Sounds and Hearing, Pt. 1

Here's a beginner's approach that explains some of the sounds that you hear, how you hear them, and how the ear works.

Considering all the time and effort audiophiles spend producing such a wide variety of amplifiers, speakers, and other devices, the hearing process is just as important as any other part of the music reproduction chain—from the recording to the ear mechanism and nerve impulses sent to the brain to be interpreted. Depending on an individual's conditioning, a sound can produce an entire spectrum of emotions and physical reactions, such as the quickening of a heartbeat, laughter, and tears. Sound triggers movement and speech, stirs the memory, and brings forth long-forgotten mental pictures. To accomplish this, many different ways of making sounds and music have evolved over the years, and now, sound recording and reproduction are ways to store and play these valuable sounds.

STEREO SOUND

You can think of a simple stereo recording as being made with two microphones spaced apart and played back through two speakers that are similarly spaced apart. That is all there is to it! However, because there are no standards for stereo recordings, this is not the case for every recording, particularly popular music. Recordings of individual instruments and vocalists can even be made at different times and different places and combined later. The distinct sound of the instruments can be recorded through the use of a microphone or two within a few feet of each instrument or group of instruments. This might explain why some recordings sound much brighter than what you might hear in the audience, unless that live sound, too, is altered to account for the difference[1].

Each instrument or group can be panned with the mixing console—as mono or stereo—to appear at different places between the speakers. It can bring the instruments into your room and produce the effect of being there. Mixers can add artificial reverberation and alter the response. The only people who truly hear the final production exactly the way it was intended are, perhaps, the recording engineer and producer.

Stereo recordings have several limitations when played through speakers. One occurs when an audience is involved and the microphones are located between the orchestra and the audience. With music played through speakers, a cough in the audience, for example, may sound as though it is coming from a spot in the orchestra instead of behind you.

Many years ago, I made a stereo recording at the edge of the road in front of Sonotone, where I used to work. When I played the recording back on a pair of speakers, a car that had passed by continued to remain in the right speaker and did not sound like it had gone to the extreme right as it really had done when it went farther down the road.

In addition, there is no vertical component. This became obvious when I made a stereo recording of a Wurlitzer pipe organ in an old movie theater. The pipe lofts were located to the left and right and above the height of the stage. When played back on speakers, the sound appeared to come only from the speakers and the space between them, but not up in the air as the organ lofts actually were.

STEREO ATTEMPTS

It was my hope that 4-channel stereo of the 1970s would enable us to be surrounded with stereo, at least in the two dimensions of the horizontal plane. In this configuration, speakers were typically located in four corners of the room and the listener sat in the center. The sound occasionally accomplished some interesting effects. However, technology was limited and fidelity was not always the greatest.

Although mentioned in an article in *Hi-Fi News* from the UK in the 1960s, tetrahedral stereo never caught on. This consists of two front speakers and one centered at the rear. These form an equilateral triangle. The listener sits in the center. A fourth speaker is located above the center of the triangle forming a tetrahedron. If recordings were made with corresponding microphone positions, a 3-dimensional sound space could be produced. Despite some limitations, very convincing sound could be reproduced. Even more interesting effects could be made with electronic enhancements, particularly for very creative composers.

This process was referred to as Ambisonics. I remember seeing the closely spaced tetrahedral microphone array when visiting with Mark Davis at MIT in Cambridge, Mass., in the 1970s. Calrec was the first manufacturer to develop this array. The idea never caught on, although today, with high quality multichannel capability, it could be easily accomplished.

Photo 1 shows a way to produce a 3-dimensional environment. "One extreme was carried out by the Air Force at Wright-Patterson Air Force Base. To create 3-D sound, researchers used a 14′ geodesic sphere of high-grade aluminum alloy, incorporating 272 speakers at regular intervals over the surface. Up to 15 speakers work at a time, emitting the same or different sounds. Sound can move around the sphere at up to two revolutions per second—the same rate that a pilot can roll an F-16 or T-38. In entertainment, 3-D sound achieves greater realism; in the same way that

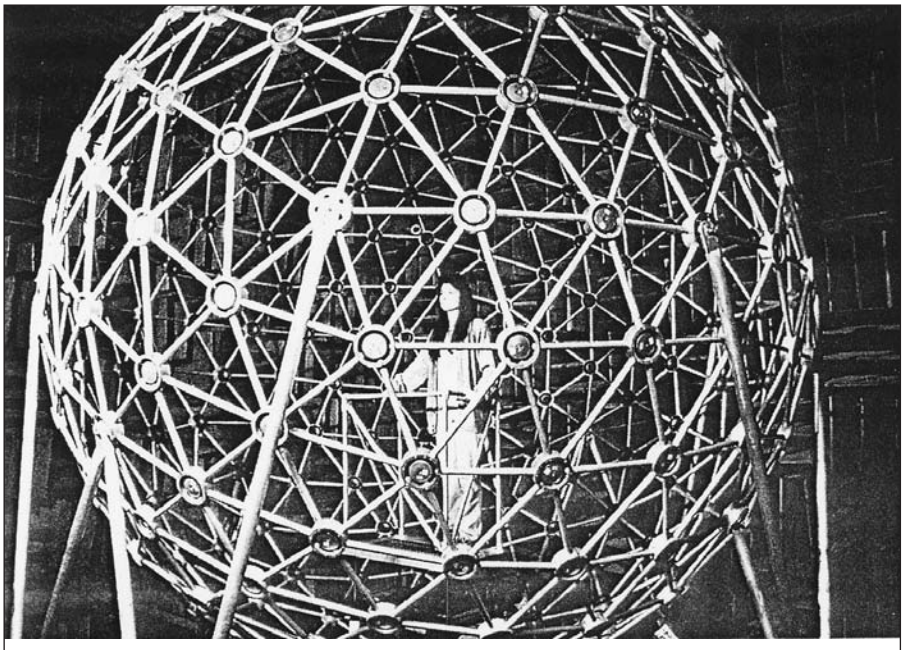

PHOTO 1: Sound system for pilots.

stereo improves on the mono system of the 1950s. Sound is generated in many locations, and when you turn your head, sound changes just as it does in real life."[2]

Despite the methods for recording stereo, even a simple stereo recording has problems with playback through speakers. First, a straightforward monophonic drum recording arrangement in **Fig. 1** can be used to play back through a single speaker without any difficulty. However, when two microphones are used to record stereo, as in **Fig. 2**, stereo crosstalk from each speaker to both ears interferes with what would be heard if

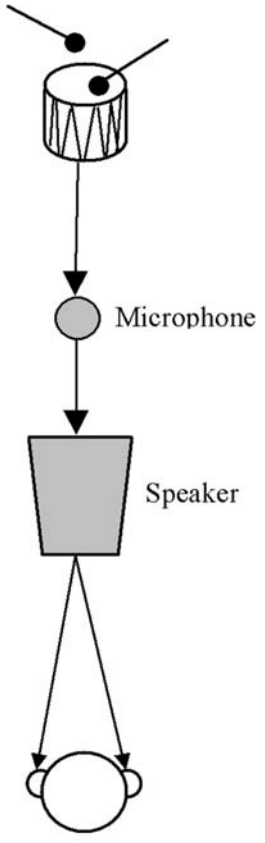

FIGURE 1: Monophonic recording.

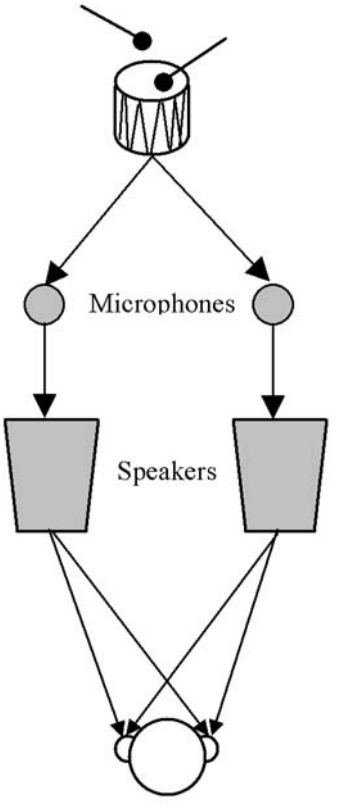

FIGURE 2: Stereo recording.

the drum were heard directly. I will discuss the mechanics of this later. A partial solution is listening with headphones, but this is not entirely satisfactory.

BINAURAL SOUND

Binaural recording and playback are very different approaches. After all, you really hear binaurally. It enables an even greater degree of perception and eliminates speakers, room acoustics, and stereo speaker crosstalk. Sounds can be located in all directions similar to how you really hear. On the other hand, it presents other problems such as headphone quality, outer ear compatibility, and skin perception.

In order to present binaural sound correctly to the ears, you need only a two-channel system and headphones. Ideally, a binaural recording should be able to let you perceive sound as being outside and from any direction. To achieve this, several recording methods are used with a microphone spacing of about 7″ to simulate the distance between the ears. A dummy head is commonly utilized to provide the necessary conditions for binaural recordings.

The Kunstkopf and Neumann KU-81i artificial heads have a small microphone located in each ear canal and include artificial pinna [outer ear]. The head and ears are made of appropriate acoustic materials. Another method is to process signals electronically to simulate ear-related transfer functions (HTRF).

A further approach is to incorporate a pair of tiny microphones that can be fitted inside the ears of the person making the recording. If recordings are made this way or with an artificial head that also includes the pinna, playback with headphones involves hearing through a double set of pinna. However, eliminating the pinna in the dummy heads or using earbuds to bypass some of the outer ear could help. Of course, there is still a compromise because the pinna on the artificial head are not the same as yours and affect sound differently from what you hear in the real world.

Another problem is that the listener is not free to turn his/her head to locate sounds more accurately. In addition, particularly for loud sounds, body perception comes into play. Skin perception provides additional information and is useful to detect the impact of a shock wave, such as a drum strike or other percussive sounds. Even loud infrasonic sounds can be felt by the whole body.

Despite the drawbacks of the various recording methods, stereo and binaural sounds both produce some very pleasing effects. In addition, the recording medium has improved tremendously from what it was 50 or 60 years ago, but there is still a lot of room for improvement to reproduce a truly realistic experience in which you can move around and hear sound like you do in real life.

How do I know what good sound is? I played the violin for many years in grade school and then in the high school orchestra. I attended many live concerts—the kind without speakers to enhance the sound. These examples could be convincing but memory fades over time and playing in an orchestra with instruments surrounding you is not the same as sitting in the audience.

Better yet are the sounds heard around us every day. You normally take hearing for granted and awareness of sound location and frequencies are only in the background of our attention. The worst instance for me was when earwax had built up in one of my ears and I could hear only with the other ear (monaural sound—bad). It is at times like this when you can appreciate all that is missing from normal binaural hearing. It was unbearable. I went to the emergency room and had the wax removed. It was like being reborn!

THE HAAS EFFECT

The Haas Effect[3], also known as the Precedence Effect, plays an important role in how you hear. When two identical sounds come from two different sources, but one is delayed to simulate an echo, the interpretation of what you hear is a unique characteristic of the ear/brain combination. After hearing the initial sound, the brain will suppress any later sound, such as an echo, for up to 30 milliseconds (some sources report up to 40 milliseconds). In effect, this causes you to literally not hear the delayed sound for that short time.

This inhibition is called time or temporal masking. The first sound to arrive

defines the direction. If the arrival time of the delayed sound is longer, then two distinct sounds are heard, even if the second arrival is as much as 10dB higher. In addition, if the initial burst is from one speaker but continues as a steady tone from another speaker in a different location, the source will still be identified as coming from the speaker with the initial burst.

Even in an enclosed environment, such as a living room that has several reflective surfaces, you can reliably localize the initial source from the signals that reach your ears first. However, localizing a steady tone in a room is nearly hopeless.

Reflected sounds from the wall, floor, and ceiling form many different reinforcements and cancellations throughout the room. You can hear them as peaks and nulls as you move around. There are no clues for direction. When viewed on a survival basis, this is very important to locate any danger, and not be confused with immediate echoes that can come from other directions.

One of the classic demonstrations of temporal masking is to stand in front of a hard surface such as a brick wall, away from other reflecting surfaces that might interfere with what you hear. By clapping your hands and gradually moving back, you will reach a point where you can begin to discern two separate sounds—your initial handclap and the echo.

When I tried this, I measured 17´ to the wall. The total distance, representing the time it takes for the sound to travel to the wall and back again, was 34´ or 30.6 milliseconds. The temperature was 53° F. You can even tell whether the distance is great or small depending on how long the reflected sound takes to return. The longer the delay is, the greater the distance you are from the wall.

Accurate onset of transients is essential for initial perception and is of utmost importance for locating sounds. The leading edge or wavefront of a transient provides higher-frequency content that permits localization. In this way, the ear effectively acts as a waveform analyzer. One author suggests that this method of detection may be the way human hearing has evolved, what our continued existence is all about, and sine waves may not be what you respond to at all[4].

MONOPHONIC SOUND WITH STEREO SPEAKERS

Sound originating from one point in space, such as a single speaker that handles the entire frequency range, can be called monophonic sound. It has no dimensions at all and requires only one channel. There is a difference between monophonic and monaural sound. Monophonic means only one channel as in a recording, and monaural means hearing with only one ear. Binaural is hearing with both ears. With both ears, you can easily locate the source of a monophonic sound (**Fig. 3**).

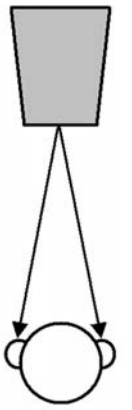

FIGURE 3: Monophonic source diagram.

When two stereo speakers (**Fig. 4**) are separated by several feet and fed with the same monophonic signal, a phantom image appears to come from straight ahead. However, the phantom image is not the same as the separate monophonic center channel in **Fig. 3** because the sound from each speaker travels to both ears resulting in a crosstalk of sounds from the left speaker into the right ear and vice versa. Crosstalk alters the am-

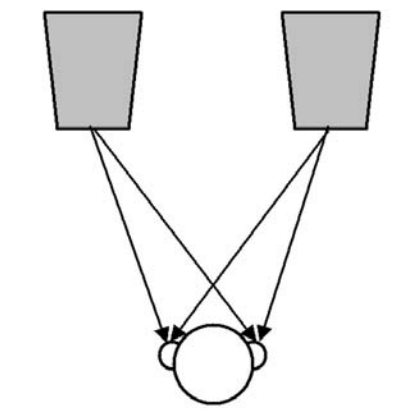

FIGURE 4: Stereo crosstalk diagram.

plitude and phase response of the phantom center image and is not the way you would hear a sound from a single-point source.

The resulting response for a monophonic phantom image is referred to as a comb filter because the nulls resemble a comb. The nulls shift to lower frequencies as you move further to one side. The curve in **Fig. 5** was made with two identical speakers placed about 2´ apart and fed the same signal. The microphone was located 30˚ off to the left side of the left speaker. Although each speaker by itself had a very flat response, comb filtering can drastically alter the combined response.

Comb filtering can be easily heard by playing FM inter-station hiss or pink noise in the mono mode. As you move your head from side to side from the center position, you can hear the hiss increase to a maximum and then decrease again. This effect may be more audible with some speakers than others depending on their directional characteristics and surrounding room reflections. However, when using random phased pink noise, which means a separate generator for each channel, this effect is eliminated. The speakers then essentially do not radiate the same frequency at the same time.

With home theater material, voice is almost always in mono and is directed to the center channel speaker. However, with only two speakers playing home theater and the listener sitting at the center position, the mono voice essentially appears as a center phantom in spite of the missing frequency effects of comb filtering. Nevertheless, the image will exhibit the same problems with waveform (amplitudes and phase) and follow you as you move all the way to one side. On the other hand, a separate mono center channel does not suffer from comb filtering or waveform distortion, and the sound continues to appear to be coming from the center speaker even when you move to the side.

Some recordings may have a mix of stereo and mono, such as a vocalist, piano, or other instruments that appear in the exact center. Listening to a piano that appears to have no width or several instruments at the exact same location is not becoming for good stereo sound, but if a single channel and speaker are used for each instrument, comb filtering and the apparent narrow-width effect are not present. Some stereo recordings with much random spatial content and no mono added will not exhibit this characteristic nearly as much and might be more like the behavior of random phased noise. Monophonic recordings would seem to be more accurately played using only one speaker.

SOUND LOCALIZATION

Human survival still depends on the ability to locate sounds accurately, and two ears do hear more effectively than one. Most investigations for sound perception have been limited to the horizontal dimension. This includes not only the sound source but also assumes that the listener's ears are in the same horizontal plane. Tests to locate direct sounds are normally conducted in a reflection-free environment.

The ability to locate a sound in real life is determined by the path length difference between the ears. The path length is the actual distance the sound must travel around the head, including obstructions such as the contours of the pinna. The time difference to reach each eardrum is referred to as the inter-aural time difference, which is significant because the path to the farther ear is not only longer but also curved. Higher frequencies tend to propagate in straight lines and do not follow curvatures as well. Sound reaching the farther ear is reduced in amplitude producing an inter-aural amplitude difference. The results are shown in **Fig. 6** for several different angles[5].

The ear closest to the source determines the side the sound comes from, and the amount of inter-aural time difference defines the direction. Sounds that arrive at both ears at the same time are interpreted as coming from straight ahead, but you can easily tell whether the sound comes from a different direction. Sounds near the front can be located with an accuracy of 2 or 3˚, but accuracy decreases as the angle to the source increases. This invites turning the head to more accurately locate the source.

The head is a significant influence for

frequencies above 600Hz[5] but with a discontinuity in the area of 2kHz. The same area of the brain also processes sounds that come from the rear or are inclined above or below but are modified differently by different portions of the head and pinna. The shape of the pinna is different for different arrival directions. The brain can interpret all these clues to locate the origin of the sound.

STEREO CROSSTALK CANCELLATION

Referring again to **Fig. 4** and the arrangement of two speakers, the right ear hears some of the sound coming from the left speaker, and vice versa. To compensate for stereo crosstalk and to locate an image at a particular place, you can add electronic compensation to the left and right channels so the speakers radiate the sound with the inter-aural time delay that the brain would expect if the sound were actually coming from that place directly to your ears. Normally, if the spot is to the listener's right, the sound from the right speaker should reach the right ear before the left speaker sound reaches the left ear. But the sound from the right speaker typically travels around the head and enters the left ear before the left speaker sound can arrive.

The problem is further complicated by the Haas effect. After hearing a signal, the brain suppresses any similar signals that occur within about 30ms. In this example, the delayed sound from the left speaker is blocked if it arrives too soon after the same sound from the right speaker.

The crosstalk cancellation concept was patented by Atal and Schroeder at the Bell Telephone Laboratories in 1966. It was based on reproducing binaural recordings through speakers that are located relatively close together. Several different commercial versions were developed over the years. One sophisticated device (**Photo 2**) was called the Sound Retrieval System (SRS), which was refined by Arnold Klayman at Hughes in the 1980s.

The SRS system does not have any time delay processing, artificial reverberation, harmonic generation, artificial phase correction or alteration, and does not require any prior encoding of program material. It was designed to

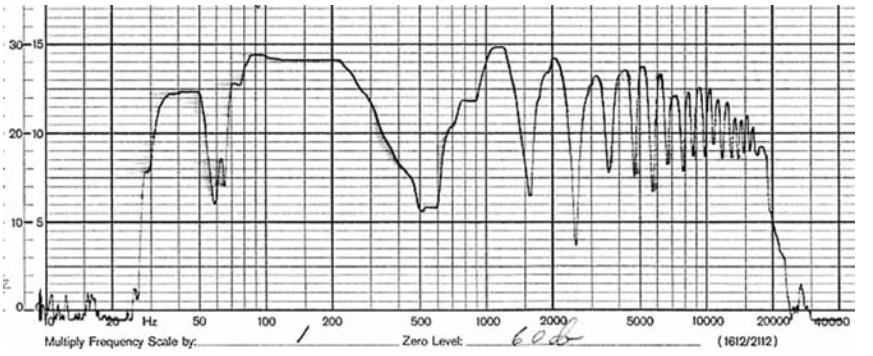

FIGURE 5: Comb filtering response.

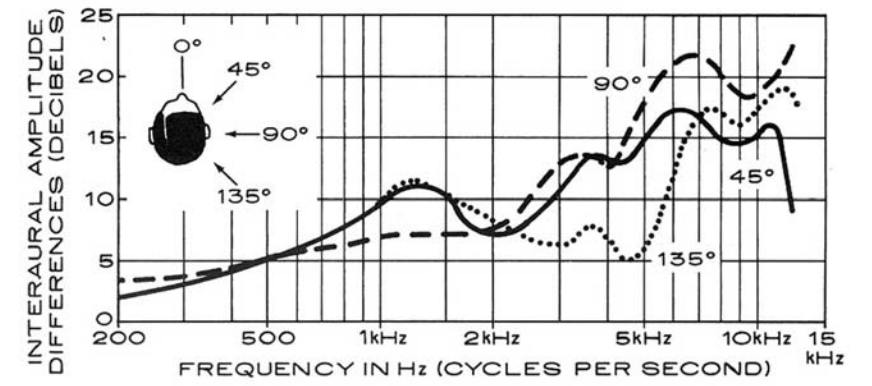

FIGURE 6: Inter-aural hearing curves.

PHOTO 2: Hughes AK-100 Sound Retrieval System.

simulate inter-aural differences between the ears and to be played through closely spaced stereo speakers. (Inter-aural effects and hearing are discussed further on.)

Although Klayman intended it to be for use in 747 airplanes and was awarded four patents, it never went into production. Later, the design and patents were purchased from Hughes by SRS Labs headed by Klayman and Jim Lucas. It was produced and sold as the model AK-100 Sound Retrieval System.

A demonstration at McIntosh Lab by the Hughes representatives utilized two small speakers placed about 12″ apart on the floor. A stereo recording was played through the AK-100 and then through a power amplifier and speakers. It sounded surprisingly spacious and pleasing. The center control can be adjusted to fill in the hole-in-the-middle as desired and the spaciousness can also be adjusted.

Although the effects are sometimes interesting, they vary from recording to recording requiring readjustment of the controls and there is no accurate setting other than personal preference. The arrangement assumes you will be at the center between the speakers and facing toward the front. Sounds can be heard well beyond the space between the speakers. However, when the AK-100 is used with binaural recordings, a better sense of space and direction can be achieved, again depending on what is in the recording. Sometimes the spacing between the speakers needs to be adjusted as well.

For speakers with wide spacing, the effect is obscured and response appears to be altered when seated back in the listening area. In addition, it is claimed that the AK-100 can add interest to monophonic recordings. When you are close to the wall where the speakers are, you can hear sounds coming from the sides as well as the center, but as soon as you turn your head to locate the sounds at the side, the illusion disappears.

The concept still exists today. It was refined by the Cooper-Bauck technology to make the best listening area about the same as you would expect with a normal stereo system with a relatively narrow sweet spot. But the claims are that this sound envelops the listener in acoustic space. It is used by several different companies and modified for use with surround sound, car audio, computers, games, and so on[6]. One such company is Harman International.

Ambiophonics (not to be confused with Ambiphonics) claims a different approach. An out-of-phase component is injected into each channel to cancel out the inter-aural time delay. However, the cancellation component from each speaker is heard by both ears.

Ambiophonics claims that further computer processing produces very realistic effects even with two widely spaced front speakers. In surround recordings, side and rear channels can exhibit similar crosstalk, and it was found that simulated

algorithms for concert halls were a satisfactory substitute. These have no time relationship to the front channels, and many surround speaker locations around the sides and back are preferred.

Although these concepts produce an exciting and engaging sound field, they still do not reproduce a real life experience, only different or alternate realities. In addition, a real concert hall provides other sensory input such as viewing the musicians on the stage that can help locate the sources of sound and at the same time offer a visual distraction as well.

MASKING

If the ear is exposed to two different sounds at the same time and one of the sounds is very loud, the second sound is "drowned" and cannot be heard. This is termed masking and the very loud sound masks the other sound. **Figure 7** shows the threshold for various sound levels at a frequency of 1200Hz[7]. The threshold tends to follow a class III one-third octave shape but is skewed toward the upper frequencies.

This effect was particularly noticeable for me when listening for distortion. When doing a sine-wave sweep for distortion in a tweeter, I was able to hear a third harmonic distortion product just above the masking threshold very clearly, but if the distortion was only a dB or so less, it was clearly not audible.

LOUDNESS

Unlike many test signals and test equipment used today, the ear responds very differently. For instance, Fletcher-Munson curves[7] of perceived equal loudness in **Fig. 8** show that the ear is much less sensitive to lower frequencies at lower levels. One good reason is that if you have ever tried to make a recording outdoors on a windy day, the wind noise will easily overdrive the recorder because the microphone responds to low frequencies at low levels—something you barely hear and which, again, benefits your survival.

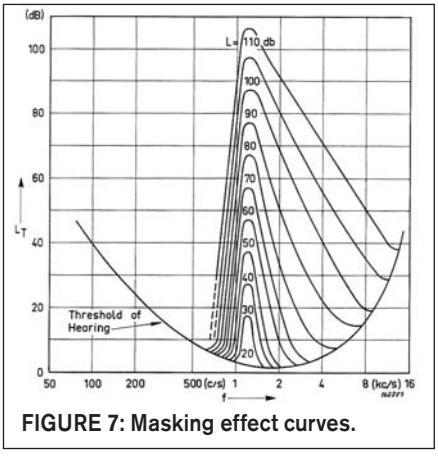

FIGURE 7: Masking effect curves.

The inverse of the loudness curves indicates that the most sensitive area of hearing at lower levels is in the range of about 400 to 8000Hz and is greatest in the 3000 to 4500Hz range.

Part 2 takes a look at how our hearing mechanism works.

REFERENCES

1. "O, O, O, That Concert Hall Realism Rag," Craig Stark, *Stereo Review,* Nov. 1971, p. 75.
2. "Sound System for Pilots Turns Heads," *Design News*, Dec. 17, 1990.
3. "The Influence of a Single Echo on the Audibility of Speech," Helmut Hass, *Journal* of the Audio Engineering Society, Vol. 20 (Mar. 1972), pp. 145-159.
4. "Sineward Distortion," John W. Campbell, Jr., *Audio,* Nov. 1961, p. 52.
5. "Stereo Hearing," Floyd Toole, *Stereo Review*, Jan. 1973, p. 75.
6. "Two Speakers are Better than 5.1," Alan Kramer, IEEE *Spectrum*, May 2001, pp. 71-74.
7. "Acoustic Noise Measurements," Bruel & Kjaer, Copenhagen, Denmark, Jan. 1971, p. 27.

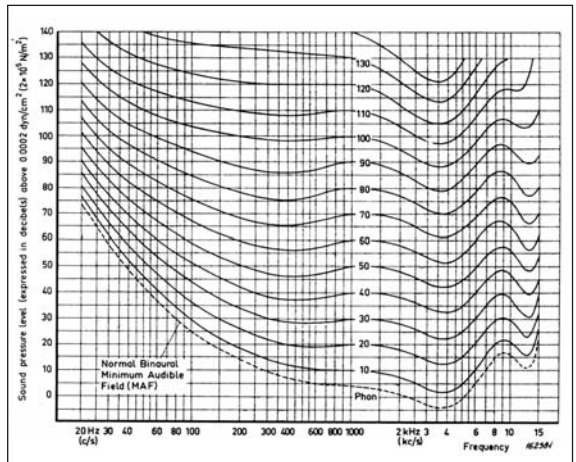

FIGURE 8: Loudness contours.