Handbook of Sound Reproduction

EDGAR M. VILLCHUR*

Chapter 8. Elements of Fidelity in Sound Reproduction

A discussion of the characteristics of a reproducing system which must be considered in evaluating the performance in terms of its fidelity.

THE CHARACTERISTICS of a (monaural) reproducing system that determine its ability to reproduce sound faithfully may be categorized as:

Frequency response
Transient response
Harmonic distortion
Intermodulation distortion
Power capability
Noise level
Dynamic range

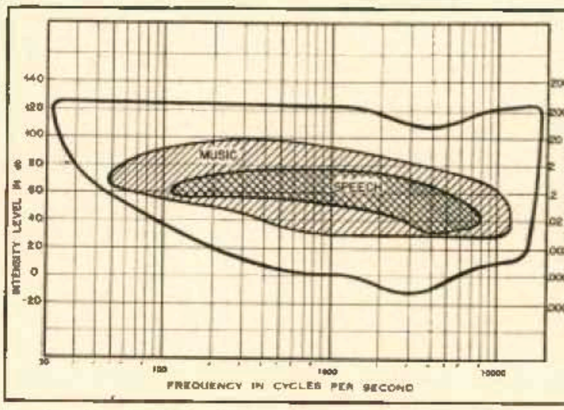

Fig. 8—1. Frequency and intensity ranges of speech and music. The solid line represents the limits of normal hearing. (Courtesy Bell Laboratories Record, June, 1934)

Frequency Response

Frequency response is the most widely discussed factor in sound reproduction, and is often the first technical expression that the audio novitiate learns, although he is frequently led to accept a very incomplete meaning for the term. The frequency response of a system refers to the relative amplitude with which sounds or components of sound, varying only in that they are of different frequency, are reproduced. This response is typically such that a certain band of the frequency spectrum is reproduced with a given degree of uniformity, with output beyond the ends of the band dropping off. When the drop is sharp it is called *cut-off;* when it is gradual it is referred to as *roll-off*.

Thus there are two aspects of frequency response: the range of frequencies covered by the reproduced band with a specified minimum response, and the type and degree of variation which occurs within the band. Except where the range is very inadequate, the second of these aspects is usually the more important. So-called frequency response ratings which merely record the two frequency extremes of the reproduced spectrum may have little relationship to the quality of sound to be expected. Two loudspeakers with the same response at 40 and 15,000 cycles may produce entirely different tonal qualities due to the dips and peaks of acoustical output at less extreme frequencies.

Uniform or "flat" response over the band of reproduced frequencies is desirable when two conditions are satisfied. These are: (1) that the reproduced sound is at the same intensity level (produces the same number of acoustical watts per square centimeter) as the original sound at a normal listener's position, and (2) that acoustical conditions in the room in which the sound is reproduced create more or less the same frequency discrimination as conditions in the original hall. When these two conditions do not exist, as is often the case, a rising bass characteristic to compensate for the Fletcher-Munson effect (see Chapter 6), and some tonal adjustment for discriminatory acoustical conditions should be provided by variable tone controls.[1]

We have seen that intensity differences on the order of a fraction of a db are just discernible at most frequencies and intensity levels, but this perceptive sensitivity is for single, pure tones heard under laboratory conditions. The degree of non-uniformity of frequency response that can be tolerated without effect on the quality of reproduction is undoubtedly somewhat greater, probably on the order of plus or minus one to two db. The indicated standard for maximum variation in frequency response is easily met in electronic circuits, more difficult to achieve in pickups, and, up to the present, an impossible ideal for loudspeaker systems, the very best of which reproduce their spectra with a variation of plus or minus five db. When sections of the frequency band are reproduced with ten db greater amplitude than others, tonal coloration must be expected.

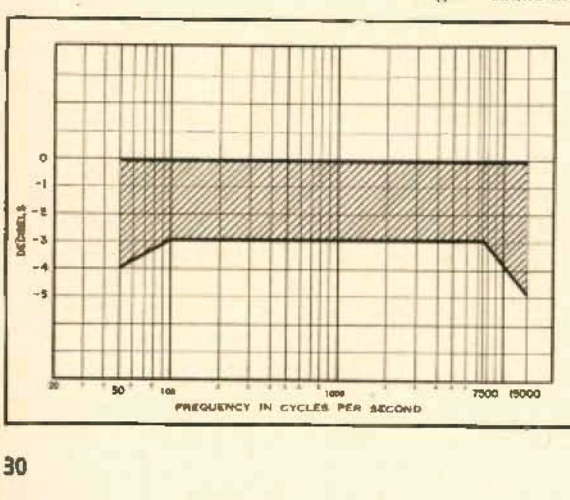

Fig. 8—2. FCC limits of frequency-response variation permitted FM broadcast stations (corrected for pre-emphasis). The range from 100 to 7,500 cps is restricted to a variation of ±1.5 db, while the standards beyond this band are not as strict. (The minimum frequency response standard for AM stations is 100-5,000 cps, ±2 db.)

* *Contributing Editor,* AUDIO ENGINEERING.

[1] The tone controls of an amplifier also provide compensation for associated equipment and program characteristics, but this discussion is limited to the desirable frequency response of the complete system.

The range of frequency response required to reproduce orchestral music without such coloration, and to include all significant overtones, transient effects, and noise, is not a matter for theorizing but for experimental verification. We are not concerned with reproducing all of the acoustical energy of the orchestra for its own sake, but only that part of it which has significance in the perception of quality.

Figure 8-1 is a chart reproduced from the *Bell Laboratories Record*, which includes information on the frequency ranges required for the reproduction of music and speech. This chart agrees closely with the data appearing in Snow's graph of audible frequency ranges for speech and music (see Chapter 5). We may consider the frequency range for perfect apparent fidelity of speech reproduction to be from 100 to 8,000 cps (Snow uses a slightly higher upper limit for female speech), and the range for perfect apparent reproduction of orchestral music as 40 to 15,000 cps.[2] A study in detectable band-width differences[3] has indicated that this last upper limit must be reduced to about 11,000 cps before the change becomes noticeable. It may be inferred that a corresponding liminal (just noticeable) change in the bass limit would probably be to 60 cps or higher. These latter figures may therefore be taken as reasonable standards for a high-fidelity reproducing assembly one step down from perfection as far as frequency range is concerned.

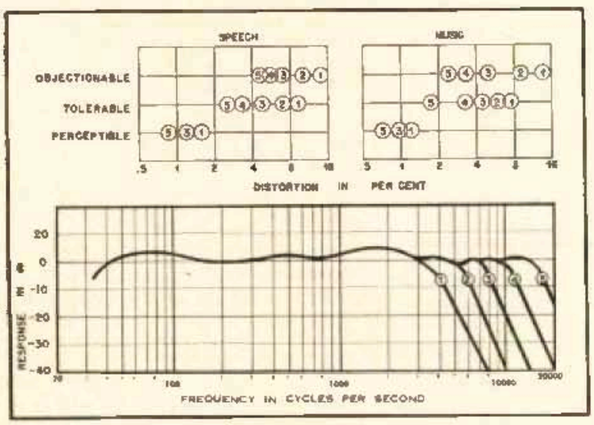

Fig. 8—4. Objectionable, tolerable (for low-fidelity units), and perceptible total rms harmonic distortion in music (produced by triode), for various frequency cut-offs. From Olson, Elements of Acoustical Engineering; courtesy D. Van Nostrand Company, Inc.

The only standard orchestral instruments whose fundamental frequency ranges go below 40 cps are the piano, the harp, the pipe organ, and the contra bassoon. It has been experimentally established that the fundamental energy in the very low notes of the piano is so small as to have no effect on perceived quality, and it is logical to assume that the same is true of the harp. The extreme low-frequency contributions of the two other instruments will rarely be a factor in the tonal value of reproduced music. The fact that there are a few organs in the world which produce subsonic 8-cps fundamentals should not unduly influence the design of a reproducing assembly made for home entertainment.

Sometimes too great a concern with extending the frequency limits of reproduction to extreme values has led to rationalizations of the need for such extension. It has been stated, for example, that sound reproducing equipment must be capable of transmitting supersonic and subsonic sums and differences between the frequency components of music, or that a band-pass extending into the subsonic region is necessary to reproduce properly a sound whose intensity is changing at a subsonic rate. We have seen that sum and difference frequencies of an intermodulatory nature can only exist when the signal is passed through a non-linear device such as the ear; whatever such intermodulation products will be formed in the ear of the concert listener will also be formed in the ear of the listener at home, without direct transmission of these products. As for variations of sound intensity at subsonic frequencies produced by vibrato, beats, or other causes, such effects are not tones in themselves but amplitude modulations of the wave envelope. Just as the coupling circuits of broadcast AM receivers transmit the wave envelope of the r.f. carrier without having any direct response to the audio frequencies involved, the audio reproducing system does not require subsonic response in order to transmit and reproduce amplitude variations of subsonic frequency. A mathematical analysis of the varying signal finds the modulating frequency conspicuously absent.

The frequency-response range of stages of the reproducing system within a feedback loop must be extended far beyond the audible limits in order to prevent phase shift and regeneration. Phase shift in the overall system, however, has little effect on perceived quality (see Chapter 6), and is, in any case, normally introduced by the reactive elements of tone controls or equalization networks.

Figure 8-2 shows the minimum frequency-response range and maximum variation allowed FM broadcast stations by the FCC.

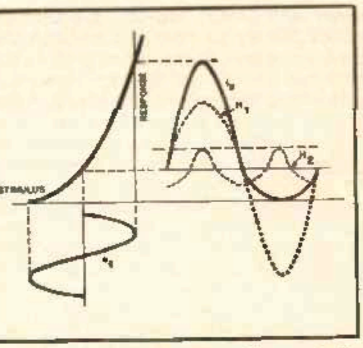

Fig. 8—3. Generation of second harmonic distortion by a non-linear transfer characteristic. The actual output, i_p, may be analyzed into fundamental and second harmonic components, shown as H_1 and H_2. e_g is the input signal. (After Reich, Theory and Applications of Electron Tubes; courtesy McGraw Hill Book Co., Inc.)

Balance

Tonal balance refers to the symmetry of frequency response deficiencies relative to the mid-point of the audio-frequency spectrum. Since there are an equal number of useful octaves on each side of the geometric mean, which is about 800 cps, this frequency constitutes the perception mid-point of the spectrum.

It is generally considered desirable to balance losses in the treble region with more or less corresponding losses in the bass, and vice versa, even though some sacrifice from absolute fidelity is involved. This view may be explained by an analogy in the field of color reproduction of pictorial material. Inadequate reproduction of the cold blues and greens, giving a picture a warm reddish cast, could be offset by a corresponding reduction in the intensity of the warm oranges and reds, a reduction which, while it involved a loss of absolute fidelity, might serve to give the picture a less artificial character.

Aural balance is often discussed in terms of the relationship between the

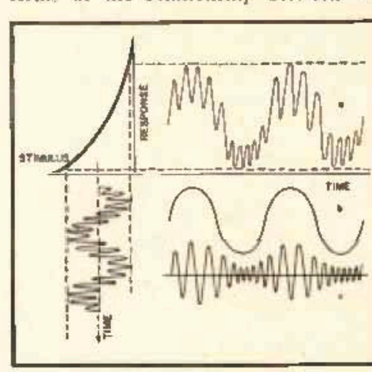

Fig. 8—5. Intermodulation produced by non-linear transfer characteristic and attendant distortion. (a) is the output signal, (b) and (c) are the distorted low-frequency component and the amplitude modulated high-frequency component, respectively. From Reich, Theory and Applications of Electron Tubes; courtesy McGraw Hill Book Co., Inc.

[2] These are the figures used in *Musical Engineering*, by H. F. Olson.

[3] D. K. Gannett and I. Kerney, "The discernibility of changes in program band width," *Bell System Technical Journal*, Jan. 1944.

upper and lower "cut-off" frequencies, which this writer considers unfortunate. Such an approach is valid only when the response curves are flat up to certain points and then cut off sharply. A more typical condition involves gradual losses over a fairly large band at the end of the spectrum, which would be balanced by similar losses symmetrical (on a logarithmic scale) to the mid-point. An additional difficulty in achieving balance by selection of cut-off frequency is that a change of frequency range involving only the bass extreme of the spectrum will leave the general character of the sound unchanged during the major part of most music, when there is no significant energy present in this frequency region. One conclusion about which there seems to be agreement, however, is that there is a large amount of latitude permissible before the response becomes audibly unbalanced. Most tone control circuits are capable of making the necessary corrections.

All of the above discussion of frequency response refers to the effect on a listener at one particular position. It is important, of course, that the response be as uniform as possible over as wide an angle as is required to cover all listening positions.

Transient Response

Proper reproduction of the instantaneous wave forms associated with the starting and stopping of sound vibrations requires a much more extended high-frequency response than is called for by the steady-state tone. The initial percussive or plucked impulse contains many high-frequency partials which are not part of this steady tone and which disappear after a short time.

Inadequate high-frequency response deadens the quality created by transient components. But transients suffer from accentuated response as well. The existence of a resonant response peak increases the tendency of the system, when stimulated by an impulse (especially at or near the frequency of the peak), to continue to oscillate after the stimulating impulse has stopped. This tendency is directly proportional to the height of the peak and to the extent to which it is confined to a narrow band of frequencies, that is, to the "Q", and is inversely proportional to the degree of damping. The effect is called *hangover* in the bass; in the treble ranges it is often referred to as *singing*. We thus see that transient response is directly related to both the range and the uniformity of frequency response.

Distortion

The output of a device is said to be distorted when the instantaneous response is not directly proportional to the instantaneous stimulus at all times. This distortion has several synonymous titles: it is called *non-linear, amplitude*, or *harmonic*. *Figure* 8-3 illustrates how the wave form distortion produced by a non-linear transfer characteristic[4] may be analyzed into partials consisting of the fundamental and second harmonic.

The spurious harmonics that are created in this way do not have as great a direct irritation value as might be supposed. They are, after all, harmonically related (in a musical sense) to the fundamental, and they may serve to intensify or cancel natural harmonics which already exist. Yet harmonic distortion is very unpleasant, and we are sensitive to very small amounts of it, as indicated in *Fig.* 8-4.

The explanation lies largely in an *effect* of harmonic distortion called intermodulation, which was touched upon in the discussion of subjective tones. When a complex tone consisting of both low- and high-frequency components is passed through a device with the transfer characteristic of *Fig*. 8-3, the high-frequency signal will be reproduced with increased amplitude during most of the first half of the cycle, and with reduced amplitude during most of the second half of the cycle. This is because during these periods the device has, in turn, exaggerated and diminished response to all stimuli. The amplitude modulation created by this non-linear transfer characteristic is illustrated in *Fig*. 8-5.

Certain wave forms, such as those produced by combinations of fundamentals and harmonics, are easily derived graphically and understood intuitively. Others, more complex, do not have components that appear quite as obvious. For example, the wave form that results from combining, in a linear system, signals of 940 cps, 1000 cps, and 1060 cps is a 1000-cps wave, amplitude modulated at 60 cps. It is a little difficult to see 940 cps and 1060 cps in this last wave form, but an application of the principles of beat phenomena will indicate that every one-sixtieth of a second the phase of the 940-cps signal will have advanced, and the phase of the 1060-cps signal will have retreated, by 360 deg. relative to the 1000-cps signal. The three signals thus return to the same phase relationship every sixtieth of a second, and a 1000-cps wave pulsating in amplitude at the rate of 60 cps implies the existence of the two sum and difference frequencies, which are called *sidebands* in r.f. parlance. (It must be emphasized that this modulated wave form does *not* result from mixing 60 and 1000 cps in a linear system.[5] The reference to beat phenomena is used here to indicate how an amplitude modulated wave implies the existence of sideband components, and not to explain intermodulation, which is *not* a beat phenomenon.)

When the amplitude modulated wave form of *Fig*. 8-5 is created through non-linear distortion, therefore, two new sideband frequencies are also created. If the input signals of *Fig*. 8-4 are 60 and 1000 cps,[6] the output will contain, in addition to these two frequencies, the intermodulation products 940 and 1060 cps. Both are discordant to 1000 cps.

When the amplitude distortion is such that the positive and negative halves of the cycle are symmetrical (odd orders of harmonics), the complete

[*Continued on page 60*]

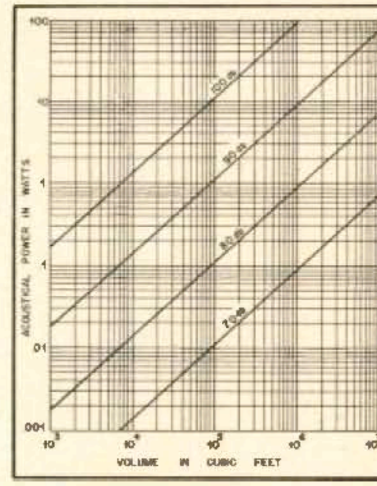

Fig. 8—7. Acoustical power required to create saundand intensity levels of 70 to 100 db. To convent to electrical power, divide by efficency of the loudspeaker.

AM STATIONS

MAXIMUM HARMONIC DISTORTION AT LESS THAN 84% MODULATION	MAXIMUM HARMONIC DISTORTION 85% TO 95% MODULATION	MAXIMUM HUM AND NOISE, 150-5000 cps	MAXIMUM HUM AND NOISE, OTHER FREQS.
5 %	7.5 %	50 db BELOW MAX AUDIO OUTPUT	40 db BELOW MAX AUDIO OUTPUT

FM STATIONS

MAXIMUM HARMONIC DISTORTION 50 - 100 cps	MAXIMUM HARMONIC DISTORTION 100-7500 cps	MAXIMUM HARMONIC DISTORTION 7500-15000 cps	MAXIMUM HUM AND NOISE, 50-15000 cps
3.5 %	2.5 %	3 %	60 db BELOW MAX AUDIO OUT

Fig. 8—6. Maximum harmonic distortion and noise permitted broadcast stations by the FCC.

[4] A transfer characteristic is a graph which plots instantaneous response against instantaneous stimulus. For example, the transfer characteristic of a vacuum tube is normally plotted as output plate current *vs.* stimulating grid voltage; of a pickup, output voltage *vs.* stimulating velocity, etc.

[5] It is this fact which makes it necessary to operate the heterodyning "mixer" stage of a super-het receiver in Class C.

[6] These two frequencies are harmonically related; 1000 cps is the tenth harmonic of 100 cps, which is separated from 60 cps by a musical interval of a major sixth.

SOUND HANDBOOK

(*from page 32*)

amplitude modulation occurs twice per low-frequency cycle. The intermodulation frequencies in this case will consist of the sum of and difference between the high-frequency signal and double the frequency of the low-frequency signal, the latter figure representing the rate of modulation of the envelope.

Harmonic distortion is described quantitatively as the ratio, in percentage, between the r.m.s. amplitudes of the spurious harmonics and of the fundamental. Each order of harmonic may be listed separately, or the total distortion may be expressed in a single percentage representing the graphical addition of all component orders. The total will thus be less than the arithmetic sum of its various parts. Higher orders of harmonic distortion have an especially great irritation value.

Figure 8-6 is a table listing the maximum harmonic distortion allowed broadcast stations by the FCC. Modern high-quality recording equipment, pickups, and amplifiers in particular are able to do considerably better than the FM performance standards. These standards are still too high for loudspeakers, however. Speaker distortion data is rarely published, but when such data does appear it reveals the fact that the very best and expensive loudspeakers create distortion, at certain lower portions of the frequency spectrum and at moderately high power levels, of the order of 5 per cent.

Intermodulation distortion is also indexed as a percentage ratio, that between the amplitude of the unmodulated high-frequency carrier and the amplitude of the low-frequency modulating envelope. This is the same figure as the percent of modulation referred to in AM transmitter work. The ratio between harmonic and intermodulation percentage is not fixed, and depends upon factors such as the signal frequencies, their relative intensities, and the type of non-linear distortion involved, but usually the intermodulation distortion is between three and four times the harmonic distortion. It will be seen that multichannel systems, in which the low and high frequencies are not allowed to pass simultaneously through the same distorting system, discriminate against the intermodulatory effects of harmonic distortion.

Power Capability

The generally accepted goal of a home reproducing system, as far as power capability is concerned, is to be able to create intensity levels of sound in the living room equal to the intensity level of sound at a good seat in the concert hall. This is not as overpowering an intensity as it might at first seem; the relatively quiet coughs and sneezes of one's neighbor at the concert hall may be heard with ease during the performance.

The acoustical output of the reproducing system is spread over a much smaller area, and is therefore able to recreate concert hall intensity levels with much smaller amounts of power. The acoustical power of a seventy-five piece orchestra may approach peak instantaneous values of 75 watts, which, with a typical home speaker system of 5 per cent efficiency, would require an instantaneous electrical power capability of 1,500 watts to reproduce. We are obviously not concerned with absolute values of power, however, but with sound intensity, in watts per square centimeter (watts/cm^2.)

The walls and other surfaces of a room are partially reflecting, so that sound directed against these surfaces will not disappear immediately after the source stops, but will be reflected back and forth, losing part of its energy with each reflection. The sound will take a certain amount of time to die away; the time required for the steady-state intensity to *decay* to a value one millionth of the original (60 db down) is called the *reverberation time*. It is directly proportional to the volume of the room and inversely proportional to the total absosption present.

The acoustical power required to create a given intensity level is inversely proportional to this reverberation time, because new sound energy being pumped into the room finds some of the old sound energy still working. Thus the absorptive materials in a room require a source of sound to radiate greater power for the same intensity level. *Figure* 8-7 is a chart indicating the acoustical power required to produce intensity levels of 70 to 100 db (above threshold) for various auditorium or room volumes with typical reverberation characteristics. Since the desired concert hall levels may approach peaks as high as 100 db (*Fig.* 8-1), a room 10 × 15 × 20 ft. would require a maximum acoustical output of 0.45 watts. With a loudspeaker system of 5 per cent efficiency, this output would call for an amplifier with a maximum power capability of 9 electrical watts. Larger rooms, a great deal of sound absorbent furnishings, and less-efficient speakers increase this requirement.

The power rating is determined by the number of watts that can be drawn from the system without distortion exceeding a specified amount. This rating tends to drop at the frequency exremes. *Figure* 8-1 shows that such a deficiency may be offset by the reduced intensity levels of orchestral music in the end frequency regions.

Dynamic Range

The dynamic range required for the reproduction of orchestral music with perfect subjective fidelity is about 70 db (see *Fig.* 8-1). The upper limit of the dynamic range of reproducing equipment is determined by the power capability of the system (in the case of disc recording by the maximum groove deviation allowable); the lower limit is determined by the level of noise, which would mask low-intensity sounds. Older records used a dynamic range severely limited by considerations of over-cutting of grooves and of surface noise, and the dynamic range of AM broadcasts is limited at each end by modulation percentage and signal-to-noise ratio respectively. These program sources therefore have a restricted range, often 40 db or less.

The use of expander circuits, which increase the gain of the amplifier to loud signals and restore some of the dynamic range lost in the original compression, was fairly common a few years ago. Although the characteristics of the expansion could not be perfectly matched to those of the original compression, the better circuits afforded considerable over-all improvement. Today, because of the extended dynamic range of records and of FM broadcasts, approaching the full requirement, the use of expanders has decreased.

Noise Level

The spurious noise components introduced into the output of reproducing systems include hum, thermal noise, microphonics, record surface noise, and turntable rumble. The annoyance value of noise is not determined exclusively by its relative amplitude to the signal, but also by other of its characteristics, particularly frequency content (due to the increased hearing sensitivity at certain parts of the spectrum described by the Fletcher-Munson effect). Noise of very low frequency may be inaudible but may create intermodulatory products in interaction with the signal. Noise level is normally described in terms of the ratio to the maximum signal amplitude, as this rating is more accurately indicative of annoyance value in a given situation than a rating in terms of absolute power would be. *Figure* 8-6 shows the maximum power level of noise allowed broadcast stations by the FCC. When the noise level is 60 db below the maximum signal power the absolute value of the hum in microwatts is equal to the value of the signal output in watts.

Due to the influence of acoustical conditions, hearing characteristics, and other such factors on the irritation value of noise, a listening test may in some situations be superior to a quantitative measurement, unless the quantitative results are properly weighted.

Binaural Reproduction

Sound which originates from an orchestra approaches the listener from several directions. The path length to each ear is therefore not the same, and in addition the waves are diffracted around the head, so that the version of the sound which each ear receives is slightly different in time, phase, intensity, and even timbre (this last due to the frequency discrimination of diffraction). Experience has enabled our perceptive mechanism to interpret these differences in terms of directional location of the source or sources. Monaural reproduction from a single reproducing source eliminates this factor of realism, and has a definite effect on the illusion created, although there may be differences of opinion on the importance of the musical values associated with the loss.

The lost illusion may be restored, to a greater or lesser degree depending upon the method used, by stereophonic or binaural reproduction. A practical binaural system that is perfect can only work through earphones. Two recording microphones are mounted as though each were one human ear; the sound received by each is recorded separately, and each recorded channel is played back into its corresponding left or right ear-piece. In this way each ear receives a version of the music with just those characteristics that the left or right-side version would have contained in the concert hall.

Since earphones are much less convenient than loudspeakers, most practical stereophonic systems are designed for speakers. Each channel no longer represents the sound received by each ear, but the sound coming from the left or the right of the hall, all of which is received by both ears. This arrangement, for accuracy of stereophonic effect, would require many channels and a careful arrangement of the speaker array calibrated to the original layout of instruments. Such a procedure is obviously impractical except for the theater, and typical stereophonic systems use two channels and two permanently mounted loudspeakers. The results are a compromise but nevertheless afford an appreciable increase in "liveness."

Binaural demonstrations with earphones are always much more dramatic than those with louspeakers. The reason for the greater illusion is described above, but there is another factor that increases the contrast between the monaural and binaural sound. Single-channel reproduction through earphones creates an artificial and impossible acoustical condition; each ear receives the identical sound at the same instant. Because of the fact that the normal left-right differences associated with diffraction around the head and reflection from room surfaces are eliminated when the ears are covered, such reproduction is unusually "flat", even when compared to that of a single loudspeaker. The contrast is thus widened by a particularly weak monaural standard.