Sound for Film and Television

2nd Edition

Tomlinson Holman

Focal Press

2002

Boston Oxford Auckland Johannesburg Melbourne New Delhi

Chapter 3: Audio Fundamentals

Audio defined

Audio is the representation of sound, electrically or by various methods on media, not sound itself. We usually say *audio tape*, not *sound tape*, for instance.[1] The advertising tag line that was tried some time ago, "It's audio that surrounds you" is also wrong, because you would be wrapped up in electrical signals or in tape, not in sound, by this image. *Sound* is the term used for acoustical energy, whereas *audio* applies to electrical signals, and magnetic and optical recordings.

Sound is the input for audio processes by way of microphones, which are *transducers*, turning sound energy into electrical energy, at that point called generically a *signal*. *Recording* is the process of converting the electrical signal into a form that is stored on a medium, from which it may be played back and converted to an electrical signal. *Mixing* involves a variety of processes to manipulate audio, and in this way winds up indirectly manipulating sound. Finally, for conversion from electrical signals back to sound, loudspeaker transducers are used.

We have already seen one primary consequence of the overall process of sound track preparation—mixing together sounds from various audio sources. For all intents and purposes, this means that they cannot be separated again by technical means, but rather the parts are separated by the final listener into a variety of auditory objects or streams, using perceptual processes. This causes us a lot of trouble and makes for a specific method of working in post production mixing, which we discuss later.

Tracks and channels

Technically, the word *track* refers to the space on the medium for the audio represen-

1. Although of course we do say, "sound track." Are we talking about the physical track on the film representing sound or what we hear? This is usually ambiguous because a mixer will say, "That's a great sound track," and an engineer will say, "The density of the sound track shows that it is underdeveloped." They are both right, thus it is the term itself that is ambiguous.

tation of sound. Thus, we say that we use *24-track* tape recorders because there are 24 parallel stripes of area on a 2 in. wide piece of tape on which we can record separate signals. The sound track on a piece of film is the area devoted to sound and the recording made on that area. In nontechnical usage, the term applies more broadly to everything recorded and its overall effect, as in "That was a dramatic track."

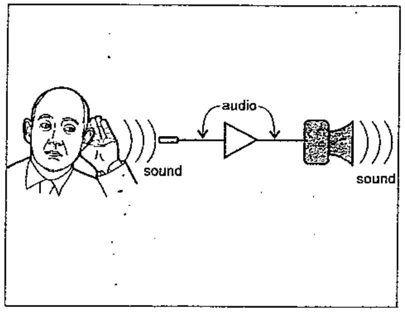


Fig. 3.1 Audio is distinguished from sound by its being the electrical representation of sound, a physical phenomenon.

The word *channel*, on the other hand, does not apply to the representation on the medium but is a more abstract term describing a signal path. It may describe a pathway for signals inside a piece of equipment, as in, "Assign the first input channel of the console to the output," or in a broader sense, as in "The transfer channel is not working at this time" (meaning that transfers from one medium to another cannot be made now).

Signals: analog and digital

There are two primary ways to represent sound as audio, by analog and by digital means. Each is important today, and each has its own strengths and weaknesses. As a trend, it has to be said that digital techniques are growing much more rapidly than analog ones at this time, but nevertheless there is a huge market for both methods. Furthermore, certain parts of the chain from microphone to loudspeaker will probably remain dominated by analog techniques for some years to come because digital techniques that will do the same thing are much more expensive, at the least.

Both analog and digital methods may also be practiced well or poorly, and may result in good or bad sound at the end of the day, although there are large differences in what the potential problems are. Probably the popular notion is that digital audio is good and analog is passé, due to the large success of the compact disc as a music medium and its obvious improvements over the phonograph record. Still, this view is naive when it comes to wider uses in professional audio, where each type of technology is seen as having particular areas where it excels.

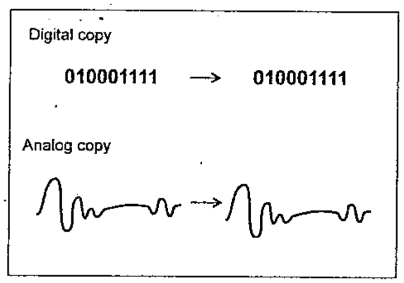


Fig. 3.2 The difference between digital copying and analog copying is the difference between copying numbers and accurately representing a waveform in a medium. This gives the advantage to digital copying, for if the copy can be read at all, it can be restored to its original values.

For instance, the strongest suit of digital technology is replication. The nature of digital recording and distribution is such that it is *potentially* far more impervious to outside influences than are analog recording and distribution chains. The word "potentially" is emphasized because an underlying medium impervious to outside influences is also needed to realize the most important benefit of digital—its permanence. This is what makes digital superior at distribution technologies, such as digital sound tracks for films and the compact disc, than competing analog technologies.

The fundamental difference between analog and digital ways of representing sound occurs in the amplitude domain. In analog, the amplitude of the audio waveform is represented as an *analogy* to the original waveform, whether that is electrically in wires or recorded on a medium. That is, there is strict proportionality between the original sound waveform, its electrical representation, and its amplitude[2] in the medium of choice. These analogies can take many forms, such as the displacement of the diaphragm of a microphone, the consequent strength of a magnetic field on audiotape, the electrical voltage in a console, or the motion of a loudspeaker cone turning audio back into sound, all looked at moment by moment.

Note that for both analog and digital methods there are two dimensions that describe a signal. The instantaneous amplitude is one dimension, and time is the other. Amplitude varying over time describes a waveform.

The fundamental difference between analog and digital signals is that the amplitude domain in digital is *quantized.* Imagine a set of bins stacked on top of one another. Quantization is the process of measuring the practically instantaneous amplitude of a waveform and "binning" it, assigning a particular numerical value by using the number of the bin closest to it in height. The number is what is stored instead of a signal strictly proportional to the waveform itself. The number is no longer strictly analogous to the waveform, since analog-to-digital conversion has been performed.[3]

2. In the strictest sense, the amplitude of the audio waveform may be represented by analog means other than by amplitude in a medium. Schemes such as FM (frequency modulation) convert amplitude variations to frequency variations for transmission, then back to amplitude variations for the output, and are still considered analog methods.

3. Although we could say that the numbers are still proportional to the waveform, the proportionality is no longer strict, since in the act of binning the amplitude, there is a range of possible waveform amplitudes that still fit within one bin before the quantizing device "snaps" to the next bin.

The beauty of storing the amplitude as numbers is the "ruggedness" with which numbers can be stored; that is, despite many corrupting influences, the numbers can still be read. By adding powerful *error codes* to the quantized numbers, especially the Reed-Solomon code,[4] the numbers are made highly robust, since most errors in their transmission can be corrected, and when uncorrectable that fact is known to the equipment so that a variety of tactics can be brought into play, such as guessing what the original waveform was, ultimately giving way, in the worst case, to no sound. Here is the Achilles' heel of digital audio: It can easily lull one into thinking that all is well, when what is actually happening is that we are operating closer and closer to the edge of not working at all, and there is no way to tell that from the sound quality.

Analog systems, on the other hand, tend to have a more gentle curve of failure, that is, they often sound increasingly bad before outright failure, giving some time to take corrective action before no sound is heard.

One primary difference between professional and consumer digital audio equipment is that professional equipment usually gives some indication of how much error correction is occurring to indicate to the user how solid things really are.

> The first strategy that digital equipment uses on finding an error is *error correction.* In error correction, the error decoder completely restores the original numbers, and there is no change in the sound whatsoever. On the other hand, if there are many occurrences of error correction it means that the medium is potentially damaged, which is likely to lead to more serious problems later.
>
> The next strategy used, after the error correction mechanism is overwhelmed, is *error concealment.* Here the playback decoding circuitry makes an educated guess about what the numbers were, based on what came before and what comes after the missing data. The sound cannot be said to be identical to that produced by the

4. This was a key technology to make the compact disc, among other digital audio media, possible and was developed by Professor Irving Reed of USC.

original, but may be good enough for all but mastering purposes. If error concealment occurs on a digital audio master, the master is usually remade. Finally, when both error correction and error concealment can no longer be used because the data are so corrupt, most equipment "mutes," that is, switches to silence.

The analog method of representing sound as analogies to the amplitude of the original waveform has the problem that the analogy is only a representation of the original, not the waveform itself. The difficulty that this causes is that with multiple-generation copying, something is inevitably lost during each generation, ultimately resulting in audible quality problems. Distortion and noise, to be discussed later in this chapter, increase from generation to generation, sometimes by tolerable amounts, but they nonetheless do inevitably increase. The popular term for this is *generation loss.*

I spent a great deal of time on *Return of the Jedi* (1983) trying to sort out why the 70mm prints sounded more distorted than the master in brief passages. Naturally the first place we worked was on the printing process itself, since that seemed to be indicated. In the end, it was found that this stage was not at fault because it was essentially equal in quality to all of the other generations, but rather it was simply that we had accumulated distortion from generation to generation to the point of audibility. (This affected only a few moments of the movie.) On the other hand, by improving the foregoing generations before the printing process, more difficult scenes had inaudible distortion by the next year in *Indiana Jones and the Temple of Doom* (1984).

Digital copying and transmission do not rely on making an analogy each generation, but rather the correct copying of numbers, a far easier task because even if the numbers are "blurry," they can still be read. So all other things being equal, and with certain assumptions, a 10th generation digital copy is indistinguishable from the first generation, whereas a 10th generation analog copy will certainly show audible defects.

Paradigms: linear vs. nonlinear

In film and television production the term *linear* and *nonlinear* are used to describe the means of access to the portions of a program to be edited or mixed. Linear in this usage means that the material is recorded along the length of a medium, which could be an analog or a digital tape. Examples include 24-track analog and digital tape machines. In order to get from one part of a program to another, winding the tape or other medium is required over the intervening portions of the tape, and this can take a considerable amount of time. nonlinear means that access to the material is available by jumping over all of the intervening material. An example is a phonograph record, where you can lift the tone arm and jump from cut one to cut ten with reasonable ease. More importantly, computers store their files in a way that permits nonlinear access. Digital audio workstations generally operate in nonlinear way, able to jump from one part of the program to another, saving time.

Some picture and sound editors, notably Walter Murch and Randy Thom, have pointed out that there are drawbacks to the enormous speed advantage of nonlinear editing, because viewing the material at high speed, such as on a flat bed editing table, can yield ideas for editors, but nonetheless the sheer speed gains of nonlinear systems make them increasingly the method of choice.

In editing systems, linear storage devices are rapidly giving way to nonlinear ones, because if a decision over the length of a shot is changed on a linear editing system, all the subsequent work in that particular reel will have to be re-done. This problem for linear systems, and the rapid access to material of the nonlinear ones, means that nonlinear systems are growing rapidly, while linear ones are shrinking in importance. The terms linear and nonlinear mean something different when describing the audio quality of a product or system, which we will take up later.

Level

The amplitude dimension of a waveform may be represented in a variety of ways, such as:

- An electrical voltage, for example, at the output of a microphone.
- The strength of a stored magnetic field on an audiotape.
- Numerically, as in digital audio.

Since program[5] waveforms constantly change amplitude with time, the value of each of these representations is also constantly changing. This constant motion makes thinking about the relative level in the various parts of the system difficult, so the idea of level is usually simplified to that which corresponds to a simple sine wave. For sound in air, we have already seen that 0 dB SPL was set as a reference at about the threshold of hearing, and practically all acoustical measurements are referenced to this "level," giving the scale to acoustical measurements shown in Chapter 1.

Microphone level

To characterize the output of a microphone, though, a reference at 0 dB SPL is inconvenient because it is difficult to obtain spaces quiet enough to make a sound of 0 dB SPL without masking by room noise, and it is nowhere near the SPL typically seen by the microphone. Thus, for most microphone measurements it is commonplace to choose as a reference sound pressure level 94 dB SPL. The microphone is rated for *sensitivity,* delivering a specific voltage at 94 dB SPL. Conventional microphones may deliver anywhere between 2 and 60 mV[6] under these conditions, depending on their type. This is a 30 dB possible range from one microphone type to another, a very large difference.

Although 94 dB SPL is a relatively high sound pressure level, the electrical voltage is still quite small, so microphones are routinely connected to *microphone preamplifiers* which amplify the output of the microphone to an electrical level that is more useful, with the amount of amplification depending on system requirements. The wide range of output levels from various microphones means that preamplifiers must be matched to the microphone type. We say that the low output level of microphones is *mike level,* and that the output of the microphone preamplifier is at *line level.* A typical microphone level taken as a snapshot on speech is 2 mV, and a microphone preamplifier may boost this to a 1.2 V line level.

Line level

Line level signals are those that are routinely interchanged within a studio environment, used for connecting different pieces of equipment together for signal processing, for instance. The method of routing such line-level signals may be by cables connected to the input and output connectors on pieces of equipment, or by way of patch bays (which look like the old-fashioned telephone operator switchboards on which they were based) or electronic switches. The electrical voltage of line level varies depending on the studio and whether the equipment is professional or consumer grade (Table 3.1). The most common professional reference line level is +4 dBu, which is 1.23 V, called here Pro 1. Pro 2 is an older reference level still found in a few broadcast applications. The most common consumer equipment line level is –10 dBV, which is 0.316 V.

Many problems in professional audio relate to interfacing equipment intended for different line levels. Patching consumer equipment into a professional studio, for instance, is difficult due to the large level difference in the reference levels of the different equipment. Typical consequences of patching in equipment without compensation for the reference level include excessive distortion, or noise. A proper method to employ such equipment uses a "pad" at the input to the consumer equipment to attenuate the higher studio level down to the consumer equipment level, and an amplifier on its output to restore the level. Boxes are available to make both these changes for the input and output of the consumer equipment. They are called *match boxes.*

A common problem in audio is connecting a microphone to an input jack on a portable recorder, such as a Betacam SP unit, and setting the level, forgetting that there is a switch

5. *Program* is derived from broadcasting practice, where *program* means the desired material to be heard by the listener, as opposed to test tones or leaders that may also be on the same medium.
6. mV is one-one thousandth of a Volt.

that sets the input sensitivity of that jack to mic or line level. Connecting a microphone to a line level input will usually result in excessive noise. Connecting a line level source to a microphone level input will usually result in gross distortion.

Speaker level

The third level used in a recording chain is speaker level. It is higher than line level and is provided by power amplifiers capable of delivering up to hundreds of watts to loudspeakers. A typical speaker voltage level is 4 V to produce 85 dB SPL in a theater, a typical reference sound pressure level in post production.

Level comparison

There are thus three principal levels used in a system—mic, line, and speaker—which correspond to a few milliVolts, around 1 Volt, and a few volts (and with high-power capability), respectively. While speaker connections would rarely be confused with line and mic connections, being of different connector types, that still leaves mic level and line level connections to be confused with each other. Perhaps no problem is so prevalent in audio as mixing up these two levels, with gross consequent distortion or noise. Often, the two levels may be presented even on the exact same connector, such as at the output of a mixer, and on the input to a professional video camera, by means of switches. If the output of a mixer is set to mic level, and the camera input is set to line level, the result will be excessive noise. Conversely, if the output of a mixer is set to line level, and the camera input is set to mic level, gross distortion at least on signal peaks will almost certainly be the result.

Interconnections

Signals are usually routed from microphones to preamplifiers, from consoles to tape machines, and from power amplifiers to loudspeakers over conductive wiring, usually copper. There are two principal ways to do this, by balanced lines, used in mostly professional applications, and by unbalanced lines, used mostly in consumer applications. Due to the differences in balanced versus unbalanced conditions, different connectors are used, generally distinguishing the types. An alternative to conductive wiring sometimes used is optical fiber interconnections among pieces of digital equipment.

Table 3.1: Line Levels

Application	Level (dB re stated reference)[1]	Voltage, (rms Volts)
Consumer and semi-pro equipment	–10 dBV	316 mV
Pro 1	+4 dBu	1.228 V
Pro 2	+8 dBu	1.946 V

1 dBV reference is deciBels relative to 1 Volt. dBu reference is deciBels relative to 0.775 Volts.

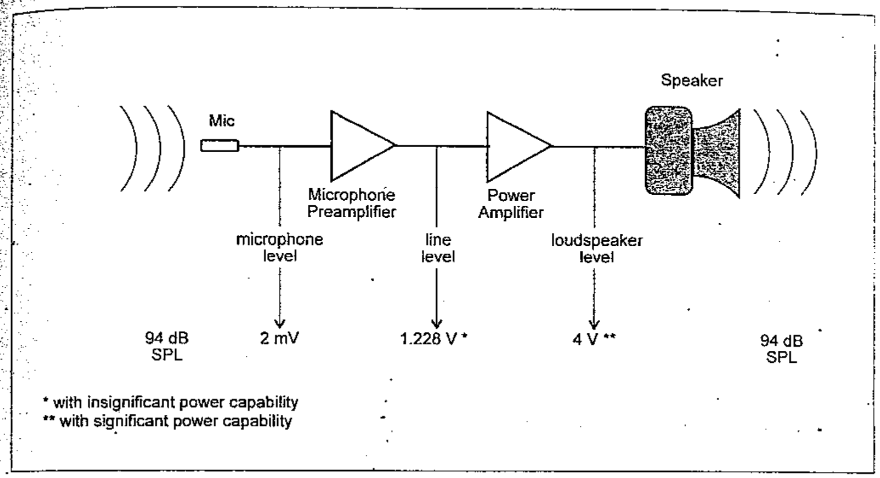


Fig. 3.3 An audio system diagram showing typical voltage levels at different points in the system. Note that although line level and loudspeaker level are similar in voltage, the loudspeaker level comes with far higher power capacity than is available at line level.

Balanced vs. unbalanced lines

Conventional home high fidelity system wiring is unbalanced. That is, there is one signal conductor contained inside a shield, and then the whole cable is wrapped in an outer insulator. The outer shield serves as an electrical ground, as well as providing shielding to prevent electrostatically induced hum.

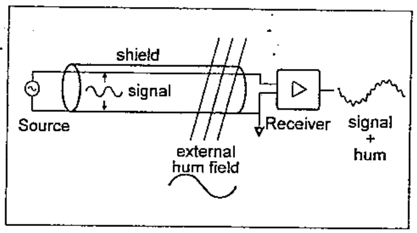
The difference between unbalanced and balanced wiring schemes occurs when there is interference from external *magnetic* sources, such as the magnetic field set up around lighting cables on a set by virtue of their carrying large amounts of current. In an unbalanced connection, the magnetic field induces a voltage in the principal conductor, which the receiving equipment sees as the same as the desired signal; thus, hum may be heard at the end of the chain.

Balanced wiring provides a means to reject hum caused by stray magnetic fields. It uses two signal conductors contained within a common conductive shield. Thus, there are three conductors altogether. At the instant when one of the two conductors has a positive-going signal voltage on it, the other one will have an equal and opposite negative-going signal voltage on it. External magnetic fields induce a voltage in the conductors, but it will be equal in magnitude and polarity in both conductors. The receiving equipment is deliberately made sensitive only to the *difference* between the two conductors, and because the voltage induced by the magnetic field is the same in both conductors, it will be rejected by the receiving equipment.

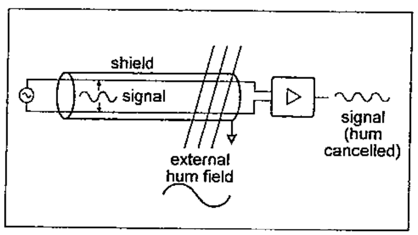
> These two modes of signals in balanced wiring are called the *differential mode* for the desired signal that is in opposite polarity in the two wires, and the *common mode* for the induced hum that is in phase in the two wires. The measure used to quantify this effect is called the *common mode rejection ratio*, a metric that compares a deliberate differential mode signal to a common mode signal and expresses the difference in deciBels. A good common mode rejection ratio is 80 dB, providing a reduction in hum of a factor of 10,000:1.

Balanced wiring is considered essential for microphone wiring, since the signal voltages are low at mike level and the cables are often in hostile environments, leading to high induced noise. Balanced wiring is also used in professional studios in between pieces of equipment, although the added expense is sometimes unneeded when the signals stay relatively local, such as within one rack of equipment, and unbalanced wiring is sometimes used in these instances.

There is one problem possible with balanced wiring compared with unbalanced: If the two signal leads inadvertently become interchanged, through wrong cable wiring, for example, then the signal will be inverted and yet still work. This condition is called *polarity reversal* or more commonly, *phase reversal.*



Unbalanced



Balanced

Fig 3.4 An unbalanced system is susceptible to hum pickup by magnetic fields being converted into voltages. The balanced system is much less susceptible because a balanced input is sensitive to the differences between the two conductors, both of which see more or less identical magnetic fields.

If the same inversion occurs in unbalanced wiring, the signal is shorted out, since the outer shield is grounded at its ends. The inadvertent "polarity reversal" of a miswired balanced line may be inaudible for some purposes, but if two microphones cover a scene and the performer is equidistant from the two, then the subsequent addition of the signals in mixing will cause the two signals to at least partially cancel. Thus the polarity of all cables must be observed in balanced wiring to prevent such cancellation.

Impedance: bridging vs. matching

Today, most audio wiring proceeds in a manner familiar to anyone who has ever plugged in multiple lamps on one electrical circuit: no matter how many are plugged in, the same voltage is delivered to the lamps (they don't get dimmer as more are plugged in), up to the capacity of the circuit breaker. So audio signals can be routed freely, and a single source can feed multiple devices, up to a reasonable limit. Such a system is called *bridging* because each of the devices connected to a source is said to *bridge* across the output of the source.

> The technical description of this condition involves a concept not yet presented—the idea of impedance. In the case of bridging systems, we say the source impedance is low and the load impedances are high, which is the same condition as exists with electrical generators supplying house wiring. This means that the source will, practically speaking, maintain the same voltage despite the number of loads bridged across it.

The alternative system of matching impedance is a system where each source is *terminated* in a specific design load impedance. It is principally useful today in very long lines, such as transmitting audio over telephone lines, but has little utility in studio environments. Still, there are some hold-over applications in broadcasting where matching is employed, because it was commonplace in earlier eras.

Connectors

Unfortunately, there are a great many connectors used for audio (Table 3.2). It is useful to know the names and area of application of audio connectors because they so frequently must be interconnected that even for the simplest jobs one must often specify an adapter for connecting two types, and this

must be done by naming the types correctly. In professional use, connectors coming from a source, such as the output of a microphone, are usually equipped with male plugs, while connectors accepting signals are usually female thus giving an immediate indication of the direction of the signal, although there are exceptions to this rule. Male connectors are those equipped with pins for the conductors and are called *plugs*, while female connectors are equipped with receptacles to accept the pins and are called *jacks*, so the "sex" of a connector is determined by the conductors, not the outer shell.

In consumer use, it is common to use hi-fi-style interconnects, where both ends of cables use male connectors and chassis connectors are female. Thus the "direction" of the signal is not indicated by the sex of the connectors. A common problem that arises as a result is finding jacks labeled "Tape Out" on the back of a receiver. What is meant in this instance is actually "Out to Tape," that is, a signal destined for the *input* of an external tape recorder. One could easily think the converse, that Tape Out should be connected to the output of the tape recorder, but that would be wrong.

Table 3.2: Some Connectors typically used for Audio[1]

Name(s)	Conductors	Photo	Usage
XLR Canon	3 – 5		Most widely used connector in professional audio, for microphone and line level analog signals and digital AES 3 (two channels on one cable). Pin 2 positive signal, pin 3 negative signal, and pin 1 shield ground.
1/4" Mono Phone	2		Mono headphones, other monaural uses such as microphones and line level signals. Tip positive signal, sleeve shield ground.
1/4" Stereo Phone 1/4" Balanced Phone	3		Stereo headphones with the tip conductor the left channel, the ring the right channel, and the sleeve ground; balanced inputs on some equipment with the tip positive signal, ring negative signal, sleeve ground.
1/4" TRS Patch Bay	3		TRS=tip, right, sleeve. For patching balanced lines in patch bays (note the tip diameter is smaller than on a conventional 1/4" plug: the two are interchngeable only in some jacks). Tip positive signal, ring negative signal, sleeve ground.
TT Tiny-T	3		A smaller version of a balanced patch bay connector

Table 3.2: Some Connectors typically used for Audio[1] (Continued)

Name(s)	Con-duc-tors	Photo	Usage (Continued)
3.5 mm Mini Mono Plug	2		Mono consumer headphones, microphones (may be called $^1/_8$").[2]
3.5 mm Mini Stereo Plug	3		Stereo consumer headphones, microphones (may be called $^1/_8$").[2]
2.5 mm Micro Mono Plug	2		Miniature recorder input plug.[2]
2.5 mm Micro Stereo Plug	3		Miniature recorder input plug.[2]
Phono Pin Plug RCA Cinch	2		Common unbalanced hi-fi system interconnects at consumer levels including for both analog and digital signals.
BNC	2		Professional audio test equipment, video, some professional digital audio use especially in video facilities.
Banana	1 per lead		Nagra tape recorder outputs, test equipment
Tuchel	2–8		Nagra portable recorders

1 many other types are used for specific circumstances, such as multi-pin connectors for multi-channel use, special connectors for radio microphones, etc.

2 The variation in body style among these four is typical of varaitions found among connectors in the field, and some large diameter connector bodies prevent full insertion into corresponding jacks due to obstructions. Also, variations in the precise dimensions, particularly the diameter of mating surfaces, vary from jack to jack. Thus if a large diameter plug has been plugged into a jack, that jack may subsequently not make good contact with a smaller diameter plug, within the tolerances of parts found in the field.

Quality issues

Audio equipment is assessed by means of measurements and listening tests. Measurements can quickly tell if certain factors are optimum in a given piece of equipment, and whether problems found are likely to be above or below audibility. Nevertheless, it is also necessary in the final analysis to conduct proper listening tests to check for factors that may escape conventional measurements. Generally, when something is heard in such listening tests, a measurement can be devised to quantify the effect.

Dynamic range: headroom and noise

One difficulty facing film and television production equipment is the very wide volume range, from the softest sound to the loudest, present on a typical set. The background noise may be at 20 dB SPL, and an actor shouting can easily reach over 120 dB SPL. The consequent more than 100 deciBel volume range is a challenge to record. There are two limitations to the ability of any item in the audio chain to reproduce the volume range from soft to loud. At the bottom end of the range, noise is the limiting factor, and at the top end distortion is the limit.

Noise is inevitable in microphones, amplifiers, tape, analog-to-digital and digital-to-analog conversion, and ancillary electronic equipment. It arises from the underlying randomness associated with the electrons comprising the signal-carrying mechanism. Even the simplest dynamic microphone (defined in the next chapter), containing no electronics and sitting in a vacuum, produces noise. This noise is caused by the Brownian motion of the electrons in the conductor comprising the voice coil and wiring of the microphone at room temperature. The only way to eliminate this noise is to cool the microphone to absolute zero temperature, where all motion ceases. Thus, at practical temperatures a noise floor is established right at the microphone, below which desired signals are masked. Ultimately, then, the way to achieve minimum noise is to capture lots of signal by using high-sensitivity microphones close to the source.

Unfortunately, using high-output microphones close to the source leads to a potential problem at the other end of the dynamic range: So much signal will be picked up when the source is loud that it may easily overload or clip the electronics or saturate the tape and cause severe distortion. The term *clipping* means that the signal peaks are literally truncated by the electronics, grossly changing the waveform and producing distortion that is quite likely to be audible.

Choosing a reference level for the electrical level in a studio, or the recorded level on tape, is like using a gray card in photography: Making an exposure based on a gray card puts midrange brightness tones in an image in the middle of the exposure range of the film. Likewise, reference levels in audio are used to represent an average "exposure" or recorded level of a program. From this reference level, the two limits to the dynamic range are measured. The dynamic range from the reference level to the noise (such as tape hiss) is called the *signal-to-noise (s/n) ratio*, and from the reference level to the maximum undistorted level is called the *headroom*. Adding the headroom in deciBels to the signal-to-noise ratio, also in deciBels, results in a single number for the dynamic range of the device or medium, which is certainly one of the most important aspects of performance.

The signal-to-noise ratio is so named because it is an expression in deciBels relating the reference level (the signal) to the noise. The signal-to-noise ratio may be *unweighted*, that is, the measurement instrument may respond to all frequencies equally, or it may be *weighted*, that is, the meter is made to respond more like a human listener by emphasizing those frequency regions where human hearing is most sensitive, with decreasing sensitivity in regions where the ear is less sensitive.

Headroom is the amount above the reference level that a system can produce without severe distortion. Headroom may be flat with frequency, with frequency components anywhere in the audio spectrum overloading at the same level, or it may not be flat with frequency, often overloading or clipping first at the frequency extremes.

An example is given by the performance of open reel analog tape recorders operating at the studio speed of 15 in./sec compared with cassette decks operating at $1\frac{7}{8}$ in./sec. Although a cassette recorder can do a credible job most of the time, one primary difference lies in the ability to record high levels of high frequencies; we say that the open-reel machine has much more *high frequency headroom*, and thus a cymbal crash, consisting mostly of high frequencies, recorded on the open-reel machine is undistorted, while recorded at the same relative level on the cassette machine it is distorted.

Audible distortion is more often a problem in film and television production than is audible noise caused by a source such as tape hiss. That is because the relatively high acoustic noise levels present on most sets thoroughly mask the noise from technical sources. (However, the noise from technical sources may become important in very quiet recording situations.) That leaves distortion as the most obvious manifestation of dynamic range problems in most situations.

There are many points in an audio chain where the signal may get to be so large that clipping or distortion occurs:

- In the microphone, especially in those equipped with their own electronics.
- In the microphone preamplifier.
- On the analog tape, or in the analog-to-digital conversion for digital recording.
- In subsequent signal processing to "improve" the sound quality.
- When multiple sources are summed together in a mixing console: One might not be too great a signal, but many added together may be too great.
- On any intermediate recording stage.
- At the final recording stage to the release medium.

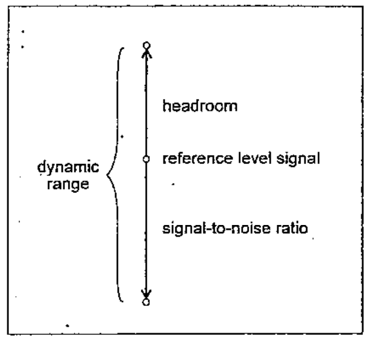


Fig. 3.5 The relationship among dynamic range, headroom, signal-to-noise ratio, and reference level.

Mixers call optimizing each of these stages for the best compromise between distortion and signal-to-noise ratio *gain staging*. Suffice it to say at this point that the signal in each stage in the chain should be optimized for level so that the widest dynamic range is preserved. This is like optimizing the exposure of film, not only on the negative, but on subsequent interpositive, internegative, and release-print stages so that the "signal," the desired picture, does not become under- nor over-exposed at any point in the chain.

Distortion added at any point in the chain can for all practical purposes not be undone at a later stage of the chain, and thus it is important to keep it low in every stage for the final result to sound undistorted. Likewise, noise accumulates from stage to stage, so under-recording is not the solution to distortion.

Linear and nonlinear distortion

Linear distortion: frequency response, amplitude, and phase

Any change in a waveform constitutes distortion in the broadest sense, but some distortions under this broad definition are benign or even beneficial, while others are quite detrimental. The first class of distortion is called *linear distortion*. These distortions change the waveform, but the effect of the change can be "undone" by equal and opposite signal processing. If, for instance, the treble is boosted a few deciBels by a piece of equipment, an equal and opposite treble cut introduced subsequently in the audio chain will restore the waveform to its original shape precisely, and essentially nothing is lost.

A linear distortion then is a change in what is called the frequency response of the system. A misnamed term, it would probably better be called "amplitude response with respect to frequency," but that being too long, the term has been shortened to the more familiar one, *frequency response*. It means how much one part of the audio frequency range or spectrum is accentuated or attenuated. For instance, a typical frequency response rating is ± 1 dB, 20 Hz to 20 kHz, meaning that there is no more than a 2 dB variation from the minimum to the maximum of the response across the range.

> Frequency response actually has two parts, the amplitude response and the phase shift, both with respect to frequency. The two together completely describe the linear distortion that the waveform undergoes.

Many sound qualities are attributed to frequency-response variations, some of them very far away from the expected definition. For instance, midrange sounds around 2 kHz have a larger effect on the perception of distance than do other frequencies. Boosting this frequency range makes sound sources seem to be closer, while cutting it makes them seem further away. Some console manufacturers, knowing this, have gone so far as to label equalization knobs for this frequency range (tone controls affecting a narrow frequency range only) in the boost condition presence and in the cut condition absence. There are many other examples of frequency response variations being ascribed subjective effects perhaps well beyond the expected range of a "tone" control.

In the most general sense, what we most often seek is a flat frequency response (sometimes called a *linear frequency response*) from most items in the audio chain. For instance, a tape recorder that discriminated against bass frequencies and boosted treble frequencies would not be desirable because all sounds going through the recorder would be affected. Although certain sounds might "sound better" with such a nonflat response, the overall average of sounds would not be improved. Thus, flat response is generally desired in most parts of the audio chain, and this is surely one of the most important specifications of any piece of equipment. In fact, in careful experiments, a deviation of as little as $\frac{1}{2}$ dB over several octaves of frequency range is an audible change. For equipment that is not supposed to change the sound quality, specifications on this order of magnitude ought to be considered necessary, especially when considering the multigenerational nature of film and television production in which each sound is recorded and played an average of six times before it reaches the listener, and thus errors accumulate.

Exceptions to the requirement for flat response include the deliberate response changes made by equalizers and filters to improve timbre and reduce noise, which are covered in Chapter 10. Microphones are most often distinguished audibly by two factors: their frequency response and how this varies with the angle to the microphone. There are reasons to make microphones nonflat, which will be taken up in the next chapter.

Nonlinear distortion

Nonlinear distortions are a class in which the waveform cannot be restored to its original shape by equal and opposite compensating equalization. In nonlinear distortions, new tonal components are added to the original ones, and no ordinary process can re-move these added components. One of the most egregious examples of nonlinear distortion is *clipping distortion*. In clipping distortion, an amplifier or other device is driven beyond its capacity, with the result being literal clipping off of the peaks of the waveform. Clipping a sine wave, for example, which is by definition only a single frequency tone, results in a great many overtones being generated because the waveform is changed dramatically. Clipping distortion is sometimes heard in production sound recordings because it is difficult to exercise adequate control over all the gain-staging[7] factors in a production sound mixer and recorder. It can usually be completely avoided by correct settings on the various pieces of equipment at hand, including the microphone itself, the mixer, and the recorder. Clipping distortion results in added frequency components at harmonics of the original tone. If the clipping is perfectly symmetrical for positive and negative excursions, the resulting harmonics are the third, fifth, seventh, etc.

Distorting analog tape also creates harmonics, although the brick-wall clipping effect is not so pronounced. Analog tape tends to distort more and more as the level is driven higher and higher above the reference level, with the ultimate limit being complete *saturation* of the magnetic oxide, equivalent to the brick wall of clipping, but with consequences well before the hard limit is reached. For this reason, it is conventional practice to modulate tape so that the peaks of the program reach only a certain distortion, without going all the way to saturation, except on nontonal sounds such as gunshots, where the added distortion is generally inaudible.

The measure usually used for the distortion described is *total harmonic distortion* (THD). THD is the sum of the energy in all of the harmonics compared with that in the fundamental, expressed as a percentage. A typical maximum level for THD in an analog tape recording system is 3% on the maximum peaks of program material. However, THD is not an adequate measure for all types of audible distortion.

To measure various distortion mechanisms beyond simple clipping or tape saturation, *intermodulation* (IM) distortion measures are used. IM distortion comes in a variety of types, but all are distinguished from harmonic distortion by the fact that the test signal contains more than one frequency, and it is the mutual effects of one frequency tone on another that are examined.

SMPTE intermodulation distortion, for example, two tones at a low and a high frequency are used to drive the system being tested, with the low-frequency tone being 12 dB greater in amplitude than the high-frequency tone. What is looked for is changes in the high-frequency tone as a result of the larger low-frequency tone being present in the system. In a perfect system, the high-frequency tone would be unaltered by the presence of the low-frequency tone, but in a practical system, intermodulation distortion makes the high-frequency tone change level over the cycle of the low-frequency tone. Changes in level of the high-frequency tone at the low-frequency rate are heard as a "roughening" of the high-frequency tone, a kind of gurgle effect.

High-frequency difference tone distortion tests send two relatively closely spaced high-frequency tones into a system and measure the resulting *difference tone intermodulation* at the difference frequency between the two tones. For example, 19 and 20 kHz tones mixed in a level ratio of 1:1 are sent into a system, and the amount of 1 kHz resulting from distortion coming out of the system is measured.

Generally speaking, the most audible of these distortions is clipping, which must be avoided for all but the briefest instants in order to remain inaudible; the next most audible is harmonic distortion on over-recorded analog tape or film; and the least significant typically is intermodulation distortion. Still, there are special cases in which each one of these distortions can come to prominence in film and television production. For example, early in the history of tape recording, an ornithologist found that bird song recorded on a tape recorder contained lots of audible distortion in the form of low-frequency thumps accompanying the high-frequency bird song.

7. A term used to describe all of the various level controls and switches in a recording chain.

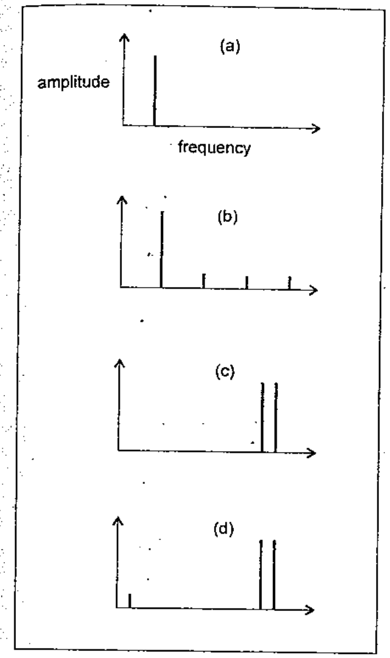


Fig. 3.6 (a) represents the sine wave input to a device under test; (b) the distorted output showing the fundamental plus harmonic distortion; (c) represents the two sine wave inputs for a difference-tone intermodulation distortion test; and (d) the distorted output showing the original sine waves plus distortion at the difference frequency.

What was found was a design problem with difference-tone intermodulation distortion in the recorder model used that had been overlooked by the designers because conventional test signals did not stimulate the effect.

Wow and flutter

Analog mechanical tape and film transports, and phonograph records, are subject to pitch variations as the speed of the mechanism varies slightly around the normal playback speed. After all, off-speed transfers are often made of sound effects in order to make them seem something other than what they are, so one can see that running a tape machine at half-speed produces frequencies that are

one-half the original on playback. It should thus not be surprising that speed variations result in pitch variations. Human hearing is particularly attuned to pitch variations and is able to distinguish a very small fraction of 1% variation under optimum conditions.

Analog tape machines, film transports, and analog optical playback from projectors in theaters and home videotape machines are all subject to wow and flutter, which are pitch variations arising from speed variations. Wow and flutter were originally distinguished as *wow* having to do with once around variations in the speed of phonograph players and *flutter* with higher speed variations. Today, the two phenomena are lumped together and are thought of as one, wow and flutter.

Wow and flutter measurements are standardized differently in different parts of the world, so the numbers derived in Europe, the United States, and Japan are not necessarily comparable. Reliable measures of this statistical phenomenon are also hard to make. Unfortunately, like noise, wow and flutter is something that accumulates over generations, and performance that is fine for one generation may well be audible when accumulated. Wow and flutter is typically most audible on music, including solo instruments such as oboes and piano.

Humans are most sensitive to wow and flutter at a rate of around four variations in frequency per second, and are less sensitive both below and above that frequency. For this reason, most wow and flutter measurements are made with a weighting curve emphasizing this frequency range.

Another form of speed modulation occurs at even higher frequencies than conventional flutter, *scrape flutter*. This is what happens when the stretched tape acts like a violin string and vibrates quite quickly. This causes *modulation noise*, which is a form of noise that is not present in quiet but occurs only when a signal is present. You can hear this easily on a Nagra recorder by recording the reference-level oscillator while listening to the playback in headphones. Stopping the roller closest to the record head with your finger will raise the scrape flutter to audible levels, while letting it move freely will essentially eliminate the audible noise.

Digital audio specific problems

All of the dynamic range and distortion measures outlined earlier apply to digital audio systems. Wow and flutter, however, can be made vanishingly small, due to the nature of digital recording. Any speed variations in tape transports can be eliminated in playback by storing the digits coming from the unevenly played tape in an electronic buffer and withdrawing them at an even rate. Digital systems also provide the potential for no generation loss. These properties constitute some of the best features of digital audio systems, where they are unequivocally better than analog ones.

On the down side, digital systems also come with their own peculiar distortion problems. Unlike analog systems, digital systems quantize the amplitude dimension. This function may cause difficulties, especially when compromises are deliberately introduced to save space on a medium. Particularly in fitting audio to the capacity of computer uses, such as CD-ROM or game boards for PCs, these compromises are likely to be so great that they become audible to the casual listener. Thus, the informed user should know what problems may arise when such measures are invoked.

The digital audio-specific problems discussed here apply to the most common digital audio method of representation, called linear *pulse code modulation* (PCM). In PCM digital audio, quantization occurs by comparing the amplitude of the waveform to the height of a series of stair steps. Each of the stair steps is of equal height. The quantizer (the heart of the analog to digital converter) compares the amplitude of the waveform to the height of the stair steps and assigns a number corresponding to the number of the nearest step.

There are other problems that occur in practical equipment; the ones outlined here are problems inherent in the basic PCM digital method. For example, an analog-to-digital converter in which all of the steps are not ascending in order (with, e.g., one step missing) is clearly defective.

Resolution

Resolution is the number of bits being used to represent the amplitude dimension, the number of steps in the stairs. For the compact disc, this number is 16 bits of binary (0 or 1) information, or about 65,000 steps. Sixteen-bit representation yields a dynamic range of nearly 96 dB,[8] since each "bit" of resolution buys 6 deciBels of dynamic range ($16 \times 6 = 96$). On the other hand, many low-end computer boards, programs, and CD-ROM recordings are made at only 8 bits, for a dynamic range of 48 dB. This produces audible noise accompanying almost any program material, since there is practically no program material that will mask noise only 48 dB below the maximum signal level. If we were to assign a reference level 8 dB below the maximum, then there is only a signal-to-noise ratio of 40 dB. A 40 dB signal to noise ratio means that the noise will be $^1/_{16}$ as loud as a signal at reference level, and thus clearly audible. Eight deciBels of headroom is also very little to accommodate louder sounds.

In addition, there is an inherent problem with complex productions using digital audio. For the result to be of a certain resolution, for example, 16 bits, if only one track goes into the production, a 16-bit source will do. As the number of source tracks grows, however, the noise from each of the sources will add, and the result will be a decrease in resolution. So a 16-bit multitrack recorder has an inherent design problem: With any degree of mixing it is impossible to deliver an output that has a 16-bit dynamic range. Although a formula can be given for determining the number of additional bits necessary to produce the desired resolution, the formula does not take into account varying levels among the channels, equalization, etc. One emerging trend in digital audio is towards greater resolution in professional equipment, to 20- and 24-bit, which may be a useful improvement in large-scale production.

8. *Nearly* is an important word here. Nearly for two reasons, that no practical device reaches the theoretical and due to the need for dither, which is explained later.

Sampling and aliasing distortion

Although quantizing is at the heart of digital audio, another process must occur beforehand, sampling. The procedure is to measure the audio signal so many times per second that all of the nuances of the signal that are in the audio frequency band are captured. The sample rate required is a little more than twice the highest frequency in the desired bandwidth. With 20 kHz usually considered the highest audible frequency, the sample rate for the compact disc became 44.1 kHz, and other professional audio uses 48 kHz. (A lower rate of 32 kHz is used in some broadcasting applications, and some available equipment uses 96 kHz sampling, coming down on the side of extending the bandwidth into what is generally considered to be the ultrasonic domain.)

In order to save space on computer discs and CD-ROMs so that more audio can be stored, it is common to sample at lower rates, usually submultiples of 44.1 kHz, such as 22.05 kHz, 11.025 kHz, etc. The problem encountered with sampling at these lower rates is that frequency components at more than one half of the sampling frequency are likely to be in the signal. The sampling process "confuses" these with lower frequency signals and produces an output tone that is different in frequency from the input one. This is called *aliasing distortion*.

One half of the sample rate is called the *folding frequency* because aliasing distortion "folds" the signal frequencies around one-half the sample rate. So any input frequency above one half of the sample rate will alias and appear as a new frequency in the output. For instance, if the sample rate is 11 kHz and the signal frequency is 7 kHz (speech recorded for a CD-ROM will contain such a frequency in an "s" sound), one half of the sample rate is 5.5 kHz and the 7 kHz signal will be seen as a 4 kHz signal (7 kHz is 1.5 kHz above the folding frequency, and 4 kHz is 1.5 kHz below the folding frequency).

An example of aliasing distortion in motion pictures is a shot of wagon wheels appearing to run backwards when photographed, or sampled, at 24 fps. Up to a certain speed of the wagon, photography renders an accurate representation of the speed of the wheels. But at the point where the shutter is open, then closed, then reopened just as the spokes have moved, for example, $^1/_6$ of a rotation for a six-spoke wheel, the film "sees" a stationary image of a moving object, which is aliasing distortion. At other relative rates between the photography and speed of the wheels, they even appear to run "backwards," which is clearly an artifact.

In audio, the sound of aliasing distortion is distinctive. On speech as a signal, it sounds like a chirp that accompanies "ess" and other high-frequency sounds in speech. It may be fairly benign, or very nasty, depending on the strength of "ess-es" in the speech and the sample rate (lower ones are worse). The way to avoid aliasing distortion is to use a filter that removes the frequencies higher than one-half the sample rate. This filter is called an *anti-aliasing filter*. Unfortunately, most low-sample-rate analog-to-digital converters are not equipped with anti-aliasing filters, and plainly audible aliases are the result on much program material.

Jitter

An issue in sampling and in digital-to-digital interfaces is called *jitter*. This is the variation in time from sample to sample from the calculated time due to imperfections, and can result in pure tones becoming "rougher" sounding, much like with *scrape flutter*, only generally jitter is smaller in effect than scrape flutter. Jitter can even be present in digital-to-digital interfaces. The input of properly designed digital audio equipment "re-clocks" incoming jittered signals and restores them.

Quantizing distortion

The process of quantizing the amplitude dimension of the signal also carries with it the potential for a particular kind of distortion, called *quantizing distortion*. Quantizing distortion arises even in a perfect linear PCM system when the amplitudes of the signals are very small, unless measures are taken to prevent it (see *Dither*, below).

Let us say that a signal is just slightly larger than one step of the staircase and is a pure sine-wave tone. It will be converted as it first crosses above and then below the level of the first tread of the staircase. The digital representation of the sine wave will simply alternate between the higher and the lower bit levels. The result, upon conversion back to analog, will be a square wave, not a sine wave. The reason is that the converter is coarse at these low levels and cannot discriminate the waveform, only the fact that the one tread has been alternately crossed. Thus, what came into the analog-to-digital conversion process as a sine wave is actually converted as a square wave, producing a very significant distortion. Furthermore, a just slightly lower level signal is not converted at all, because it never crosses the tread of the staircase. Thus any lower level signals than the smallest step (called the *least significant bit*) are discarded, another important distortion. There is a way around these distortions and it is called *dither.*

Dither

The effects of quantizing distortion can be fully eliminated by the addition of some deliberately added random noise, which may sound like hiss. Although noise is usually considered to be a detriment in any system, adding dither noise in a digital audio conversion process randomly "agitates" the signal so that even the smallest signals cross the treads of the staircase. The noise pushes the signal up and down randomly, causing it to cross the threshold often. By averaging the signal plus the dither over time, like the hearing mechanism does, signals even far below the threshold of the smallest step can be perceived. The noise effectively smears out the steps in the staircase, turning the stair steps into a linear ramp when averaged over time. The amount of dither is about equal to one step in the staircase, so the added noise to produce these benefits is quite small, although it may become noticeable when many channels are added together.

Another important role for dither is in reducing the number of bits of resolution when a source has more resolution than a copy. If 16 bit recordings are made in the studio, and then transferred to 8 bit for release on CD-ROM, very significant audible distortion results. These effects can be minimized by adding the proper amount and type of dither. With 20 bit storage capability on digital videotape, and 20 bit converters growing in use, conversion to the final consumer format of 16 bits, for example, must add dither in the conversion so that the coarser steps of 16 bits compared with the 20 bit original do not become a source of quantizing distortion.

Dither may be added in a number of forms. The noise needed may "hidden" in the less audible parts of the spectrum, by shaping the frequency response of the noise to be like the equal-loudness contours, with little noise in the 2 to 3 kHz most-sensitive region of hearing, and increasing it above 10 kHz, where sound is less audible. Called *noise shaped dither*, and also by trade names such as Super Bit Mapping™, this process yields an audible improvement in dynamic range over the restrictions of the output resolution. Perceptibly, a 16 bit system can be made to sound as though it has the dynamic range of a 19 bit system in this way.

A digital audio system

There are a number of blocks in making a digital audio system, caused by the needs outlined earlier. In addition, it should be pointed out that we are dealing with only one kind of digital audio in this exposition, and that is linear pulse code modulated digital.

A PCM system has the following parts to its block diagram, in the order that a signal encounters them:

- Anti-aliasing filter
- Dither noise generator
- Summer for the signal and the noise
- Analog-to-digital converter
- Digital circuitry to add error-protecting codes and to condition the signal for recording or transmission
- Medium for storage or transmission of the digital bits
- Digital circuits to decode the error codes and correct for errors
- Digital-to-analog converter
- Reconstruction filter, the equivalent on the output of the anti-aliasing filter on the input, which "smooths" the small steps remaining in the signal from the digitization process into a continuous signal

Specialized application areas, such as digital release prints, employ other techniques. Linear PCM is the most common method of recording and storing audio today, but it consumes a lot of digital audio storage for a given time of track time. For instance, sampling at 16 bits with a 48 kHz clock results in 720,000 bits per second per channel. Due to this high number, lower sample rates and resolution are used for CD-ROM and other such purposes to save space. Another solution is perceptual coding, which works by "throwing" away signals that would be masked by human hearing mechanisms, achieving more than ten times bit-rate reduction with potentially few audible artifacts.

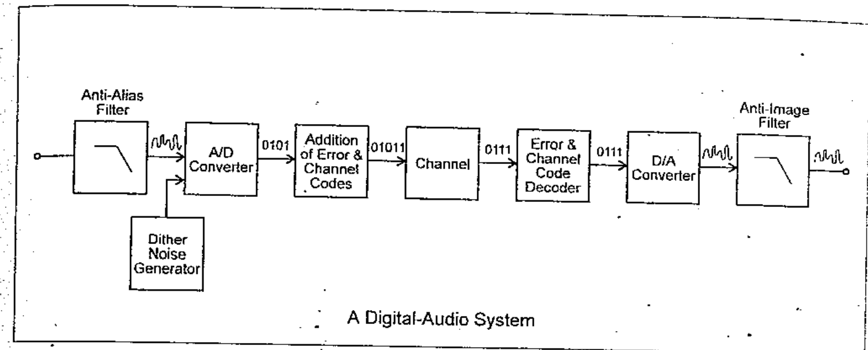


Fig. 3.7 The parts of a linear PCM digital-audio system from analog input to output.

Oversampling

Two dimensions of an audio signal are amplitude and frequency. It is possible to trade one of these off against the other, since the information-carrying capacity of a channel is related to the product of these two factors. Radio microphones, for instance, convert the audio bandwidth and dynamic range from their microphone to a wider bandwidth and a smaller dynamic range for transmission in the radio-frequency channel. The signal is represented by a different means (FM radio), but the information content is still the same. The two dimensions are represented in digital audio by quantizing and sampling respectively. Oversampling can be compared with the process used in radio microphones, trading off bandwidth and dynamic range. By sampling at a higher than normal sample rate, less dynamic range in the A/D and D/A converters is needed to produce results associated with a wider dynamic range in the audio channel.

Oversampling is a process one frequently hears about in terms of the number of times of oversampling that a particular circuit employs, from "4 times oversampling" to "128 times oversampling." The number of times refers to the sample rate, for example, a 4 times oversampled professional audio system samples at 192 kHz (4 x 48 kHz). The basic idea behind oversampling is that sampling at higher rates spreads the noise of the analog-to-digital conversion process out over a wider than audible frequency range. The portion of the noise that is ultrasonic is inaudible, so the more spreading the better (the higher the sample rate or number of times of oversampling). When converted at the other end of the process back to audio, just the audible frequency range noise counts, so it is possible for an oversampled system to come closer to the ideal than one using conventional sample rates. Oversampling also simplifies the requirements of the anti-aliasing filters, because they need not filter so steeply because the sample frequency is raised so much. On the other hand, there are practical problems involved in oversampling, such as the fact that *jitter*, small variations in the time that the samples are taken from one to the next due to imperfections in the process, becomes relatively more important than in a lower sampled system. So it is by no means clear that the system with the highest oversampling rate is necessarily the best.

Chapter 3: Audio Fundamentals
Questions

Answer the following questions in your books.

1. What is audio?

2. What are transducers?

3. What do the terms track and channel refer to?

4. How many types of audio signals are there? What are they called?

5. Which type of signal is impervious to out side influences?

6. What does the term Amplitude refer to?

7. What is the difference between professional and consumer digital audio equipment?

8. What is the difference between analogue and digital recordings when you want to make a copy, of a copy, of a copy etc?

9. What is linear and non linear recording?

10. How would the amplitude dimensions of a wave form be represented?

11. How many types of levels do we have on a mixing desk? What are they?

12. What is the difference between *microphone level* and *line level*?

13. What problems occur when interfacing professional and consumer audio equipment?

14. What are the differences between unbalanced and balanced wiring schemes?

15. What are XLR connection usually used for?

16. What are ¼’’ TRS connection usually used for?

17. How are audio equipment assessed?

18. What is Brownian motion and how can it be eliminated?