From:

The New Stereo Soundbook

F. Alton Everest and Ron Streicher

1992

3
Stereo and the auditory system

PERCEPTION OF THE STEREO EFFECT DEPENDS ENTIRELY ON THE BRAIN'S processing of the acoustical cues received by the ears. The cues generally received by one ear are slightly different than those received by the other. The signals at the two ears are incoherent, to some degree. The signals differ in intensity and timing. The stereophonic edifice rests squarely on these differences, although they are usually extremely small and often difficult to isolate.

What the brain does and how it does it is the province of experimental psychology and physiology. Even a superficial understanding of the way our auditory system interprets stereo cues requires venturing into the psychological and physiological ongoing research and continually refined concepts.

Subjective/objective distinctions

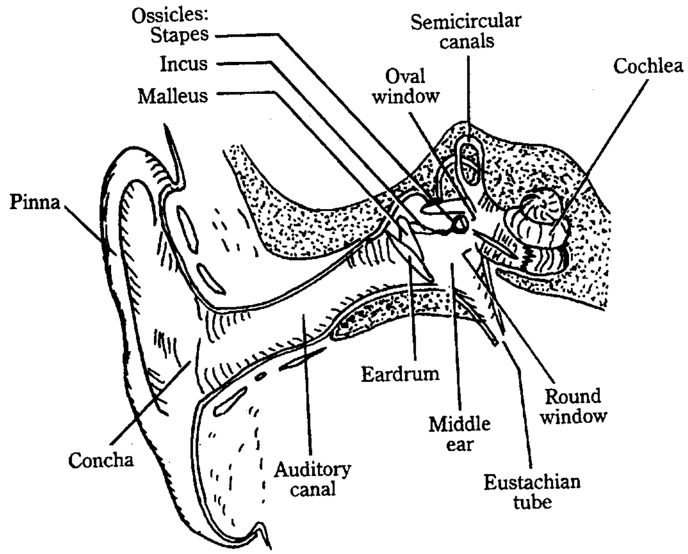
If a rock tumbles down from the heights in the middle of a great uninhabited wilderness, does sound exist if no ear is there to hear it? The physicist would certainly say yes because sound waves can be measured, even from a distance, as they travel out from the point of impact. The psychologist might say no because no human auditory system is present to receive the sound waves. Therefore, it is a matter of definition. Sound could be considered either as a stimulus or a sensation. The stimulus is a physical event, but a living auditory system is necessary to transform the stimulus into a sensation.

Loudness is a subjective term. In the objective realm of the physicist, its *alter ego* would be sound pressure. The two are related, but they are not equal. A musician's pitch has a physical counterpart called frequency. Again the two are related, but they cannot be equated. The musician uses the word *timbre* to describe the richness of tones produced by musical instruments; however, the physicist might prefer the word *spectrum*. Though the subjective and objective concepts might seem to overlap, the distinction between the two outlooks should not be forgotten.

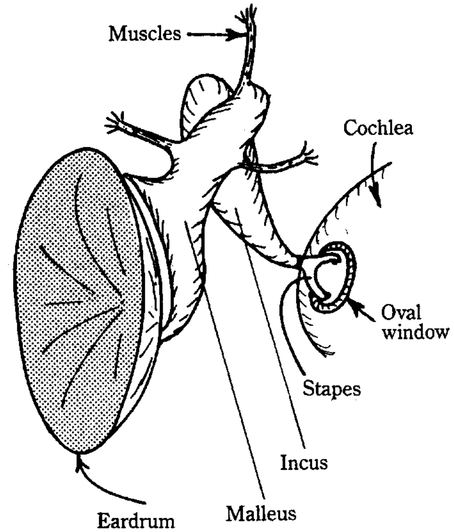
Structure of the ear

Though knowledge of the human auditory system is imperfect, the tantalizing glimpse we have of it is enough to cause a sense of awe and wonder. Even a partial knowledge of the ear's intricate workings is valuable to the practicing musician or engineer.

A highly simplified sketch of the parts of the ear is shown in Fig. 3-1. Sound in the form of vibrations of the air is gathered by the outer ear, or *pinna*. After traveling through the auditory canal the sound-pressure fluctuations cause the eardrum, or *tympanic membrane* to vibrate. In the middle ear, the vibrations of the eardrum activate the three ossicles (*malleus, incus,* and *stapes*) into rocking movement (Fig. 3-2). The mechanical vibrations of the ossicles are transmitted to the fluid-filled inner ear, or *cochlea* (Fig. 3-3). Membrane vibrations of the cochlea agitate the hair cells, which send verve impulses to the brain via the auditory nerve. The membrane, hair cells, and the nerve fibers provide a high degree of selectivity in analyzing the sound as they convert the mechanical movements of the middle ear to neural activity.

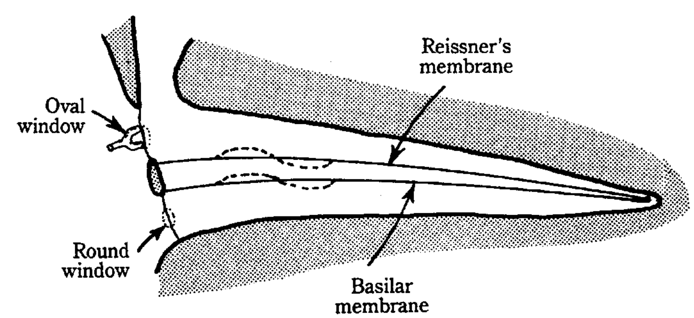


3-1 The sound entering the auditory canal actuates the eardrum. The eardrum movement is conveyed mechanically to the oval window of the cochlea by a linkage of three tiny bones, the ossicles. The movement of the oval window sets the fluid of the inner ear vibrating, which in turn, stimulates the sensitive hair cells that are projections of the auditory nerve.



3-2 The three ossicles, the malleus, the incus, and the stapes, mechanically transduce the vibrations of the eardum to the oval window of the inner ear. In essence, the action of these tiny bones is to provide maximum transfer of energy from the air to the liquid of the inner ear.

Cochlea (unrolled)

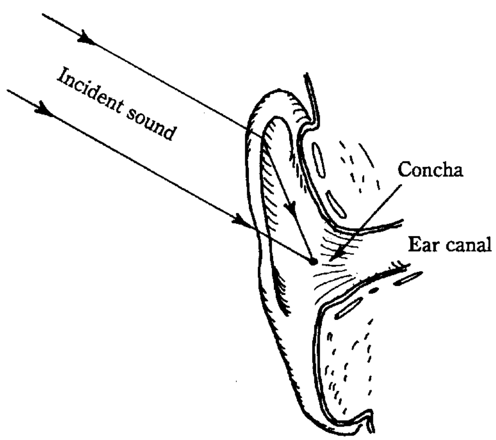


3-3 The vibrations of the oval window by the ossicles give rise to traveling waves in the fluid of the inner ear. These traveling waves stimulate the hair cells on the membranes that transmit impulses to the brain via the auditory nerve.

The pinna

The pinna used to be thought of as a vestigial organ with little or no practical value, but this idea seems obsolete today. The pinna gathers sound and differentiates to some extent between sounds from the front and sounds from the rear. This has slight localization value at frequencies important to speech. Suspicions that the pinna provides more significant localization cues than this were aroused when experimenters discovered that one-eared listening could reveal the direction from which sounds were coming.

Figure 3-4 illustrates how a wavefront, represented by the two sound rays, might enter the ear canal—one directly and the other by reflection off a fold of the pinna. The direct and reflected components come together at the entrance to the ear canal. This combining is called *interference* and results in *comb filtering*, which will be considered in detail in chapter 8. This constructive and destructive interference at the entrance to the ear canal results in significant alterations in sound pressure at the eardrum. The interference effect is not the whole story, however. Resonances within the concha are selectively excited by sounds from specific directions. Resonances produce directional cues in the form of sound pressure changes at the eardrum.



3-4 The sound pressure at the eardrum is greatly affected by interference and resonance effects at the pinna and concha. Sound arriving from different directions results in sound pressure variations at the eardum, which the brain interprets as directional cues.

The auditory canal

The auditory canal, or *auditory meatus*, could be imagined as a twisted pipe with a changing cross-sectional shape, but acoustically it acts like an organ pipe that is about 3 cm long and 0.7 cm in diameter. The canal is terminated by the eardrum, or *tympanic membrane*. At the frequency at which the canal is a quarter wavelength long, a significant acoustical amplification occurs at the eardrum. This amplification is one of sound pressure only—there is no increase in energy level—but it is useful to pressure actuated devices like the eardrum. Compared to the sound pressure at the entrance to the auditory canal, the sound pressure at the eardrum is increased by this pipe resonance by about 10 dB, or 3-fold in the 2 – 4 kHz region. A further amplification effect occurs due to the diffraction of sound waves around the head. Adding the effects of head diffraction and resonance of the ear canal, a total acoustical amplification of sound pressure at the eardrum could be as much as 20 dB, or 10-fold (Reference 3-1).

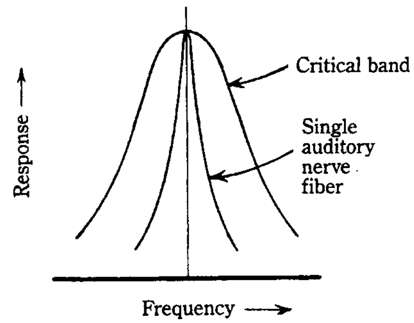
The middle ear

The malleus is attached to the eardrum (Fig. 3-2). The vibrations of the eardrum are transmitted to the three ossicles, which transmit them in turn to the oval window of the cochlea. This mechanical action of the hammer, anvil, and stirrup (*malleus, incus,* and *stapes*) transmits the vibrations of the air in the auditory canal to the liquid of the inner ear with maximum efficiency. If airborne sound impinged directly on the oval window of the cochlea, 99.9% of the energy would be reflected, and only 0.1% would reach the fluid of the cochlea. The reason is because air is tenuous and compressible, but the waterlike fluid of the cochlea is dense and incompressible. This remarkable ossicular arrangement, along with other factors such as difference in areas of eardrum and oval window, solves the difficult energy-transfer problem. To the technical person it is a matter of matching the grossly different acoustical impedances of air and liquid by a mechanical system. The person interested in loudspeakers can learn that the eardrum operates as an *acoustic-suspension* system. The *eustachian tube* equalizes static air pressure on both sides of the eardrum. It is normally essentially closed. Therefore the movement of the eardrum is opposed by the springiness of the air trapped in the middle ear.

The inner ear

The stirrup bone, or *stapes* of the middle ear is attached to the oval window of the cochlea, which is part of the inner ear (Fig. 3-2). Vibrations of the eardrum transmitted by the ossicles cause traveling waves to build up on the basilar membrane in the cochlea (Fig. 3-3). The position of the maxima of these traveling waves is a function of frequency. Movements of the basilar membrane excite the sensitive haircells attached to it. When stimulated, the hair cells, or extensions of auditory nerves, send electrical impulses to the brain.

The cochlea is a sound-analyzing mechanism capable of amazing pitch and frequency discrimination. The ear can differentiate between a tone of 1,000 Hz and a tone of 1,003 Hz (a discrimination of 0.3%). This precise distinction between frequency components of music or speech signals requires a narrow peak response with steep sides as shown in Fig. 3-5.



3-5 The cochlea is the sound analyzer of the auditory system. The finest analysis is that provided by the sensitive hair cells. A *tuning curve* of a single hair cell is here compared to that of a so-called *critical band*, which is the auditory filter involved in masking of one sound by another.

Masking and the auditory filters

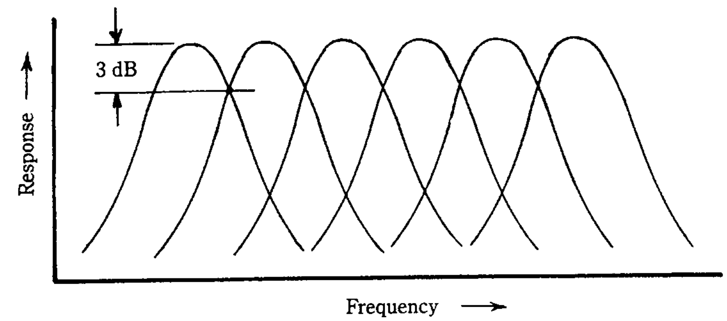
"I can't hear you when the water's running" is a statement about sound masking. Water sound obscures the sound of the voice, road noise in an automobile can drown out low-level radio music, factory noise can make conversation difficult, and excessive noise from a poorly designed air-conditioning system can degrade the intelligibility of dramatic speech from the stage of a theater.

The American Standards Association has defined masking in two ways:

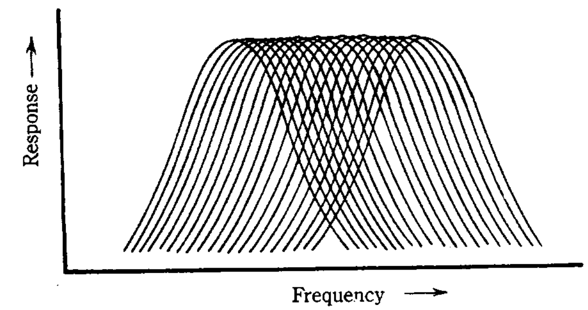
- The *process* by which the threshold of audibility of one sound is raised by the presence of another masking sound.
- The *amount* by which the threshold of audibility of a sound is raised by the presence of another masking sound.

In early masking experiments it was found that a signal was most easily masked by a noise having significant energy at the same frequency as the signal. H. Fletcher introduced the *critical band* concept suggesting that the auditory system behaves as if it contains a bank of bandpass filters with continually overlapping center frequencies (Reference 3-2). The common audio equalizer has a limited number of bandpass filters spaced in frequency so they overlap at the half-power points or −3 dB as shown in Fig. 3-6A. The filters of the auditory system are continuous, however, with a complete filter for every audible frequency, suggested in Fig. 3-6B. Selected auditory filters vary in width according to frequency, shown in Table 3-1.

For example imagine a desired signal is a 260 Hz tone, which is close to middle C on the piano, and an interfering (or masking) sound is *white noise* (which has energy uniformly distributed throughout the entire audible band). If an auditory bandpass filter is centered on the 260 Hz tone, according to Fletcher's theory, only



(A)



(B)

3-6 (A) The characteristics of a standard 1/3 octave filter set of the type used in acoustical measurements. Adjacent filters cross over at the half-power point or at −3 dB. (B) The filter set of (A) is coarse and crude compared to the auditory filters (critical bands) of the human hearing system. Auditory filters are continuous, i.e., a sound of any frequency would encounter an auditory filter centered on it.

the white noise energy passed by that filter will be effective in masking the tone. By increasing the level of the noise, a point will eventually be reached at which the tone is masked and becomes inaudible. At this point all that can be heard is noise.

Binaural unmasking

Binaural interactions are important because the stereophonic illusion depends on the differences of sounds falling on the two ears. A psychoacoustic experiment

Table 3-1.
Auditory filter bandwidths

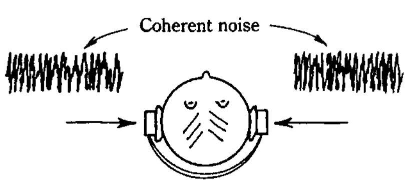
Center frequency Hz	Width of critical band* Hz
100	38
200	47
500	77
1,000	128
2,000	240
5,000	650

*Calculated equivalent rectangular bandwidths as proposed by Moore and Glasberg, (Reference 3-3)

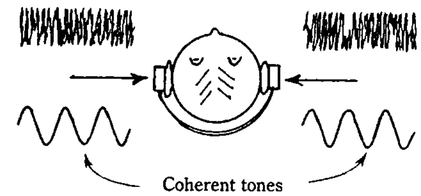
demonstrating *binaural unmasking* might shed some light on interaural incoherency. In this experiment the listener wears headphones that shut out all sounds except those that pertain to the experiment. Coherent noise produced from one noise signal generator is sent to both the left and the right ears (Fig. 3-7). Then coherent tones set to a frequency of 260 Hz (close to middle-C) are added to the noise and the combination is adjusted to a comfortable volume level (Fig. 3-8). Both the tone and the noise are heard in each ear clearly. Next the level of the tone is decreased until it is masked by the noise. The subject then hears only the noise.

The noise in the left ear is coherent to the noise heard in the right ear because they come from the same noise generator. The tones heard by the two ears are also coherent because they too come from the same tone generator.

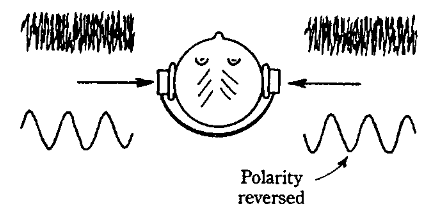
If the two leads from the tone generator to the right ear are reversed, the tones in the two ears are made incoherent, but the noise sent to each ear remains coherent (Fig. 3-9). The frequency and level of the tones to each ear remain the same, but reversing the polarity of the tone to the right ear makes the tones incoherent. At the instant the polarity reversal of the right ear tone takes place, the right ear tone, which was previously masked and inaudible, suddenly springs into clear audibility. Making the tones to the two ears somewhat different or incoherent, actually improves the audibility.



3-7 Steps in the study of binaural unmasking; coherent noise from the same signal generator is first applied to both left and right headphones.



3-8 Coherent tones of 260 Hz (close to middle C) are added to the noise of Fig. 3-7 so that both tones and noise are at a comfortable level. The level of the tone is now decreased until it is just masked by the noise.



3-9 With the tone just masked by the noise (Fig. 3-8), the two leads from the tone generator to the right ear are simply reversed. As this is done, the tone to the right ear (which was masked) suddenly becomes audible. Introducing an incoherency by reversing the leads improves the audibility of the tone.

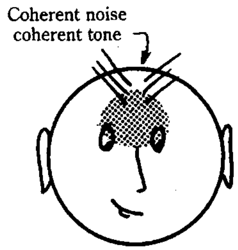
The tone in this experiment represents any desired signal—such as speech or music. The noise represents any signals that interfere with hearing the desired speech or music. This experiment would apply to music interfered by air-conditioner noise as the masker, but measurements would be more difficult due to the fluctuation of the music signals.

The amount of binaural unmasking can be measured. Going back to the combined tone and white noise signals sent to the two ears in Fig. 3-8, the tone level again is reduced until the tone is just masked. When the wires connecting the tone generator leading to the right ear are reversed, the tone suddenly becomes audible. Then the tone level is decreased again until it is masked, but this time the amount of the decrease is measured. To overcome the effects of incoherency, the tone had to be reduced about 15 decibels, called the *Binaural Masking Level Difference*, (or BMLD).

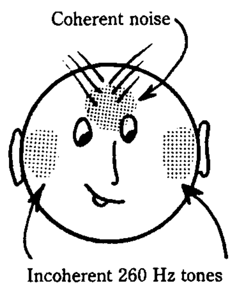
This improvement in audibility results simply from introducing a difference in the tones to the two ears, which is astounding. Investigations like this can help us understand how slight differences between the signals to the two ears (interaural incoherencies) are interpreted by the brain to give us stereo image perception.

Perceptual space

Another way of looking at the binaural unmasking effect is to observe that the coherent white noise appears to be located in the center of the head (Fig. 3-10). When the coherent 260 Hz tone is added to the coherent noise, both seem to be located in the center of the head. Again the level of the tone is decreased until it is masked. As the polarity of the tone to the right ear is reversed, the tone suddenly becomes audible and its perceptual image seems to split, appearing in the vicinity of both the left and right ears. At this point the coherent noise image is in the center of the head, and the incoherent tonal image is at both ears as in Fig. 3-11 (Reference 3-4).



3-10 A mixture of coherent white noise and coherent tones in both ears is perceived in the center of the head.



3-11 A mixture of coherent noise and incoherent tones is perceived as coherent noise in the center of the head, and the incoherent tones near both ears.

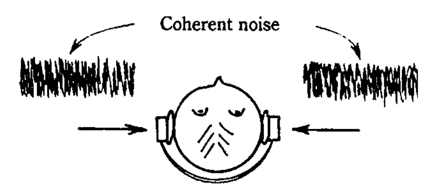
Therefore masking appears to be more effective when the masking noise and the masked tone are localized in the same perceptual space within the head. Unmasking only takes place when the images of the two appear in different parts of the head. Coherent sounds seem to be conceptualized in the center of the head, but incoherent sounds form near the ears. Thus the important relationship between coherency and masking can be visualized.

The complete relationship between stereo and the perceptual location of sounds in different parts of the head is obscure. The stereo illusion is based on specific localizable images and highly correlated sounds at the two ears, perhaps differing only in timing or amplitude.

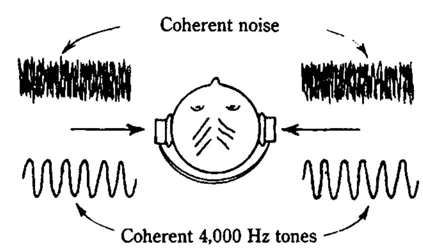
Binaural unmasking vs. frequency

Binaural unmasking works well at 260 Hz, but does it work at other audible frequencies? The results of the same experiment at 4,000 Hz, a high treble frequency near the top of the piano keyboard, indicate that frequency does have a major effect on binaural unmasking. Again white noise is introduced to both left and right earphones (Fig. 3-12). Then a coherent tone with a frequency of 4,000 Hz is introduced to both ears, clearly audible in the white noise (Fig. 3-13). Slowly the level of the tone is reduced until it is masked in both ears. The polarity of the tone to the right ear is reversed, but this time the tone does not spring into audibility; in fact it barely becomes audible (Fig. 3-14). The 15 dB unmasking at 260 Hz shriveled to 2 or 3 decibels at 4,000 Hz.

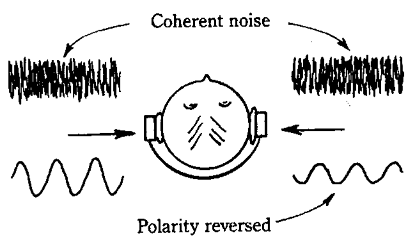
Also resulting from the change to higher frequency, the coherent noise and the incoherent 4,000 Hz tone localized in the same perceptual space—the center of the head (Fig. 3-15). The loss of the binaural unmasking effect is associated with the loss of the incoherent tone's image shift to the ears.



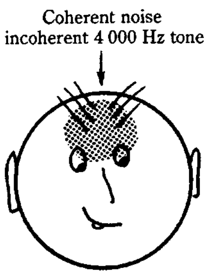
3-12 A repeat of the experiment on binaural unmasking, this time with a frequency of 4,000 Hz. Once again the coherent white noise is applied to both earphones.



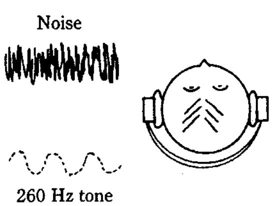
3-13 Coherent 4,000 Hz tones are applied to both earphones at a comfortable level and then reduced until they are just masked by the noise.



3-14 As the tone is made incoherent by reversing the two leads, it is made audible (i.e., unmasked), but just barely so. Binaural unmasking, so effective at 260 Hz, is of minor effect at the frequency of 4,000 Hz.

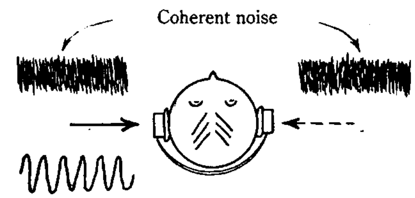


3-15 Interestingly, as binaural unmasking fails at 4,200 Hz, both the tone and the noise are perceived in the center of the head, indicating essential coherency for both.



3-16 In an experiment on the unmasking phenomenon taking place in a single ear, the noise and a 260 Hz tone are applied to the left ear only.

Unmasking in a single ear can be triggered binaurally. This time the noise and 260 Hz tone are fed only to the left ear (Fig. 3-16). The level of the tone is reduced until it is masked by the noise. The noise is introduced only to the right ear (Fig. 3-17). As the coherent noise is introduced to the right ear, the tone suddenly becomes clearly audible in the left ear. This is further evidence of the interaction between the two ears.



3-17 In the unmasking in a single ear (Fig. 3-16) the 260 Hz tone is reduced until it is just masked by the noise. As the same coherent noise is introduced to the right ear, the tone suddenly becomes clearly audible in the left ear. This indicates a definite binaural interaction.

The binaural masking effect seems to work better for the lower frequencies. In chapter 2 other low and high frequency spatial localization effects were noted in stereo. At low frequencies our ears respond to time differences, but at high frequencies they compare both intensity and time differences as shown in Fig. 2-20. It is possible that stereo spatial localization and binaural unmasking are related.

Wide band random noise was used as the masking signal for simplicity, but the mask does not have to be a wideband to demonstrate binaural unmasking. The noise in a critical band centered on the frequency of the tone works just as well. At 260 Hz and 4,000 Hz the widths of the critical bands are about 50 Hz and 500 Hz respectively.

Understanding speech in high background noise

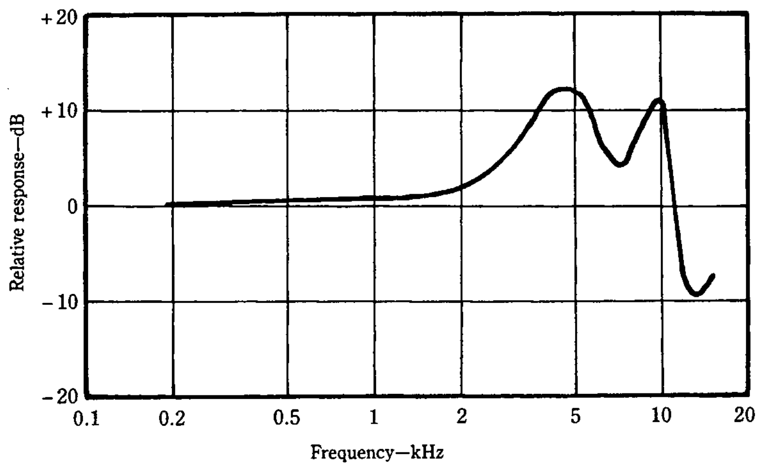
Specialists in the field suspect that the same processes that release a desired signal from being masked also help in understanding speech in situations with high background noise. At a party, the human auditory system is capable of picking out and understanding a single voice from a confusing babble of other voices. When facing the talker, the desired speech is coherent in the ears of the listener, while the background noise coming from all directions tends to be incoherent. Thus, the desired signal and the interfering masking noise are perceptually localized in different places in the head. The time and level differences between them trigger the release from masking, which makes the speech intelligible.

When someone speaks in high background noise, attention is focused on what is being said. Subconsciously the two-eared, coherent and incoherent, correlated and uncorrelated magic occurs. It is ready to work at any moment, and it is scarcely thought about.

Transfer functions

The human brain has only sound pressure cues acting on the eardrum to work with when giving a perception of stereo images, localization of sound sources, etc. The eardrum pressure is made up of two components—one fixed and one variable. The fixed component is the auditory canal itself, and the variable component is the sound entering the auditory canal, which carries all of the directional information.

The resonances of the auditory canal, or the fixed component, are described by the curve of Fig. 3-18. The resonances could be called the frequency response of the auditory canal, but research literature uses the term, *transfer function*, which also includes time of arrival information. The difference between the sound pressure transfer function measured at the entrance of the ear canal and that at the eardrum yields the transfer function of the ear canal alone (Fig. 3-18). A first resonance at about 5 kHz and a second resonance at about 10 kHz result from the physical parameters of the ear canal alone. Because these are fixed the transfer function of the ear canal also remains a fixed entity.

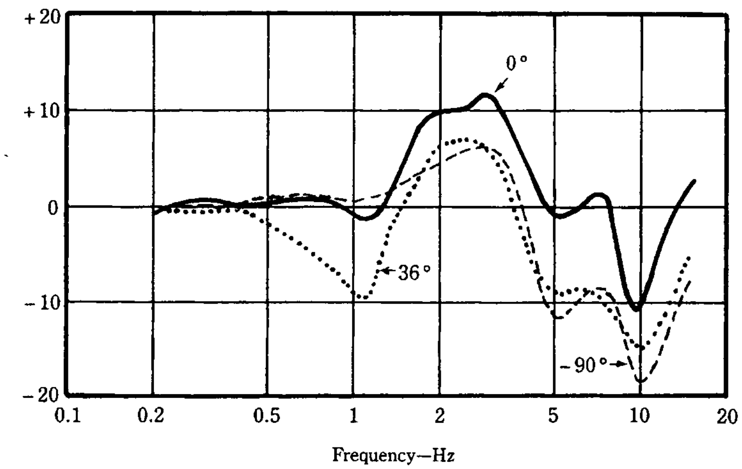


3-18 The auditory canal acts as a resonant pipe. By measuring the transfer function (frequency response) both at the entrance to the canal and at the eardrum and subtracting one from the other, the transfer function of the auditory canal alone is found. The transfer function of the auditory canal is a fixed entity that is superimposed on each of the host of highly variable transfer functions of the outer ear containing directional cues. After Mehrgardt and Mellart, Reference 3-5.

Typical transfer functions of the outer ear are shown in Fig. 3-19, as reported by Mehrgardt and Mellart (Reference 3-5). The sounds falling on the pinna result in a myriad of transfer functions images of different shapes combining at the entrance of the ear canal. Sound source directional information is encoded in these contorted transfer functions.

Sound arriving at the outer ear from directly in front of the listener yields the heavy line transfer function of Fig. 3-19. Sound arriving from 36 degrees to the left of the straight-ahead position yields the transfer function for the left ear labeled *36°*. Sound arriving directly from the left gives the transfer function marked *90°* for the left ear. Combining these highly variable transfer functions at the entrance of the ear canal with the fixed transfer function of the ear canal yields the total sound pressure at the eardrum, complete with all directional cues. This subject is treated in greater detail in chapter 6.

The *0-degree* curve of Fig. 3-19 explains the common practice of *equalizing for presence*. Intelligibility of speech with musical background is improved by an equalization boost in the 2 to 5 kHz region, making the voice more up front and distinct from the musical background.

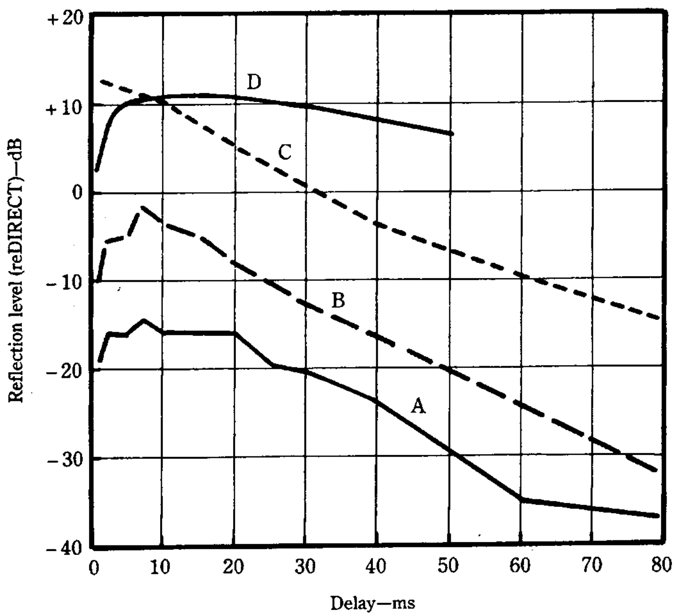


3-19 Typical transfer functions of the outer ear measured between the free field and the entrance to the ear canal. The transfer function for sound arriving from directly in front of the subject (0 degrees), is different from that for sound arriving 36° to the left or from the left (90°). Directional information seems to be encoded in these directionally unique *frequency responses*. After Mehrgardt and Mellart, Reference 3-5.

Effects of reflected sound

Sound travels from a source to a human ear over a direct path. Reflections of the same sound could arrive at the ear over numerous reflected paths. The direct sound arrives first, and a host of reflections arrive later. Each reflection is delayed by an amount of time determined by the distance traveled between the source and the listener's ears. Thousands of reflections are spread out in time and arrive after the direct sound. The human reaction to these reflections depends on the magnitude and the delay. Sometimes the reflections are helpful, but other times they are disastrous.

Many investigations were conducted to study the perceived effect of delayed or reflected sound on a desired signal. The results of some of these studies are summarized graphically in Fig. 3-20 (References 3-6, 3-7, 3-8). All of these experiments used an arrangement of two loudspeakers and a listener—similar to the standard stereo arrangement. The angle between the loudspeakers varied from 45 to 90 degrees. Speech was used as the test signal for these experiments. The studies were conducted under anechoic conditions to minimize extraneous effects.



3-20 The results of several investigations on the effects of lateral reflections on the perception of the direct sound in a simulated stereo arrangement. Type of sound: speech, environment: anechoic, angle of lateral reflections: 45 – 90°. (A) Absolute threshold detection of reflection. After Olive and Toole, Reference 3-6. (B) Image shift threshold. After Olive and Toole, Reference 3-6. (C) Lateral reflection perceived as discrete echo, After Meyer and Schodder, Reference 3-7. (D) Equal loudness of lateral reflection and direct. After Haas, Reference 3-8.

One loudspeaker radiated the direct speech signal. Another loudspeaker, representing the lateral reflection, radiated the same signal delayed by various amounts of time. The effects were documented as the level of the lateral reflection was increased from below threshold.

The level at which the observers first heard something is shown in curve (A) of Fig. 3-20. This was the threshold at which the presence of the reflection was first detectable. A sènse of spaciousness was imparted as the reflection level was increased. The effect of lateral reflections in music halls is a sensation of spaciousness added to the music. In the experiment observers experienced an impression of spaciousness, even though they were in an anechoic room. No particular sense of direction was associated with this threshold perceptual effect.

As the level of the reflection was increased by about 10 dB above the threshold of (A), the observers became aware of a change in size or position of the primary auditory image. Curve (B) is a threshold of this new effect.

Curve (C) delineates the threshold of a distinct echo apart from the primary image. This echo was heard along with the direct sound.

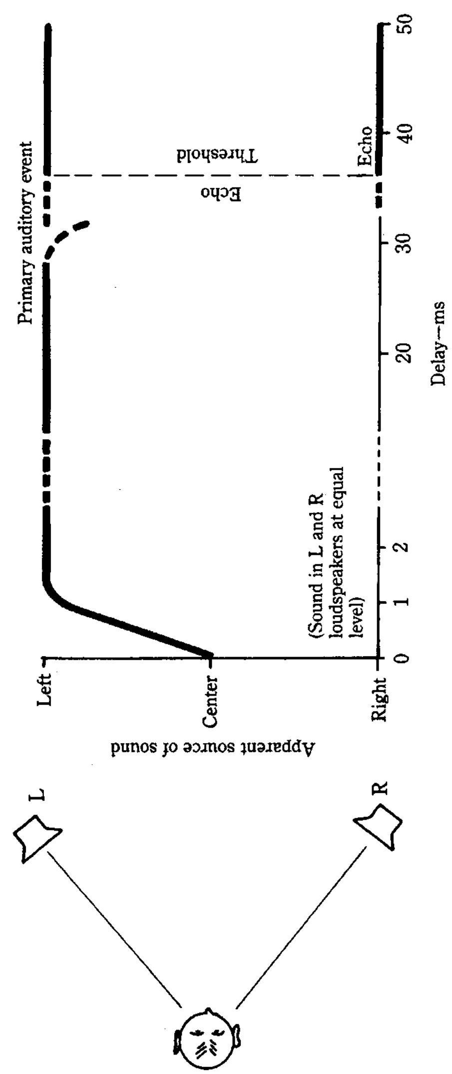
Then reflection levels were increased until reflected and direct sounds were equally loud shown in curve (D), which is the well-known Haas curve. Haas reported some interesting observations during this investigation on the effect of a single reflection on speech (Reference 3-8). He found that the level of the delayed sound could be up to 10 dB higher than the direct sound without sounding louder than the direct sound, although the intelligibility of speech was degraded. This effect persisted for delays up to the vicinity of 30 ms. At the same time increased loudness was observed as well as a change in timbre, image shift, and a sense of spaciousness.

Summarizing, Fig. 3-20 pictures the range of perceptual effects resulting from progressive increases in the level of the reflection. Above the threshold of (A) reflections cause a sense of spaciousness. Increasing reflection level even more, another threshold (B) of changes in the auditory image comes into play along with spaciousness. Further increase in reflection level reaches the threshold of discrete echoes at (C). Higher reflection levels reach curve (D), at which the reflection and the direct sound are of equal loudness. These threshold curves are spaced by roughly 10 dB.

Various timbral effects occurred as the reflection level was increased, especially at higher reflection levels, but the spatial and directional perceptions dominate.

Law of the first wavefront

The domination of the earlier sound has been called *the law of the first wavefront*, by Cremer. This concept is extremely important when listening in enclosed spaces filled with a welter of sound reflections. The direct sound from a source in an enclosed space arrives at the listener's ear earlier than sound reflected from surfaces. Therefore the *direction* to the sound source is perceived clearly—even in reverberant spaces. Delays within a 0 – 1 ms range move the sound from the center to the earlier loudspeaker as shown in Fig. 3-21. The law of the first wavefront can be the working equivalent of the precedence effect.



3-21 As the sound to the right loudspeaker is delayed from 0 to 1 ms, the image shifts from the center to the left loudspeaker illustrating the law of the first wavefront. As the delay exceeds 35 ms, the sound appears to come from both loudspeakers with a discrete echo.

Echoes

As the sound in one loudspeaker is delayed approximately 30 ms, the delayed sound begins to take the form of a discrete echo. There is a transitional region in which the incipient echo is evident, but not annoying (Fig. 3-21). The echo does not become fully discrete and really annoying until the delay reaches 40 or 50 ms. Beyond this point the echo is definitely both discrete and annoying. This echo threshold is the upper limit of validity of the law of the first wavefront.

A common problem in an auditorium results from sound from the main loudspeakers being reflected from the rear wall or balcony face. The reflected sound often reaches those sitting in seats near the front with a sufficient delay to create a definite echo with accompanying intelligibility problems. For people sitting further back, however, the reflected energy could actually enhance their sound.

Often sound from local supplemental loudspeakers cause the main loudspeakers to sound like an echo because of the differential path length between the two sources and the listener. By delaying the signal to the local loudspeakers, it is possible to achieve an increase in level without the echo.

Effect of reflections on music and speech

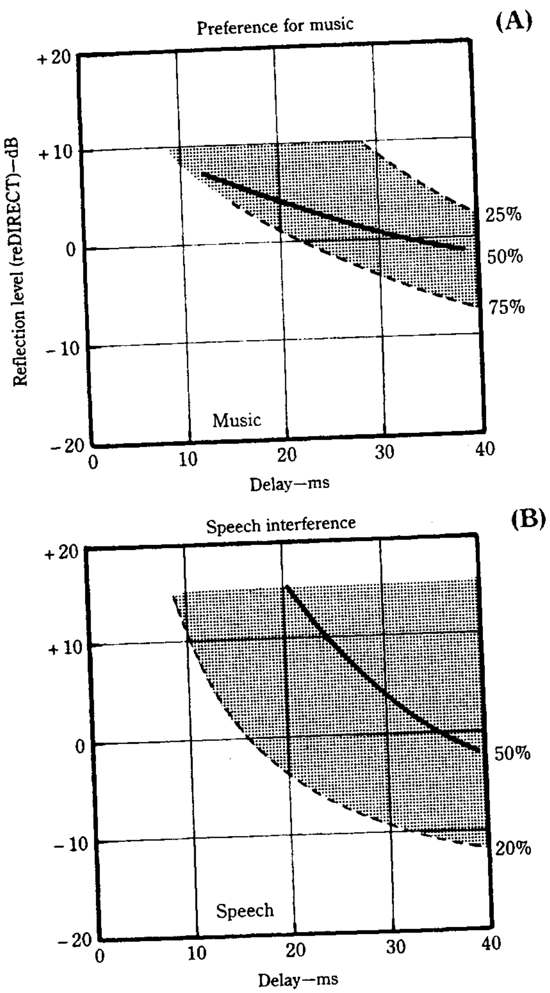
Early lateral reflections can help or harm the perception of desired sound. The results of Y. Ando's study documented in *Concert Hall Acoustics*, (1985) considers the effects of lateral reflections on the enjoyment of music, shown in Fig. 3-22A (Reference 3-9). Observers were asked to state whether they preferred the music with or without the spaciousness provided by a single lateral reflection. Half the observers preferred the reflection level and delay indicated by the heavy line. The upper broken line is for 25% and the lower broken line for 75% of the observers preferring reflections of the specified level and delay.

H. Muncie approached a similar problem by asking the observers whether the single lateral reflection interfered with speech (Reference 3-11). The response of half of the observers is represented by the heavy line in Fig. 3-22B. The broken line gives the response of 20% of the observers. The 80% line is off this graph.

The coordinates of Fig. 3-22A and B are identical, inviting the reader to imagine one superimposed on the other for a comparisón of the effect of spaciousness on music and speech. The shaded area for music and that for speech interference partially overlap, indicating that some listeners prefer the spaciousness provided by a single lateral reflection for music, but not for speech.

In the average listening room interest in stereo reproduction is focused on reflections delayed in the approximate range of 10 – 20 ms at a level roughly 5 – 10 dB below the direct sound, depending on the geometry and reflecting properties of the surfaces. These figures describe an area of Fig. 3-22A somewhat below the 75% curve. If a 100% curve existed, it might actually include this area, which would indicate that everyone would like the early reflections in this listening room.

Normal reflections from the walls may be ideal for the sound of a live musician in the listening room but intolerable for reproduction of music having spaciousness already recorded with the music.



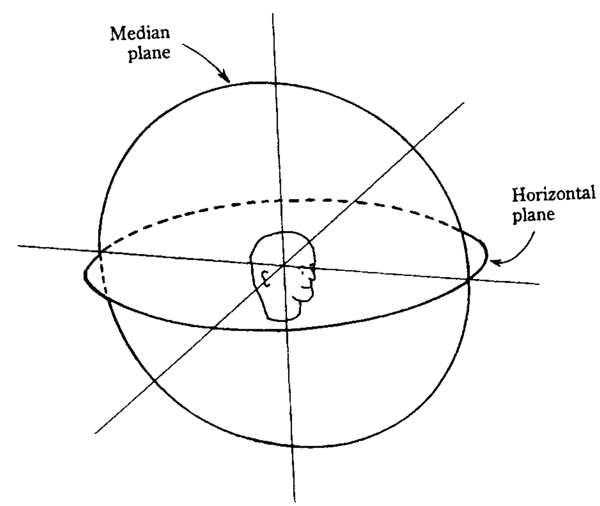
3-22 (A) The results of a study showing the preference for a single lateral reflection on the sound of music. The percentages indicate the proportion of the listeners who preferred the lateral reflection. After Ando, Reference 3-9. (B) The results of a study showing the disturbance of speech by a single lateral reflection. The percentages indicate the proportion of the listeners disturbed. After Muncie, Reference 3-11.

The spaciousness perceived in these tests result from the comb filters produced by combining of lateral reflections with the direct sound. *Combing* is basically a steady-state effect. Music is mostly transients; however, combing of transients results in something interpreted as spaciousness. Ando has demonstrated that people prefer music with spaciousness.

Localization of sound in the median plane

A major emphasis has been the left-to-right localization of sound sources in the horizontal plane. The next natural issue to address is the possibility for a stereo system to yield perceptions of sound arriving from overhead and back or any other point on the hemisphere. The human auditory system is capable of perceiving sound arriving from any direction in the hemisphere. For this to occur appropriate transfer functions must carry the directional information.

The *median plane* is that imaginary vertical plane passing through the head and nose, at right angles to the horizontal plane (Fig. 3-23). The localization of sound perceived by humans is different in the median plane than it is in the horizontal plane. Sound sources in the median plane result in essentially identical signals in the two ears. No time/intensity cues or interaural dissimilarity aids in perception, which characterizes the horizontal plane. Cues resulting from changes in the spectrum of the sound source are evident, however.

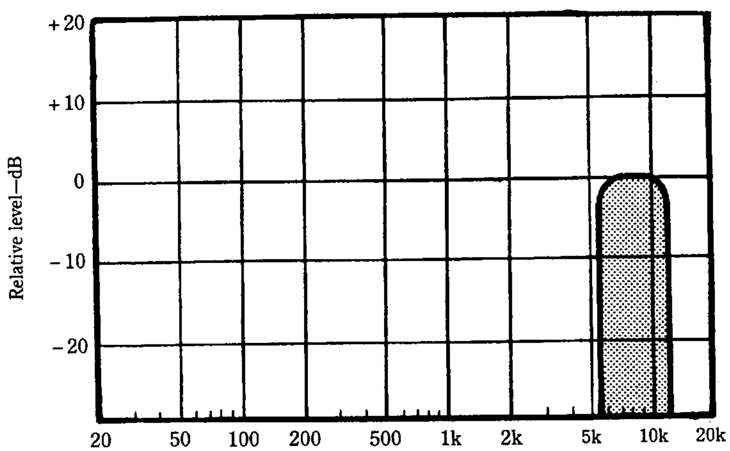


3-23 The median plane is that imaginary vertical plane passing through the center of the head, at right angles to the horizontal plane.

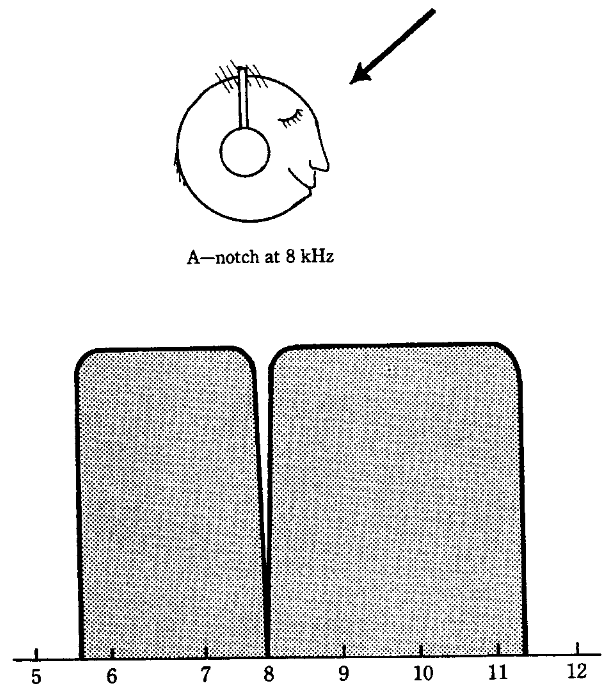
Vertical cues are produced by changes in the signal spectrum introduced by diffraction at head and pinnae, like the localization of sound in the horizontal plane. Certain frequency ranges of the incoming signal are emphasized or depressed by head diffraction or pinna effects, depending on the direction of arrival. For any 3-dimensional sound reproducing system to work, physical cues must faithfully be recorded and reproduced.

The effectiveness of spectral changes to convey perceptions of sound arriving from points outside the horizontal plane was demonstrated by P.J. Bloom (Reference 3-12). It was suspected that deep notches in the transfer functions at about 10 kHz were cues for specific directions of arrival of the sound (Fig. 3-19). An octave band of noise centered on 8 kHz could simulate sound at the eardrum (Fig. 3-24). A deep and narrow notch was introduced that could be moved about. When listening on headphones to this band of noise and the sharp notch at 8 kHz, the noise appears to come down from above (Fig. 3-25). Shifting the notch to 7.2 kHz changes the apparent direction of the source to a horizontal direction (Fig. 3-26). Moving the notch to a frequency of 6.3 kHz makes the sound appear to come from below (Fig. 3-27). The brain seems to use the position of the notch as a directional cue, apparently confirming a relationship between pinna reflections and the perception of direction.

Blauert summarized many experiments conducted on sound localization in the median plane in his book, *Spatial Hearing*, (Cambridge, 1983) (Reference 3-13).



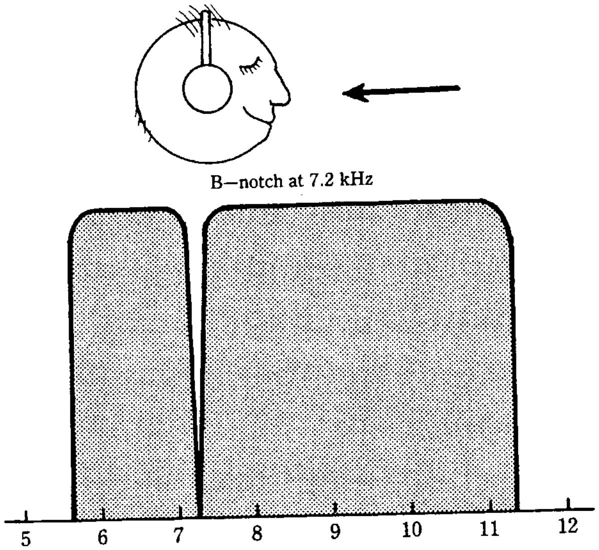
3-24 Spectral changes convey directional information to the ear. To demonstrate this, an octave band of noise is used, which is centered on a frequency of 8 kHz.



3-25 Listening with headphones to the octave band of noise with a sharp notch at 8 kHz gives the impression that the sound comes from above.

His studies show that the direction of the auditory event depends primarily on the frequency of the signal—not on the direction of the sound source. Figure 3-28 illustrates the correlation between the positions in the median plane and the frequency. Certain redundancy was revealed, for example overhead precepts result from regions near both 500 Hz and 8 kHz, and rear percepts are generated from regions near both 1 and 10 kHz.

In Blauert's characteristically thorough fashion, the subject of sound localization in the median plane was studied formally with many subjects. The signals used were 1/3 octave noise pulses. To simplify an already complex study, the subjects were asked only to judge if the sound appeared to come generally from the foreward, overhead, or rear directions. A statistical study of the data gave the per-



3-26 A notch at 7.2 kHz in the octave band of noise gives the impression that the sound comes from a horizontal source.

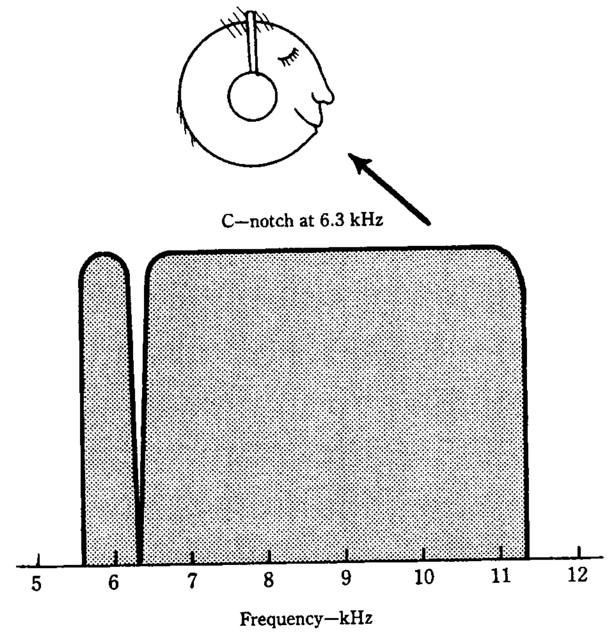
cent of responses corresponding to each frequency. The results of this study are shown graphically in Fig. 3-29 and can be summarized:

Approximate Frequency Ranges For Directional Bands, Hz

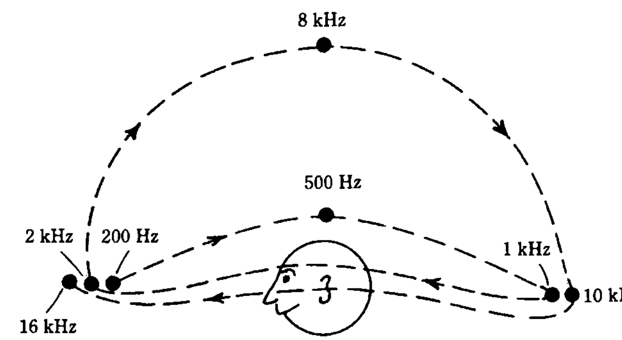
Front	300 - 600	3,000 - 6,000
Overhead	8,000 - 9,000	
Rear	800 - 1,800	9,000 - 15,000

This study suggests the presence of two front, two rear, and one overhead directional bands, though other bands could be revealed with refined measurements. The direction of the auditory event might be correlated with these directional bands on the basis of signal power at specific frequencies.

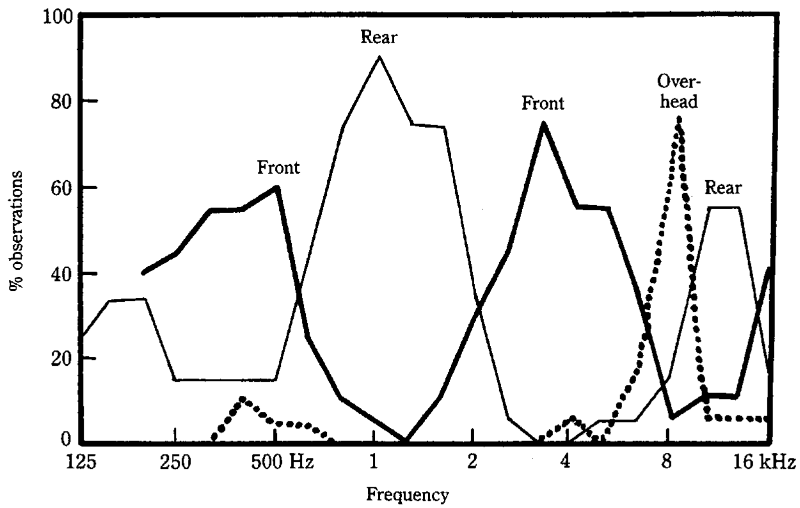
All of these experiments point to the fundamental importance of the transfer functions. Ample reason now exists to hope for a sound recording and reproducing system that will eventually approximate the ideal and ultimate 3-dimensional meaning of the word *stereo*. It will involve accurately recording and faithfully reproducing the transfer functions of signals from all directions.



3-27 A notch at 6.3 kHz gives the impression that the sound comes from below.



3-28 The apparent direction of tonal signals from loudspeakers in the median plane depends on the frequency of the tone. After Blauert, Reference 3-13.



3-29 Localization of 1/3 octave pulses of noise in the median plane revealed two front, two rear, and one overhead directional bands. After Blauert, Reference 3-5.

This optimistic view, however, neglects the significant individual differences in the auditory transfer functions. Head, pinna, and ear canal shapes are like fingerprints—no two individuals are identical. Precise pickup, recording, and reproduction techniques are of no avail if the characteristics of the consumer's auditory apparatus vary within wide limits.