HRTF 3D Positional Audio Technology Overview

1. Preface.
2. The difference between HRTF 3D Positional Audio and 3D Surround.
3. HRTF 3D positional audio technology.
4. How to measure HRTF library?
5. How to synthesize 3D positional audio effect by means of HRTF library?
6. Is it applicable to have two speakers replace a pair of headphones in terms of 3D listening?
7. What is the limit of sweet spot?
8. HRTF 3D positional audio in application.
9. One present mainstream 3D positional audio source and its bloom in Taiwan: CRL 3D Audio.
10. Conclusion.

1.Preface

You must have shared the same experience: in a hot summer night, irritating mosquitoes are flying to and fro right besides your ears. At this moment, nothing is more urgent than getting rid of the nuisance from you, isn't it? To do this, you don't have to keep your eyes wide open in order to find out where it is. Just depend on the keen sense of hearing, you can likewise locate the mosquito's position, giving the mosquito a precise and fatal attack in total darkness.

Standing on a street corner, we sometimes can see blind people walk along the crowded street all by themselves. It is considered unsafe for blind people to do so; however, their act exemplifies one of the amazing body endowments--hearing, which when being exercised to a full scale, can give human beings a big helping hand.

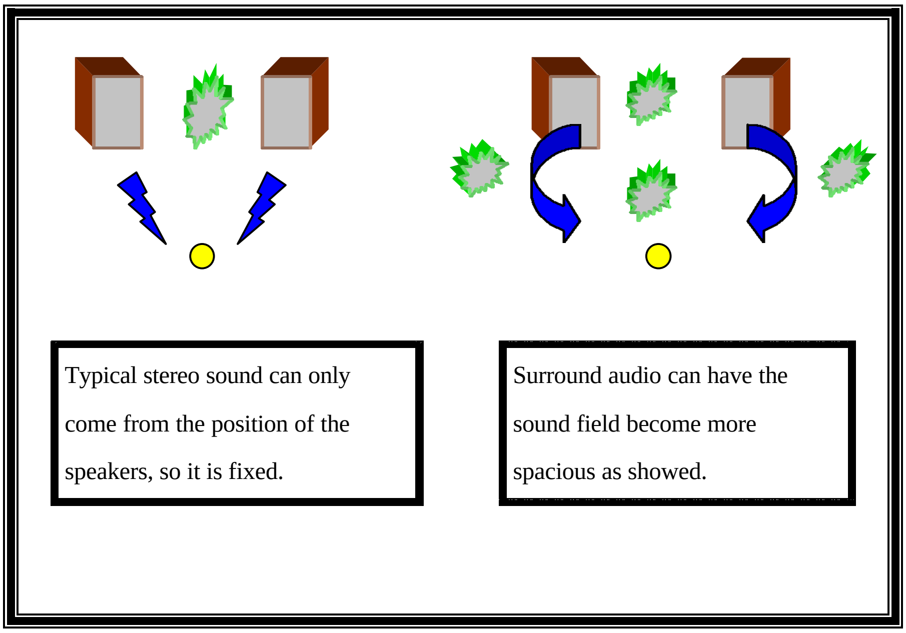
In fact, not only our eyes can differentiate objects located in three dimensions, but our ears as well. In our daily life, seeing and hearing complement each other. Therefore, if one day we are deprived of our eyesight, our sense of hearing will automatically amend that gap. Try to close your eyes and feel, don't you think your sense of hearing is becoming sharper than before?

It has been years since human beings started doing research in 3D graphics, therefore, all the relevant technologies have reached their maturity. It is not difficult to reproduce things existing in the real and dimensional world, and afterward apply the reproduction onto the PC. Nevertheless, we are no longer satisfied merely with the visual reproduction. What concerns us right now is: how to utilize the imitation-reproduction technology in an audio level? In other words, what kind of audio technology can help our brain feel the three dimensional effect? The answer is--HRTF(Head Related Transfer Function) 3D Positional Audio.

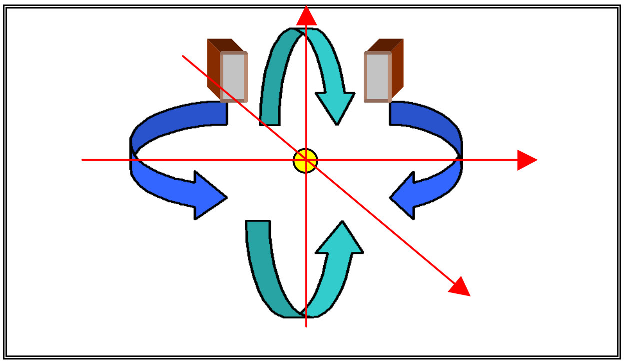
2. The Difference between HRTF 3D Positional Audio and 3D Surround

In 1996, there came up with a new term describing a new feature of PC sound cards--3D sound, which claimed to have 3D sound effect only by adding an additional IC on the circuit. This seemingly magical effect is in fact another kind of surround presentation, not a genuine manifestation of 3D audio effect. Basically speaking, this technology utilizes some simple delayed circuits and filters. Besides, it mixes the left channel with the right one. Hence, human brain got the impression that the audio field has become more spacious. Since it does not require any software to activate the 3D audio effect, it is very convenient for people to edit and compile CD music in an advanced and professional level. However, this technology has its drawback. That is, the processing circuit mixes sound sources coming from the left and the right channels. In that regard, the original source is distorted, and such a distortion can not be tolerated by professional stereo lovers.

Typical Surround sound technology can not exhibit sound sources coming from above or below, nor can it by means of software locate the sound sources in a preferred position in the virtual world, or calculate the comparative sound relationship in 3D games in real-time. In a strict sense, this kind of technology can not be called 3D sound manifestation; it can only be called 2.5D, to the most, because it improves the limitation of enjoying the sound only from the comparative direction of the speakers.(See the picture below).

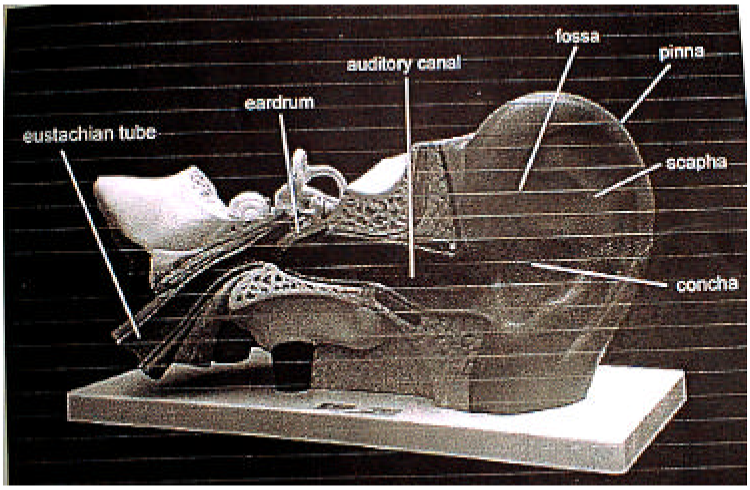


Things will be different when HRTF 3D technology is applied. HRTF uses the frequency vibration perceived by human ears and brain to synthesize 3 dimensional audio effect, so via high speed DSP calculation, HRTF can in real-time process the sound sources in a virtual world to 3D sound render engine. When the sound chip calculates the 3D-included sound wave, by way of headphones, human brain can feel a real sense of location, such as sounds coming from front/back, above/below, and anywhere in a three-dimensional space.



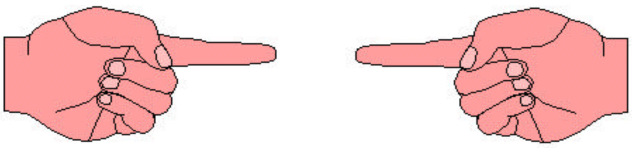
3. HRTF 3D Positional Audio Technology

To understand the 3D positional technology, first, it is necessary to know how human brain uses ears to tell the position of the sound source. Let us begin with the ear structure:



The Human ear can be divided into three parts: outer ear(pinna), mid ear(auditory canal), and inner ear(eardrum). When sounds are perceived by the pinna, they will be transmitted to eardrum through the auditory canal. At this time, rear eardrum will transformed the mechanic energy into biological and electronic energy, which will then be transmitted to human brain via the nerve system. All of the nerve cells in our brain have accumulated whatever audio data coming to us since the day we were born, so those cells are capable of distinguishing the difference between the contents as well as the directions of sound wave in a very quick speed. It only takes one ear to analyze the contents of the sound wave, yet it takes both ears to analyze the directions of the sound wave. This holds the same for seeing, as one eye can not tell the multi-facetedness of the object.

Try to close one of your eyes, and point out both of your index fingers. Intentionally miss some distance, and reach your left index finger to the right one in a slow movement:

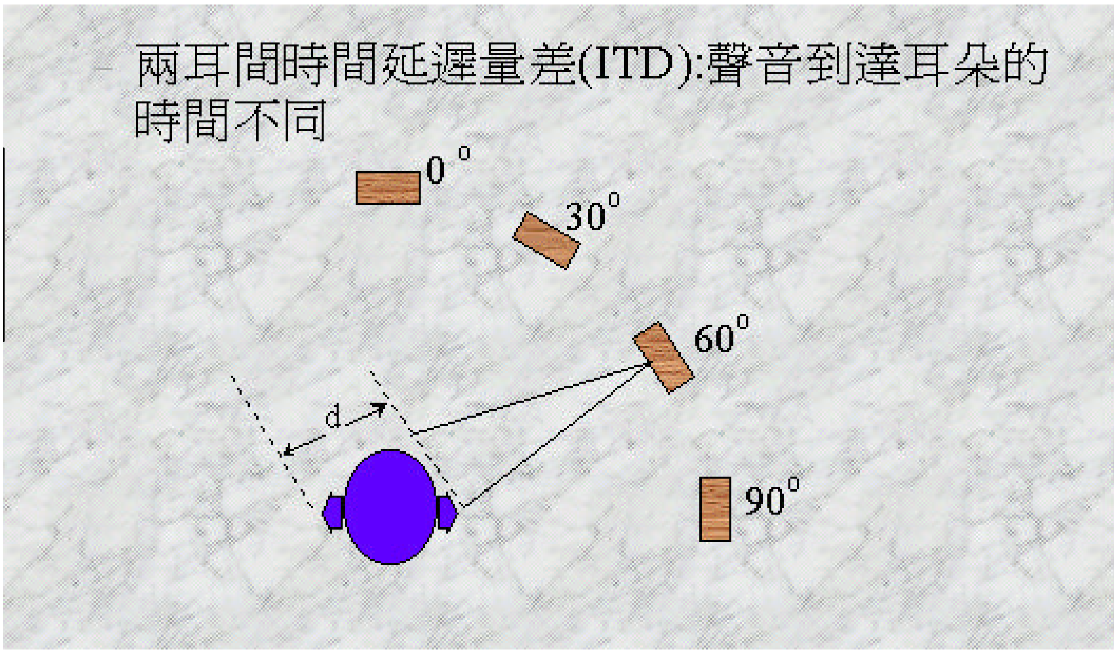


Did you left index fingertip reach the right one? Isn't it amazing that you think you can have both fingertips meet without any difficulties, yet they miss with each other by a "finger" distance! In fact, seeing with one eye won't prevent you from judging what stretch before you are fingers, but seeing with only one eye deprives you of the sense of dimension. Right now, you can open your other eye, in case you are exhausted. Whereas seeing can differentiate objects in different position, hearing can, too. So, there comes one question: what parts of our ears are in charge of locating the sound source?

Elements Regarding Telling the Location by the Sound Source

1.ITD(Inter Aural Time Delay):

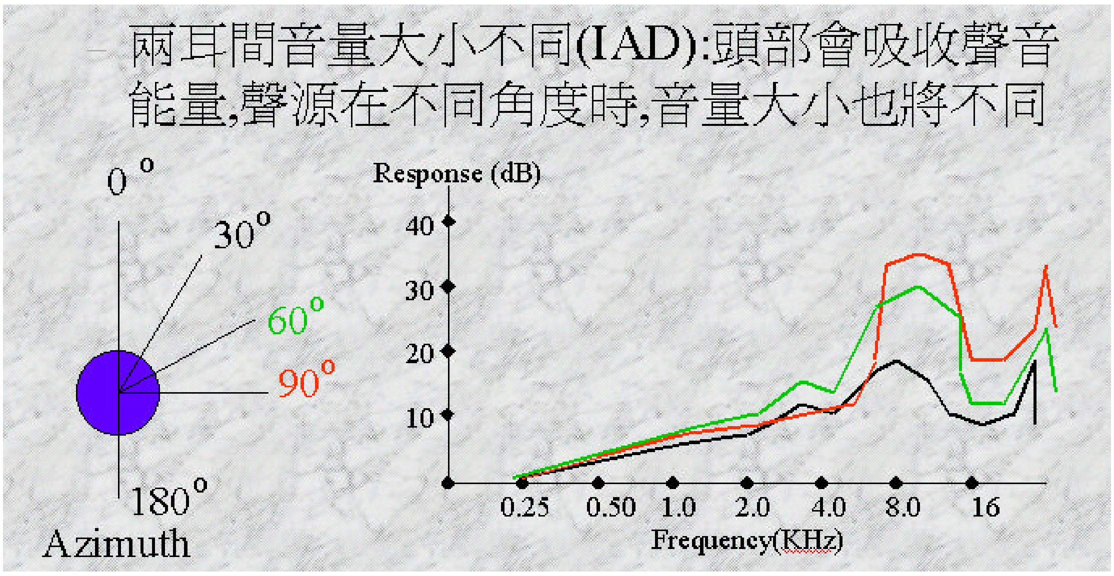
Sound wave is transmitted in the air in a speed of 345 meters per second. Let us assume that the distance between our ears is 20 centimeters, and the sound source comes from our left. It would be no doubt that the sound wave will first reach our left ear, and in 580us(time for sound wave to travel for twenty centimeters), the sound wave will reach our right ear. If the sound source comes right in front of us, then it will reach our both ears at the same time. As to sound sources coming from other angles, it is easy to calculate them by means of trigonometric function. Therefore, human brain has no difficulty in distinguishing different directions of the sound sources via ITD.



(Sounds coming from different angles will definitely take different time spans to reach the listener's ears.)

2. IAD (Inter Aural Amplitude Difference)

Every one of us must have the same experience: if sound is blocked by objects, it will turn out to be lower in volume when it reaches us. Try to imagine that the sound comes exactly from our left, then what our left ear perceives will retain its original volume, whereas what perceives by our right ear will be decreased in volume because our head absorbs some part of the volume. Theoretically speaking, it is feasible that the amplitude reached human ears from any point in a 360 degree space can be measured, and its relative relationship can be figured out.



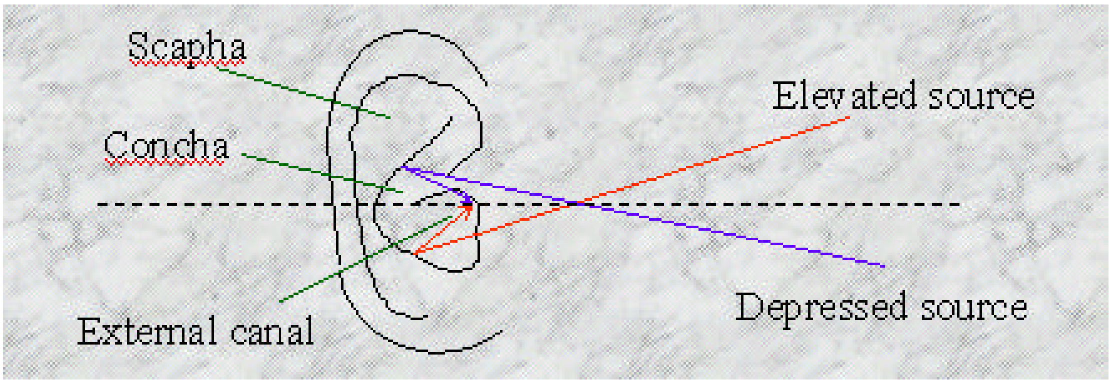
(The human head will absorb some part of the sound volume. When sound sources coming from different angles, they vary in the volume.)

2.The Pinna Diffractive Effect and the Frequency Vibration of Aural Canal to Sound Wave

In terms of helping our brain distinguish only the direction of the sound source, ITD and IAD will be sufficient. Yet, ITD and IAD can not tell the difference between sound sources coming from our straight front and our straight back, in that both values are almost the same. This also happens when the sound sources are from our straight top and our down bottom. To make it simple, let us suppose that right in front of us lies a ball surface which is 0 degree, then if we measure its ITD and IAD, we will see both values are very close. Accordingly, depending solely on ITD and IAD will not be good enough. To solve this, our pinna plays a very vital role.

Sound wave bounces back when it hits objects. Our ears are oval-shaped and hollow inside; therefore, sound waves with different kinds of length will correspondingly produce diffractive effects in the outer canal. From the perspective of frequency analysis, when sound sources come from different angles, they will definitely produce different frequency vibrations to the eardrum. It is because of the existence of the external canal that sounds coming from our front and our back are quite distinctive.

There is a 2-centimeter canal between the aural canal and the eardrum, and its hollow co-vibration will have the greatest enhancement to the 5Khz signal, which is exactly the highest limit of human acoustic frequency. Therefore, we got to be really careful when we whisper or mutter in secrecy, as our voice is extremely audible to others in this way.



Now, let's we have another experiment. Seal up the hollow part of your pinna with clay, and leave the auditory canal open. Shake up a string of keys in front of you, and you will see your ability of telling the difference between up and down is severely impaired. Again, try to pull your pinna back to your head closely, you will also see your sense of direction is much different than before. Nevertheless, our brain is very smart in terms of telling the direction. Therefore, all the discrepancies described above will be bridged by means of seeing and the room reflection effect. This is what people call "telling the direction by sound" in psychology. For instance, when we hear a helicopter flying, we will look above to seek for the sound source. Hence, psychological anticipation plays an integral part in audio positioning.

3.Reflection and Absorption

Room or environmental reflection effect is also an important parameter. The reflected objects have their own sound wave absorption value; for example, tiles and plywood boards have distinctive reflective values. With your eyes closed, you will not have any difficulty in telling whether you are in a bathroom or in a Japanese bedroom, will you?

Therefore, if we want to measure the 3D positional audio effect, we had better take all the factors such as room size, shape, and building material, into consideration, so as to enhance the accuracy of the sound quality.

4.The Psychological Anticipation to the Sound Source

Other factors such as the shoulder reflection wave, and the psychological anticipation to the sound source(for example, we know the plane is on the air, and the cricket in on the ground) are also useful factors for our brain to tell the location.

To sum up, we got three major elements:
ITD, IAD, and the pinna frequency vibration. All these three elements are HRTF--Head Related Transfer Function parameters. Other elements such as room reflection and absorption can be processed by audio physics.

How to Measure HRTF Library?

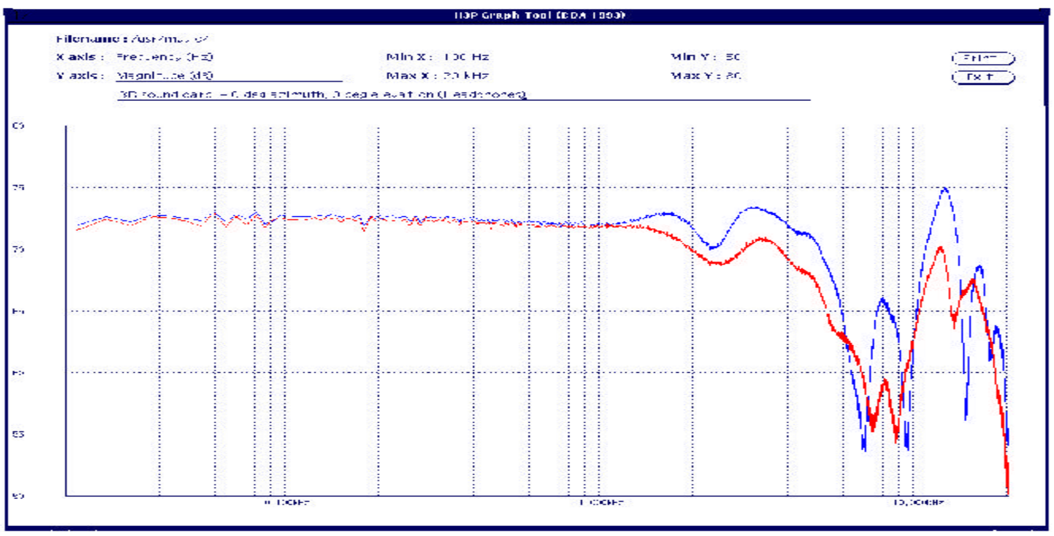
After knowing the three elements of sound(ITD, IAD, and Pinna Effect), our question will be: how to measure HRTF library?

AT first thought, we would come up with an idea of putting a microphone inside a person's ear, and we think human HRTF parameter can be measured this way. However, if the microphone is put in the concha(the opening of the ear), then the frequency vibration effect of the auditory canal will be sacrificed, and the measured parameter will not be accurate enough. Therefore, using an artificial head will be the best solution to this problem.

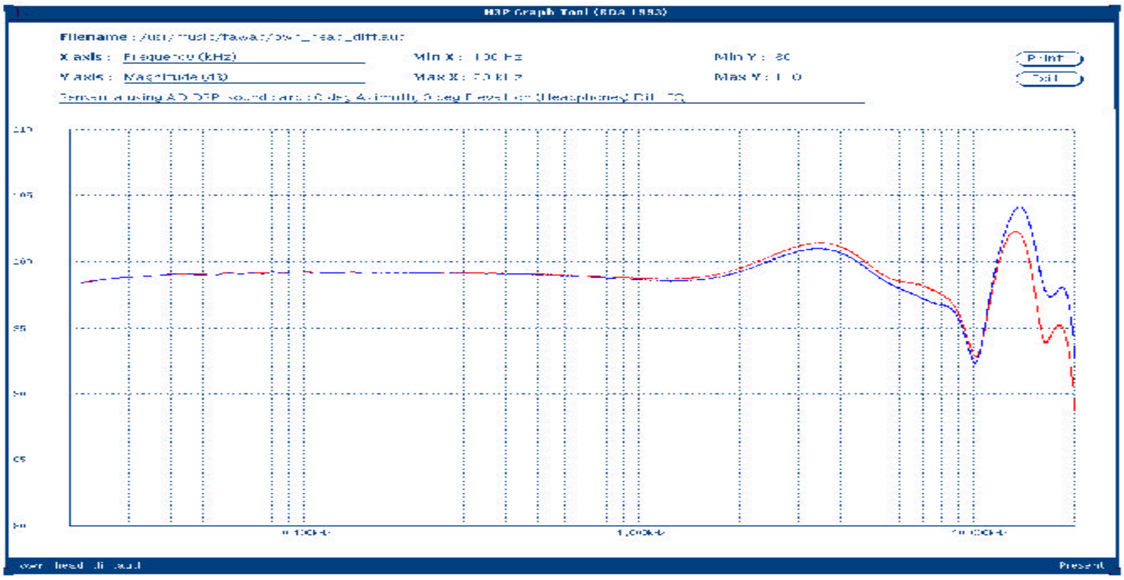
In fact, the shape of the artificial ear is even more important in terms of measuring HRTF library. Some research institutes use plastic material to produce the ear mold, while there are others, in a more advanced way, use digital CAD / CNC to reconstruct the best computer-emulated ear mold, fitting it into the artificial head. Once a high-quality microphone is put inside the artificial head, the measuring work would begin. This type of technique is developed by Britain-based Central Research Laboratories(CRL), called "Digital Ear." Digital Ear can help researchers measure a HRTF parameter which is much more sufficiently accurate than using other commercially available solutions.

The following pictures are the sound wave measured from the straight front horizontally. Theoretically considered, the frequency vibrations of a sound source coming from the straight front should produce to both ear the same effect. That is, the red and the blue curves should be very close, or even overlapping. It is quite obvious that CRL's Digital Ear is capable of measuring a more accurate frequency vibration than others.

3rd Party HRTF curves Azimuth = 0° Elevation=0°



HRTF curves generated from the Digital Ear Azimuth=0 ¢ Elevation=0 ¢ X



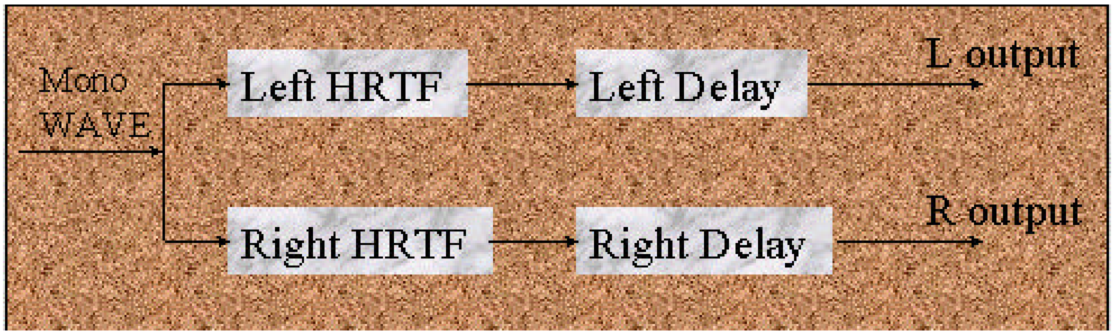
Having a perfect artificial head in hand, the next step would be starting with the measuring work. The measuring environment must be echo-free, in that echo will increase difficulties in extracting our preferred sound characteristics from the collected data. The echoic chamber is just an echo-free, and therefore suitable environment for this. Put the artificial head in a closed room inside which numerous pieces of cone-shaped sponge are placed all over the ceiling and the floor, prepare a sound source which can randomly position at any spot in a three-dimensional space, and then the measuring work would begin. Generally speaking, a complete set of parameter, including afterward adjustment, will take months to complete.

When different types of sound wave ranging from 20hz to 20Khz are recorded at different positions in a three-dimensional room, the complete HRTF raw data is said to be collected. The raw date, coupled with the math formula-generated HRTF parameter complemented by the accurately calculated EQ in case the measuring equipment may not be sufficient enough, are what we want--HRTF library. Usually, the highest quality, 48Khz frequency, is chosen in the measuring work. In order to meet the hardware cost requirement, two other types of parameters, 44.1khz and 22khz, are also implemented in the same way. It goes without saying that the higher the chosen frequency, the better the quality. However, higher frequency requires higher taps in the digital filter. Therefore, it depends on the cost and the quality concern while choosing the preferred frequency.

How to Synthesize 3D Positional Audio Effect by Means of HRTF Library?

HRTF audio synthesis emulation circuit can be divided into two major parts: Digital Filter and Time Delay.

1. The application program will first send the preferred synthesis coordinate to HRTF library in order to find out the appropriate parameter, and then it will send the parameter back to the Digital filter. Now the filter of the left and the right channels will back-calculate the preferably-heard frequency vibration curve of the brain, which is IAD degree, and the pinna effect merged in the frequency domain.
2. Since the characteristics of 3 dimensions have been emulated in the frequency domain, the only thing left is ITD. It is much easier to synthesize ITD compared to digital filter: only by digital delayed circuit to reproduce the estimated time delay, ITD will be available.
3. Other characteristics like reflection, wall absorption effect, distance, and Doppler effect can be processed in the application program interface.



Now, if the sound wave is transmitted directly to the eardrum inside the inner ear, it is a 3D virtual audio effect for real, capable of fooling the human brain around. But how to transmit it? Yes, by a pair of headphones. A pair of good frequency-responding headphones will be the fittest equipment; plug-in earphones used in the walkman also have fairly good effect. Don't sacrifice 3D features generated by time-and-energy-consuming calculation by buying a poor performance headphones only for economic concern, because the effects will be devastatingly impaired.

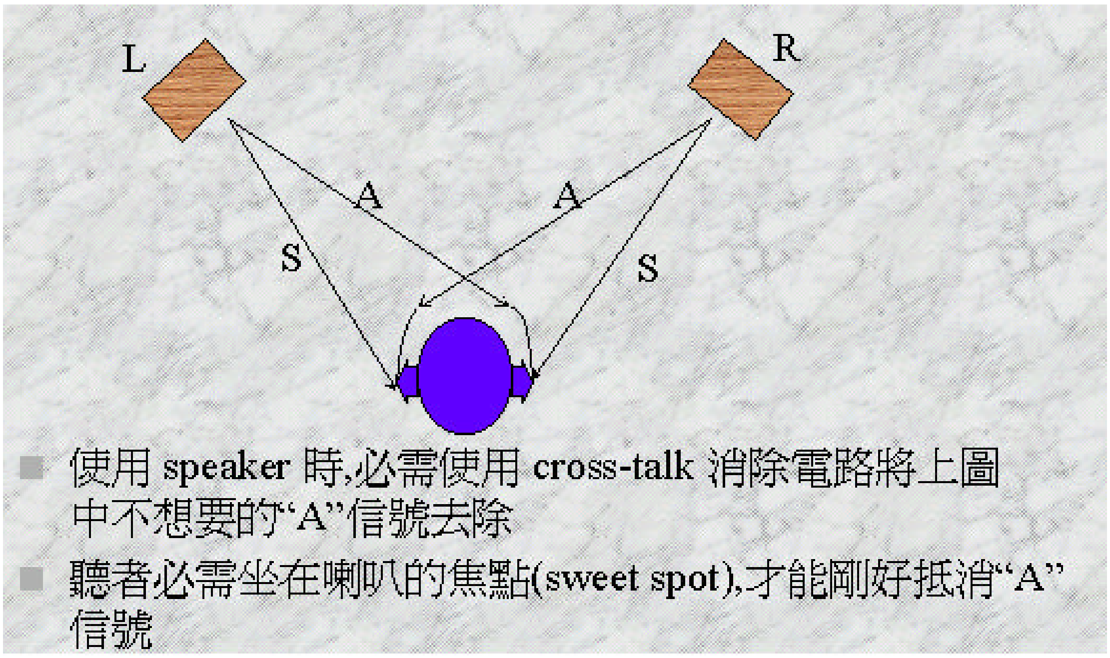
Is it Applicable to Have Two Speakers Replace a Pair of Headphones in Terms of 3D Audio Listening?

It won't be a beautiful thing to wear a pair of headphones while playing video games in a hot summer. Is there anything we can do to improve this? For example, can we have two speakers replace headphones? The answer is: it is possible under certain circumstances.

Regarding listening, the biggest problem for using speakers is the cross-talk interference: how can we have the left ear hear sounds coming from the left speaker, and the right ear hear sounds coming from the right, between them exists no interference? As the picture below shows, both ears will definitely hear sounds coming from both speakers. Moreover, if the listener walks to and fro, then the interference will become even worse.

However, if the listener is seated between the two speakers, right at the point of an equilateral triangle as showed below, there won't be any interference. That is, we can send the pre-calculated cancellation signal to reach the listener's ears at the same time to cancel the cross-talk interference, and this circuit is the so-called cross-talk cancellation.

A wholesome cross-talk cancellation formula can make sure the sound quality in the frequency domain is smooth, and the low pitch will not distort. It has to be very careful in this, as the



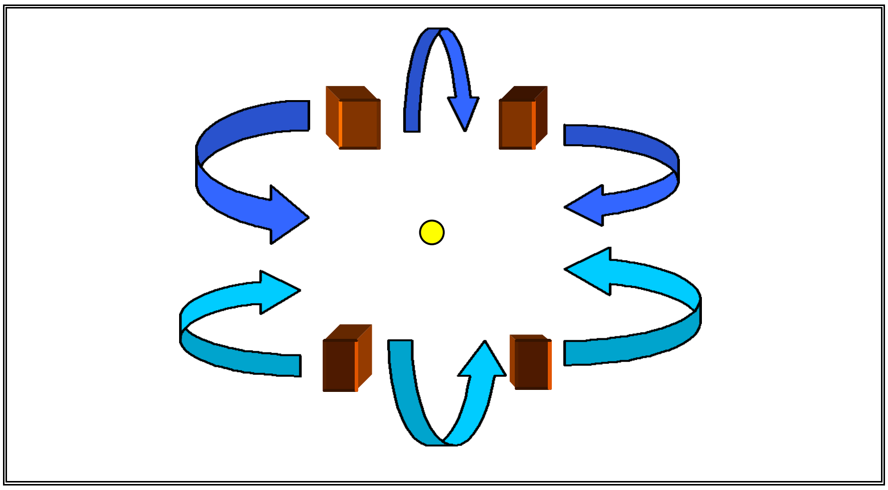
- When two speakers are used, cross-talk cancellation circuit has to be applied in order to cancel the undesired signal A.
- The listener has to be seated at the focus of the speakers(sweet spot) in order to cancel signal A.

What is the Limit of Sweet Spot? Is There Any Way to Eliminate it?

In application, the limit of cross-talk cancellation includes: the listener has to be seated at the point of a triangle formed by two speakers and himself, and the point is what we know as the sweet spot. At the same time, the listener can not turn his head randomly, otherwise the ITD will not be accurate enough. Furthermore, the position of putting the speakers is also very important: the speakers have to be at the same height of the listener's ears, and neither in the listener's near front nor his close back can put any reflective materials, such as partitions, walls, etc., or the reflected sound wave will disturb the direction-telling ability of the brain. Of course, the quality of the speakers is also very important. Speakers with frequency vibration distortion and poor left/right distinction are not considered appropriate in terms of 3D listening.

Therefore, without critical preparation beforehand, it would be quite difficult for the general public to have access to enjoying a perfect 3-dimensional sound effect. Even if people spend much money buying all the necessary equipment, there is the biggest limit: the listener can not turn his head or leave the sweet spot.

Wherever there is a problem, there is a solution. When the listener's head leaves the sweet spot, human brain can not perceive the front/back distinction, still the left/right difference remains. Therefore, so long as there is a way to enhance the front/back distinction, the lost part can be complemented. The easiest way is to add an extra pair of rear speakers. That is, when the audio chip is doing the process with the rear coordinate, it sends the calculated sound data to the rear speakers. In other words, it is to have the front speakers process with the front half of the data, and the rear speakers with the rear half. In this architecture, the sweet spot is wider than before, allowing the listener to move his head in a reasonable sphere. At the same time, the listener does not have to choose the best-quality speakers in order to enjoy the special sound effect; the speakers with general multimedia sound effect will be sufficient enough to complement the rear characteristics. C-Media's CMI-8338A/C3DX PCI single-chips offer 4CH integration playback function, so without increasing any buying budget, fairly good 3- dimensional sound effect is accessible through this kind of sound chips.



HRTF 3D Positional Audio in Application

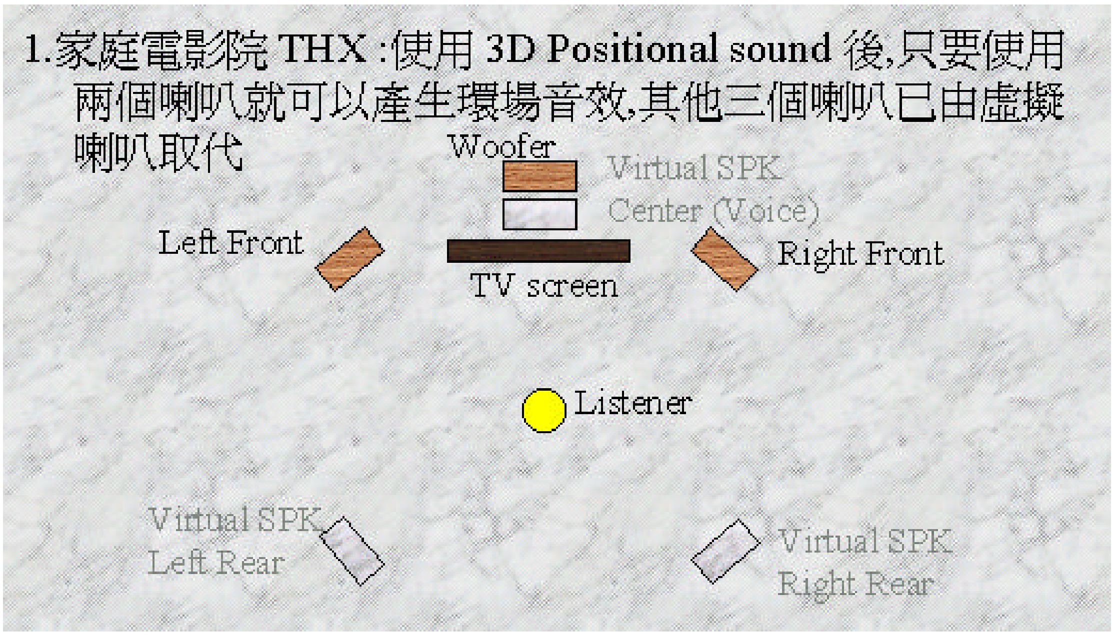
There is no doubt that the most prominent HRTF 3D positional audio application is in multimedia games on PC. So far, there are almost a hundred kinds of video games supporting 3D positional audio effect. You may visit the Web Site below to find out the latest games with 3D audio effect:

http://www.3dair.com

http://www.aureal.com/tech/a3ddevs.html

Most of the games at this site provide downloadable demos, and if you are interested, you can play those games and enjoy the genuine audio effect while competing with contestants in a car racing game, or feel the power of battlefield while involving yourself in a war game. Provided you got a 3DFX graphic acceleration card, you got to avail yourself to the sensational 3D positional audio technology.

Besides video games, HRTF 3D positional audio can also be applied in DVD. DVD provides 5-channel output powers: left front, right front, left rear, right rear, center(film dialog) and sub woofer. When HRTF 3D positional audio is applied, listeners can enjoy advanced audio effect which is quite similar to that of a theater.



(Home Theater THX: When 3D Positional audio technology is applied, using only a pair of speakers can generate surround effect, as the other three speakers are replaced by the virtual ones.)

EMI Record Company, U.K. utilized CRL's 3D positional audio technology to produce many records which create a special listening experience for the listener: while listening to those records, the listener can have a feeling as if he were standing at the conductor's position, and listening to a grand orchestra play right in front. Besides, EMI also designed a program in which the singer can move randomly around the listener's head to sing.

In Japan, there is a big amusement park inside which built a "Haunted House" which also applied 3D positional audio technology in its entertainment design. Put in a dark room with headphones on, tourists can hear ghosts flying all around. Furthermore, the floor moves in accordance with the sound, so the effect amazingly sensational.

The United States NASA uses 3D positional technology to enhance the flying emulator effect, so the astronauts can be better prepared in confronting all possible situations. If one day the astronauts are asked to perform certain mission in the outer space, they can do it without any difficulties.

One Present Mainstream 3D Positional Audio Source and Its Bloom in Taiwan: CRL 3D Audio

CRL(Central Research Laboratories), UK is a subsidiary of Britain-based EMI Record Company, and it has a 68-year history. CRL is well-known for its constant advances of leading-edge technology: its invention of CAT scanner is so outstanding that it won a Nobel Prize at its time. In order to open a new era in music production, CRL spent 6 to 7 years researching 3D positional audio technology, which then reached its maturity in 1995. At that time, MOEA(Ministry of Economic Affairs)Taiwan, Taipei Computer Association, and III(Institute for Information Industry) were operating the first national technology transfer project sponsored by Taiwan government. In this project, there was a sub category called "3D sound," and potential technology transfer companies were Crystal River(later Aureal) and CRL. CRL stood out for its excellent audio technology and complete technology transfer plan, so it won the recognition of relating governmental and academic departments and institutes.

At that time, there were three institutes, III, C-Media, and Mitac, had their seed engineers go to Great Britain to receive relating training for half a month, and such a training helped to establish a solid foundation upon which Taiwan is able to develop high-end audio chips. One year after(1996), C-Media successfully placed its first-generation 3D positional audio chip under the spotlight in the presentation exhibition of the technology transfer project in Computex, Taipei. During that time, Yamaha Japan decided to have the same audio technology. IN 1997, sound-chip maker ESS chose also CRL 3D as its high-end 3D audio engine after choosing Qsound 3D for its low-end products.

CRL utilizes the Digital Ear technology as a base for the 3D audio research; at the same time, in order to perfect the audio technology, CRL also collaborated with several famous high-fidelity recording companies so as to make sure that CRL 3D audio technology can meet the strict quality requirements of music industry. Accordingly, CRL proves that its 3D audio technology is ahead of the rest in the professional audio field.

Conclusion

The feature of 3D positional audio has become a must-have in choosing the next generation PCI audio chips. After knowing the HRTF 3D positional audio technology, we have a clear picture about it, thereby having the knowledge of distinguishing between poor and good HRTF libraries. Integrating audio chips into motherboards has become the mainstream in PC industry. Better yet, the gap between quality and price has been narrowed down reasonably and substantially. Whereas audio chips made by name-brand companies provide high performance, audio chips designed by Taiwanese design houses also exhibit competing features, such as 4-channel output with 3D positional audio, digital stereo interface SPDIF IN/OUT, and DLS music synthesizer, etc. Therefore, don't ever equalize cost-effectiveness with low functionality. C-Media's CMI8338A/CDX PCI single-chips just prove to the world that with strong and dynamic design capabilities, Taiwanese IC design houses are capable of offering competitive as well as satisfactory audio solutions to customers.