off a listener's shoulders, face, and outer ear, before it reaches the ear drum. The pattern that is created by those reflections is unique for each location in space relative to the listener. A human brain can therefore learn to associate a given pattern with a location in space. These patterns can be summarized in a set of acoustic filters called **Head Related Transfer Functions (HRTFs)**.
- **acoustic materials:** by absorbing the sound energy at different frequencies, the material of which an object is made effects the way the sound reflects off and transmits through the object. A carpeted room sounds very different from a glass room. An object's material characteristics can be measured empirically by recording known sounds as they bounce off of materials.
- **ambient channel:** a way of displaying sounds as coming from everywhere - all around the listener. This is useful for background music or ambiance sounds such as rain.
- **atmospheric absorption:** the attenuation of sounds as they propagate through a medium. For example, in air the high frequency components of sound attenuate faster than the lower frequency components.
- **auralization:** the process of rendering audio by physically or mathematically modeling a sound field of a source in space in such a way as to simulate the binaural listening experience at any given position in a modeled space.
- **binaural:** two audio tracks, one for each ear (as opposed to stereo, which is one for each speaker). Binaural sounds are what we hear in everyday life.
- **cross talk:** the fact that when listening to a set of stereo speakers, the left ear can hear some of the right speaker, and the right ear some of the audio coming from left speaker. Cross-talk cancellation refers to technologies that remove such crossing of audio information to the “opposite” ear.
- **direct path:** the direct path from a sound source to a listener's ears (as opposed to reflections off of surfaces). The direct path allows a listener to tell where each sound is coming from, 360 degrees both in azimuth and elevation. This is the main concept of any interactive 3D sound system.
- **Doppler effect:** the change in frequency of a sound wave due to the motion of a sound source or of a listener. For example, if a car moves past a listener while sounding its horn, the listener will hear a sudden drop in pitch as the car passes.

Interactive Audio Special Interest Group
3D Audio Rendering and Evaluation Guidelines — Level One

- **extended stereo** (see stereo expansion)
- **gain**: the amplification or attenuation of a sound source, usually measured in dB (decibels). 0 dB means no amplification and no attenuation. A positive value amplifies a source, a negative value attenuates it.
- **HRTF**: HRTFs, or Head Related Transfer Functions, are a set of mathematical transformations which can be applied to a mono sound signal. The resulting left and right signals are the same as the signals that someone perceives when listening to a sound that is coming from a location in real-life 3D space. HRTFs contain the information that is necessary to simulate a realistic sound space. Once the HRTF of a generic person is captured, it can be used to create 3D sound for a large percentage of the population (most people's heads and ears, and therefore their HRTFs, are similar enough for the filters to be interchangeable). However, additional performance increases are possible through HRTF customization.
- **IASIG Interactive 3D Audio**: see “interactive 3D audio”
- **IID**: Interaural Intensity Difference, see "3D sound".
- **interactive 3D audio** summarizes technologies that allow interactive placement (positioning) of sound sources anywhere in the 3D space surrounding a listener. The **IASIG Interactive 3D Audio** category contains these technologies. The term was devised to indicate the difference between interactive positional 3D audio and other technologies (especially stereo expansion) that are sometimes labeled 3D.
- **interactive audio**: sounds that are created on-the-fly based on unpredictable user actions and story lines. The opposite of a pre-recorded soundtrack.
- **ITD**: Interaural Time Difference, see "3D sound".
- **listener**: an object in a sound space that is sampling ("listening to") sound, usually a head with associated HRTF characteristics.
- **medium**: see "atmospheric absorption" and "transmission loss".
- **mono/monophonic**: refers to a single audio signal, usually rendered on a single speaker. Mono sounds appear to originate from the speaker, or from the center of a listener's head in the case of headphones.
- **positional 3D Audio**: see “interactive 3D audio”
- **psycho-acoustics**: an area of psychology that studies the structure and performance of human auditory perception.
- **quad**: a speaker configuration where four speakers are used instead of stereo, which allows sounds to be produced in 360-degree space.
- **radiation pattern**: a sound-emitting object can radiate sound in a certain pattern (rather than uniformly all around it). For example, a head emits sounds in the direction that its nose is pointing.
- **reflection**: a sound reflection off of a surface. It gives a listener information about the listening environment and the location and motion of sound sources. See "surfaces".
- **refraction**: dispersion of sound waves as they travel around the edges and through openings of objects.
- **reverberation**: or reverb, refers to the sum of all sound reflections in a listening environment.

Interactive Audio Special Interest Group
3D Audio Rendering and Evaluation Guidelines — Level One

- **sound source:** refers to an object in 3D space that emits sound. The actual sound signal that it sends out can be a live signal, a wave file, a MIDI voice, or any other audio signal. A 3D sound device often gets rated on how many different 3D sources it can independently position at any one time. Realistic sound spaces can be created with as few as four concurrent sources, very complex spaces can have dozens of separate sounds at a time.
- **speaker arrays:** an installation of multiple speakers in a certain pattern, usually designed to create a sound field within the space defined by the speakers. Examples are stereo speakers, or surround speakers.
- **stereo/stereophonic:** refers to two audio signals, usually rendered on two separate speakers. Stereo sounds appear to originate from somewhere between the two speakers, or between the ears of a listener in the case of headphones.
- **stereo expansion:** a term that summarizes a number of techniques that involve processing of traditional stereo sounds with the goal of making them appear to originate from a range which extends beyond the physical speaker locations. The effect often produces sound that is "bigger" or more diffuse than regular stereo sound. Most extended stereo technologies are optimized for stereo speaker environments with the speaker placed close to each other (i.e. a multimedia computer environment). Fewer technologies focus on enhanced headphone reproduction.
- **surfaces:** sounds not only travel to a pair of ears on a direct path, but they also bounce off of objects in the world. Most natural listening environments contain at least a sound reflecting ground plane, such as a floor. Therefore, reflecting objects are necessary to make virtual environments sound natural and realistic. They help listeners navigate and enhance the overall effect of immersion in a virtual environment. Almost as important as reflections, is the absence of a reflection. For example, the ears can tell the change in a sound space when a reflection is removed by opening a door or window.
- **surround sound:** the term surround sound summarizes a number of audio technologies that are designed to store and deliver multi-channel audio content in efficient ways. The result is a sound field that is much larger than a stereo sound field, without the limitations associated with early multi-channel audio solutions (see quadrasonic sound). These technologies usually involve some form of analog or digital audio compression and decompression. Audio content is pre-encoded using compression techniques, delivered, and decompressed for playback.
- **sweet spot:** the location where a listener has to be placed to get the optimal effect when listening to a specific speaker setup.
- **transmission loss:** sounds get absorbed as they travel through objects such as walls (similar to atmospheric absorption in the case of traveling through a medium). Transmission loss models are needed to realistically simulate sounds outside a closed window or in the next room.
- **update rate:** the number of times that a specific instance of a sound space gets re-computed and updated per second. Each time any object moves (most often the listener), the space needs to get updated. The higher the update rate, the faster the objects can move without creating audio artifacts, such as clicking. Audio update rates generally range from a minimum of 20Hz to 100Hz. Video update rates are usually in the same range (TV signals are updated at 30Hz).
- **virtual surround sound:** technologies that employ 3D audio techniques to deliver surround sound soundtracks (see "surround sound") over a regular set of stereo speakers or headphones, therefore alleviating the need for a five speaker setup.

3D Audio Evaluation Guidelines

These guidelines are intended to assist reviewers, test labs and others interested in 3D audio with the evaluation of 3D audio products, especially “accelerators”, for PC multimedia. The emphasis is specifically on comparing 3D features and the performance of algorithms used in individual products. Therefore, more general aspects of sound system evaluation (such as signal/noise ratio, Sound Blaster compatibility, etc.) are not addressed except as they may directly affect the perception of 3D sound.

While a certain emphasis will be on what items can be objectively evaluated and quantified, it should be noted that 3D perception is largely a subjective experience. Therefore, some subjective listening tests must be performed to fully evaluate 3D audio products.

It should also be pointed out that, unlike visuals, location of objects via audio signals is not very precise for humans. With the exception of the anecdotal blind individual who is able to hear and locate sounds with extreme precision, it is exceedingly difficult for humans to establish, from sound alone, the *precise* location of an object in the complete absence of all other clues 100% of the time. Since our natural hearing is somewhat imprecise, mimicking the effect with computer processing can only go so far in fooling the listener as to the perceived location of a particular sound. It is therefore strongly recommended, that moving sound objects are tied/synchronized to visual moving objects to reinforce the audio positioning cues.

Given the subjective limitations, current 3D audio algorithms are none-the-less quite sophisticated. They provide very compelling 3D audio experiences, especially when tied to visuals as recommended. In conjunction with other sound cues such as Doppler shift and other effects of motion, the experience from current 3D audio technology can be quite dramatic. It is hoped that this document will educate you and help you better understand how to best evaluate this exciting and new technology for the PC.

General Considerations

Hearing a 3D audio effect is a *psychoacoustic* phenomenon. It is therefore harder to quantify success than something that can be measured more objectively, like the number of polygons per second a graphics system can render. As much as possible, the evaluation of psychoacoustic systems such as 3D audio should be performed by examining statistically significant numbers of test subjects in properly controlled experiments. In fact, these types of experiments are routinely performed by the creators of 3D audio algorithms. However, realizing that full scientifically valid, controlled experiments are not likely to be performed when reviewing 3D audio products, reasonable evaluations can still be made as long as the following points are considered:

Playback System

Although 3D audio products are designed to be as robust as possible over a wide range of listening environments, a poor playback system can adversely affect the quality of the 3D audio effects.

- **Use reasonable quality, matched speakers**

Any audio system is only as good as its speakers. While you may want to use a wide variety of speakers when conducting listening tests, and report how robust the 3D audio effect is over various speakers, realize that the quality of the speakers may be the limiting factor for hearing a good 3D audio effect, rather than the 3D sound card itself.

Speakers should also be set up according to the particular sound card's instructions. Typically, this will call for both speakers (in the case of 2 speaker playback) to be placed symmetrically on either side of the monitor. Each speaker is a directional emitter and the location and orientation is important for good 3D audio playback. The speakers should be facing forward and their height should match the display. The left speaker should target the left ear and the right speaker should target the right ear while looking at the screen during normal use of the computer. Where possible the location of the speakers should be near the center of the room to avoid strong first order reflections (see figures 4 and 11). Improper placement can weaken the distance cues and muddy the positional cues as well.

Nearly flat frequency response speakers are not in the PC consumer budget, however, when selecting speakers chose those with better tweeters. As a general rule, higher frequency sounds are more easily positioned and the pinnae effect is in the higher frequency as well. Although the average PC uses sub-optimal speakers, significant parts of the 3D effect will still be audible.

- **Verify Headphones, 2-Speaker and Multi-speaker Listening**

Sophisticated 3D audio processing requires different algorithms for 2-speaker, headphone and multiple speaker (more than 2) playback. Therefore, separate listening tests should be performed with each output format supported by the sound card. The selection of output format can likely be done with an application provided by the sound card manufacturer. It is very important to make sure that the sound card is set to the proper output format. For example, if the sound card is set for headphone playback, but is listened to over speakers, the 3D audio effects will be drastically reduced or eliminated. (NOTE: Some 3D audio implementations do not make a distinction between dual speaker, multi-speakers and headphone playback.)

- **Ensure only a single 3D audio process is used**

On systems that support more than one 3D audio process (such as a sound card that provides Interactive 3D audio and speakers with "3D Stereo Expansion") turn off all processes except the one being measured/evaluated.

Listeners

Since 3D audio is a subjective experience, different listeners can have widely differing experiences even when listening to exactly the same sounds.

- **Use multiple listeners when testing**

As mentioned, different people may hear the same sound and have differing opinions as to the quality of the 3D audio effect. Therefore, have as many people listen to the system as possible and record their input as part of the evaluation.

- **Beware of "adaptation"**

Adaptation is the bias often resulting from the prolonged listening of one system. When one system is heard after listening to another system, there can be a tendency to rate the first system as "better" than the second; the listener has adapted to the first system as the reference standard. Ideally, systems should be listened to, in a multi-pass pseudo-random sequence.

Sounds

- **Use a variety of sounds**

A 3D audio process may not give the same 3D effect for two differing sound sources (for example, a helicopter and glass shattering). By using a wide variety of sounds of differing types (speech, “boomy” sounds, “screechy” sounds, etc.) you will get a better idea of the 3D audio effect.

- **Use motion, with visual references**

In the complete absence of any visual or other clues, even in the “real world”, 3D hearing is not precise. For this reason, completely "static" tests (the abstract positioning of sound with no other reference) may not give a good indication of how a sound card will perform in a real application. When elements such as visual cues, motion of sounds and context are added, 3D audio effects generally increase greatly.

- **Test using multiple bit-depth and sampling rate sounds**

The perceived 3D audio effect may depend on the quality and nature of the test material, and the particular audio card may also have different latency and CPU requirements. Use test sounds with a variety of bit depths (8-bit, 16-bit) and sampling rates (11kHz, 22kHz, 44kHz, 48kHz).

- **Use the same sound set to test each system**

Since the perceived effect may depend on the particular sound and sampling rate/bit depth, it is critical that the same sound sets are used when evaluating different 3D audio systems. I.e. use exactly the same sound files when testing specific features of different sound cards.

Considerations for Controlled Experimentation

As mentioned above, evaluation of 3D audio through statistically significant numbers of test subjects in properly controlled experiments can be done, if desired, and will produce objective and quantifiable results when done with care. To achieve an acceptable margin of error with subjective audio tests, the following factors should be considered and controlled in each test and for each test subject.

Physical Biases

The physical or electroacoustical properties of the audio equipment, the program material or the listening room can bias the results. Therefore the quality of the test and testing environment can be improved by the following:

- Listening Room should be the same for all tests
- Loudspeaker position should meet manufacturers specification and be kept constant
- Listener position should be appropriate for proper use model (for example, the listener must be at the proper distance from the speakers)
- Speaker balance should be properly adjusted and repeatable (the same) in all tests
- Speaker loudness should be the same in all tests
- Program material should be appropriate and constant among all tests

Psychological/Physiological Biases

To lessen psychological and physiological issues the focus needs to be on the test methodology for bias removal. The speakers with nicer cabinets will sound better unless bias is removed through proper methodology. Screening out listeners who have bad hearing and building a methodology which trains listeners improves results. The test bias can be improved using some of the following:

- The Look - hide the product so that brand names or visual beauty can not provide bias
- The Task - take a test run to train the listener before testing the listener
- The Ability - screen out listeners who have severe hearing issues
- The Peer Pressure - keep expectations removed from the testing process
- Randomize the Test - keep recognition of the above issues to a minimum

Experimental Biases

The data and test method creation along with recording techniques are key to a good test results. Insuring that the tests produce quantifiable results that are meaningful demand careful planning. The meaning and quantification can be improved by using the following:

- Classification - carefully define categories and scales which create meaningful ranges for analysis
- Number - test enough dimensions to profile the issues; use at least 8 testers for significant results
- Adaptation - rapid testing to stop adaptation effects improve results
- Order - Follow a specific order every time.

Specific Evaluation Criteria

The following criteria should be considered when listening to and comparing Interactive 3D Audio technology.

CHARACTERISTIC	EXPLANATION
Roll-off	Does a sound properly attenuate with distance ?
Radiation	Does a sound properly play in accordance to defined radiation patterns ?
Doppler	Does sound exhibit proper Doppler shift effects ?
Coloration	Does sound maintain timbre when position is changed ?
Position:	How “good” is the 3D image 1.) left and right?, 2.) up and down?, 3. front and back?
Listener Zone:	How directional is this implementation. Is the area where you must stand/sit to hear the effect (a “sweet spot”) typical for normal use?
Total experience (real-world test):	What is the overall 3D audio experience when running the test application with 3D graphics and 3D audio?

Roll-off:

Roll-off refers to how a sound changes with distance. As a sound-emitting object moves further from the listener, the sound becomes softer and in some cases may change in other ways to give an illusion of distance.

Trait	Listening Considerations
Depth and Distance	<p>Sounds should get quiet and slightly muffled as they move away from the listener (air absorbs sound waves, high frequencies more rapidly than low frequencies a level two issue). Within the minimum distance, the sound source should remain at constant maximum volume; beyond the maximum distance, the source should get no quieter than at max. distance. Between the two values, the source should attenuate as normal.</p> <p>Listen to how well the technology works with sounds that travel towards and away from you. Does it transition smoothly? Try listening with your eyes closed, to eliminate any reinforcement by visual cues. Solutions should create a realistic illusion of distance and depth.</p>

Radiation:

When a person is talking, if they are facing towards you, the sound of their voice will be louder than if they are facing away from you. This is sometimes referred to a "sound cone." By knowing the sound cone characteristics, a subjective analysis can be made of the radiation factor by judging how the sound changes with orientation between source and listener.

- The cone allows sources to be louder in a specified cone, and quieter outside it.
- Smooth volume changes as the cone rotates are desirable, as well as a clear difference in the volume when the cone is oriented away from the listener.

Doppler:

Doppler shift is the pitch-shifting effect that occurs when there is relative movement between the emitter of a sound and the listener. One very familiar example of this effect is the change of pitch of the bell heard as a train approaches a railroad crossing. Doppler shift must be calculated from both object velocities and reproduced by the 3D audio renderer.

- A clear change in pitch should be detected as the source travels past the listener. With exaggerated passes for drama, this pitch shifting should sound better than natural.
- Noise, distortion and lack of or inaccurate Doppler behavior should also be listened for.

Coloration:

Coloration refers to how the timbre or sound quality (sometimes called tone color) changes as the position of a sound is changed. It is very important to note that some coloration of sound is inherent in the 3D audio process and is actually part of the 3D audio effect itself. As a result, the judgment of "undesirable" coloration can be very difficult.

However, as a side effect of the positioning algorithm, a sound may become brighter or darker, thinner or fuller or change in other undesirable ways. Coloration on Windows systems can be evaluated by choosing a single .wav file as a test standard, and playing the sound through the regular Windows wave player. Since the wave player does not send data through the 3D process, this represents the "true" sound of the wave file. Then play the same sound using a suitable 3D demo program (such as the Microsoft DirectSound3D Mixing test). Listen to the sound at several different positions. Note if and how the tone color of the sound compares to the "true" tone of the wave file when played with the Windows wave player, and the tone color of the sound when placed at different locations using one of the test programs. The tone of the sound should remain close to the "true" tone of the wave file as the position of the sound is changed. Any changes in tone color should still sound "natural."

Position:

Perhaps the most important characteristic to test is the quality of positioning effect; namely, how well does the technology create the appearance of a sound in 3D space when played back through an appropriate playback system. As mentioned above, this test is somewhat subjective in nature. One way to evaluate a particular system is in comparison to another, such as how good a "3D accelerator" sounds compared with the default software 3D audio effect provided by Microsoft (Windows) or another 3D audio accelerator products in static testing.

Trait	Listening Considerations
General Localization	How well does it position a sound in space? Can you close your eyes and point to the location of the sound emitter? Solutions should realistically position sound effects in 3D space, from both static sources and sources in motion.
Elevation	Listen to hear if sounds appear to be coming from above and below you. Try moving the sound up and down. Does it appear to go to the ceiling and beyond? To the floor? Is the transition smooth or does it jump? Try this with and without visual cues. Solutions should create sounds that smoothly travel up and down.

Noise, distortion and lack of or inaccurate distance behavior should also be listened for, and are negative artifacts of this process.

Listener Zone:

When listening to 3D audio over speakers there is always an optimal position for the listener. Sometimes called the "sweet spot," it represents the place for the listener's head with respect to the speakers that provides the most vivid 3D audio effect. The perceived size of the sweet spot may vary among different renderers, and will also be affected by your choice of speakers and their physical placement.

Trait	Listening Considerations
Sweet Spot – Width	Listen for the “sweetest” spot, and then see how far off axis you can be and still get the effect. Is the effect enjoyable (though not necessarily optimal) for multiple positions? Solutions should allow for any comfortable listening position.
Sweet Spot – Degradation	Move your head around to see if the image diminishes gracefully or abruptly. Do you hear a “phase-change” or “buzziness” from any listening position? The effect should smoothly degrade without dramatic drop-offs or artifacts.

For speaker playback, the 3D effect produced by sound cards is often expected to also provide an "out of speaker" effect. That is, the sound card should be able to produce the effect of a sound coming from a location that is not between the two speakers. For headphone playback, the sound card should be able to product an "out of head" effect.

Trait	Listening Considerations
Openness (“out-of-head”)	How close to your head are the sounds? Are they in your head, right next to your ear, or beyond your shoulders. This is particularly important for headphone-based listening, where “out-of-head” sounds are difficult to reproduce. A robust, airy solution will put sounds beyond your shoulders, because anything closer causes listener fatigue when used for extended periods of time.

Total Experience:

The total experience provided by IASIG Interactive 3D audio is the most important evaluation that can be made, and should be done using an application which allows the listener to move about in a 3D world and see and hear various objects. As objects move in 3D space with respect to the listener, the quality of the 3D audio effect can be determined as it would be heard in an actual application, complete with visual and movement cues.

When making your evaluation it is important to remember:

- Locational hearing is not as accurate as vision: while we can hear three-dimensionally (in the real world, or in a high-quality 3D audio simulation), we cannot pin-point locations down to the very exact position (unless our ears are trained very well, as is the case with non-sighted persons)
- Visual cues and interactivity matter: by far the best 3D audio results are achieved when 3D visual and 3D audio cues are synchronized, and when the listener can interact with the simulation in real-time.
- Playback environment matters: most positional 3D audio systems distinguish between headphone and stereo speaker playback.

You can expect the following results in a well performing system:

- Headphones: sounds are outside the head (externalization); with closed eyes the listener can clearly determine if a sound is left, right, up, down, front or back (front, up and down being slightly less pronounced than other positions).
- Stereo speakers: with listener seated in the sweet spot, sounds should extend far beyond physical speaker locations. Left/right positioning should wrap all around the listener, front cues should be very clear; back, up and down cues should be noticeable but diminished compared to headphones.
- Audio is rendered with proper Doppler and distance cues.
- Audio is synchronized with 3D visual objects (i.e. sounds are attached to flying pyramids)
- User can interact, and audio is adjusted in real-time (i.e. user can look and move around and audio reacts on the fly to user actions)

System Issues

The following two issues are not about the perception of 3D audio but can affect the quality of rendering and are important to the overall 3D experience. For correct 3D rendering, latency (time between issuing an instruction and performing it) must be low enough to allow for audio to be in synchronization with graphics. The CPU usage must also be low enough to prevent affecting the quality of visual or audio rendering while still handling audio tasks.

CHARACTERISTIC	EXPLANATION
Latency	Is sound properly synchronized with graphics
CPU Usage	What is the percentage of CPU required, per channel, comparable to other solutions.

Latency:

Latency refers to the lag in time between when positions or other updates are requested by the program and when they are actually heard. Latency is very difficult to determine objectively, since there is no way to measure it directly, and no way objective and absolute way to determine what is causing it.

However, subjective evaluation of audio/visual synchronization is still a good idea, and made possible with a simple test application. Simply moving a sound around the screen using the mouse allows subjective analysis of the system audio latency by judging how closely the sound follows the mouse

movements and visuals. The perceived location of the sound should appear to match the movements of the mouse.

If there is a noticeable lag or delay between the mouse movement and sound, the system is said to have a high latency. Conversely, if there is no apparent lag between the mouse movement and sound, the system is said to have low latency or no perceivable latency. (Latency on Windows systems can be checked by using DirectSound3D Mixing Test and moving the x, y, z position of the source or listener. Comparison of accelerated and non-accelerated DirectSound3D solutions can be done in this manner).

CPU Usage:

3D audio processing, whether in hardware or software, places a certain load on the CPU. If the load is excessive, it can slow down other operations, such as video frame rate. Therefore, a system evaluation should include some measurement of load, at least relative to other 3D audio solutions.

CPU load on Windows systems can be measured experimentally using either the Windows 95 “System Monitor”, or for a much more precise measurement, a program such as Intel's “Vtune”. Similar programs exist for other platforms.

Currently most available 3D audio solutions for the PC are hardware (or at least a combination of hardware and software), and designed to provide better 3D performance than Microsoft's own software-based 3D rendering system (DirectSound3D). Therefore, comparison of latency and CPU use against DirectSound3D is an appropriate task. As a basis for comparison, Microsoft reports that the default DirectSound3D software included with DirectX 5 requires approximately 6% per channel on a Pentium 90 (DirectSound documentation, dsound.doc, DirectX 5). Although the Microsoft default DS3D algorithm is linear (6% for 1 channel, 12 % for 2 channels, etc.), accelerated systems often are not 100% linear in their CPU utilization. For this reason, CPU consumption should be measured with several sounds playing concurrently, rather than measuring CPU usage for only a single sound and extrapolating.

IASIG Interactive 3D Audio (Level 1) Requirements

In order to be classified as an IASIG Interactive 3D Audio renderer, the following functions must be supported:

- **Playback of 8 simultaneous sources (16 streams are recommended) which support**

- a minimum sample rate of 22050Hz 16-bit output (44.1kHz recommended)
- object and listener 3D position (x, y, z)
- object and listener velocity
- listener orientation
- object orientation and radiation pattern

- **Rendering of the following effects**

- distance
- Doppler
- true 3D positions (x, y, z)
- radiation model

- **Rendering is all real-time and interactive without audible artifacts or latency**

8 Simultaneous Sources

Due to the interactive aspect of computer gaming, it is not enough to be able to render only the 3 or 4 dominant sounds of a scene, as is typically the limit in film sound tracks. Though it is unlikely that the listener will be able to hear and locate any more sounds than that, which ones will be required at any instant is probably impossible to predict in any well-written interactive game. Therefore the system must be rendering multiple sounds at all times, even if they are currently playing at lower volumes than other sounds and therefore momentarily inaudible. The more sources that can be rendered at once, the better the interactive rendering engine can perform the illusion of a realistic sound environment and the more layers of sound the closer it approaches realism.

X, Y, Z Coordinates

The audio environment should be viewed from the perspective of the listener and corresponds to the view being depicted on the computer screen. The location of the user's head and their head's orientation are key to locating the ears, needed for proper rendering of the audio content. If the listener is stationary then the movement of the sources provide key information to be tracked. If both the listener and the sources are moving then the relative distance from the listener and each source must be calculated. The positions are recorded using a three dimensional Cartesian coordinate (x, y, z) which usually corresponds to the graphical data being displayed.

Orientation and Radiation

Sources can have a directional sound source instead of being omni-directional, and act like carefully constructed loud speakers, where the volume drops when one stands at the side as compared to standing directly in front.

Distance

Distance models must support sound attenuation and velocity models properly. The environment has a maximum distance where sound volume drops to zero. Simple distance models use attenuation as a listener moves away from a sound source, the sound gets quieter. The opposite is also true: the sound gets louder as you approach the source. The basic formula comes from the observation of the surface of a sphere which is the radiation pattern of a omni-directional sound source in space (the inverse square law). When the distance is doubled from the source and listener it should be attenuated or dropped 6dB. Each doubling drops another 6dB until the sound cannot be heard.

This method is to be used under anechoic conditions and does not include early reflections which will be covered in level two of the specification. When attenuation is used as the sole mechanism for distance modeling it is less effective when reverberation is used. Reverberation in general is a level two issue. The assumption for level one is to use the simplified anechoic chamber models for 3D audio rendering and to exclude concurrent reverb. The maximum distance is the point where the source has moved far enough away that further reductions in sound levels are not necessary.

Velocity

Objects and listeners in motion are required to use a velocity model that includes a three dimensional direction and speed. When sources move relative to the listener, a Doppler effects is applied by the renderer according to the velocity data. The Doppler shift effect is layered on top of the distance effect and the frequency is shifted based on the relative velocity of the listener and the source. Movies often exaggerate this effect to provide more drama.

Velocity is relative to each sound source and the combination of the source movement and listener movement. The frequency shift is proportional to the simulated radial velocity of the source relative to the listener (Doppler effect). Due to the interactive content, the listener location, the source location and the velocity must be kept up to date for proper effects rendering.

Future Directions (Level Two)

The current generation of interactive positional 3D audio systems allow for 3D placement of listeners and sound sources. This represents a dramatic improvement of the one-dimensional placement offered by stereo audio. However, additional improvements are still possible.

For example, surfaces such as walls in the 3D environment can have a significant impact on audio realism. Walls can reflect or occlude sound waves as they travel from sound source to the listeners ears. As a result, sounds change dramatically as they disappear behind walls, appear through an open window, or move from an outside space into a room.

Future generations of interactive 3D audio systems are expected to be able to render the dozens of 3D sound streams that are necessary for the modeling of the entire audio environment.