Razer Maelstrom Audio Engine

The Evolution of Audio

Unlike video technologies, which have seen a stream of new innovations over the years (color screens, ever improving resolutions, brighter LCD and plasma screens, various 3D vision enhancements for home and cinema, etc), stereo has been, since 1931, the predominant technology used for audio reproduction.

1877 – Monophonic sound reproduction is created and the phonograph is born.

1931 – Stereo sound is invented. By using balanced volume between right and left, it is possible to give the impression of a sound source's movement on a line between the two speakers. In effect, this is a one dimensional technology (one axis).

1940 – First documented use of surround sound using a multichannel audio application comprising of three audio channels.

1985 – To render more realistic audio for cinemas, multichannel audio standards are created. This allows for the reproduction of ambient sounds, such as rain or crowd noises, to the rear of the listener.

However, the multichannel audio technology remains essentially the same as that of stereo: placing a sound between 2 speakers, and adding more and more speakers (5.1, 7.1, and even up to 12.1 today) which will add one more axis in which you can hear sound.

Hearing in 2D

Our perception of sound is a complex sense that actively seeks to detect the provenance of audio events that happen in the world around us. Since we are able to easily detect both the direction that a sound comes from, as well as its distance in relation to our position, we often use our sense of hearing as a guide to protect us from danger.

However, recent years have seen a lack of improvements in audio technology that has led to the neglect of its proper use in gaming. Traditional stereo only allows the user to locate an event that occurs to their left or right. But with the advent of 3D gaming, gameplay is no longer limited to side-to-side scrolling and jumping from one's left or right.

There is a disconnect between what a human's perception of audio is capable of and what traditional stereo or multichannel speakers are actually capable of. Because of this, the potential for creating a natural and true 3D sound system is enormous, as we are already trained to use audio cues and actively seek such cues to better orient themselves to the environment around us.

Compare what you can hear when at a live concert, and listening to a recording of the same concert. They sound nothing alike because traditional stereo flattens the sound to a line.

HRTF: Virtualized Surround Sound

The first virtualized surround sound technologies appeared in the 1980's. This technology is based off of sound processing techniques called Head Related Transfer Function ("HRTF") and it aims to reproduce and spatialize any sound, virtually anywhere.

HRTF uses the way our brain decodes the spatial localization of sounds and creates positional sound cues to recreate that effect.

For example, when a sound comes from the left of someone, the sound first hits their left ear and then their right ear (*see fig. 1*). This time difference, called ITD (Inter-aural Time Difference) is used by the brain to decode the position of a sound source in the horizontal plane. This information is also combined with ILD (Inter-aural Level Difference) between signals coming to the right or left ear due to the masking of the head (diffraction of the sound caused from your head absorbing the sound) to determine the location of a sound source.

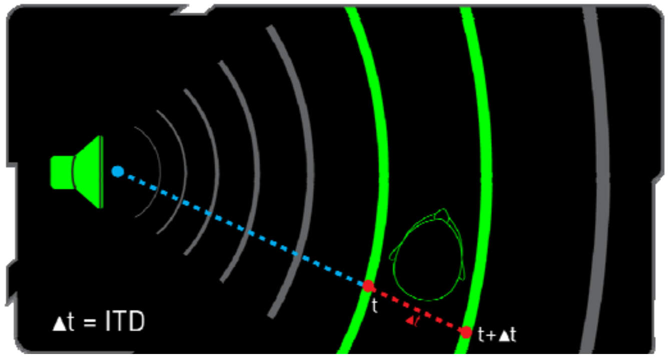

Figure 1: The coefficient t represents the amount of time required to reach the users left ear. The difference in time to reach the right ear is represented in red.

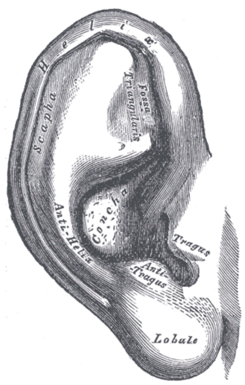

Figure 2: The pinnae is the outer part of the ear. Comprised of cartilage and skin, it directly affects the way a sound is perceived.

To explain our ability to locate sound above or under (elevation degree), the brain decodes small variations in the sound due to diffraction on the ear's pinnae *(see fig. 2)*. When a sound hits your ear from a certain elevation, it is changed by multiple reverberations on the pinnae before entering into your ear canal. These changes are different for different elevations. The brain will then decode these small changes to decode spatial cues based on elevation. All HRTF technologies are based on mimicking these spatial cues by altering sounds with filters, before broadcasting it to your ears.

Taking these principals, a mannequin or a real person's head, with microphones fitted into each ear canal, is used to measure and compare the difference between recordings of the sound source at a certain spatial position, and the recording of the sound when it finally reaches the inside of the ear canal. This direct comparison gives two filters (left and right) that can reproduce the distortion of the sounds to recreate a spatial feeling.

By using these techniques and the filters derived, HRTF technology is able to simulate a 3D sound field using just stereo headphones or stereo speakers, without the need for discrete multiple speaker systems.

Traditional problems with HRTF

Due to physiological features inherent in HRTF, there are a number of traditional problems that occur when 3D sound is generated by HRTF.

Front/Back Audio Imaging

Current HRTF technologies suffer from poor reproduction or imaging of sounds from the front and the back. For example, if a recording was positioned as if it was coming in front of the listener, the traditional HRTF algorithms would inadvertently place it at the rear of the listener. Accordingly, in gameplay, shots fired in front of a gamer may be confused as coming from the rear with traditional HRTF protocols.

Selective Audience

Traditional HRTF protocols suffer from low success rates with most listeners. A user typically has a really low chance that HRTF will work for him as intended. Most HRTF protocols today have at most a 20% chance of success rate where most users are likely not to notice any difference, or worse, actively dislike the 3D sound effect reproduced by HRTF and switch off the 3D sound function.

Loss of Fidelity

Traditional HRTF algorithms also suffer from a loss of fidelity post-processing where natural or original sound cues are lost or muted. While traditional HRTF algorithms may bring out positional audio cues, it is almost as likely to result in significant loss in fidelity resulting in a poorer gaming audio environment for listeners.

Poor Externalization of Sources Resulting in Listener Fatigue

HRTF generated sounds are typically perceived to be very close to the head of the listener as it is generally difficult to reproduce the effect of distant sounds as opposed to positional audio when HRTF algorithms are employed. Accordingly, HRTF can result in listener fatigue due to the “in-head” sound perception.

Elevation/z-Axis Sounds

HRTF is based on the psychoacoustic sound localization methods which provide only a two-dimensional sound field with headphones. Accordingly, sounds on the z-axis or on an elevated or lower plane are difficult to reproduce with traditional HRTF algorithms.

Current Generation HRTF Technologies

There are many in the audio industry that have used the theories of HRTF to create technologies that can virtualize surround sound.

Some providers of virtualized surround sound rely solely on the principals found in HRTF, and will use filters on sounds that alter the phase and gain of those sounds to give positional cues. This is a very rudimentary process and does not work on most people as certain sounds simply are louder or softer and merely seem to echo in each ear.

Others use FFT (Fast Fourier transform) convolution to process a number of variables, such as the phase, gain, as well as refractions of sounds off the user’s head, pinnae and body. This is a more accurate method, but many such technologies provide sound filters that are short and can only give a limited number of positional cues before the DSP has to begin processing other new sounds.

Both these methods will give sounds with positional cues that the user’s brain can use to determine the position of a sound, but are associated with DSPs that can only process algorithms and variables at 34 MIPS (Millions of Instructions per Second).

Most of the current generation HRTF technologies suffer from the traditional problems cited above. Accordingly, for most gamers, the current virtual surround sound that is generated in current generation HRTF technologies is not suitable for gaming, or to provide accurate surround sound environments for gaming.

Razer Maelstrom Audio Engine - The Next Generation HRTF Technology

The Razer Maelstrom Audio Engine originated as a military grade audio technology that was originally developed for fighter pilots that needed precise audio warnings for incoming missiles. The Razer Maelstrom Audio Engine is based on the same platform, but developed specifically for gaming. The Razer Maelstrom engine was also developed to address the traditional problems with HRTF algorithms so that a more convincing surround sound environment can be reproduced.

The Razer Maelstrom Audio Engine is the result of thousands of hours of audio engineering, testing and validation to create a state of the art surround sound platform based on next generation HRTF technologies.

The processing done by Razer Maelstrom algorithms is significantly more complex than traditional HRTF methods. The position of the speakers in the headset has been determined by long hours of processing, using complex mathematical and physical theories. Meanwhile, the phase, gain, refractions of sounds off the user's head, pinnae, body, have all been taken into account, but so have other factors, such as the distance of the sound from the user, reverberation, and other effects (*see fig. 3*).

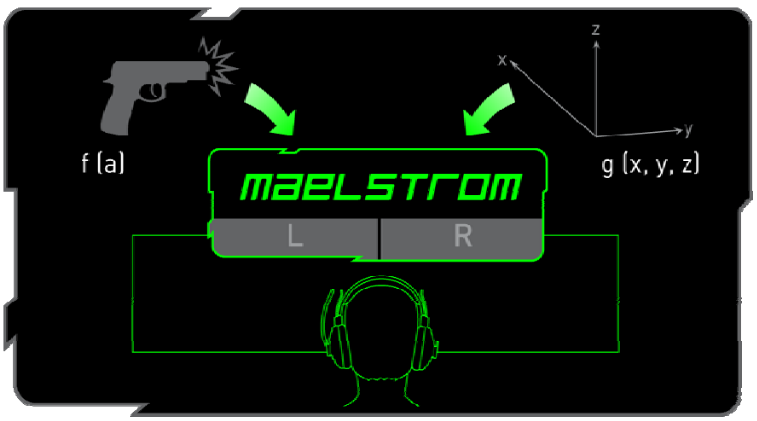

Figure 3: In addition to normal factors taken into account by traditional HRTF technologies, the Razer Maelstrom will take the original sound source, as well as the exact coordinates of where the sound occurred, and add positional cues that are processed into sounds that can be played in the left and right channels of a headset.

When the sound is replayed in the Razer Megalodon headset, the proper left and right sounds are conveyed, and the user will be able to accurately determine the position of a sound.

The enhanced processing power of the Razer Maelstrom engine results in better front/back audio imaging as compared to traditional HRTF methods. Further, the additional processing also provides for a more accurate surround sound imaging environment. The Razer Maelstrom engine also ensures that there is minimal loss of audio fidelity post processing so that gaming environments are kept to the original intent of the game developers. Likewise, the externalization of the audio cues is also enhanced with the Razer Maelstrom advanced HRTF algorithms.

Additionally, the Razer Maelstrom can virtualize the sound source and process the sound dynamically against thousands of virtualized ear pinnae to ensure that the resulting positional cues work for more of the population. This is important because the resulting positional cues of traditional HRTF technologies are based off of a single reference recording of a mannequin's head, and is the primary reason why traditional HRTF technologies are unconvincing for most of the population.

With the need for complex algorithms to be processed on the fly, one of the key aspects of the Razer Maelstrom engine is the use of a separate audio processing unit that offloads most of the processing required. This advanced processing method uses a separate, cutting edge DSP that can process up to 500 MIPs, or 400 to 800% more data than other conventional DSPs, and results in a phenomenal performance upgrade over and above current generation HRTF methods. Further, it also enhances gameplay as it offloads the processing required from the CPU to the Razer Maelstrom audio processing unit.

Current generation sound virtualization models are unable to provide for true to life positional audio in gaming. Further, traditional HRTF technologies have also relied on processing to be offloaded on the CPU, thereby creating poorer performance in-game.

The Razer Maelstrom engine was developed specifically to address the issues above in games and has been extensively tested in-game, as well as together with gamers. The Razer Maelstrom engine is the next generation virtual sound engine that provides true-to-life 7.1 Surround Sound Gaming without the need for software drivers, which results in enhanced processing – a boon for demanding games today.

With thousands of audio models created to ensure that the optimal audio environments are created in-game, the Razer Maelstrom Audio Engine is by far the most advanced audio engine system created for gamers, by gamers.

The Razer Megalodon 7.1 Surround Sound Gaming Headset

The Razer Megalodon is a 7.1 virtualized surround sound gaming headset designed specifically to incorporate the Razer Maelstrom Audio Engine, to make it the definitive gaming headset for gamers today. The Razer Megalodon is by far the most advanced gaming audio headset of its kind, utilizing next generation HRTF technologies and a superb DSP with unheard of processing power.

Because ear pinnae are incredibly complex and unique to each user, we decided to position speakers around the ear, generating a sound field that will be naturally diffracted by your own ears-drum, capitalizing on the Razer Maelstrom's extensive virtualization of pinnae. This close-to-ears sound field is the same for everyone, but the effect on your ear drums is more personal and will more accurately match your own 3D audio experience in the real world (*See fig. 4*).

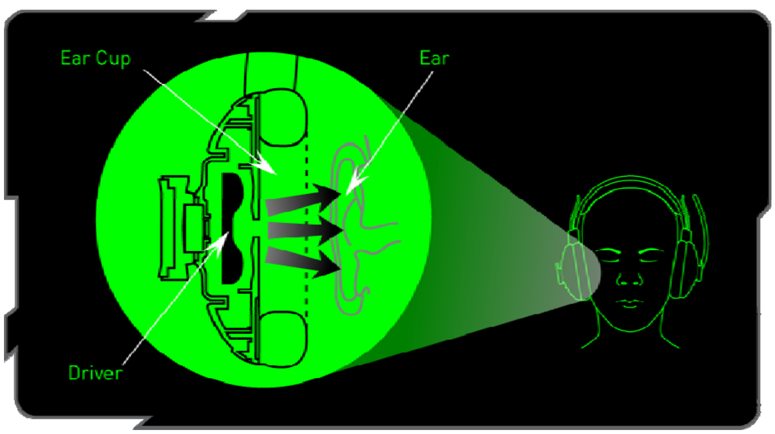

Figure 4 Megalodon headset speakers are positioned close to the ear, so the sound field is personal and diffracted directly by the ear drum.

Thus, the precise placement of speakers in the Razer Megalodon headset and the advanced sound processing techniques of the Razer Maelstrom Audio Engine allow for:

- More externalization
- More focused sound
- A more precise virtual sound image

The Razer Megalodon control pod also contains the audio processing unit that processes all the complex algorithms described above that make up the Razer Maelstrom engine. This external audio processing unit incorporates the next generation DSP that allows it to process up to 800% more information than current generation DSPs.

The final effect on the sound precision for the Razer Megalodon, and on all traditional HRTF problems, is outstanding. Save in respect of the z-axis/elevation problem, the Razer Maelstrom engine definitively addresses most of the problems traditionally associated with HRTF technologies. Suffice to say, Razer's audio engineers are actively pursuing true 3D audio imaging as 360 degree imaging is now finally at hand.

CONCLUSION

With the Razer Megalodon headset, Razer intends to demonstrate how the power of the Razer Maelstrom Audio Engine and precise placement of speakers can effectively reproduce accurate virtualized surround sound better than any other HRTF technologies currently on the market.

By processing up to 800% more information in the Razer Maelstrom Audio Engine than comparable traditional DSPs, the Razer Megalodon offers more positional sound cues that allow gamers to better visualize the entire virtual surround sound image. With these cues, gamers can easily put each and every sound in its place and will always know where a gunshot is coming from.

Razer also intends to address all problems inherent in traditional HRTF technologies in the future. Eventually showing how the Razer Maelstrom Audio Engine and precise speaker positioning in the Razer Megalodon will create a sound solution that truly allows gamers to hear in three dimensions, so no sound is ever left unheard again.