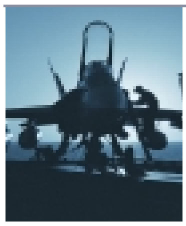


Part Two: 3D Audio Relieves Military Information Overload

An in-depth look at 3D audio and how it can improve military applications.

From the Editor files

In part one of this article, the military benefits of 3D audio were described. For example, using 3D audio to increase both situational awareness and information transmission/perception can improve communication systems in air traffic control towers and mobile dismounted infantry, simulation-based training systems, and cockpit/vehicle operator information displays. In order to fully achieve these improvements, particular attributes of 3D audio are needed. Here, features such as head-tracking, high precision and “own ear” adjustments are described.

Editor

Brook L. Eaton, Product Marketing Manager, AuSIM

3D audio can supplement visual inputs, or can by itself form “audio displays”. But not all 3D audio technology is created equal. Quality 3D audio with full interactivity, high precision and versatility removes the confusion in audio and adds back organization. With this organization, human perception is improved and more information can be offloaded from the visual domain to be absorbed aurally. The end result: No information is lost or overloaded; some is simply diverted to the aural display and overall information transmission is more effective.

3D Audio Characteristics

In order to clarify what particular features are needed from 3D audio, it’s important to clearly define it. Although there is a plethora of technology commonly referred to as “3D audio,” much of this technology does not fit a simple definition. 3D audio should be simply defined as the sound heard in a realistic 3D environment. For example, you can distinguish between sounds heard in front of you, behind you, above you and below you.

This simple definition sets a standard for what people should expect from 3D audio technology. However, there are features not contained in this simple definition that are critical to delivering sound that enables even more information transmission. For example, it is implied that a listener in the 3D environment hears the sound with his or her ears, yet everyone hears differently due to their physical characteristics, and a “one size fits all” solution just won’t work in precision military applications.

Since most military applications that need sound are interactive, the 3D audio must also be interactive and thus must support a user (listener) who can interact in an environment by moving their head without the aural world around them shifting with their movements. Instead, objects should remain virtually stationary regardless of position, direction, tilt or movement of the head. This requires positional information about the listener’s head to compute what the listener should hear. This ability to interact is critical in military applications where a soldier cannot be strapped down into one position and told not to move his head (Figure 1a and 1b).

Another quality that distinguishes different types of 3D audio is the level of overall precision. Precision in 3D audio can best be described by the accuracy with which a listener is able to localize sound. Since humans can localize to within two degrees of the actual location of a sound, delivering sound with this degree of precision is necessary. If you do not utilize 3D audio with this level of precision, the results can be drastic. The sound can be completely confusing to a listener and cause them to react in an inappropriate way.

The final characteristics that are important for 3D audio are on the system level. First, it should be noted that most 3D audio technologies are not versatile or scalable enough to be integrated into many applications. Simply put, if you cannot use the technology, you cannot gain the benefits. Scalable and versatile technology will easily integrate into every application from small wearable communication and navigation computers to large, complex and stationary networked tank battalion simulators.

Shortcomings

Most 3D audio technologies use computationally cheap methods of generating 3D sounds that simply do not come close to accurately modeling sound. They often modify the audio with approximate delays, frequency shifts and gain losses in an attempt to deliver the illusion of sound emanating from around a listener. In general, however, there are several problems with this approach. The modifications are approximations and result in a lack of 3D sound positioning accuracy. Surrounding environments are not well accounted for and many ways in which audio is uniquely shaped by the path are ignored (reflections, refraction, diffusion, diffraction).

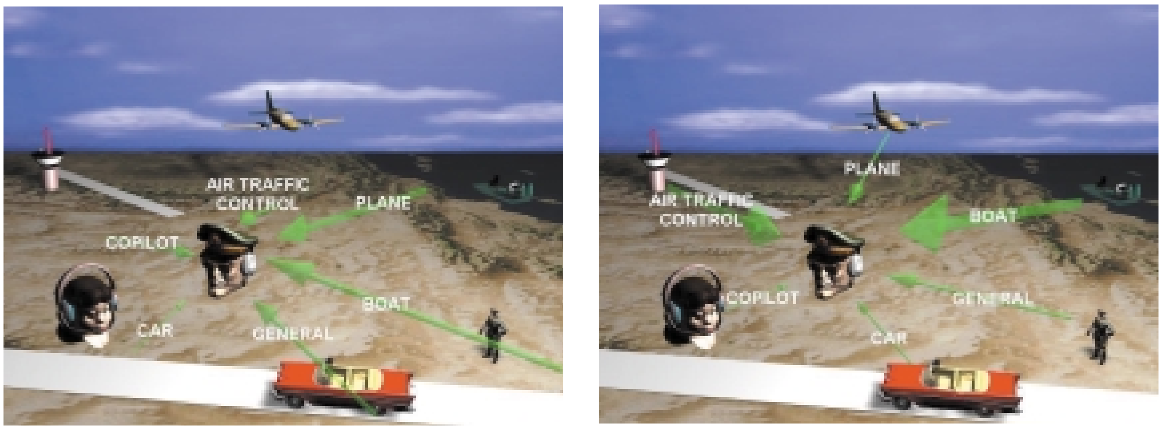


Figure 1a

Figure 1b

Figure 1a: With his head turned to the right, the pilot on the *left* illustrates the confusing nature of non-interactive 3D audio. But in Figure 1b with fully interactive 3D audio, the pilot on the *right* is able to quickly and accurately locate sounds in his environment even while moving his head.

For example, sounds coming from the back hemisphere are often simulated by cutting off the high frequencies; no sound propagation simulation is computed at all. While sounds emanating from the back hemisphere often do have slight attenuation in the higher frequencies, there is still vital information that must be simulated and presented in the higher frequencies, particularly for generating cues for the perception of sounds at varying heights. In general, the value of 3D audio is greatly diminished by inaccurate technologies and the resulting sound is often unnatural, difficult to localize, and lacking in spatial correlation and organization, leading to confusion.

In many cases, a lack of precision causes many 3D audio technologies to deliver sound that is less helpful to a listener than simple stereo or mono sound because it confuses them. A lack of spatial organization and seemingly unnatural sound can be perceived entirely incorrectly. Multiple sounds may be perceived as emanating from inside the head and come across as completely jumbled. Often, visuals help reinforce the perceived location of an object generating sound, but when not precisely spatially correlated the misleading aural cues are worse than none at all.

In mission-critical situations, the sound positioning must be exact. For example, when a fighter pilot is being approached by an aircraft from behind and slightly above, he needs to know this exact information. If his technology is limited to positioning audio without the correct elevation and places the sound directly behind him, he could make the mistake of maneuvering his plane directly into the path of an oncoming aircraft. Precise sound positioning is required to prevent confusion.

Another important shortcoming of many 3D audio technologies stems from the fact that there are varying levels of interactivity. Elementary systems assume fixed positions of both the listener and sounds. Others assume a fixed head position but simulate the motion of sounds. Fully interactive technologies account for both the movements of the head and sounds, alleviating a problem referred to as the "frozen perspective."

For example, in a search and rescue mission where a wounded soldier (relaying his position via GPS) is being sought out in dense jungle, a sound beacon can serve to guide a searcher to the wounded soldier. Without complete interactivity, as soon as the searcher moves his position the beacon would no longer represent the location of the wounded soldier. An added bonus with this technology: the searcher does not have to stop walking to study a map and compare GPS readings.

Finally, many existing systems capable of overcoming these shortcomings still remain either too costly, too voluminous or cannot scale and integrate into varying applications.

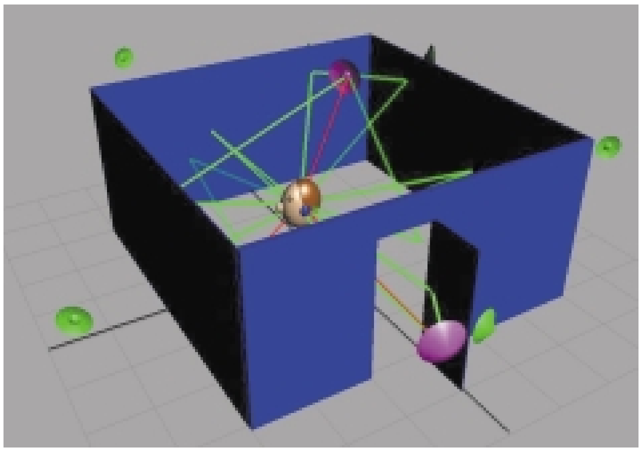


Figure 2

Illustrated here are the minimal necessary sound paths and reflections calculated during audio propagation simulation in a 4-walled room with a door. The red line shows the direct sound path from the purple cones to the listener and the green lines represent first order reflections—the perceived location of the sound due to these reflections are indicated with green cones.

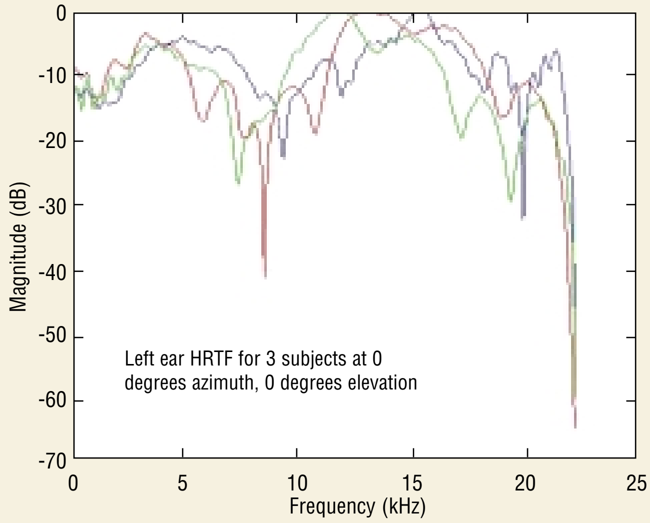
How We Hear Spatially

Agnieszka Jost, Acoustics Researcher, AuSIM

When a sound event occurs in a three-dimensional space, the sound waves that reach the listener's eardrums are affected not only by the air through which they travel but also by the interaction of the sound waves with the listener's head, torso and pinnae (outer ear). The combination of these effects, or characterizations, can be captured as a Head-Related Transfer Function (HRTF).

A sound that is equidistant from the two ears (on the *median* plane) will have very similar HRTFs—although not identical due to the asymmetries of the head. When a sound occurs to the side of a listener, the HRTF of each ear has very different characteristics. First, the sound reaching the ear that is closest to the sound source, the *ipsilateral* ear, will have a greater sound intensity than the *contralateral* ear (the ear farthest from the source). Second, the sound will arrive earlier at the ipsilateral ear. These differences between the two ears are defined as the Interaural Intensity Difference (IID) and Interaural Time Difference (ITD). A third characteristic arises due to the reflection and absorption of sound waves on the torso, head and pinnae. The upper body acts as a filter with unique frequency characteristics for each direction, with the pinnae having the primary role.

Because every individual's upper body and pinnae are shaped differently, a comparison of HRTFs between individuals reveals great spectral differences in both general shape and details. Although, with training, it is possible to learn to hear using somebody else's ears, we will get the most realistic listening experience with our own ears. There are technologies available today that measure a person's HRTFs in order to apply these characteristics to simulating 3D sound (see Figure).



Figure

Reflections, refraction, diffraction and absorption of sound due to a listener's unique head, outer ear and shoulder geometry cause each individual to hear in a different and personal way. This graph illustrates the differences heard by three individuals when listening to the same sound emanating from the same location.

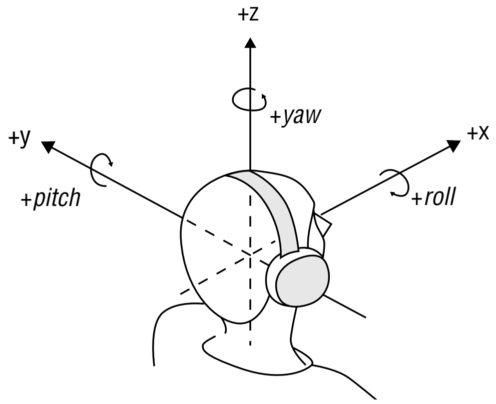


Figure 3

This diagram shows all 6 degrees of freedom that the head can move.

Solutions

There are ways to overcome the poor performance of many low-end, low-cost 3D technologies, while still maximizing information transmission in an aural display. First, environments (room or area) must be created in the virtual domain that closely match the intended or realistic circumstance. Second, the sound propagation away from the source, through the environment, and into the listener's ears must be correctly simulated to generate accurate perceptual spatial cues. Third, the sound simulation must happen in real time so as to account for both movements of the listener's head and sounds in the 3D world (Figure 2).

In order to deliver the level of precision required to avoid mixing up sounds, it is necessary to take a physics-based approach and accurately model the complex, and spectrally unique, filtering characteristics that occur in sounds emanating from different locations. Not only does this solve the really necessary problem of precisely locating sounds, but it also has the added benefit of realistically depicting sounds.

For example, the radiation pattern of a sound source is modeled differently from one sound to another. While a human voice projects more sound directly in front of them than behind them, the sound of a gunshot radiates more equally in all directions (acting like a point source). Accounting for these radiation patterns helps determine the path that the audio travels in order to reach the listener. In the environment, elements such as humidity affect sound propagation because it can vary the speed of sound and attenuates it as it travels through the environment.

When the sound finally reaches the listener, the person's physical characteristics take over. While accounting for radiation patterns and environment shaping is necessary for completely accurate sound modeling, the ability of a person to localize sound is directly affected by their physical features. Sound is reflected off the shoulders, the cheeks, the back of the ear, and so on and based on its incoming direction is uniquely shaped before entering the listener's ear canals.

In order to deliver the most accurate positional sound possible, these physical traits should be accounted for in the computation. Since it is possible to measure the filtering characteristics of each individual, these measurements can be incorporated into the model to ensure each person hears and localizes sound as they would naturally (see sidebar: How We Hear Spatially).

AuSIM 3D Technology

3D audio has been around for a while but its usefulness has increased with recent advancements. AuSIM3D is an audio simulation technology that generates extremely accurate, completely interactive and cost effective 3D audio. Perceptual location cues are calculated by physically modeling the three areas of sound propagation—source propagation, environment propagation and listener filtering. The models are updated in real time and work well with position tracking devices for head and object tracking.

The physical delivery forms of advanced 3D audio technologies have changed dramatically in the last ten years along with the quality, size and cost. The latest 3D audio technology from AuSIM is available in two COTS forms. Hardware-based client-server systems that offload all audio processing from a host typically serve as a scalable (one size fits all) form for use as development or runtime systems.

The second COTS form is a cost-effective software-based 3D audio engine that can be installed on existing computing hardware and run either directly on a host or in a client-server fashion. The versatility of this second software form proves key for integration into a plethora of applications. It enables the technology to exist in peripheral boxes, integrated subsystems, standalone systems, end-user applications, application plug-ins, SDK libraries and software resources (DLLs) with an application-programming interface (API).

In addition to using a physics-based approach to achieve precision, it is important to provide interactivity. The best way to provide interaction capability hinges on incorporating head movement with a positional tracking device placed somewhere on the listener's head. Since natural movements of the head are to be tracked, it is necessary to not only discern between up down and side to side movement of the head, but also rotation and tilt. Additionally, head movement turns out to be a key component to human localization.

The ability to turn one's head helps the brain fine tune sound source locations in order to disambiguate up/down and front/back sounds whose locations are often confused. Effectively, localization is improved when head movement is accounted for by using a head-tracking device. Capturing complete head movement provides the level of interactivity that is required by most applications (Figure 3).

Fortunately the necessary level of precision and interactivity is available in a form that is scalable and versatile enough to be cost-effectively integrated into any application. 3D audio technology from companies like AuSIM meets these requirements (see sidebar: AuSIM Technology). ▮▮

AuSIM Inc.
Mountain View, CA.
(650) 322-8746.
[www.ausim3d.com].