SECTION 20

AUDIO SYSTEMS

Although much of this section contains fundamental information on basic audio technology, extensive information on new evolving digital audio formats and recording and reproduction systems has been added since the last edition.

As noted herein, DVD-Audio, for example, is based on the same DVD technology as DVD-Video discs and DVD-ROM computer discs. It has a theoretical sampling rate of 192 kHz with 24-bit processing and can store 4.7 gigabytes on a disc with a choice of two- or six-channel audio tracks or a mix of both. Super Audio CD (SACD) has the same storage capacity. It uses direct stream digital (DSD) with 2.8 MHz sampling in three possible disc types. The first two contain only DSD data (4.7 gigabytes of data on a single-layer disc and slightly less than 9 gigabytes on the dual layer disc). The third version, the SACD hybrid, combines a single 4.7 gigabyte layer with a conventional CD that can be played back on conventional CD players.

MPEG audio coding variations continue to evolve. For example:

- MPEG-1 is a low-bit-rate audio format.
- MPEG-2 extends MPEG-1 toward the audio needs of digital video broadcasting.
- MPEG-2 Advanced Audio Coding (AAC) is an enhanced multichannel coding system.
- MP3 is the popular name for MPEG-1 Layer III.
- MPEG-4 adds object-based representation, content-based interactivity, and scalability.
- MPEG-7 defines a universal standardized mechanism for exchanging descriptive data.
- MPEG-21 defines a multimedia framework to enable transparent and augmented use of multimedia services across a wide range of networks and devices used by different communities. R.J.

In This Section:

On the CD-ROM:

The following are reproduced from the 4th edition of this handbook:

"Ambient Noise and Its Control," by Daniel W. Martin;

"Acoustical Environment Control," by Daniel W. Martin;

"Mechanical Disc Reproduction Systems," by Daniel W. Martin;

"Magnetic-Tape Analog Recording and Reproduction," by Daniel W. Martin.

CHAPTER 20.1

SOUND UNITS AND FORMATS

Daniel W. Martin, Ronald M. Aarts

STANDARD UNITS FOR SOUND SPECIFICATION[1,2]

Sound Pressure

Airborne sound waves are a physical disturbance pattern in the air, an elastic medium, traveling through the air at a speed that depends somewhat on air temperature (but not on static air pressure). The instantaneous magnitude of the wave at a specific point in space and time can be expressed in various ways, e.g., displacement, particle velocity, and pressure. However, the most widely used and measured property of sound waves is *sound pressure*, the fluctuation above and below atmospheric pressure, which results from the wave.

An *atmosphere* (atm) of pressure is typically about 10^5 pascals (Pa) in the International System of units. Sound pressure is usually a very small part of atmospheric pressure. For example, the minimum audible sound pressure (threshold of hearing) at 2000 Hz is 20 μPa, or $2(10)^{-10}$ atm.

Sound-Pressure Level

Sound pressures important to electronics engineering range from the weakest noises that can interfere with sound recording to the strongest sounds a loudspeaker diaphragm should be expected to radiate. This range is approximately 10^6. Consequently, for convenience, sound pressures are commonly plotted on a logarithmic scale called *sound-pressure level* expressed in *decibels* (dB).

The decibel, a unit widely used for other purposes in electronics engineering, originated in audio engineering (in telephony), and is named for Alexander Graham Bell. Because it is logarithmic, it requires a reference value for comparison just as it does in other branches of electronics engineering. The reference pressure for sounds in air, corresponding to 0 dB, has been defined as a sound pressure of 20 μPa (previously 0.0002 dyn/cm^2). This is the reference sound pressure p_0 used throughout this section of the handbook. Thus the sound-pressure level L_p in decibels corresponding to a sound pressure p is defined by

$$L_p = 20 \log (p/p_0) \text{ dB} \tag{1}$$

The reference pressure p_0 approximates the weakest audible sound pressure at 2000 Hz. Consequently most decibel values for sound levels are positive in sign. Figure 20.1.1 relates sound-pressure level in decibels to sound pressure in micropascals.

Sound power and sound intensity (power flow per unit area of wavefront) are generally proportional to the square of the sound pressure. Doubling the sound pressure quadruples the intensity in the sound field, requiring four times the power from the sound source.

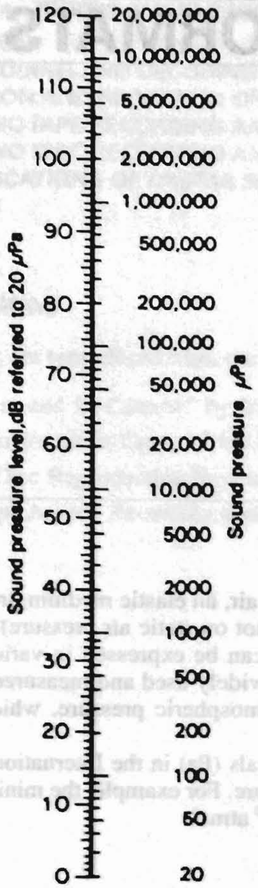

FIGURE 20.1.1 Relation between sound pressure and sound-pressure level.[1]

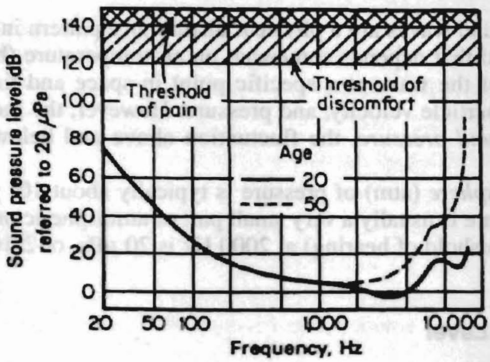

FIGURE 20.1.2 The auditory area.

Audible Frequency Range

The international abbreviation Hz (hertz) is now used (instead of the former cps) for audible frequencies as well as the rest of the frequency domain. The limits of audible frequency are only approximate because tactile sensations below 20 Hz overlap aural sensations above this lower limit. Moreover, only young listeners can hear pure sounds near or above 20 kHz, the nominal upper limit.

Frequencies beyond both limits, however, have significance to audio-electronics engineers. For example, near-infrasonic (below 20 Hz) sounds are needed for classical organ music but can be noise in turntable rumble. Near-ultrasonic (above 20 kHz) intermodulation in audio circuits can produce undesirable difference-frequency components, which are audible.

The *audible sound-pressure level range* can be combined with the audible frequency range to describe an *auditory area*, shown in Fig. 20.1.2. The lowest curve shows the weakest audible sound-pressure level for listening with both ears to a pure tone while facing the sound source in a free field. The minimum level depends greatly on the frequency of the sound. It also varies somewhat among listeners. The levels that quickly produce discomfort or pain for listeners are only approximate, as indicated by the shaded and cross-hatched areas of Fig. 20.1.2. Extended exposure can produce temporary (or permanent) loss of auditory area at sound-pressure levels as low as 90 dB.

Wavelength effects are of great importance in the design of sound systems and rooms because wavelength varies over a 3-decade range, much wider than is typical elsewhere in electronics engineering. Audible sound waves vary in length from 1 cm to 15 m. The dimensions of the sound sources and receivers used in electroacoustics also vary greatly, e.g., from 1 cm to 3 m.

Sound waves follow the principles of geometrical optics and acoustics when the wavelength is very small relative to object size and pass completely around obstacles much smaller than a wavelength. This wide range of physical effects complicates the typical practical problem of sound production or reproduction.

Loudness Level

The simple, direct method for determining experimentally the loudness *level* of a sound is to match its observed loudness with that of a 1000-Hz sinewave reference tone of calibrated, variable sound-pressure level. (Usually this is a group judgment, or an average of individual judgments, in order to overcome individual observer differences.)

When the two loudnesses are matched, the *loudness level* of the sound, expressed in *phons*, is defined as numerically equal to the sound-pressure level of the reference tone in decibels. For example, a series of observers, each listening alternately to a machine noise and to a 1000-Hz reference tone, judge them (on the average) to be equally loud when the reference tone is adjusted to 86 dB at the observer location. This makes the loudness level of the machine noise 86 phons.

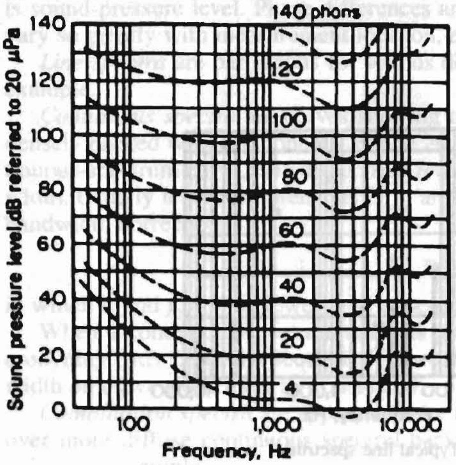

FIGURE 20.1.3 Equal-loudness-level contours.

At 1000 Hz the decibel and phon levels are numerically identical, by definition. However, at other frequencies sinewave tones may have numerically quite different sound- and loudness-levels, as seen in Fig. 20.1.3. The dashed contour curves show the decibel level at each frequency corresponding to the loudness level identifying the curve at 1000 Hz. For example, a tone at 80 Hz and 70 dB lies on the contour marked 60 phons. Its sound level must be 70 dB for it to be as loud as a 60-dB tone at 1000 Hz. Such differences at low frequencies, especially at low sound levels, are a characteristic of the sense of hearing. The fluctuations above 1000 Hz are caused by sound-wave diffraction around the head of the listener and resonances in his ear canal. This illustrates how human physiological and psychological characteristics complicate the application of purely physical concepts.

Since loudness level is related to 1000-Hz tones defined physically in magnitude, the loudness-level scale is not really psychologically based. Consequently, although one can say that 70 phons is louder than 60 phons, one cannot say *how much* louder.

Loudness

By using the phon scale to overcome the effects of frequency, psychophysicists have developed a true loudness scale based on numerous experimental procedures involving relative-loudness judgments. *Loudness*, measured in *sones*, has a direct relation to loudness level in phons, which is approximated in Fig. 20.1.4. (Below 30 phons the relation changes slope. Since few practical problems require that range, it is omitted for simplicity.) A loudness of 1 sone has been defined equivalent to a loudness level of 40 phons. It is evident in Fig. 20.1.4 that a 10-phon change doubles the loudness in sones, which means *twice as loud*. Thus a 20-phon change in loudness level quadruples the loudness.

Another advantage of the sone scale is that the loudness of components of a complex sound are additive on the sone scale as long as they are well separated on the frequency scale. For example (using Fig. 20.1.4), two tonal components at 100 and 4000 Hz having loudness levels of 70 and 60 phons, respectively, would have individual loudnesses of 8 and 4 sones, respectively, and a total loudness of 12 sones.

level in decibels as the frequency changes with a constant electric or sound input; or it may be a ratio of the output to input (expressed in decibels) as long as they are linearly related within the range of measurement.

When the response-frequency characteristic is measured with the input frequency filtered from the output, a distortion-frequency characteristic is the result. It can be further filtered to obtain curves for each harmonic if desired.

Directional Characteristics

Sound sources radiate almost equally in all directions when the wavelength is large compared to source dimensions. At higher frequencies, where the wavelength is smaller than the source, the radiation becomes quite directional.

Time Characteristics

Any sound property can vary with time. It can build up, decay, or vary in magnitude periodically or randomly. A reverberant sound field decays rather logarithmically. Consequently the sound level in decibels falls linearly when the time scale is linear. The rate of decay in this example is 33 dB/s.

REFERENCES

1. Harris, C. M. (ed.), "Handbook of Acoustical Measurements and Noise Control," Chaps. 1 and 2, McGraw-Hill, 1991.
2. Acoustical Terminology (Including Mechanical Shock and Vibration), S1.1–1994, Acoustical Society of America, 1994.
3. ANSI Standard S3.4-1980 (R1986).

CHAPTER 20.2
SPEECH AND MUSICAL SOUNDS

Daniel W. Martin, Ronald M. Aarts

SPEECH SOUNDS

Speech Level and Spectrum

Both the sound-pressure level and the spectrum of speech sounds vary continuously and rapidly during connected discourse. Although speech may be arbitrarily segmented into elements called phonemes, each with a characteristic spectrum and level, actually one phoneme blends into another.

Different talkers speak somewhat differently, and they sound different. Their speech characteristics vary from one time or mood to another. Yet in spite of all these differences and variations, statistical studies of speech have established a typical "idealized" speech spectrum. The spectrum level rises about 5 dB from 100 to 600 Hz, then falls about 6, 9, 12, and 15 dB in succeeding higher octaves.

Overall sound-pressure levels, averaged over time and measured at a distance of 1 m from a talker on or near the speech axis, lie in the range of 65 and 75 dB when the talkers are instructed to speak in a "normal" tone of voice. Along this axis the speech sound level follows the inverse-square law closely to within about 10 cm of the lips, where the level is about 90 dB. At the lips, where communication microphones are often used, the overall speech sound level typically averages over 100 dB.

The peak levels of speech sounds greatly exceed the long-time average level. Figure 20.2.1 shows the difference between short peak levels and average levels at different frequencies in the speech spectrum. The difference is greater at high frequencies, where the sibilant sounds of relatively short duration have spectrum peaks.

Speech Directional Characteristics

Speech sounds are very directional at high frequencies. Figure 20.2.2 shows clearly why speech is poorly received behind a talker, especially in nonreflective environments. Above 4000 Hz the directional loss in level is 20 dB or more, which particularly affects the sibilant sound levels so important to speech intelligibility.

Vowel Spectra

Different vowel sounds are formed from approximately the same basic laryngeal tone spectrum by shaping the vocal tract (throat, back of mouth, mouth, and lips) to have different acoustical resonance-frequency combinations. Figure 20.2.3 illustrates the spectrum filtering process. The spectral peaks are called *formants*, and their frequencies are known as formant frequencies.

The shapes of the vocal tract, simplified models, and the acoustical results for three vowel sounds are shown in Fig 20.2.4. A convenient graphical method for describing the combined formant patterns is shown in Fig 20.2.5. Traveling around this vowel loop involves progressive motion of the jaws, tongue, and lips.

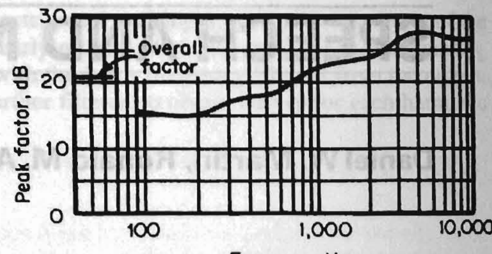

FIGURE 20.2.1 Difference in decibels between peak pressures of speech measured in short (1/8-s) intervals and rms pressure averaged over a long (75-s) interval.

Speech Intelligibility

More intelligibility is contained in the central part of the speech spectrum than near the ends. Figure 20.2.6 shows the effect on articulation (the percent of syllables correctly heard) when low- and high-pass filters of various cutoff frequencies are used. From this information a special frequency scale has been developed in which each of 20 frequency bands contributes 5 percent to a total *articulation index* of 100 percent. This distorted frequency scale is used in Fig. 20.2.7. Also shown are the spectrum curves for speech peaks and for speech minima, lying approximately 12 and 18 dB, respectively, above and below the average-speech-spectrum curve. When all the shaded area (30-dB range between the maximum and minimum curves) lies above threshold and below overload, in the absence of noise, the articulation index is 100 percent.

If a noise-spectrum curve were added to Fig 20.2.7, the figure would become an articulation-index computation chart for predicting communication capability. For example, if the ambient-noise spectrum coincided with the average-speech-spectrum curve, i.e., the signal-to-noise ratio is 1, only twelve-thirtieths of the shaded area would lie above the noise. The articulation index would be reduced accordingly to 40 percent.

Figure 20.2.8 relates monosyllabic word articulation and sentence intelligibility to articulation index. In the example above, for an articulation index of 0.40 approximately 70 percent of monosyllabic words and 96 percent of sentences would be correctly received.

However, if the signal-to-noise ratio were kept at unity and the frequency range were reduced to 1000 to 3000 Hz, half the bands would be lost. Articulation index would drop to 0.20, word articulation to 0.30, and sentence intelligibility to 70 percent. This shows the necessity for wide frequency range in a communication system when the signal-to-noise ratio is marginal. Conversely a good signal-to-noise ratio is required when the frequency range is limited.

The articulation-index method is particularly valuable in complex intercommunication-system designs involving noise disturbance at both the transmitting and receiving stations. Simpler effective methods have also been developed, such as the *rapid speech transmission index* (RASTI).

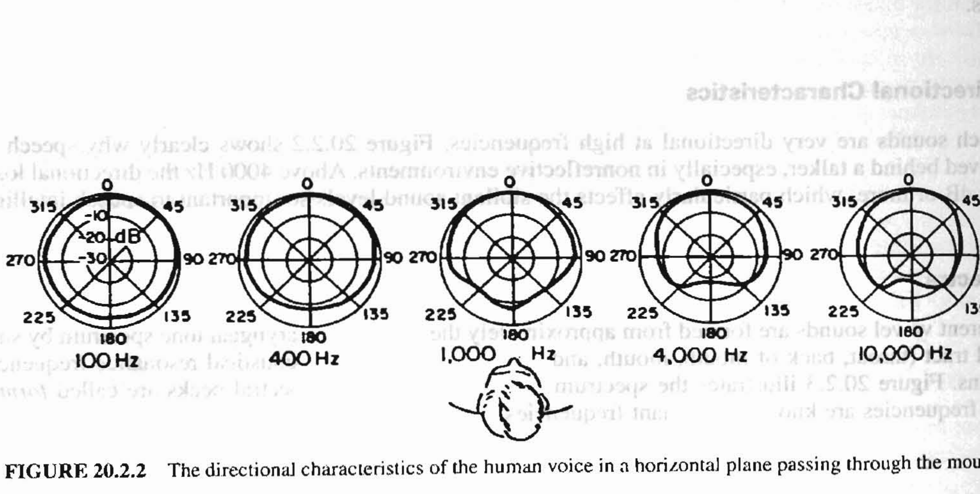

FIGURE 20.2.2 The directional characteristics of the human voice in a horizontal plane passing through the mouth.

Speech Peak Clipping

Speech waves are often affected inadvertently by electronic-circuit performance deficiencies or limitations. Figure 20.2.9 illustrates two types of amplitude distortion, center clipping and peak clipping. Center clipping, often caused by improper balancing or biasing of a push-pull amplifier circuit, can greatly interfere with speech quality and intelligibility. In a normal speech spectrum the consonant sounds are higher in frequency and lower in level than the vowel sounds. Center clipping tends to remove the important consonants.

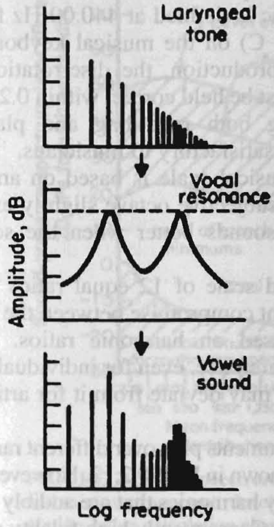

FIGURE 20.2.3 Effects on the spectrum of the laryngeal tone produced by the resonances of the vocal tract.[5]

By contrast peak clipping has little effect on speech intelligibility as long as ambient noise at the talker and system electronic noise are relatively low in level compared with the speech.

Peak clipping is frequently used intentionally in speech-communication systems to raise the average transmitted speech level above ambient noise at the listener or to increase the range of a radio transmitter of limited power. This can be done simply by overloading an amplifier stage. However, it is safer for the circuits and it produces less intermodulation distortion when back-to-back diodes are used for clipping ahead of the overload point in the amplifier or transmitter. Figure 20.2.10 shows intelligibility improvement from speech peak clipping when the talker is in quiet and listeners are in noise. Figure 20.2.11 shows that caution is necessary when the talker is in noise, unless the microphone is shielded or is a noise-canceling type.

Tilting the speech spectrum by differentiation and flattening it by equalization are effective preemphasis treatments before peak clipping. Both methods put the consonant and vowel sounds into a more balanced relationship before the intermodulation effects of clipping affect voiced consonants.

Caution must be used in combining different forms of speech-wave distortion, which individually have innocuous effects on intelligibility but can be devastating when they are combined.

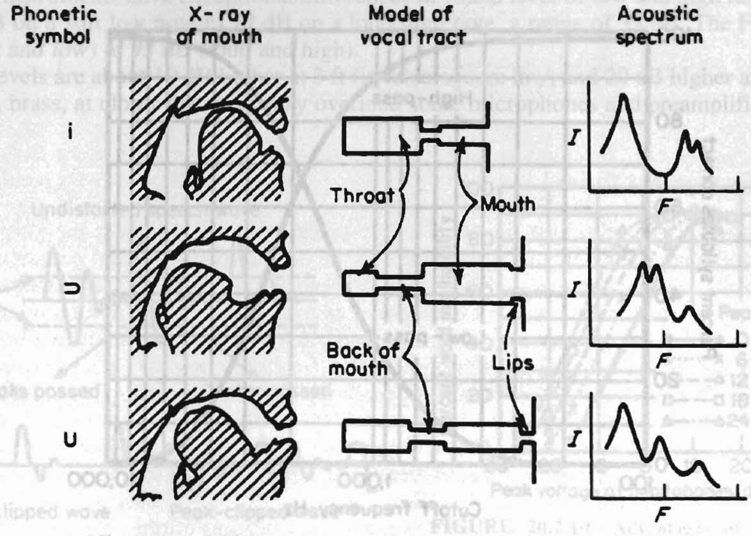

FIGURE 20.2.4 Phonetic symbols, shapes of vocal tract, models, and acoustic spectra for three vowels[6].

MUSICAL SOUNDS

Musical Frequencies

The accuracy of both absolute and relative frequencies is usually much more important for musical sounds than for speech sounds and noise. The international frequency standard for music is defined at 440.00 Hz for A_4, the A above C_4 (middle C) on the musical keyboard. In sound recording and reproduction, the disc-rotation and tape-transport speeds must be held correct within 0.2 or 0.3 percent error (including both recording and playback mechanisms) to be fully satisfactory to musicians.

The mathematical musical scale is based on an exact octave ratio of 2:1. The subjective octave slightly exceeds this, and piano tuning sounds better when the scale is stretched very slightly.

The equally tempered scale of 12 equal ratios within each octave is an excellent compromise between the different historical scales based on harmonic ratios. It has become the standard of reference, even for individual musical performances, which may deviate from it for artistic or other reasons.

Different musical instruments play over different ranges of *fundamental* frequency, shown in Fig.20.2.12. However, most musical sounds have many harmonics that are audibly significant to their tone spectra. Consequently high-fidelity recording and reproduction need a much wider frequency range.

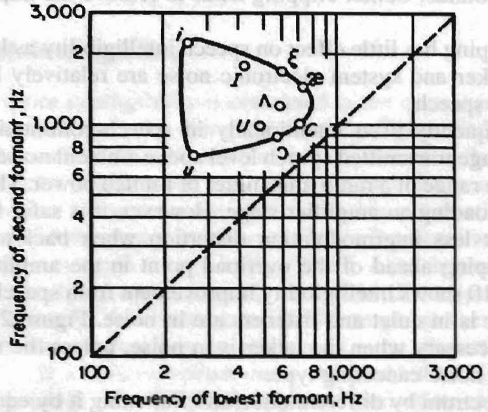

FIGURE 20.2.5 The center frequencies of the first two formants for the sustained English vowels plotted to show the characteristic differences.[7]

Sound Levels of Musical Instruments

The sound level from a musical instrument varies with the type of instrument, the distance from it, which note in the scale is being played, the dynamic marking in the printed music, the player's ability, and (on polyphonic instruments) the number of notes (and stops) played at the same time.

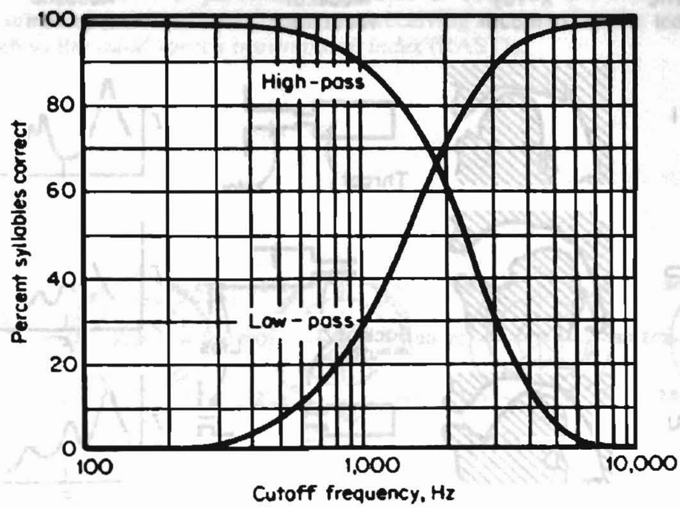

FIGURE 20.2.6 Syllable articulation score vs. low- or high-pass cutoff frequency.[8]

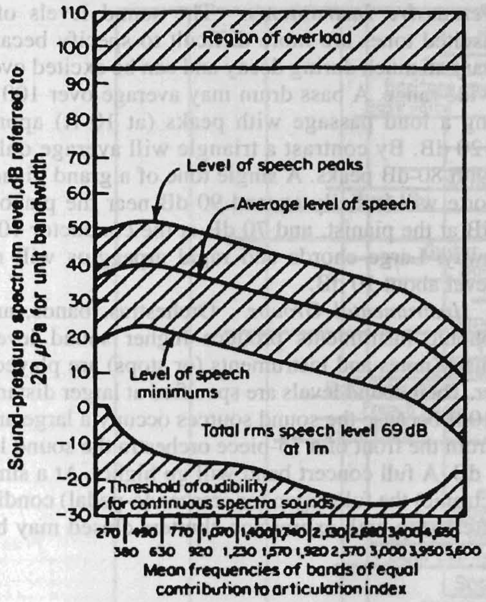

FIGURE 20.2.7 Speech area, bounded by speech peak and minimum spectrum-level curves, plotted on an articulation-index calculation chart.[9]

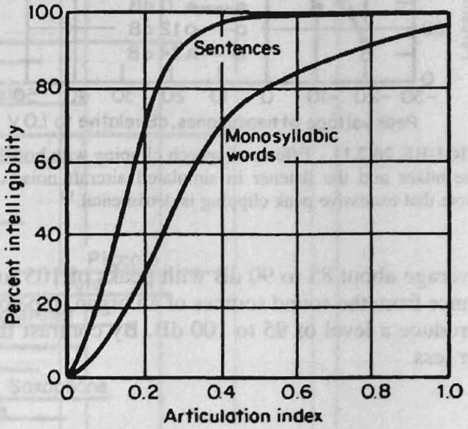

FIGURE 20.2.8 Sentence- and word-intelligibility prediction from calculated articulation index.[10]

Orchestral Instruments. The following sound levels are typical at a distance of 10 ft in a nonreverberant room. Soft (pianissimo) playing of a weaker orchestral instrument, e.g., violin, flute, bassoon, produces a typical sound level of 55 to 60 dB. Fortissimo playing on the same instrument raises the level to about 70 to 75 dB. Louder instruments, e.g., trumpet or tuba, range from 75 dB at pianissimo to about 90 dB at fortissimo.

Certain instruments have exceptional differences in sound level of low and high notes. A flute may change from 42 dB on a soft low note to 77 dB on a loud high note, a range of 35 dB. The French horn ranges from 43 dB (soft and low) to 93 dB (loud and high).

Sound levels are about 10 dB higher at 3 ft (inverse-square law) and 20 dB higher at 1 ft. The louder instruments, e.g., brass, at closer distances may overload some microphones and preamplifiers.

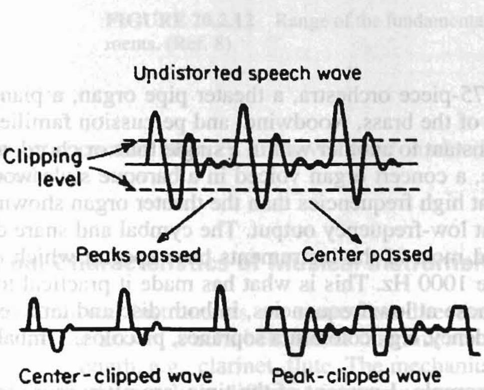

FIGURE 20.2.9 Two types of amplitude distortion of speech waveform.[5]

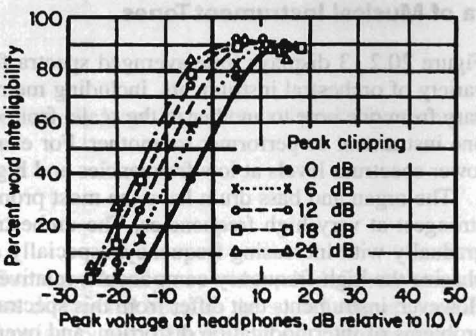

FIGURE 20.2.10 Advantages of peak clipping of noise-free speech waves, heard by listeners in ambient aircraft noise.[10]

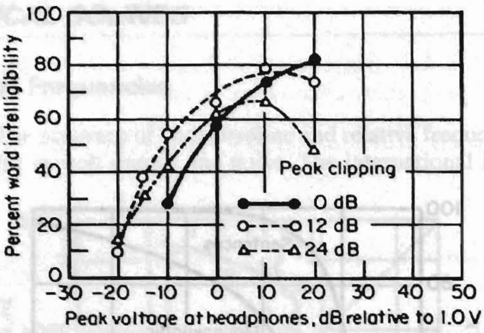

FIGURE 20.2.11 Effects of speech clipping with both the talker and the listener in simulated aircraft noise. Note that excessive peak clipping is detrimental.[10]

Percussive Instruments. The sound levels of shock-excited tones are more difficult to specify because they vary so much during decay and can be excited over a very wide range. A bass drum may average over 100 dB during a loud passage with peaks (at 10 ft) approaching 120 dB. By contrast a triangle will average only 70 dB with 80-dB peaks. A single tone of a grand piano played forte will initially exceed 90 dB near the piano rim, 80 dB at the pianist, and 70 dB at the conductor 10 to 15 ft away. Large chords and rapid arpeggios will raise the level about 10 dB.

Instrumental Groups. Orchestras, bands, and polyphonic instruments produce higher sound levels since many notes and instruments (or stops) are played together. Their sound levels are specified at larger distances than 10 ft because the sound sources occupy a large area; 20 ft from the front of a 75-piece orchestra the sound level will average about 85 to 90 dB with peaks of 105 to 110 dB. A full concert band will go higher. At a similar distance from the sound sources of an organ (pipe or electronic) the full-organ (or crescendo-pedal) condition will produce a level of 95 to 100 dB. By contrast the softest stop with expression shutters closed may be 45 dB or less.

Growth and Decay of Musical Sounds

These characteristics are quite different for different instruments. Piano or guitar tones quickly rise to an initial maximum, then gradually diminish until the strings are damped mechanically. Piano tones have a more rapid decay initially than later in the sustained tone. Orchestral instruments can start suddenly or smoothly, depending on the musician's technique, and they damp rather quickly when playing ceases. Room reverberation affects both growth and decay rates when the time constants of the room are greater than those of the instrument vibrators. This is an important factor in organ music, which is typically played in a reverberant environment.

Many types of musical tone have characteristic transients which influence timbre greatly. In the "chiff" of organ tone the transients are of different fundamental frequency. They appear and decay before steady state is reached. In percussive tones the initial transient is the cause of the tone (often a percussive noise), and the final transient is the result.

These transient effects should be considered in the design of audio electronics such as "squelch," automatic gain control, compressor, and background-noise reduction circuits.

Spectra of Musical Instrument Tones

Figure 20.2.13 displays time-averaged spectra for a 75-piece orchestra, a theater pipe organ, a piano, and a variety of orchestral instruments, including members of the brass, woodwind, and percussion families. These vary from one note to another in the scale, from one instant to another within a single tone or chord, and from one instrument or performer to another. For example, a concert organ voiced in a baroque style would have lower spectrum levels at low frequencies and higher at high frequencies than the theater organ shown.

The organ and bass drum have the most prominent low-frequency output. The cymbal and snare drum are strongest at very high frequencies. The orchestra and most of the instruments have spectra which diminish gradually with increasing frequency, especially above 1000 Hz. This is what has made it practical to preemphasize the high-frequency components, relative to those at low frequencies, in both disc and tape recording. However, instruments that differ from this spectral tendency, e.g., coloratura sopranos, piccolos, cymbals, create problems of intermodulation distortion, and overload.

Spectral peaks occurring only occasionally, for example, 1 percent of the time, are often more important to sound recording and reproduction than the peaks in the average spectra of Fig. 20.2.13. The frequency

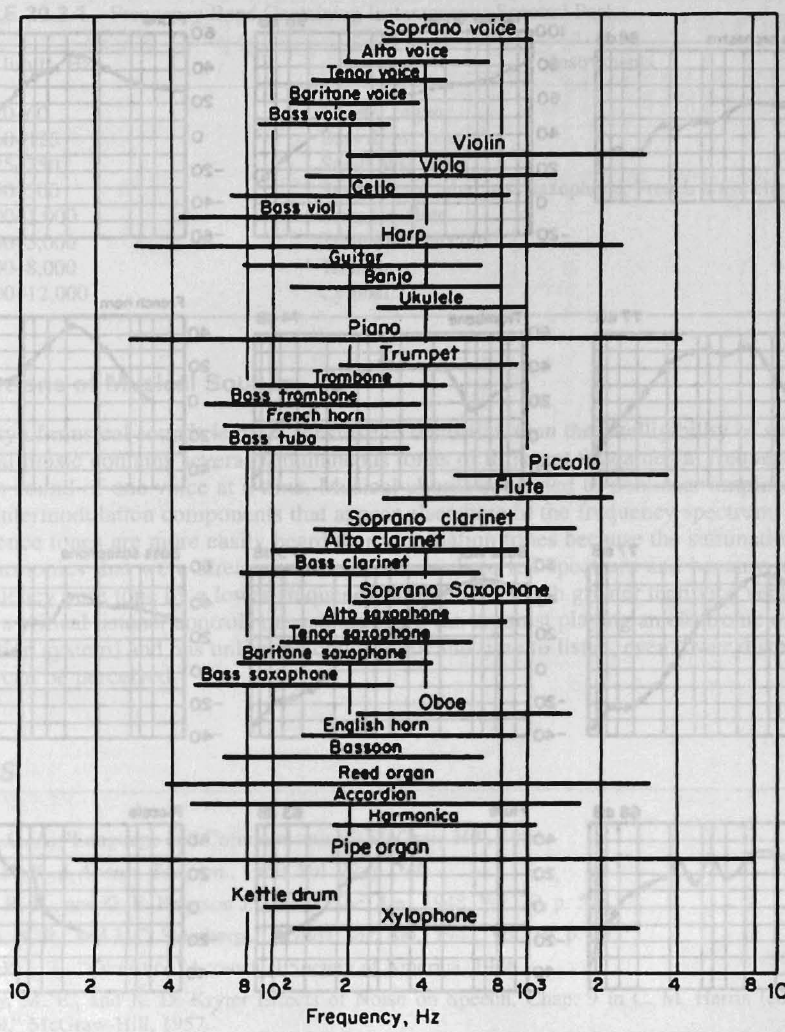

FIGURE 20.2.12 Range of the fundamental frequencies of voices and various musical instruments. (Ref. 8).

ranges shown in Table 20.2.1 have been found to have relatively large instantaneous peaks for the instruments listed.

Directional Characteristics of Musical Instruments

Most musical instruments are somewhat directional. Some are highly so, with well-defined symmetry, e.g., around the axis of a horn bell. Other instruments are less directional because the sound source is smaller than the wavelength, e.g., clarinet, flute. The mechanical vibrating system of bowed string instruments is complex, operating differently in different frequency ranges, and resulting in extremely variable directivity. This is significant for orchestral seating arrangements both in concert halls and recording studios.

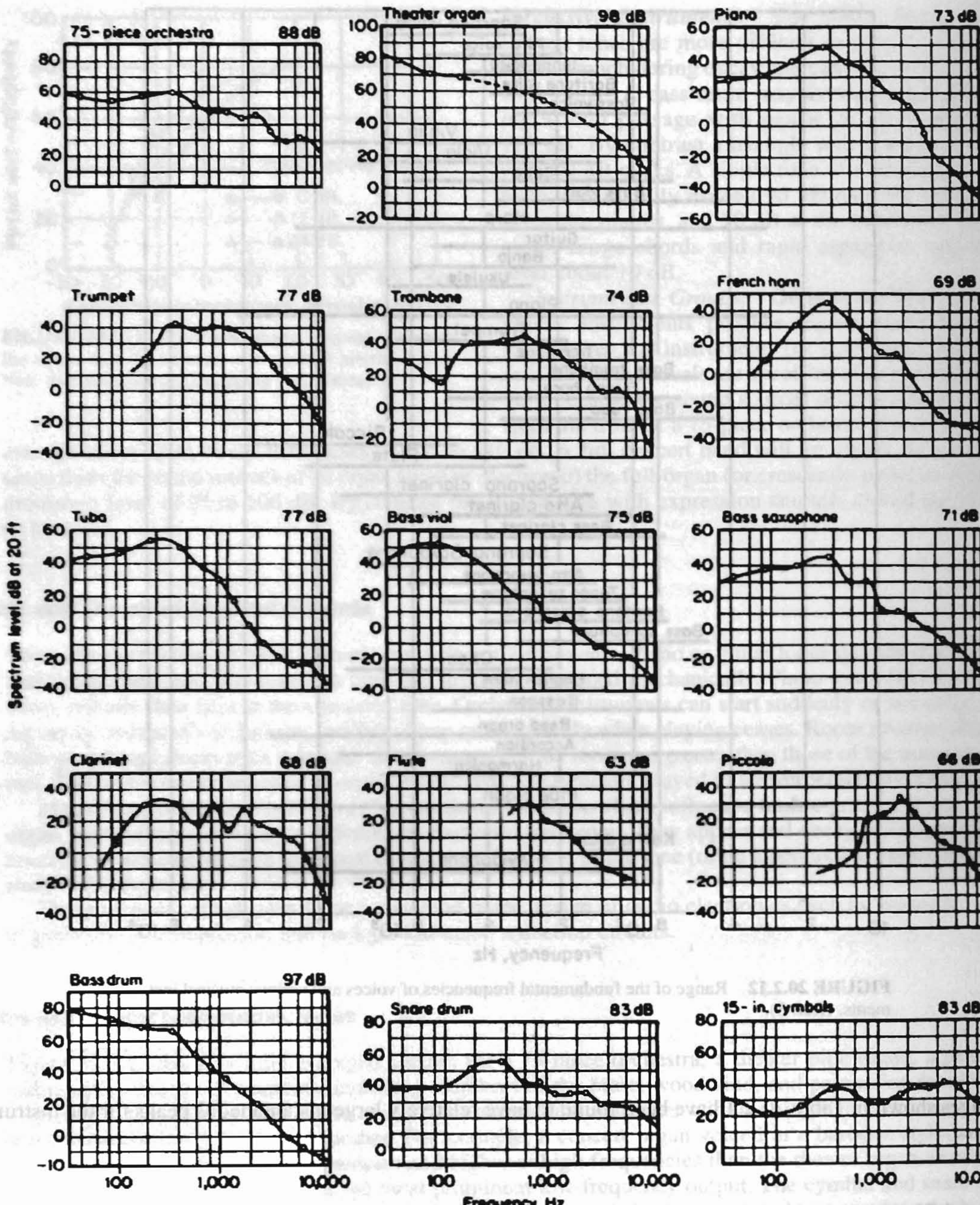

FIGURE 20.2.13 Time-averaged spectra of musical instruments.

TABLE 20.2.1 Frequency Band Containing Instantaneous Spectral Peaks

Band limits, Hz	Instruments
20–60	Theater organ
60–125	Bass drum, bass viol
125–250	Small bass drum
250–500	Snare drum, tuba, bass saxophone, French horn, clarinet, piano
500–1,000	Trumpet, flute
2,000–3,000	Trombone, piccolo
5,000–8,000	Triangle
8,000–12,000	Cymbal

Audible Distortions of Musical Sounds

The quality of musical sounds is more sensitive to distortion than the intelligibility of speech. A chief cause is that typical music contains several simultaneous tones of different fundamental frequency in contrast to typical speech sound of one voice at a time. Musical chords subjected to nonlinear amplification or transduction generate intermodulation components that appear elsewhere in the frequency spectrum.

Difference tones are more easily heard than summation tones because the summation tones are often hidden by harmonics that were already present in the undistorted spectrum and because auditory masking of a high-frequency pure tone by a lower-frequency pure tone is much greater than vice versa.

When a critical listener controls the sounds heard (an organist playing an electronic organ on a high-quality amplification system) and has unlimited opportunity and time to listen, even lower distortion (0.2 percent, for example) can be perceived.

REFERENCES

1. Miller, G. A. "Language and Communication," McGraw-Hill, 1951.
2. Dunn, H. K. *J. Acoust. Soc. Am.*, 1950, Vol. 22, p. 740.
3. Potter, R. K., and G. E. Peterson *J. Acoust. Soc. Am.*, 1948, Vol. 20, p. 528.
4. French, N. R., and J. C. Steinberg *J. Acoust. Soc. Am.*, 1947, Vol. 19, p. 90.
5. Beranek, L. L. "Acoustics,"Acoustical Society of America, 1986.
6. Hawley, M. E., and K. D. Kryter Effects of Noise on Speech, Chap. 9 in C. M. Harris (ed.), "Handbook of Noise Control," McGraw-Hill, 1957.
7. Olson, H. F. "Musical Engineering," McGraw-Hill, 1952.
8. Olson, H. F. "Elements of Acoustical Engineering," Van Nostrand, 1947.
9. Sivian, L. J., H. K. Dunn, and S. D. White *IRE Trans. Audio*, 1959, Vol. AU-7, p. 47; revision of paper in *J. Acoust. Soc. Am.*, 1931, Vol. 2, p. 33.
10. Hawley, M. E., and K. D. Kryter Effects of Noise on Speech, "Handbook of Noise Control," McGraw-Hill, 1957.

CHAPTER 20.3
MICROPHONES, LOUDSPEAKERS, AND EARPHONES

Daniel W. Martin

MICROPHONES

Sound-Responsive Elements

The sound-responsive element in a microphone may have many forms (Fig. 20.3.1). It may be a stretched membrane (*a*), a clamped diaphragm (*b*), or a magnetic diaphragm held in place by magnetic attraction (*c*). In these the moving element is either an electric or magnetic conductor, and the motion of the element creates the electric or magnetic equivalent of the sound directly.

Other sound-responsive elements are straight (*d*) or curves (*e*) conical diaphragms with various shapes of annular compliance rings, as shown. The motion of these diaphragms is transmitted by a drive rod from the conical tip to a mechanical transducer below.

Other widely used elements are a circular piston (*f*) bearing a circular voice coil of smaller diameter and a corrugated-ribbon conductor (*g*) of extremely low mass and stiffness suspended in a magnetic field.

Transduction Methods

Microphones have a great variety of transduction methods shown in Fig. 20.3.2.

The loose-contact transducer (Fig. 20.3.2*a*) was the first achieved by Bell in magnetic form and later made practical by Edison's use of carbonized hard-coal particles. It is widely used in telephones. Its chief advantage is its self-amplifying function, in which diaphragm amplitude variations directly produce electric resistance and current variations. Disadvantages include noise, distortion, and instability.

Moving-iron transducers have great variety, ranging from the historic pivoted armature (Fig. 20.3.2b) to the modern ring armature driven by a nonmagnetic diaphragm (Fig. 20.3.2*h*). In all these types a coil surrounds some portion of the magnetic circuit. The reluctance of the magnetic circuit is varied by motion of the sound-responsive element, which is either moving iron itself (Fig. 20.3.2*c* and *d*) or is coupled mechanically to the moving iron (Fig. 20.3.2*e–h*). In some of the magnetic circuits that portion of the armature surrounded by the coil carries very little steady flux, operating on differential magnetic flux only. Output voltage is proportional to moving-iron velocity.

Electrostatic transducers (Fig. 20.3.2*i*) use a polarizing potential and depend on capacitance variation between the moving diaphragm and a fixed electrode for generation of a corresponding potential difference. The *electret microphone* is a special type of electrostatic microphone that holds polarization indefinitely without continued application of a polarizing potential, an important practical advantage for many applications.

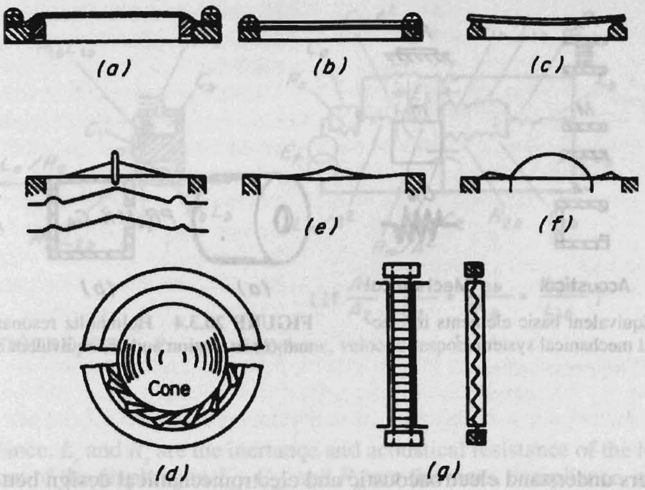

FIGURE 20.3.1 Sound-responsive elements in microphones.[1]

Piezoelectric transducers (Fig. 20.3.2*j*) create an alternating potential through the flexing of crystalline elements which, when deformed, generate a charge difference proportional to the deformation on opposite surfaces. Because of climatic effects and high electric impedance the rochelle salt commonly used for many years has been superseded by polycrystalline ceramic elements and by piezoelectric polymer.

Moving-coil transducers (Fig. 20.3.2*k*) generate potential by oscillation of the coil within a uniform magnetic field. The output potential is proportional to coil velocity.

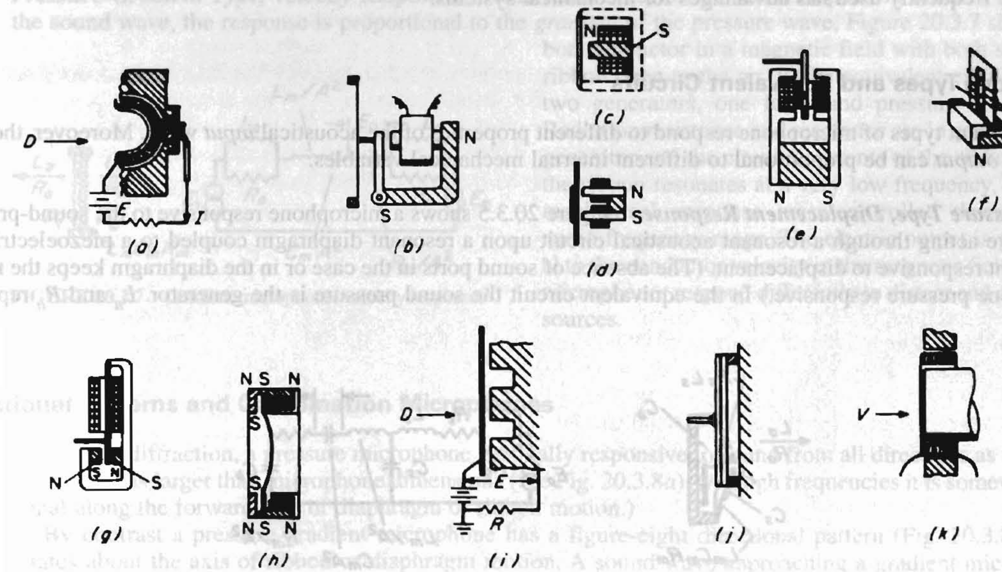

FIGURE 20.3.2 Microphone transduction methods.[1]

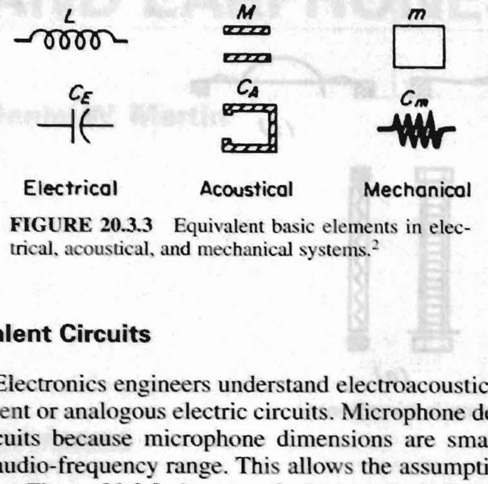

FIGURE 20.3.3 Equivalent basic elements in electrical, acoustical, and mechanical systems.[2]

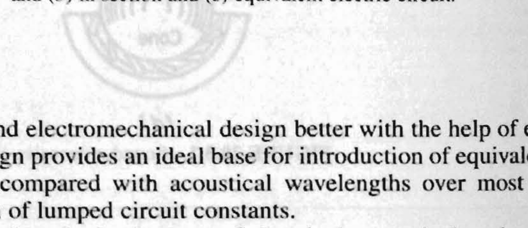

FIGURE 20.3.4 Helmholtz resonator in (*a*) perspective and (*b*) in section and (*c*) equivalent electric circuit.[2]

Equivalent Circuits

Electronics engineers understand electroacoustic and electromechanical design better with the help of equivalent or analogous electric circuits. Microphone design provides an ideal base for introduction of equivalent circuits because microphone dimensions are small compared with acoustical wavelengths over most of the audio-frequency range. This allows the assumption of lumped circuit constants.

Figure 20.3.3 shows equivalent symbols for the three basic elements of electrical, acoustical, and mechanical systems. In acoustical circuits the resistance is air friction or viscosity, which occurs in porous materials or narrow slots. Radiation resistance is another form of acoustical damping. Mechanical resistance is friction. Mass in the mechanical system is analogous to electric inductance. The acoustical equivalent is the mass of air in an opening or constriction divided by the square of its cross-sectional area. The acoustical analog of electric capacitance and mechanical-spring compliance is acoustical capacitance. It is the inverse of the stiffness of an enclosed volume of air under pistonlike action. Acoustical capacitance is proportional to the volume enclosed.

Figure 20.3.4 is an equivalent electric circuit for a Helmholtz resonator. Sound-pressure and air-volume current are analogous to electric potential and current, respectively. Other analog systems have been proposed. One frequently used has advantages for mechanical systems.

Microphone Types and Equivalent Circuits

Different types of microphone respond to different properties of the acoustical *input* wave. Moreover, the electric *output* can be proportional to different internal mechanical variables.

Pressure Type, Displacement Response. Figure 20.3.5 shows a microphone responsive to the sound-pressure wave acting through a resonant acoustical circuit upon a resonant diaphragm coupled to a piezoelectric element responsive to displacement. (The absence of sound ports in the case or in the diaphragm keeps the microphone pressure responsive.) In the equivalent circuit the sound pressure is the generator. L_a and R_a represent

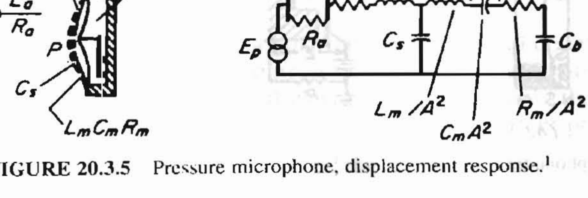

FIGURE 20.3.5 Pressure microphone, displacement response.[1]

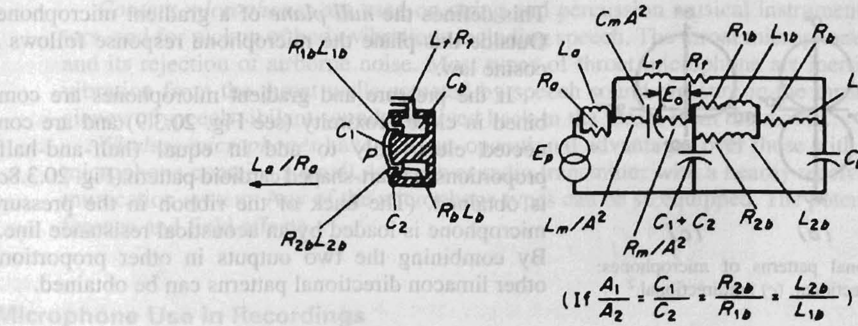

FIGURE 20.3.6 Pressure microphone, velocity response.[1]

the radiation impedance. L_s and R_s are the inertance and acoustical resistance of the holes; C_s is the capacitance of the volume in front of the diaphragm; L_m, C_m, and R_m are the mass, compliance, and resistance of the piezoelectric element and diaphragm lumped together; and C_b is the capacitance of the entrapped back volume of air. The electric output is the potential differential across the piezoelectric element. It is shown across the capacitance in the equivalent circuit because microphones of this type are designed to be stiffness-controlled throughout most of their operating range.

Pressure Type, Velocity Response. Figure 20.3.6 shows a moving-coil pressure microphone, which is a velocity-responsive transducer. In this microphone three acoustical circuits lie behind the diaphragm. One is behind the dome and another behind the annular rings. The third acoustical circuit lies beyond the acoustical resistance at the back of the voice-coil gap and includes a leak from the back chamber to the outside. This microphone is resistance-controlled throughout most of the range, but at low frequencies its response is extended by the resonance of the third acoustical circuit. Output potential is proportional to the velocity of voice-coil motion.

Pressure-Gradient Type, Velocity Response. When both sides of the sound-responsive element are open to the sound wave, the response is proportional to the *gradient* of the pressure wave. Figure 20.3.7 shows a ribbon conductor in a magnetic field with both sides of the ribbon open to the air. In the equivalent circuit there are two generators, one for sound pressure on each side. Radiation resistance and reactance are in series with each generator and the circuit constants of the ribbon. Usually the ribbon resonates at a very low frequency, making its mechanical response mass-controlled throughout the audio-frequency range. The electric output is proportional to the conductor velocity in the magnetic field. Gradient microphones respond differently to distant and close sound sources.

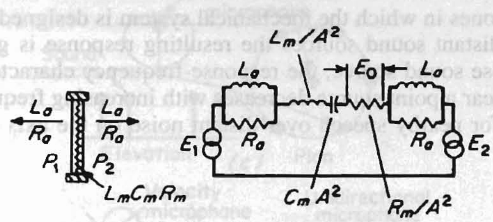

FIGURE 20.3.7 Gradient microphone, velocity response.[1]

Directional Patterns and Combination Microphones

Because of diffraction, a pressure microphone is equally responsive to sound from all directions as long as the wavelength is larger than microphone dimensions (see Fig. 20.3.8*a*). (At high frequencies it is somewhat directional along the forward axis of diaphragm or ribbon motion.)

By contrast a pressure-gradient microphone has a figure-eight directional pattern (Fig. 20.3.8*b*), which rotates about the axis of ribbon or diaphragm motion. A sound wave approaching a gradient microphone at 90° from the axis produces balanced pressure on the two sides of the ribbon and consequently no response.

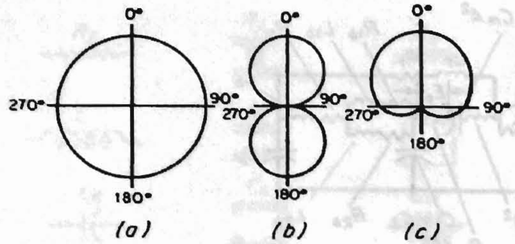

FIGURE 20.3.8 Directional patterns of microphones: (*a*) nondirectional; (*b*) bidirectional; (*c*) unidirectional.[2]

This defines the *null plane* of a gradient microphone. Outside this plane the microphone response follows a cosine law.

If the pressure and gradient microphones are combined in close proximity (see Fig. 20.3.9) and are connected electrically to add in equal (half-and-half) proportions, a heart-shaped cardioid pattern (Fig. 20.3.8*c*) is obtained. (The back of the ribbon in the pressure microphone is loaded by an acoustical resistance line.) By combining the two outputs in other proportions other limacon directional patterns can be obtained.

Phase-Shift Directional Microphones

Directional characteristics similar to those of the combination microphones can also be obtained with a single moving element by means of equivalent circuit analysis using acoustical phase-shift networks. Figure 20.3.10 shows a moving-coil, phase-shift microphone and its simplified equivalent circuit. The phase-shift network is composed of the rear-port resistance R_2 and inertance L_2, the capacitance of the volume under the diaphragm and within the magnet, and the impedance of the interconnecting screen. The microphone has a cardioid directional pattern.

Special-Purpose Microphones

Special-purpose microphones include two types that are superdirectional, two that overcome noise, and one without cables.

Line microphones use an approximate line of equally spaced pickup points connected through acoustically damped tubes to a common microphone diaphragm. The phase relationships at these points for an incident plane wave combine to give a sharply directional pattern along the axis if the line segment is at least one wavelength.

Parabolic microphones face a pressure microphone unit toward a parabolic reflector at its focal point, where sounds from distant sources along the axis of the parabola converge. They are effective for all wavelengths smaller than the diameter of the reflector.

Noise-canceling microphones are gradient microphones in which the mechanical system is designed to be stiffness-controlled rather than mass-controlled. For distant sound sources the resulting response is greatly attenuated at low frequencies. However, for a very close sound source, the response-frequency characteristic is uniform because the *gradient* of the pressure wave near a point source decreases with increasing frequency. Such a microphone provides considerable advantage for nearby speech over distant noise on the axis of the microphone.

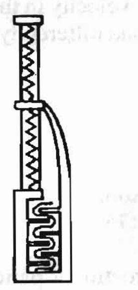

FIGURE 20.3.9 Combination unidirectional microphone.[1]

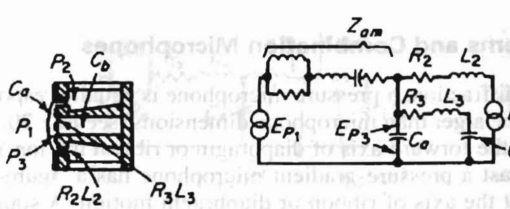

FIGURE 20.3.10 Phase-shift unidirectional microphone.[1]

Contact microphones are used on string and percussion musical instruments, on seismic-vibration detectors, and for pickup of body vibrations including speech. The throat microphone was noted for its convenience and its rejection of airborne noise. Most types of throat microphone are inertia-operated, the case receiving vibration from the throat walls actuated by speech sound pressure in the throat. The disadvantage is a deficiency of speech sibilant sounds received back in the throat from the mouth.

Wireless microphones have obvious operational advantages over those with microphone cords. A wireless microphone contains a small, low-power radio transmitter with a nearby receiver connected to an audio communication system. Any of the microphone types can be so equipped. The potential disadvantage is in rf interference and field effects.

Microphone Use in Recordings

The choice of microphone type and placement greatly affects the sound of a recording. For speech and dialogue recordings pressure microphones are usually placed near the speakers in order to minimize ambient-noise pickup and room reverberation. Remote pressure microphones are also used when a maximum room effect is desired.

In the playback of monophonic recordings room effects are more noticeable than they would have been to a listener standing at the recording microphone position because single-microphone pickup is similar to single-ear (monaural) listening, in which the directional clues of localization are lost. Therefore microphones generally need to be closer in a monophonic recording than in a stereophonic recording.

In television pickup of speech, where a boom microphone should be outside the camera angle, unidirectional microphones are often used because of their greater ratio of direct to generally reflected sound response.

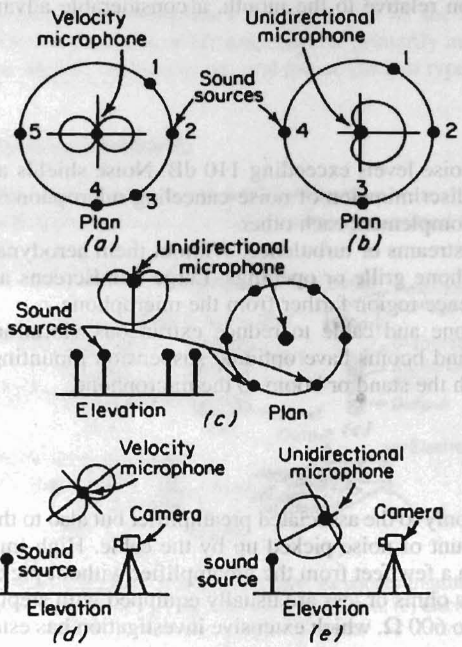

FIGURE 20.3.11 Use of directional microphones.[2]

Both velocity (gradient) microphones and unidirectional microphones can be used to advantage in broadcasting and recording. Figure 20.3.11*a* shows how instruments may be placed around a figure-eight directivity pattern to balance weaker instruments 2 and 5 against stronger instruments 1 and 3 with a potential noise source at point 4. In Fig. 20.3.11*b* source 2 is favored, with sources 1 and 3 somewhat reduced and source 4 highly discriminated against by the cardioid directional pattern. In Fig. 20.3.11*c* an elevated unidirectional microphone aimed downward responds uniformly to sources on a circle around the axis while discriminating against mechanical noises at ceiling level. Figure 20.3.11*d* places the camera noise in the null plane of a figure-eight pattern, and Fig. 20.3.11*e* shows a similar use for the unidirectional microphone. Camera position is less critical for the cardioid microphone than for the gradient microphone.

Early classical stereo recordings used variations of two basic microphone arrangements. In one scheme two unidirectional microphones were mounted close together with their axes angled toward opposite ends of the sound field to be recorded. This retained approximately the same arrival time and phase at both microphones, depending chiefly on the directivity patterns to create the sound difference in the two channels.

In the second scheme the two microphones (not necessarily directional) were separated by distances of 5 to 25 ft, depending on the size of the sound field to be recorded. Microphone axes (if directional) were again directed toward the ends of the sound field or group of sound sources. In this arrangement the time of arrival and phase differences were more important, and the effect of directivity was lessened. Each approach had its advantages and disadvantages.

With the arrival of tape recorders having many channels a trend has developed toward the use of more microphones and closer microphone placement. This offers much greater flexibility in mixing and rerecording, and it largely removes the effect of room reverberation from the recording. This may be either an advantage or a disadvantage depending on the viewpoint. Reverberation can be added later.

In sound-reinforcement systems for dramatic productions and orchestras the use of many microphones again offers operating flexibility. However, it also increases the probability of operating error, increased system noise, and acoustical feedback, making expert monitoring and mixing of the microphone outputs necessary.

An attractive alternative for multimicrophone audio systems is the use of independent voice-operated electronic control switches in each microphone channel amplifier, in combination with an automatic temporary reduction of overall system gain as more channels switch on, in order to prevent acoustical feedback. Automatic mixers have been devised to minimize speech signal dropouts, and to prevent the inadvertent operation of channel control switches by background noises.

Microphone Mounting

On podiums and lecterns microphones are typically mounted on fixed stands with adjustable arms. On stages they are mounted on adjustable floor stands. In mobile communication and in other situations where microphone use is occasional, handheld microphones are used during communication and are stowed on hangers at other times. For television and film recording, where the microphone must be out of camera sight, the microphones are usually mounted on booms overhead and are moved about during the action to obtain the best speech-to-noise ratio possible at the time. In two-way communication situations which require the talker to move about or to turn his head frequently, the microphone can be mounted on a boom fastened to his headset. This provides a fixed close-talking microphone position relative to the mouth, a considerable advantage in high-ambient-noise levels.

Microphone Accessories

Noise shields are needed for microphones in ambient noise levels exceeding 110 dB. Noise shields are quite effective at high frequencies, where the random-noise discrimination of noise-canceling microphones diminishes. Noise shields and noise-canceling microphones complement each other.

Windscreens are available for microphone use in airstreams or turbulence. Without them aerodynamically induced noise is produced by turbulence at the microphone grille or openings. Large windscreens are more effective than small ones because they move the turbulence region farther from the microphone.

Special sponge-rubber mountings for the microphone and cable to reduce extraneous vibration of the microphone are often used. Many microphone stands and booms have optional suspension mounting accessories to reduce shock and vibration transmitted through the stand or boom to the microphone.

Special Properties of Microphones

The source impedance of a microphone is important not only to the associated preamplifier but also to the allowable length of microphone cable and the type and amount of noise picked up by the cable. High-impedance microphones (10 kΩ or more) cannot be used more than a few feet from the preamplifier without pickup from stray fields. Microphones having an impedance of a few ohms or less are usually equipped with stepup transformers to provide a line impedance in the range of 30 to 600 Ω, which extensive investigation has established as the most noise-free line-impedance range.

The microphone unit itself can be responsive to hum fields at power-line frequencies unless special design precautions are taken. Most microphones have a hum-level rating based on measurement in a standard alternating magnetic field.

For minimum electrical noise balanced and shielded microphone lines are used, with the shield grounded only at the amplifier end of the line.

Microphone linearity should be considered when the sound level exceeds 100 dB, a frequent occurrence for loud musical instruments and even for close speech. Close-talking microphones, especially of the gradient type, are particularly susceptible to noise from breath and plosive consonants.

Specifications

Microphone specifications typically include many of the following items: type or mode of operation, directivity pattern, frequency range, uniformity of response within the range, output level at one or more impedances for a standard sound-pressure input (for example, 1 Pa or 10 dyn/cm^2), recommended load impedance, hum output level for a standard magnetic field (for example, 10^{-3} G), dimensions, weight, finish, mounting, power supply (if necessary), and accessories.

LOUDSPEAKERS

Introduction

A loudspeaker is an electroacoustic transducer intended to radiate acoustic power into the air, with the acoustic waveform equivalent to the electrical input waveform. An earphone is an electroacoustic transducer intended to be closely coupled acoustically to the ear. Both the loudspeaker and earphone are receivers of audio-electronic signals. The principal distinction between them is the acoustical loading. An earphone delivers sound to air in the ear. A loudspeaker delivers sound indirectly to the ear through the air.

The transduction methods of loudspeakers and earphones are historically similar and are treated together. An overview of loudspeaker developments of the closing 50 years of the last millennium is given by Gander.[3] However, since loudspeakers operate primarily into radiation resistance and earphones into acoustical capacitance, the design, measurement, and use of the two types of electroacoustic transducers will be discussed separately.

Transduction Methods

Early transducers for sound reproduction were of the mechanoacoustic type. Vibrations received by a stylus in the undulating groove of a record were transmitted to a diaphragm, placed at the throat of a horn for better acoustical impedance matching to the air, all without the aid of electronics. Electro-acoustics and electronics introduced many advantages and a variety of transduction methods including moving-coil, moving-iron, electrostatic, magnetostrictive, and piezoelectric (Fig. 20.3.12).

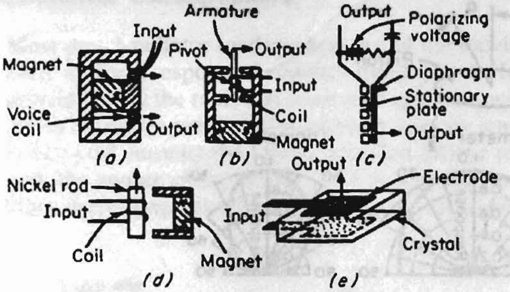

FIGURE 20.3.12 Loudspeaker (and earphone) transduction methods: (*a*) moving-coil; (*b*) moving-iron; (*c*) electrostatic; (*d*) magnetostrictive; (*e*) piezoelectric.[4]

Most loudspeakers are moving-coil type today, although moving-iron transducers were once widely used. Electrostatic loudspeakers are used chiefly in the upper range of audio frequencies, where amplitudes are small. Magnetostrictive and piezoelectric loudspeakers are used for underwater sound. All the transducer types are used in earphones except magnetostrictive.

Moving-Coil. The mechanical force on the moving coil of Fig. 20.3.12*a* is developed by the interaction of the current in the coil and the transverse magnetic field disposed radially across the gap between the magnet cap and the iron housing, which completes the magnetic circuit. The output force along the axis of the circular coil is applied to a sound radiator.

Moving-iron transducers reverse the mechanical roles of the coil and the iron. The iron armature surrounded by the stationary coil is moved by mechanical forces developed within the magnetic circuit. Moving-iron magnetic circuits have many forms. As an example in the balanced armature system (Fig. 20.3.12*b*) the direct magnetic flux passes only transversely through the ends of the armature centered within the two magnetic gaps. Coil current polarizes the armature ends oppositely, creating a force moment about the pivot point. The output force is applied from the tip of the armature to an attached sound radiator. In a balanced-diaphragm loudspeaker the armature is the radiator.

Electrostatic. In the electrostatic transducer (Fig. 20.3.12*e*) there is a dc potential difference between the conductive diaphragm and the stationary perforated plate nearby. Audio signals applied through a blocking capacitor superimpose an alternating potential, resulting in a force upon the diaphragm, which radiates sound directly.

Magnetostrictive transducers (Fig. 20.3.12*d*) depend on length fluctuations of a nickel rod caused by variations in the magnetic field. The output motion may be radiated directly from the end of the rod or transmitted into the attached mechanical structure.

Piezoelectric transducers are of many forms using crystals or polycrystalline ceramic materials. In simple form (Fig. 20.3.12*e*) an expansion-contraction force develops along the axis joining the electrodes through alternation of the potential difference between them.

Sound Radiators

The purpose of a sound radiator is to create small, audible air-pressure variations. Whether they are produced within a closed space by an earphone or in open air by a loudspeaker, the pressure variations require air motion or current.

Pistons, Cones, Ports. Expansion and contraction of a sphere is the classical configuration but most practical examples involve rectilinear motion of a piston, cone, or diaphragm. In addition to the primary direct radiation from moving surfaces, there is also indirect or secondary radiation from enclosure ports or horns to which the direct radiators are acoustically coupled.

Attempts have been made to develop other forms of sound radiation such as oscillating airstreams and other aerodynamic configurations with incidental use, if any, of moving mechanical members.

Directivity. Figure 20.3.13 shows the directional characteristics of a rigid circular piston for different ratios of piston diameter and wavelength of sound. (In three dimensions these curves are symmetrical around the axis of piston motion.) For a diameter of one-quarter wavelength the amplitude decreases 10 percent (approximately

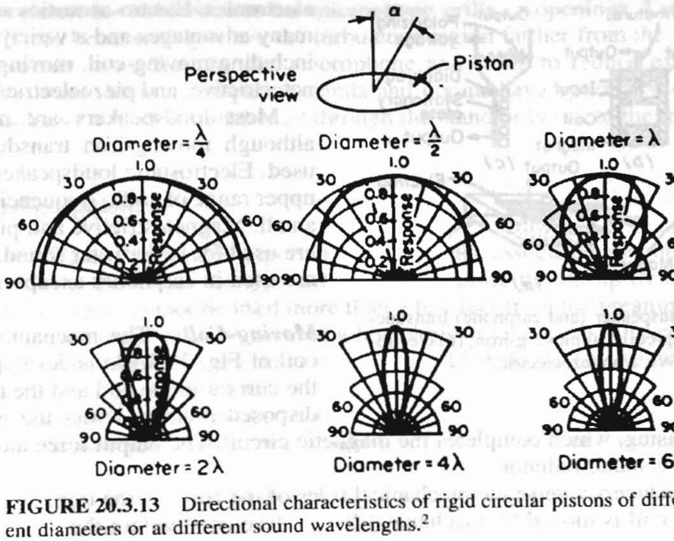

FIGURE 20.3.13 Directional characteristics of rigid circular pistons of different diameters or at different sound wavelengths.[2]

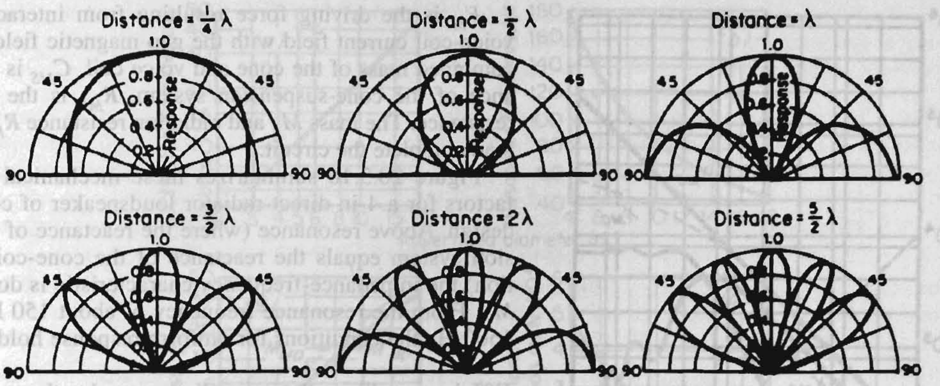

FIGURE 20.3.14 Directional characteristics of two equal small in-phase sound sources separated by different distances or different sound wavelengths.[2]

1 dB sound level) at 90° off axis. For a four-wavelength diameter the same drop occurs in only 5°. (The beam of an actual loudspeaker cone is less sharp than this at high frequencies, where the cone is not rigid.) Note that all the polar curves are smooth when the single-source piston vibrates as a whole.

Radiator Arrays. When two separate, identical small-sound sources vibrate in phase, the directional pattern becomes narrower than for one source. Figure 20.3.14 shows that for a separation of one-quarter wavelength the two-source beam is only one-half as wide as for a single piston. At high frequencies the directional pattern becomes very complex. (In three dimensions these curves become surfaces of revolution about the axis joining the two sources.)

Arrays of larger numbers of sound radiators in close proximity are increasingly directional. Circular-area arrays have narrow beams which are symmetrical about an axis through the center of the circle. Line arrays, e.g., column loudspeakers, are narrowly directional in planes containing the line and broadly directional in planes perpendicular to the line.

Direct-Radiator Loudspeakers

Most direct-radiator loudspeakers are of the moving-coil type because of simplicity, compactness, and inherently uniform response-frequency trend. The uniformity results from the combination of two simple physical principles: (1) the radiation resistance increases with the square of the frequency, and hence the radiated sound power increases similarly for constant velocity amplitude of the piston or cone; (2) for a constant applied force (voice-coil current) the mass-controlled (above resonance) piston has a velocity amplitude which decreases with the square of the frequency. Consequently a loudspeaker designed to resonate at a low frequency combines decreasing velocity with increasing radiation resistance to yield a uniform response within the frequency range where the assumptions hold.

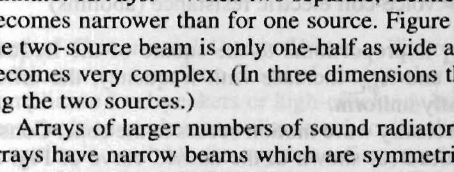

FIGURE 20.3.15 (*a*) Structure, (*b*) electric circuit, and (*c*) equivalent mechanical circuit for a direct-radiator moving-coil loudspeaker in a baffle.[4]

Equivalent Electric Circuits. Figure 20.3.15 shows a cross-sectional view of a direct-radiator loudspeaker mounted in a baffle, the electric voice-coil circuit, and the equivalent electric circuit of the mechanoacoustic system. In the voice-coil circuit e is the emf and R_{EG} the resistance of the generator, e.g., power-amplifier output, L and R_{EC} are the inductance and resistance of the voice coil. Z_{EM} is the motional electric impedance from the mechanoacoustic system.

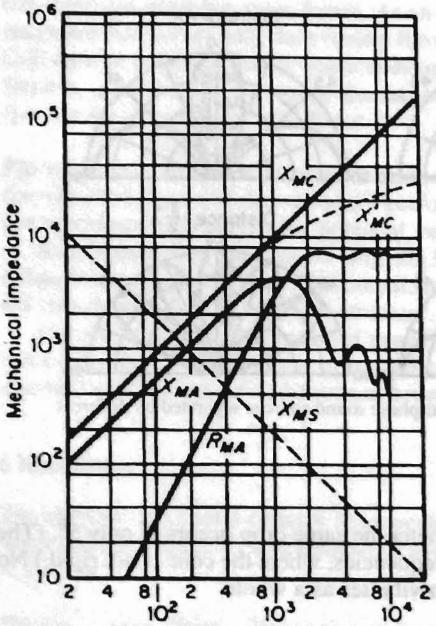

FIGURE 20.3.16 Components of a mechanical impedance of a typical 4-in loudspeaker.[4]

F_M is the driving force resulting from interaction of the voice-coil current field with the gap magnetic field. M_C is the combined mass of the cone and voice coil. C_{MS} is the compliance of the cone-suspension system. R_{MS} is the mechanical resistance. The mass M_A and radiation resistance R_{MA} of the air load complete the circuit.

Figure 20.3.16 summarizes these mechanical impedance factors for a 4-in direct-radiator loudspeaker of conventional design. Above resonance (where the reactance of the suspension system equals the reactance of the cone-coil combination) the impedance-frequency characteristic is dominated by M_C. From the resonance frequency of about 150 Hz to about 1500 Hz the conditions for uniform response hold.

Efficiency. Since R_{MA} is small compared to the magnitudes of the reactive components, the efficiency of the loudspeaker in this frequency range can be expressed as

$$\text{Efficiency} = \frac{100(Bl)^2 R_{MA}}{R_{EC}(X_{MA} + X_{MC})^2} \text{ percent} \qquad (1)$$

where B = gap flux density (G)
l = voice-coil conductor length (cm)
R_{EC} = voice-coil electric resistance (abohms)

Since R_{MA} is proportional to the square of the frequency and both X_{MA} and X_{MC} increase with frequency, the efficiency is theoretically uniform.

All this has assumed that the cone moves as a whole. Actually wave motion occurs in the cone. Consequently at high frequencies the mass reactance is somewhat reduced (as shown in the dashed curve of Fig. 20.3.16), tending to improve efficiency beyond the frequency where radiation resistance becomes uniform.

Magnetic Circuit. Most magnets now are a high-flux, high-coercive permanent type, either an alloy of aluminum, cobalt, nickel, and iron, or a ferrite of iron, cobalt, barium, and nickel. The magnet may be located in the core of the structure or in the ring, or both. However, magnetization is difficult when magnets are oppositely polarized in the core and ring.

Air-gap flux density varies widely in commercial designs from approximately 3000 to 20,000 G. Since most of the reluctance in the magnetic circuit resides in the air gap, the minimum practical voice-coil clearance in the gap compromises the maximum flux density. Pole pieces of heat-treated soft nickel-iron alloys, dimensionally tapered near the gap, are used for maximum flux density.

Voice Coils. The voice coil is a cylindrical multilayer coil of aluminum or copper wire or ribbon. Aluminum is used in high-frequency loudspeakers for minimum mass and maximum efficiency. Voice-coil impedance varies from 1 to 100 Ω with 4, 8, and 16 Ω standard. For maximum efficiency the voice-coil and cone masses are equal. However, in large loudspeakers the cone mass usually exceeds the voice-coil mass. Typically the voice-coil mass ranges from tenths of a gram to 5 g or more.

Cones. Cone diameters range from 1 to 18 in. Cone mass varies from tenths of a gram to 100 g or more. Cones are made of a variety of materials. The most common is paper deposited from pulp on a wire-screen form in a felting process. For high-humidity environment cones are molded from plastic materials, sometimes with a cloth or fiber-glass base. Some low-frequency loudspeaker cones are molded from low-density plastic foam to achieve greater rigidity with low density.

So far piston action has been assumed in which the cone moves as a whole. Actually at high frequencies the cone no longer vibrates as a single unit. Typically there is a major dip in response resulting from quarter-wave

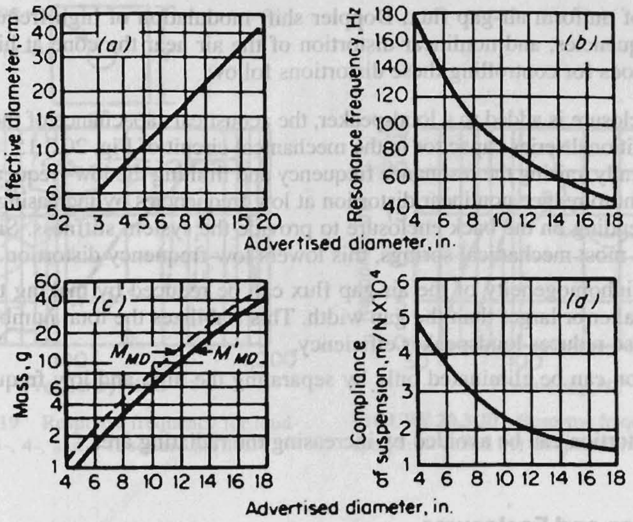

FIGURE 20.3.17 Typical cone and coil design values.[5]

reflection from the circular rim of the cone back to the voice coil. For loudspeaker cones in the range of 8 to 15 in. diameter this dip usually occurs in the range of 1 to 2 kHz.

Typical Commercial Design Values. Figure 20.3.17 shows typical values for several cone and voice-coil design parameters for a range of loudspeaker diameters. These do not apply to extreme cases, such as high-compliance loudspeakers or high-efficiency horn drivers. The effective piston diameter (Fig. 20.3.17*a*) is less than the loudspeaker cone diameter because the amplitude falls off toward the edges. A range of resonance frequencies is available for any cone diameter, but Fig. 20.3.17*b* shows typical values. In Fig. 20.3.17*c* typical cone mass is *M* including the voice coil and *M′* excluding the voice coil. Figure 20.3.17*d* shows typical cone-suspension compliance.

Impedance. A major peak results from motional impedance at primary mechanical resonance. Impedance is usually uniform above this peak until voice-coil inductance becomes dominant over resistance.

Power Ratings. Different types of power rating are needed to express the performance capabilities of loudspeakers. The large range of typical loudspeaker efficiency makes the acoustical power-delivering capacity quite important. The electrical power-receiving capacity (without overload or damage) determines the choice of power amplifier.

Loudspeaker efficiencies are seldom measured but are often compared by measuring the sound-pressure level at 4 ft on the loudspeaker axis for 1-W audio input. High-efficiency direct radiators provide 95 to 100 dB. Horn loudspeakers are typically higher by 10 dB or more, being both more efficient and more directional.

Loudspeakers are also rated by the maximum rms power output of amplifiers which will not damage the loudspeaker or drive it into serious distortion on peaks. Such ratings usually assume that the amplifier will seldom be driven to full power. For example, a 30-W amplifier will seldom be required to deliver more than 10 W rms of music program material. Otherwise music peaks would be clipped and sound distorted.

However, in speech systems for high-ambient-noise levels the speech peaks may be clipped intentionally, causing the loudspeaker to receive the full 30 W much of the transmission time. Then the loudspeaker must handle large excursions without mechanical damage to the cone suspension and without destroying the cemented coil or charring the form.

Distortion. Nonlinear distortion in a loudspeaker is inherently low in the mass-controlled range of frequencies. However, distortion is produced by nonlinear cone suspension at low frequencies, voice-coil motion beyond the limits of uniform air-gap flux, Doppler shift modulation of high-frequency sound by large cone velocity at low frequencies, and nonlinear distortion of the air near the cone at high powers (particularly in horn drivers). Methods for controlling these distortions follow.

1. When a back enclosure is added to a loudspeaker, the acoustical capacitance of the enclosed volume is represented by an additional series capacitor in the mechanical circuit of Fig. 20.3.15. Insufficient volume stiffens the cone acoustically, raising the resonance frequency and limiting the low-frequency range of the loudspeaker. It is convenient to reduce nonlinear distortion at low frequencies by increasing the cone-suspension compliance and depending on the back enclosure to provide the system stiffness. Since an enclosed volume is more linear than most mechanical springs, this lowers low-frequency distortion.
2. Distortion from inhomogeneity of the air-gap flux can be reduced by making the voice-coil length either considerably smaller or larger than the gap width. This stabilizes the total number of lines passing through the coil, but it also reduces loudspeaker efficiency.
3. Doppler distortion can be eliminated only by separating the high and low frequencies in a multiple loudspeaker system.
4. Air-overload distortion can be avoided by increasing the radiating area.

Loudspeaker Mountings and Enclosures

Figure 20.3.18 shows a variety of mountings and enclosures. An unbaffled loudspeaker is an acoustic doublet for wavelengths greater than the rim diameter. In this frequency range the acoustical power output for constant cone velocity is proportional to the fourth power of the frequency.

Baffles. In order to improve efficiency at low frequencies it is necessary to separate the front and back waves. Figure 20.3.18*a* is the simplest form of baffle. The effect of different baffle sizes is given in Fig. 20.3.19. Response dips occurring when the acoustic path from front to back is a wavelength are eliminated by irregular baffle shape or off-center mounting.

Enclosures. The widely used open-back cabinet (Fig. 20.3.18*b*) is noted for a large response peak produced by open-pipe acoustical resonance. A closed cabinet (Fig. 20.3.18*c*) adds acoustical stiffness at low frequencies where the wavelength is larger than the enclosure. At higher frequencies the internal acoustical resonances create response irregularities requiring internal acoustical absorption.

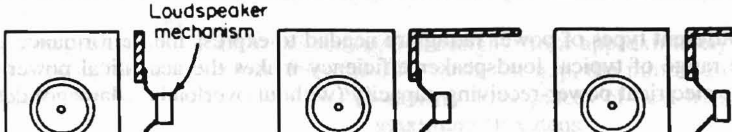

FIGURE 20.3.18 Mountings and enclosures for direct-radiator loudspeaker: (*a*) flat baffle; (*b*) open-back cabinet; (*c*) closed cabinet; (*d*) ported closed cabinet; (*e*) labyrinth; (*f*) folded horn.[4]

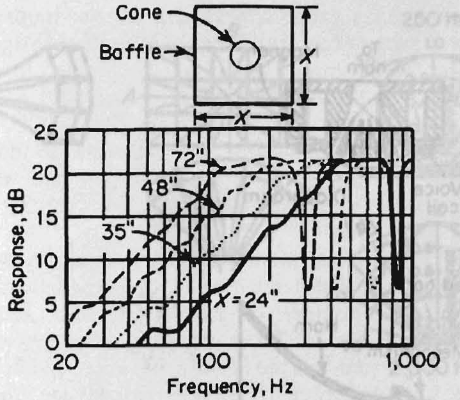

FIGURE 20.3.19 Response frequency for loudspeaker in 2-, 3-, 4-, and 6-ft square baffles.[6]

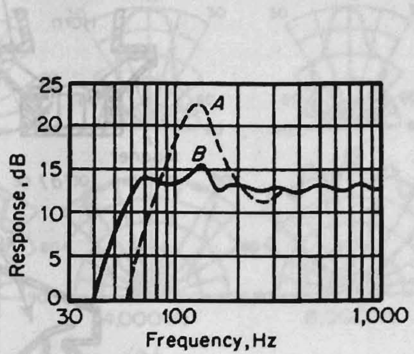

FIGURE 20.3.20 Response frequency for loudspeaker in closed (A) and ported (B) cabinets.[6]

Ported Enclosures (Fig. 20.3.18d). Enclosure volume can be minimized without sacrificing low-frequency range by providing an appropriate port in the enclosure wall. Acoustical inertance of the port should resonate with the enclosure capacitance at a frequency about an octave below cone-resonance frequency. B, Fig. 20.3.20, shows that this extends the low-frequency range. This is most effective when the port area equals the cone-piston area. Port inertance can be increased by using a duct. An extreme example of ducting is the acoustical labyrinth (Fig. 20.3.18*e*). When duct work is shaped to increase cross section gradually, the labyrinth becomes a low-frequency horn (Fig. 20.3.18*f*).

Direct-radiator loudspeaker efficiency is typically 1 to 5 percent. Small, highly damped types with miniature enclosures may be only 0.1 percent. Transistor amplifiers easily provide the audio power for domestic loudspeakers. However, in auditorium, outdoor, industrial, and military applications much higher efficiency is required.

Horn Loudspeakers

Higher efficiency is obtained with an acoustic horn, which is a tube of varying cross section having different terminal areas to provide a change of acoustic impedance. Horns match the high impedance of dense diaphragm material to the low air impedance. Horn shape or taper affects the acoustical transformer response. Conical, exponential, and hyperbolic tapers have been widely used. The potential low-frequency cutoff of a horn depends on its taper rate. Impedance transforming action is controlled by the ratio of mouth to throat diameter.

Horn Drivers. Figure 20.3.21 shows horn-driving mechanisms and straight and folded horns of large- and small-throat types. A large-throat driver (Fig. 20.3.21*a*) resembles a direct-radiator loudspeaker with a voice-coil diameter of 2 to 3 in. and a flux density around 15,000 G. A small-throat driver (Fig. 20.3.21*b*) resembles a moving-coil microphone structure. Radiation is taken from the back of the diaphragm into the horn throat through passages which deliver in-phase sound from all diaphragm areas. Diaphragm diameters are 1 to 4 in. with throat diameters of $^1/_4$ to 1 in. Flux density is approximately 20,000 G.

Large-Throat Horns. These are used for low-frequency loudspeaker systems. A folded horn (Fig. 20.3.21*c*) is preferred over a straight horn (Fig. 20.3.21*d*) for compactness.

Small-Throat Horns. A folded horn (Fig. 20.3.21*e*) with sufficient length and gradual taper can operate efficiently over a wide frequency range. This horn is useful for outdoor music reproduction in a range of 100 to 5000 Hz. Response smoothness is often compromised by segment resonances. Extended high-frequency range requires a straight-axis horn (Fig. 20.3.21*f*).

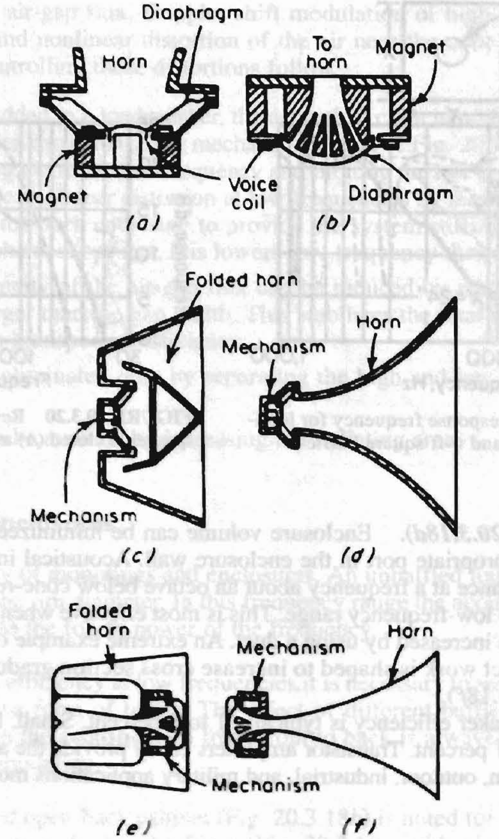

FIGURE 20.3.21 Horns and horn drivers: (*a*) large-throat driver; (*b*) small-throat driver; (*c*) folded large-throat horn; (*d*) straight large-throat horn; (*e*) folded small-throat horn; (*f*) straight small-throat horn.[4]

Horn Directivity. Large-mouth horns of simple exponential design produce high-directivity radiation that tends to narrow with increasing frequency (as in Fig. 20.3.13). In applications requiring controlled directivity over a broad angle and a wide frequency range, a horn array (shown in Fig. 20.3.22*a*) can be used, with numerous small horn mouths spread over a spherical surface and throats converging together. Figure 20.3.22*b* shows the directional characteristics. Single sectoral horns with radial symmetry can provide cylindrical wavefronts with smoother directional characteristics which are controlled in one plane. Recent rectangular or square-mouth "quadric" horns, designed by computer to have different conical expansion rates in horizontal and vertical planes, provide controlled directivity in both planes over a wide frequency range.

Special Loudspeakers

Special types of loudspeakers for limited applications include the following.

Electrostatic high-frequency units have an effective spacing of about 0.001 in. between a thin metalized coating on plastic and a perforated metal backplate. This spacing is necessary for sensitivity comparable to moving-coil loudspeakers, but it limits the amplitude and the frequency range. Extension of useful response to the lower frequencies can be obtained with larger spacing, for example, 1/16 in., with a polarizing potential of several thousand volts. This type of unit employs push-pull operation.

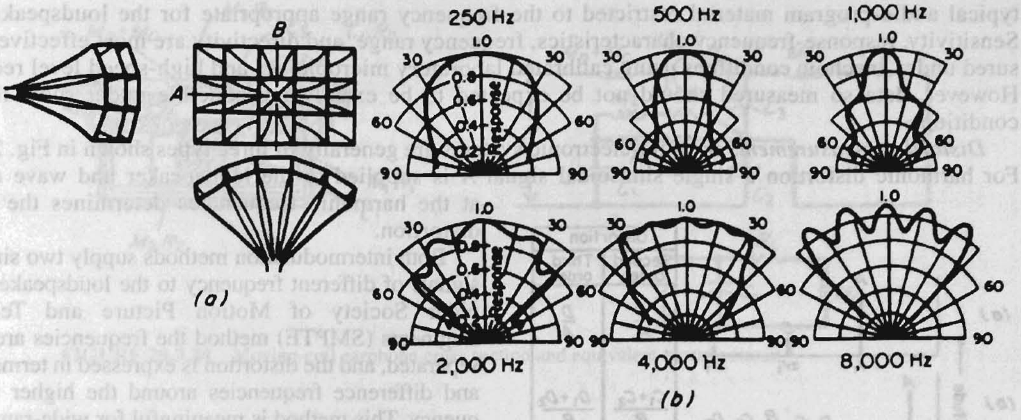

FIGURE 20.3.22 Horn array (cellular) and directional characteristics: (*a*) array; (*b*) horizontal directional curves.[6]

Modulated-airflow loudspeakers have an electromechanical mechanism for modulating the airstream from a high-pressure pneumatic source into a horn. Low audio power controls large acoustical power in this system. A compressor is also needed. Nonlinear distortion in the air and reduced speech intelligibility have been limitations of this high-power system.

Loudspeaker Specifications and Measurements

Typical loudspeaker specifications are shown in Table 20.3.1 for a variety of loudspeaker types.

Loudspeaker impedance is proportional to the voltage across the voice coil when driven by a high-impedance constant-current source. Continuous power ratings are obtained from sustained life tests with

TABLE 20.3.1 Characteristics of a Variety of Loudspeaker Types

Company	Altec	Altec	Bozak	RCA
Model no.	775C	1505B horn 290D driver	CM-109–23	LC1B
Type	Direct radiator	Cellular horn (3 × 5)	Three-way column	Duo-cone
Sensitivity (at 4 ft for 1 W), dB	95	110	106	95
Frequency range, Hz	40–15,000	300–8,000	65–13,000	25–16,000 (± 4 dB)
Impedance, Ω	8	4	8	15
Power rating, W	15	100	200	20
Distribution angle, deg	90	105 horizontal 60 vertical	90 horizontal 30 vertical	120
Voice-coil diameter, in.	2	2.8	(3 sizes)	(2 cones)
Cone resonance, Hz	52	…	(3 sizes)	22
Crossover frequency, Hz	…	500	800, 2,500	1,600
Diameter, in.	$8^3/_8$	$18^1/_2$ high $30^1/_2$ wide	57 in. high $22^3/_4$ wide	17
Depth, in.	$2^1/_4$	30	$15^3/_4$	$7^1/_2$
Weight, lb	$3^3/_4$	43	250	21

typical audio-program material restricted to the frequency range appropriate for the loudspeaker type. Sensitivity, response-frequency characteristics, frequency range, and directivity are most effectively measured under anechoic conditions using calibrated laboratory microphones and high-speed level recorders. However, data so measured should not be expected to be exactly reproducible under room-listening conditions.

Distortion measurements in audio-electronic systems are generally of three types shown in Fig. 20.3.23. For harmonic distortion a single sinusoidal signal A is supplied to the loudspeaker and wave analysis at the harmonic frequencies determines the percent distortion.

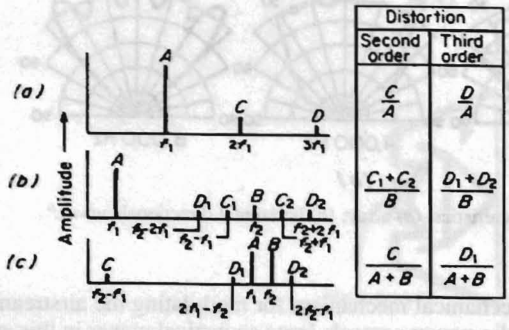

FIGURE 20.3.23 Methods of measuring nonlinear distortion: (*a*) harmonic; (*b*) intermodulation method of SMPTE; (*c*) intermodulation method of CCIF.[7]

Both intermodulation methods supply two sinusoidal signals of different frequency to the loudspeaker. In the older Society of Motion Picture and Television Engineers (SMPTE) method the frequencies are widely separated, and the distortion is expressed in terms of sum and difference frequencies around the higher test frequency. This method is meaningful for wide-range loudspeaker systems.

The CCIF (International Telephone Consultative Committee) method is applicable to narrow-range systems and loudspeakers receiving input at high frequencies. It supplies two high frequencies to the loudspeaker and checks the low difference frequency.

Transient intermodulation distortion, resulting from nonlinear response to steep wavefronts, is measured by adding square-wave (3.18-kHz) and sine-wave (15-kHz) inputs, with a 4:1 amplitude ratio, and observing the multiple sum- and difference-frequency components added to the output spectrum.

EARPHONES

The transduction methods are the same as for loudspeakers. Telephone and hearing aid receivers are usually moving-iron. Most military headsets are now moving-coil. Piezoelectric, moving-coil, and electrostatic types are used for listening to recorded music.

Equivalent Electric Circuits

Figure 20.3.24 shows a cross section of a moving-coil earphone and the equivalent electric circuit. The voice-coil force drives the voice coil and diaphragm. (Mechanical resonance of earphone diaphragms occurs at a high audio frequency in contrast to loudspeakers.) Diaphragm motion creates sound pressure in several spaces behind the diaphragm and the voice coil and between the diaphragm and the earcap. Inertance and resistance of the connecting holes and clearances combine with the capacitance of the spaces to add acoustical resonances. Z is the acoustical impedance of the ear.

Idealized Ear Loading

The ear is approximately an acoustical capacitance. However, acoustical leakage adds a parallel resistance-inertance path affecting low-frequency response. At high frequencies the ear canal-length resonance is a factor.

Since the ear is a capacitance, the goal of earphone design is a constant diaphragm amplitude throughout the frequency range. This requires a stiffness-controlled system or a high-resonance frequency. The potential across the ear is analogous to sound pressure within the ear cavity. This sound pressure is proportional to

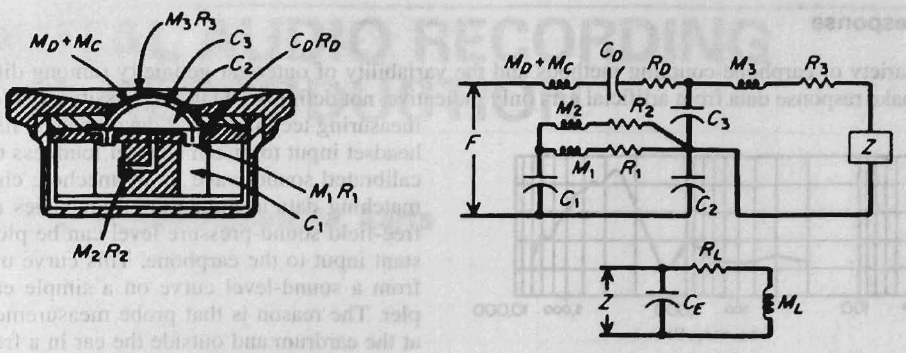

FIGURE 20.3.24 Moving-coil earphone cross section and equivalent electric circuit.[8]

diaphragm area and inversely proportional to enclosed volume. Earphone loading conditions are extremely varied for different types of earphone mountings.

Earphone Mountings

The most widely used earphone is the single receiver unit on a telephone handset. It is intended to be held against the ear but is often tilted away, leaving considerable leakage.

Headsets provide better communication than handsets because they supply sound to both ears and shield them.

A remote earphone can drive the ear canal through a small acoustic tube. The length may be an inch or two for hearing aids and several feet for music listening on aircraft.

Efficiency, Impedance, and Driving Circuits

Moving-iron earphones and microphones can be made efficient enough to operate as sound-powered (batteryless) telephones. Efficient magnet structures, minimum mechanical and acoustical damping, and minimum volume of acoustical coupling are required for this purpose. In some earphone applications overall efficiency is less critical, and wearer comfort is important.

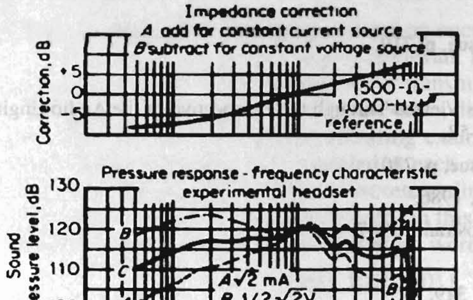

FIGURE 20.3.25 Effect of source impedance upon earphone response curve: (*a*) constant current; (*b*) constant voltage; (*c*) constant power.[9]

Insert earphones need less efficiency than external earphones because the enclosed volume is much smaller; however, they require moderate efficiency to save the amplifier batteries.

Circumaural earphones are frequently driven by amplifiers otherwise used for loudspeakers. Here efficiency is less important than power-delivering capacity.

Typically 1 mW of audio power to an earphone will produce 100 to 110 dB in a standard 6-cm^3 coupler. The same earphone will produce less sound level in an earmuff than in an ear cushion and more when coupled to an ear insert.

The shape of the enclosed volume also affects response. The farther the driver is from the eardrum the lower the frequency of standing-wave resonance. Small tube diameters produce high-frequency attenuation.

The response-frequency characteristic of moving-iron or piezoelectric earphones is quite dependent on source impedance. A moving-iron earphone having uniform response when driven at constant power will have a rising response (with increasing frequency) at constant current and a falling response at constant voltage (Fig. 20.3.25).

Real-Ear Response

The variety of earphone-coupling methods and the variability of outer-ear geometry (among different listeners) make response data from artificial ears only indicative, not definitive. Out of necessity a real-ear response-measuring technique was developed. A listener adjusts headset input to match headset loudness to an external calibrated sound wave in an anechoic chamber. From matching data at numerous frequencies an equivalent free-field sound-pressure level can be plotted for constant input to the earphone. This curve usually differs from a sound-level curve on a simple earphone coupler. The reason is that probe measurements of sound at the eardrum and outside the ear in a free field differ because of ear amplification and diffraction about the head (Fig. 20.3.26).

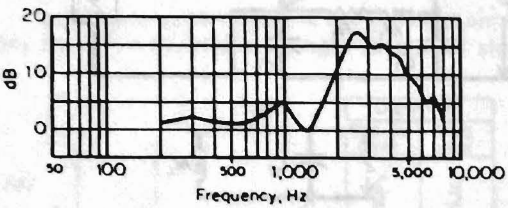

FIGURE 20.3.26 Relative level of sound pressures at the listener's eardrum and in the free sound field.[10]

Acoustic attenuation by earphones can be measured either by threshold shift or by matching the loudness of tones heard from an external loudspeaker, with and without the headset on. The sound-level difference is plotted as attenuation in decibels.

Monaural, Diotic, and Binaural Listening

A handset earphone provides monaural listening. Diotic listening with the same audio signal in both earphones localizes sound within the head. This is not unpleasant and may actually be an aid to concentration. In natural-binaural listening the ears receive sound differently from the same source unless it is directly on the listening axis. Usually there are differences in phase, arrival time, and spectrum (because of diffraction about the head).

Recordings provide true binaural effects only if the two recording microphones are on an artificial head. Stereophonic microphones are usually separated much farther, so that headset listening gives an exaggerated effect. For some listeners this is an enhancement.

REFERENCES

1. Bauer, B. B. *Proc. IRE*, 1962, Vol. 50, 50th Anniversary Issue, p. 719.
2. Olson, H. F. "Musical Engineering," McGraw-Hill, 1952.
3. Gander, M. R. Fifty Years of Loudspeaker Developments as Viewed Through the Perspective of the Audio Engineering Society, *J. Audio Eng. Soc.*, 1998, Vol. 46, No. 1/2, pp. 43–58.
4. Olson, H. F. *Proc. IRE*, 1962, Vol. 50, 50th Anniversary Issue, p. 730.
5. Beranek, L. L. "Acoustics," Acoustical Society of America, 1986.
6. Olson, H. F. "Elements of Acoustical Engineering," Van Nostrand, 1947.
7. Beranek, L. L. *Proc. IRE*, 1962, Vol. 50, p. 767.
8. Anderson, L. J. *J. Soc. Motion Pict. Eng.*, 1941, Vol. 37, p. 319.
9. Martin, D. W., and L. J. Anderson *J. Acoust. Soc. Am.*, 1947, Vol. 19, p. 63.
10. Wiener, F. M., and D. A. Ross *J. Acoust. Soc. Am.*, 1946, Vol. 18, p. 401.

CHAPTER 20.4
DIGITAL AUDIO RECORDING AND REPRODUCTION

Daniel W. Martin, Ronald M. Aarts

INTRODUCTION

A digital revolution has occurred in audio recording and reproduction that has made some previous techniques only of historical interest. Although analog recording and reproduction systems have been greatly improved (Fig. 20.4.1), their capabilities are still short of ideal. For example, they could not provide the dynamic range of orchestral instrument sounds (e.g., from 42 dB on a soft low flute note to 120 dB for a bass drum peak), plus a reasonable ratio of weakest signal to background noise. Mechanical analog records are still limited by inherent nonlinear distortions as well as surface noise, and magnetic analog recording is limited by inherent modulation noise.

Digital audio signal transmission, recording, and playback have numerous potential advantages, which, with appropriate specifications and quality control, can now be realized as will now be shown.

DIGITAL ENCODING AND DECODING

There is much more to digital audio than encoding the analog signal and decoding the digital signal, but this is basic. The rest would be largely irrelevant if it were not both advantageous and practically feasible to convert analog audio signals to digital for transmission, storage, and eventual retrieval.

A digital audio signal is a discrete-time, discrete-amplitude representation of the original analog audio signal. Figure 20.4.2 is a simple encoding example using only 4 bits. The amplitude of the continuous analog audio signal wavetrain *A* is sampled at each narrow pulse in the clock-driven pulse train *B*, yielding for each discrete abscissa (time) value a discrete ordinate (voltage) value represented by a dot on or near the analog curve. The vertical scale is subdivided (in this example) into 16 possible voltage values, each represented by a binary number or "word." The first eight words can be read out either in parallel

1000, 1010, 1011, 1011, 1010, 1000, 0110, 0101, ...

on four channels, or in sequence

10001010101110111010100001100101 ...

on a single channel for transmission, optional recording and playback, and decoding into an approximation of the original wavetrain. Unless intervening noise approaches the amplitude of the digit 1, the transmitted or played-back digital information matches the original digital information.

The degree to which digitization approximates the analog curve is determined by the number of digits chosen and the number of samplings per second. Both numbers are a matter of choice, but the present specifications for digital audio systems generally use 16 bits for uniform quantization (65,536 identifiable values),

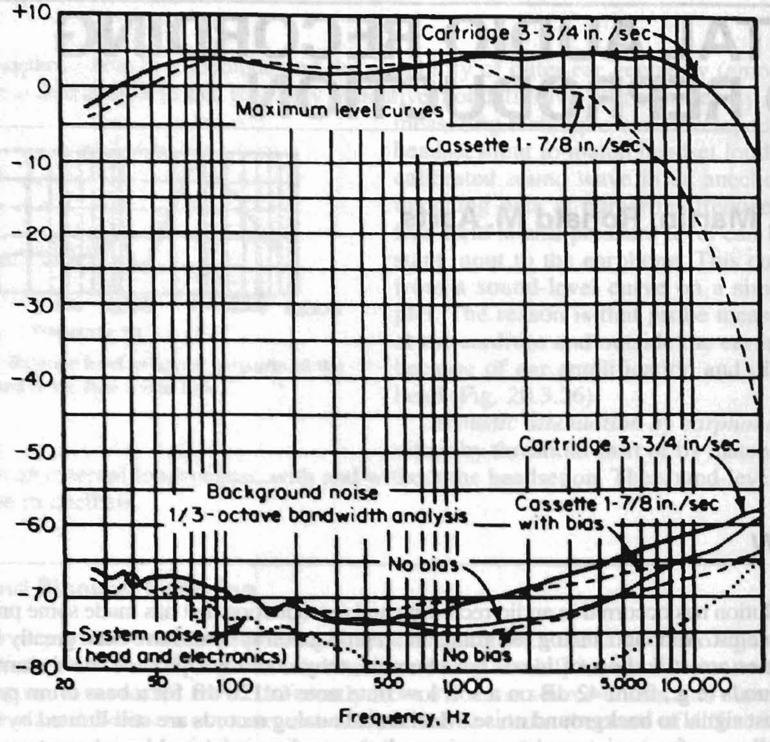

FIGURE 20.4.1 Dynamic range of analog tape cartridges and cassettes. *(After Ref. 1)*

corresponding to a theoretical dynamic range of 16(6 dB) = 96 dB. The sampling frequency, according to the Nyquist criterion, must be at least twice the highest audio frequency to be transmitted or recorded. Three different sampling frequencies are being used, 48 kHz "for origination, processing, and interchange of program material"; 44.1 kHz "for certain consumer applications"; and 32 kHz "for transmission-related applications."

Figure 20.4.3 shows the main electronic blocks of a 5-bit digital system for encoding and decoding audio signals for various transmitting and receiving purposes. The digital audio signal may be transmitted and

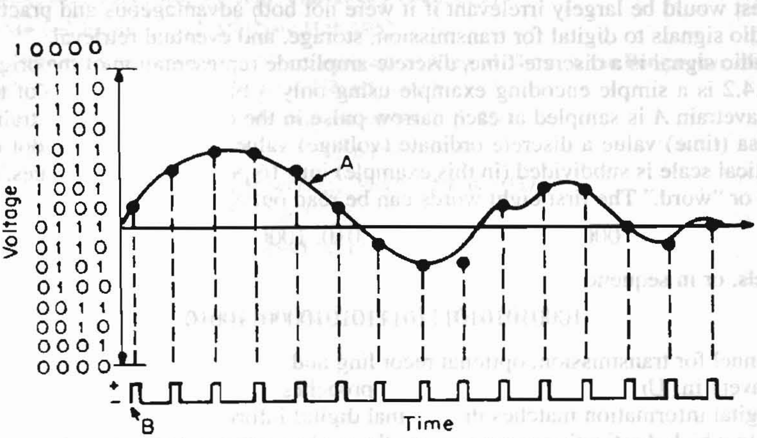

FIGURE 20.4.2 Digital encoding of an analog waveform: (*a*) continuous analog signal wavetrain; (*b*) clock-driven pulse train. At equal time intervals, sample values are encoded into nearest digital word.

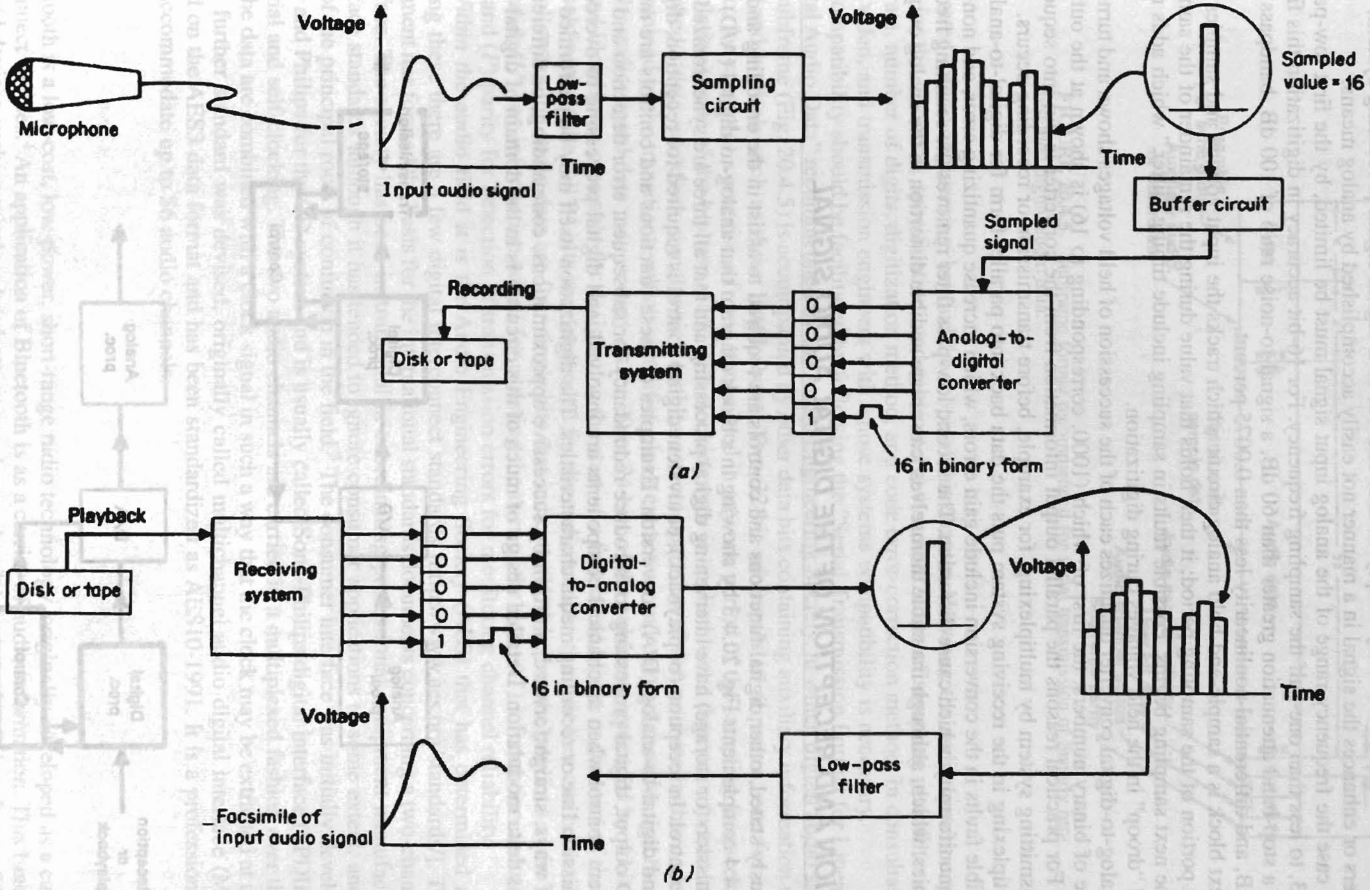

FIGURE 20.4.3 The basic electronic system components for encoding and decoding digital audio signals for (*a*) transmitting (or recording) and (*b*) receiving (or playback). (Ref. 2)

received conductively or electromagnetically. Alternatively, it may be stored and later retrieved by recording and playback. Or transmission may simply be into and out of a digital signal processing system, which purposely alters or enhances the signal in a manner not easily accomplished by analog means.

In any case the frequency range of the analog input signal must be limited, by the first low-pass filter of Fig. 20.4.3, to less than one-half the sampling frequency. For 16-bit accuracy in digitization this filter probably needs a stop-band attenuation greater than 60 dB, a signal-to-noise ratio of 100 dB, bandpass ripple less than 0.2 dB, and differential nonlinearity less than 0.0075 percent.

The next block is a sample-and-hold analog circuit which tracks the input voltage and samples it during a very short portion of the sampling period; it then holds that value during the remainder of the sampling period until the next sampling begins. Possible faults in sampling include timing "jitter," which adds modulation noise, and "droop" in the held voltages during digitization.

The analog-to-digital converter quantizes each of the succession of held voltages shown and turns them into a sequence of binary numbers, the first of which (1000, corresponding to 16) is shown at the outputs of the converter. For practical reasons the parallel output information from the converter is put into sequential form in the transmitting system by multiplexing, for example, before transmission or recording occurs.

Demultiplexing in the receiving system puts the data back into parallel form for digital-to-analog conversion. Possible faults in the conversion include gain errors, which increase quantizing error, and nonlinearity or relative nonuniformity, which cause distortion. The second low-pass filter removes the scanning frequency and its harmonics, which, although inaudible themselves, can create audible distortion in the analog output system.

TRANSMISSION AND RECEPTION OF THE DIGITAL AUDIO SIGNAL

As previously stated, other digital functions and controls are required to assist in the encoding and decoding. Figure 20.4.4 complements Fig. 20.4.3 by showing in a block diagram that analog-to-digital (A/D) conversion and transmission (or storage) have intervening digital processing and that all three are synchronized under digital clock control. In reception (or playback), equivalent digital control is required for reception, digital reprocessing, and digital-to-analog (D/A) conversion. Examples of these functions and controls are multiplexing of the A/D output, digital processing to introduce redundancy for subsequent error detection and correction, servo system control when mechanical components are involved, and digital processing to overcome inherent transmission line or recording media characteristics. The digitization itself may be performed in any of a number of ways: straightforward, uniform by successive approximations, companding, or differential methods such as delta modulation. Detailed design of much of this circuitry is in the domain of digital-circuit and

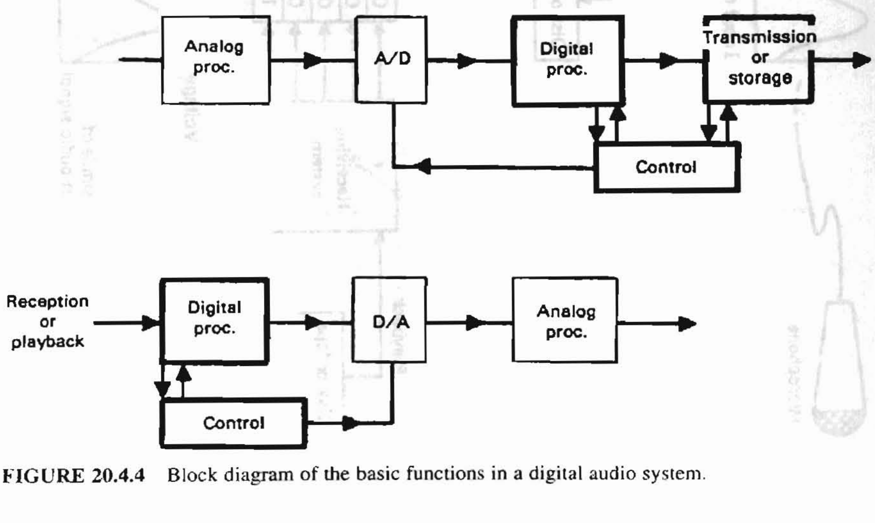

FIGURE 20.4.4 Block diagram of the basic functions in a digital audio system.

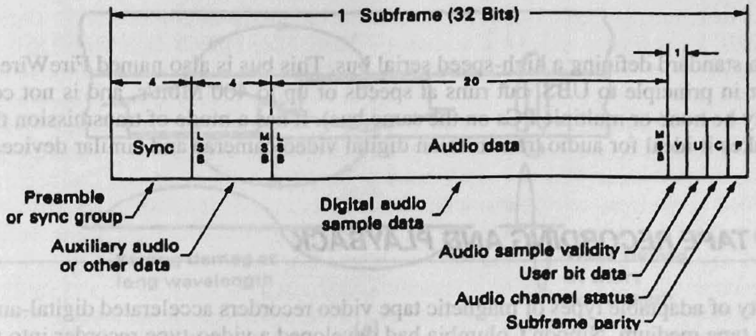

FIGURE 20.4.5 Subframe format recommended for serial transmission of linearly represented digital audio data.

integrated-circuit engineering, beyond the scope of this chapter. However, the audio system engineer is responsible for the selection, specification, and (ultimately) standardization of sampling rate, filter cutoff frequency and rate, number of digits, digitization method, and code error-correction method, in consultation with broadcast, video, and transmission engineers with whose systems compatibility is necessary.

Compatibility should be facilitated by following the "Serial Transmission Format for Linearly Represented Digital Audio Data" recommended by the Audio Engineering Society, in which digital audio sample data within a subframe (Fig. 20.4.5) is accompanied by other data bits containing auxiliary information needed for functions and controls such as those listed above. Two 32-bit subframes in sequence, one for each channel (of stereo, for example), comprise a frame transmitted in any one period of the sampling frequency. A channel (or modulation) code of the biphase mark, self-clocking type is applied to the data prior to transmission, in order to embed a data-rate clock signal which enables correct operation of the receiver. In this code all information is contained in the transitions, which simplifies clock extraction and channel decoder synchronization.

The audio signal data may occupy either 20 or 24 bits of the subframe, preceded by 4 bits of synchronizing and identifying preamble for designating the start of a frame and block, or the start of the first or the second subframe. If the full 24 bits are not needed for the audio sample, the first four can be auxiliary audio data.

Following the audio data are four single bits that indicate (V) whether the previous audio sample data bits are valid; (U) any information added for assisting the user of the data; (C) information about system parameters; and (P) parity for detection of transmission errors for monitoring channel reliability.

Within the audio field it is the Audio Engineering Society (AES) that has determined many standards. Among these there are a few digital interconnect standards [http://www.aes.org/standards/]. The AES30-1985 document has formed the basis for the international standards documents concerning a two-channel digital audio interface. The society has been instrumental in coordinating professional equipment manufacturers' views on interface standards although it has tended to ignore consumer applications to some extent, and this is perhaps one of the principal roots of confusion in the field.[3] The consumer interface was initially developed in 1984 by Sony and Philips for the CD system and is usually called Sony-Philips digital interface (SPDIF). The interface is serial and self-clocking. The two audio channels are carried in a multiplexed fashion over the same channel and the data are combined with a clock signal in such a way that the clock may be extracted at the receiver side.

A further standard was devised, originally called multichannel audio digital interface (MADI), which is based on the AES3 data format and has been standardized as AES10-1991. It is a professional interface that can accommodate up to 56 audio channels.

Bluetooth

Bluetooth is a low-cost, low-power, short-range radio technology, originally developed as a cable replacement to connect devices.[4] An application of Bluetooth is as a carrier of audio information. This functionality allows to build devices such as wireless headsets, microphones, headphones, and cellular phones. The audio quality provided by Bluetooth is the same as one would expect from a cellular telephone.

IEEE1394

IEEE1394 is a standard defining a high-speed serial bus. This bus is also named FireWire or i.Link. It is a serial bus similar in principle to UBS, but runs at speeds of up to 400 Mbit/s, and is not centered around a PC (i.e., there may be none or multiple PCs on the same bus). It has a mode of transmission that guarantees bandwidth that makes it ideal for audio transmission digital video cameras and similar devices.

DIGITAL AUDIO TAPE RECORDING AND PLAYBACK

The availability of adaptable types of magnetic tape video recorders accelerated digital-audio-recording development in the tape medium. Nippon Columbia had developed a video-type recorder into a PCM tape recorder for eight channels of audio information with each channel sampled at 47.25 kHz. Now numerous manufacturers produce audio tape recorders for professional recording studios, and some large recording companies have developed digital master tape recording systems including digital editors and mixers.

An inherent disadvantage of digital recording and playback, especially in the tape medium, has been dropout caused by voids or scratches in the tape. Some dropouts are inevitable, so protective or corrective means are used such as interlacing the encoded signal with redundancy, or reserving and using bits for error-detection schemes, e.g., recording sums of words, for comparison with sums simultaneously calculated from playback of the words. Such error detection can trigger the substitution of adjacent data, for example, into the dropout gap.

Digital audio tape recorders are of two different types, helical-scan and multitrack using rotary and stationary heads, respectively. Helical-scan systems already had the needed bandwidth, but improved recording densities and multitrack head stacks allowed multitrack systems to become competitive. A variety of tape formats has been developed. Table 20.4.1 shows part of the specifications for a multitrack professional digital recorder, the Sony PCM-3324.

Two new modes of recording on magnetic tape have permitted large increases in lineal density of recording and signal-to-noise ratio, both great advantages for digital magnetic recording. Perpendicular (or vertical) recording (see Fig. 20.4.6*a*) uses a magnetic film (e.g., CoCr crystallites), which has a preferred anisotropy normal to the surface. In contrast to conventional longitudinal magnetic recording, demagnetization is weak at short wavelengths, increasing the signal amplitude at high frequencies. Another advantage is that sharp transitions between binary states is possible. Vector field recording (Fig. 20.4.6*b*) with isotropic particles and microgap heads has also led to higher bit densities.

TABLE 20.4.1 Specifications for the PCM-3324

Number of channels (one track per channel):
digital audio 24, analog audio 2, time code 1, control 1; total 28
Tape speed, sampling rate:
70.01 cm, 44.1 kHz } with ±12.5% vernier
76.20 cm/s, 48.0 kHz }
(selectable at recording, automatic switching in playback)
Tape: 0.5-in. (12.7-mm) digital audio tape
Quantization: 16-bit linear per channel
Dynamic range: more than 90 dB
Frequency response: 20 Hz to 20 kHz, +0.5, −1.0 dB
Total harmonic distortion: less than 0.05%
Wow and flutter: undetectable
Emphasis: 50 μs/15 μs (EIAJ format and compact disc compatible)
Format: DASH-F (fast)
Channel coding: HDM-1
Error control: cross-interleave code

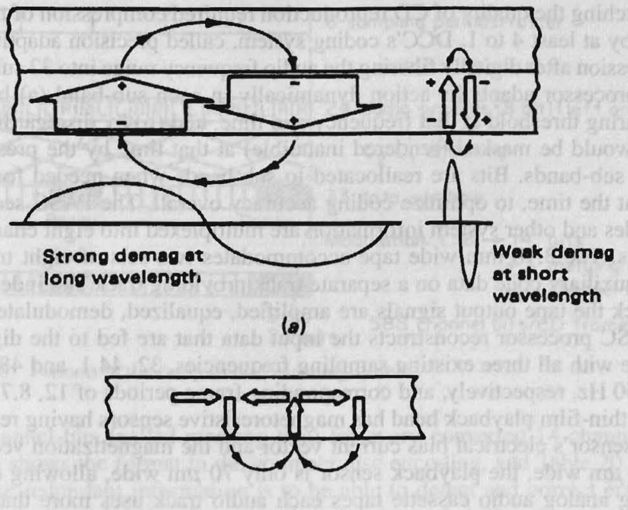

FIGURE 20.4.6 New recording modes: (*a*) perpendicular recording (CoCr); adjacent dipole fields aid; (*b*) vector field recording (isotropic medium): longitudinal and perpendicular fields aid at short wavelength.

Digital Compact Cassette (DCC)

After intensive research on digital audio tape (DAT), Philips built on this research to develop the digital compact cassette (DCC) recorder (Fig. 20.4.7). To make DCC mechanically (and dimensionally) compatible with analogue cassettes, and their tape mechanisms, the same tape speed of 4.76 cm/s ($1^7/_8$ in./s) was adopted. At

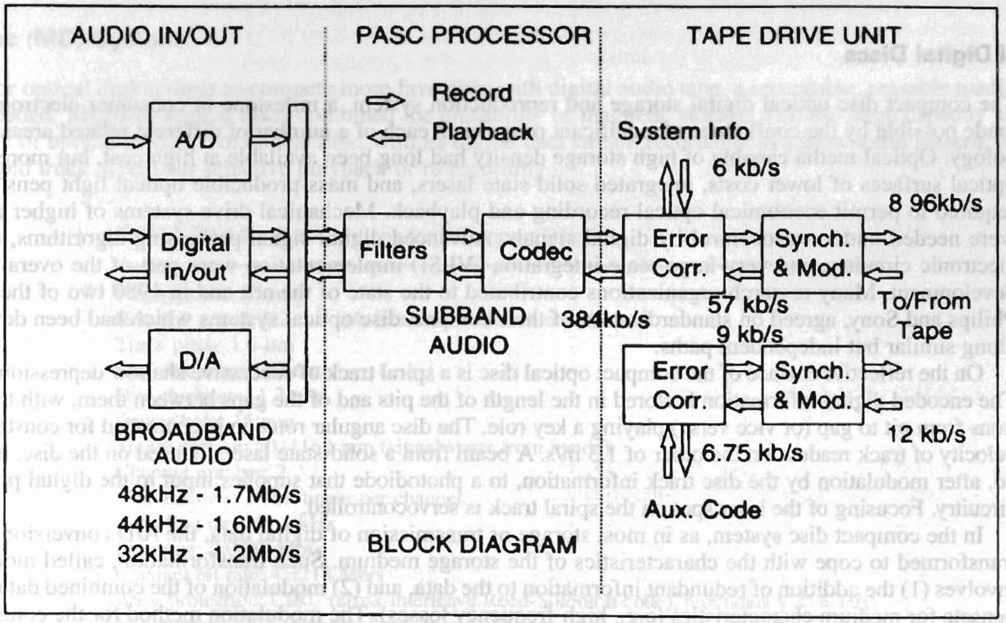

FIGURE 20.4.7 DCC recorder block diagram (Ref. 5).

that tape speed, matching the quality of CD reproduction required compression of the digital audio data stream (1.4 million bits/s) by at least 4 to 1. DCC's coding system, called precision adaptive sub-band coding (PASC) achieves the compression after digitally filtering the audio frequency range into 32 sub-bands of equal bandwidth. The PASC signal processor adapts its action dynamically in each sub-band (a) by omitting data for sounds lying below the hearing threshold at that frequency and time, and (b) by disregarding data for weak sounds in any sub-band that would be masked (rendered inaudible) at that time by the presence of stronger sounds in adjacent or nearby sub-bands. Bits are reallocated to sub-bands when needed for accuracy, from sub-bands where not needed at the time, to optimize coding accuracy overall. The PASC sequential data, together with error correction codes and other system information are multiplexed into eight channels for recording on eight 185μ m-wide tracks. The 3.78 mm wide tape accommodates two sets of eight tracks, one forward and one reverse, alongside auxiliary code data on a separate track providing track and index numbers, time codes, and so forth. In playback the tape output signals are amplified, equalized, demodulated, with error detection and correction. The PASC processor reconstructs the input data that are fed to the digital-to-analogue converter. PASC is compatible with all three existing sampling frequencies, 32, 44.1, and 48 kHz with sub-band widths of 500, 690, and 750 Hz, respectively, and corresponding frame periods of 12, 8.7, and 8 msec.

The 18-channel thin-film playback head has magnetoresistive sensors having resistance that varies with the angle between the sensor's electrical bias current vector and the magnetization vector. Although the recorded digital track is 185 μm wide, the playback sensor is only 70 μm wide, allowing considerable azimuth tolerance. When playing analog audio cassette tapes each audio track uses more than one sensor, as shown, to improve S/N ratio.

DIGITAL AUDIO DISC RECORDING AND PLAYBACK

As in video discs, two general types of digital audio discs were developed. One type recorded binary information mechanically or electrically along a spiral groove that provides guidance during playback for a lightly contacting pickup. The second type used optical laser recording of the digital information in a spiral pattern and optical playback means which track the pattern without contacting the disc directly. The optical type now appears to be dominant.

Optical Digital Discs

The compact disc optical digital storage and reproduction system, a milestone in consumer electronics, was made possible by the confluence of significant progress in each of a number of different related areas of technology. Optical media capable of high storage density had long been available at high cost, but more durable optical surfaces of lower costs, integrated solid-state lasers, and mass-producible optical light pens were all required to permit economical optical recording and playback. Mechanical drive systems of higher accuracy were needed under servocontrol by digital signals. Advanced digital signal processing algorithms, complex electronic circuitry, and very large-scale integration (VLSI) implementation were part of the overall system development. Many research organizations contributed to the state of the art, and in 1980 two of the leaders, Philips and Sony, agreed on standardization of their compact disc optical systems which had been developing along similar but independent paths.

On the reflective surface of the compact optical disc is a spiral track of successive shallow depressions or pits. The encoded digital information is stored in the length of the pits and of the gaps between them, with the transitions from pit to gap (or vice versa) playing a key role. The disc angular rotation is controlled for constant linear velocity of track readout on the order of 1.3 m/s. A beam from a solid-state laser, focused on the disc, is reflected, after modulation by the disc track information, to a photodiode that supplies input to the digital processing circuitry. Focusing of the laser spot on the spiral track is servocontrolled.

In the compact disc system, as in most storage or transmission of digital data, the A/D conversion data are transformed to cope with the characteristics of the storage medium. Such transformation, called modulation, involves (1) the addition of redundant information to the data, and (2) modulation of the combined data to compensate for medium characteristics (e.g., high-frequency losses). The modulation method for the compact disc system, called eight-to-fourteen modulation (EFM), is an 8-data-bit to 14-channel-bit conversion block code

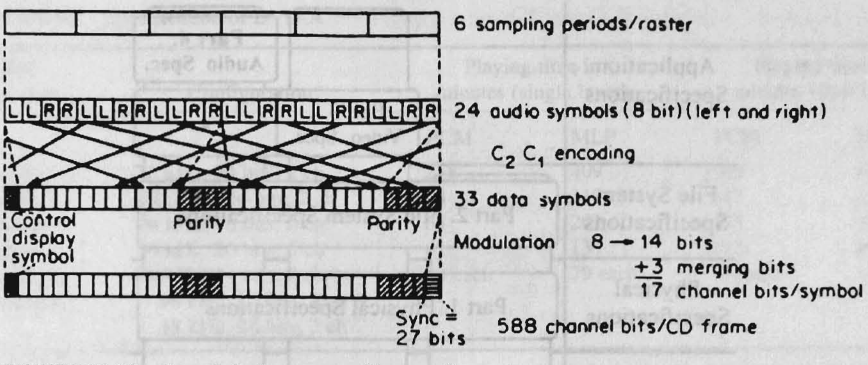

FIGURE 20.4.8 Formats in the compact disc encoding system.

with a space of 3 channel bits (called merging bits) for every converted 14 channel bits for connecting the blocks. Figure 20.4.8 shows the format in the compact disc encoding, and Table 20.4.2 the disc specification.

The purpose of the redundant information is to be able to detect and correct errors that occur because of storage medium imperfections. It is important to minimize the probability of occurrence of such imperfections. The use of optical noncontacting readout from a signal surface protected by a plastic layer allows most of the signal errors at the surface to be reduced to random errors of several bits or larger burst errors. The error-correcting code, the cross-interleave Reed-Solomon code (CIRC), adopted in the standardization provides highly efficient detection and correction for errors of these types. It happens that the EFM modulation method and the CIRC error-correction method used in the compact disc system are well matched. This combination is credited with much of the system's success.

Between tape mastering and replication lies a complex and sophisticated disc mastering process which gets the information into the CD standard format and onto the surface of the CD disc master. Optical disc preparation, recording, development, electroplating, stamping, molding, and protection film coating are the major steps in the highly technological production process.

MiniDisc (MD) System

For optical digital discs to compete more favorably with digital audio tape, a recordable, erasable medium was needed. Magneto-optical discs combined the erasability of magnetic storage with the large capacity and long life of optical storage. An optical disc, with its digital data stream recordable in a tight spiral pattern, provides rapid track access for selective playback or re-recording.

TABLE 20.4.2 Specifications for a Compact Disc

Playing time: 75 min
Rotating speed: 1.2–1.4 m/s (constant linear velocity)
Track pitch: 1.6 μm
Disc diameter: 120 mm
Disc thickness: 1.2 mm
Center hole: 15 mm
Signal surface: 50–116ϕ mm (signal starts from inside)
Channel number: 2
Quantization: 16-bit linear per channel
Sampling rate: 44.1 kHz
Data rate: 2.0338 Mb/s
Channel bit rate: 4.3218 Mb/s
Error protection: CIRC (cross-interleave Reed-Solomon code), redundancy 25% (4/3)
Modulation: EFM (eight-to-fourteen modulation)

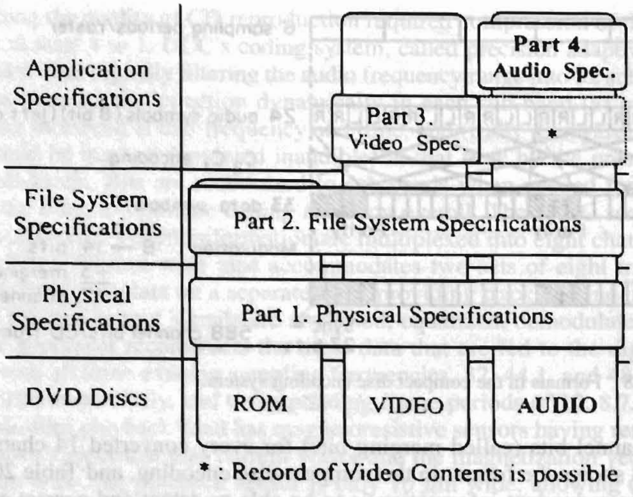

FIGURE 20.4.9 DVD specifications for DVD-ROM, DVD-video, and DVD-audio read only disks, parts 1 to 4.

On the blank disc is a thin film of magneto-optic material embedded within a protective layer, with all of the magnetic domains pointing north pole down (a digital zero). The magnetic field needed for reversal of polarity (to convert from zero to one) is very temperature dependent. At room temperature reversal requires a very strong magnetic field. However, at about 150°C only a small coercive force, provided by a dc magnetic bias field, is needed. During recording, a high-power laser beam, modulated by the digital data stream, heats microscopic spots on the rotating magneto-optic surface (within nanoseconds) to temperatures that allow the dc magnetic bias field to convert zeroes to ones. When the laser beam is off, the spots on the medium cool very rapidly, leaving the desired pattern of magnetic polarity. Erasure can be effected by repeating the procedure with the dc bias reversed.

Playback uses a low-power laser beam, which, because of the Kerr magneto-optic effect, has its plane of polarization rotated one way or the other depending on the magnetic polarity of the recorded bit. An opto-electronic playback head senses the polarization and delivers the digital playback signal.

The Sony magneto-optic MiniDisc is 6.4 cm (2½ in.) in diameter, half that of a CD. To compensate for reduced recording area, a digital audio compression technique called *adaptive transform acoustic coding* (ATRAC) is used. The analog signal is digitized at 44.1 kHz sampling frequency with 16-bit quantization. Waveform segments of about 20 ms and 1000 samples are converted to frequency components that are analyzed for magnitude by the encoder and compressed. Threshold and masking effects are used as criteria for disregarding enough data for an overall reduction of about 5 to 1. During playback, the ATRAC decoder regenerates an analog signal by combining the frequency components recorded on the magneto-optic disc. An added feature of the compression circuit is storage of 3 s of playback time when potential interruptions could occur owing to system shock or vibration.

Digital Versatile Disc-Audio (DVD-A)

DVD-Audio is a HiFi music format based on the same DVD technology as the DVD-Video discs and DVD-ROM computer discs, see Fig. 20.4.9. The disc structure is basically the same.

Recorded with current CD recording methods (PCM), DVD-Audio has a theoretical sampling rate of 192 kHz with 24-bit processing. Like Super Audio CD (SACD) and normal DVD-Video and DVD-Data formats, DVD-Audio discs can store 4.7-GB with a choice of 2-channel and 6-channel audio tracks or a mix of both (see Table 20.4.3).

Like SACD, information such as track names, artists' biographies, and still images can be stored. The format is supported by DVD-Video players made after about November 2000. Manufacturers are making audio machines compatible with playing both types of disc. Titles are available in both Dolby digital mix (so they are compatible on all DVD-Video players) and specific DVD-Audio (requiring the separate player).

TABLE 20.4.3 Specification of DVD-A

Audio combination	Configuration	Playing time in minutes (single layer)		Playing time in minutes (dual layer)	
		PCM	MLP	PCM	MLP
2 channels	48 kHz, 24 bits, 2 ch	258	409	469	740
2 channels	192 kHz, 24 bits, 2 ch	64	119	117	215
6 channels	96 kHz, 16 bits, 6 ch	64	201	117	364
5 channels	96 kHz, 20 bits, 5 ch	61	137	112	248
2 channels & 5 channels	96 kHz, 24 bits, 2 ch + 96 kHz, 24 bits, 3 ch & 48 kHz, 24 bits, 2 ch	43 each	79 each	78 each	144 each

Note: MLP is an acronym for Meridian Lossless Packing, a lossless coding scheme (see Lossless Coding section).

Super Audio CD

Super Audio CD is a new format. It uses direct stream digital (DSD) and a 4.7-GB disc with 2.8 MHz sampling frequency (i.e., 64 times the 44.1 kHz used in CD) enabling a very high quality audio format. Technical comparison between conventional CD and SACD is detailed in Table 20.4.4.

The main idea of the hybrid disc format (see Fig. 20.4.10) is to combine both well-known technologies, CD and DVD, respectively, to keep compatibility with the CD players in the market, and to use the existing DVD-video process tools to make a two-layer disc, i.e., to add a high-density layer to a CD reflective layer. As shown in Table 20.4.4, the storage capacity of the high-density layer is 6.9 times higher than the storage capacity of a conventional CD.

Direct Stream Digital

The solution came in the form of the DSD signal processing technique. Originally developed for the digital archiving of priceless analog master tapes, DSD is based on a 1-bit sigma-delta modulation together with a fifth-order noise-shaping filter and operates with a sampling frequency of 2.8224 MHz (i.e., 64 times the 44.1 kHz used in CD), resulting in an ultrahigh signal-to-noise ratio in the audio band.

TABLE 20.4.4 Comparison Between Conventional CD and SACD

	Conventional compact disc	Super Audio CD
Diameter	120 mm (4–3/4 in.)	120 mm (4–3/4 in.)
Thickness	1.2 mm (1/20 in.)	2 × 0.69 mm = 1.2 mm (1/20 in.)
Max. substrate thickness error	+/–100 μm	+/–30 μm
Signal sides	1	1
Signal layers	1	2: CD-density reflective layer + high-density semitransmissive layer
Data capacity		
Reflective layer	680 MB	680 MB
Semitransmissive layer	—	4.7 GB
Audio coding		
Standard CD audio	16-bit/44.1 kHz	16-bit/44.1 kHz
Super Audio	—	1-bit DSD/2.8224 MHz
Multichannel	—	6 channels of DSD
Frequency response	5–20,000 Hz	DC(0)–100,000 Hz (DSD)
Dynamic range	96 dB across the audio bandwidth	120 dB across the audio bandwidth (DSD)
Playback time	74 min	74 min
Enhanced capability	CD text	Text, graphics, and video

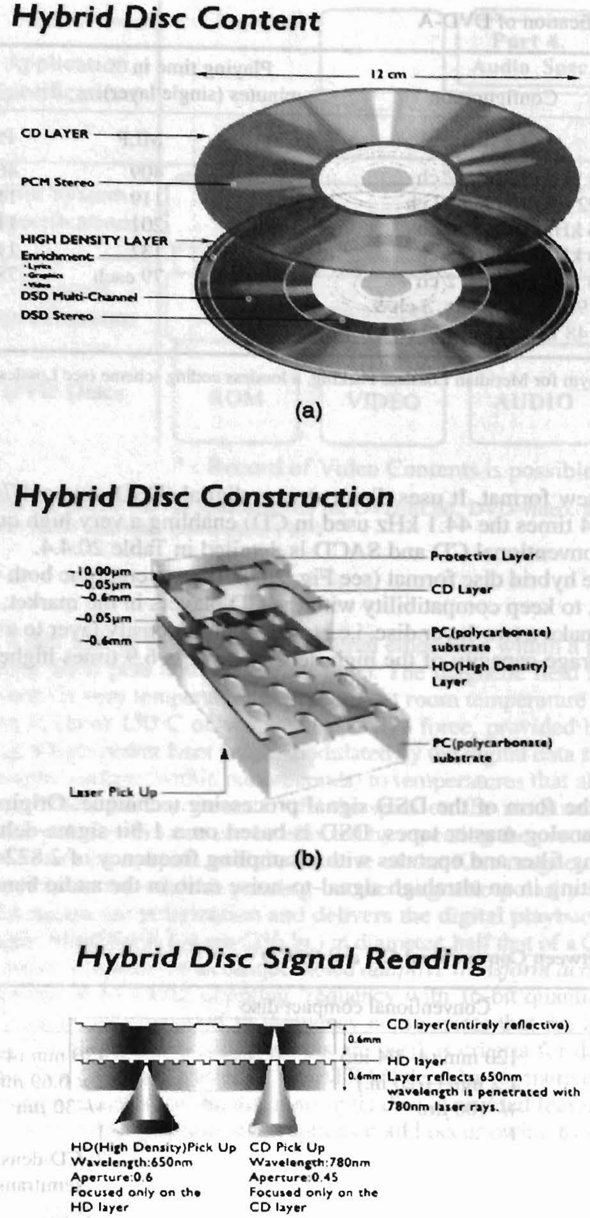

FIGURE 20.4.10 Hybrid disc content of the Super Audio CD (*a*), hybrid disc construction (*b*), and hybrid disc signal reading (*c*).

The Three Type of Super Audio CD

The SACD standard, published by Philips and Sony in March 1999, defines three possible disc types (see Fig. 20.4.10). The first two types are discs containing only DSD data; the single layer disc can contain 4.7 GB of data, while the dual layer disc contains slightly less than 9 GB. The third version—the SACD Hybrid—combines a single 4.7 GB layer with a conventional CD that can be played back on standard CD players. (For more information see http://www.sacd.philips.com/)

RDAT

Rotary head digital audio tape (RDAT) is a semiprofessional recording format, an instrumentation recorder, and a computer data recorder.[6] Mandatory specification are:

- 2 channels (optional more)
- 48 or 44.1 kHz sampling rate
- 16 bits quantization
- 8.15 mm/s tape speed
- 2 h playing time (13 μm tape)
- The cassette has a standardized format of 73 × 54 × 10.5 mm, which is rather smaller than the compact cassette.

OTHER APPLICATIONS OF DIGITAL SIGNAL PROCESSING

The main applications of audio DSP are high-quality audio coding and the digital generation and manipulation of music signals. They share common research topics including perceptual measurement techniques and knowledge and various analysis and synthesis methods.[7] Chen[8] gives a review of the history of research in audio and electroacoustics, including electroacoustic devices, noise control, echo cancellation, and psychoacoustics.

Reverberation. For some years digital processing of audio signals has been used for special purposes (e.g., echo and reverberation effects) in systems that were otherwise analog in nature. The possibility was suggested by computer-generated "colorless" artificial reverberation experiments. When high-quality A/D and D/A conversion became economical, digital time-delay and reverberation units followed. Figure 20.4.11*a* is a block diagram of a digital audio reverberation system in which the complete musical impulse sound reaching a listener (Fig. 20.4.11*b*) consists of slightly delayed direct signal, followed by a group of simulated early reflections from a tapped digital delay line and a "reverberant tail" added to its envelope by a reverberation processor using multiple recursive structures to produce a high time density of simulated reflections.

Dither is used to prevent perceptually annoying errors like quantizers. It is a random "noise" process added to a signal prior to its (re)quantization in order to control the statistical properties of the quantization error.[9,10] A common stage to perform dithering is after the various digital signal processing stages just ahead of the quantization before storing the signal or sending it to a digital-to-analog converter (DAC).

A special topic in signal processing for sound reproduction is overcoming the limitations of the reproduction set-up, e.g., reproduction of bass frequencies through small loudspeakers.[11] Another limitation is the distance between the two loudspeakers of a stereophonic setup. If one likes to increase the apparent distance, frequency dependent cross talk between the channels can be applied.[12]

Lossless Coding. Lossless compression is a technique to recode digital data in such a way that the data occupy fewer bits than before. In the PC world these programs are widely used and known under various names such as PkZip. For digital audio these programs are not very well suited, since they are optimized for text data and programs. Figure 20.4.12 shows a block diagram representing the basic operations in most lossless compression algorithms involved in compressing a single audio channel.[13]

All of the techniques studied are based on the principle of first removing redundancy from the signal and then coding the resulting signal with an efficient coding scheme. First the data are divided into independent frames of equal time duration in the range of 13 to 26 ms, which results in a frame of 576 to 1152 samples if a sampling rate of 44.1 kHz is used. Then the bits in each frame are decorrelated by some prediction algorithm as shown in Fig. 20.4.13.

The value of a sample $x[n]$ is predicted using the preceding samples $x[n-1], x[n-2], \ldots$, by using the filters A, B and quantizer Q. The error signal $e(n)$ that remains after prediction is in general smaller than x, and will therefore require fewer bits for its exact digital representation. The coefficients of the filters A and B are transmitted as well, which makes an exact reconstruction of $x[n]$ possible.

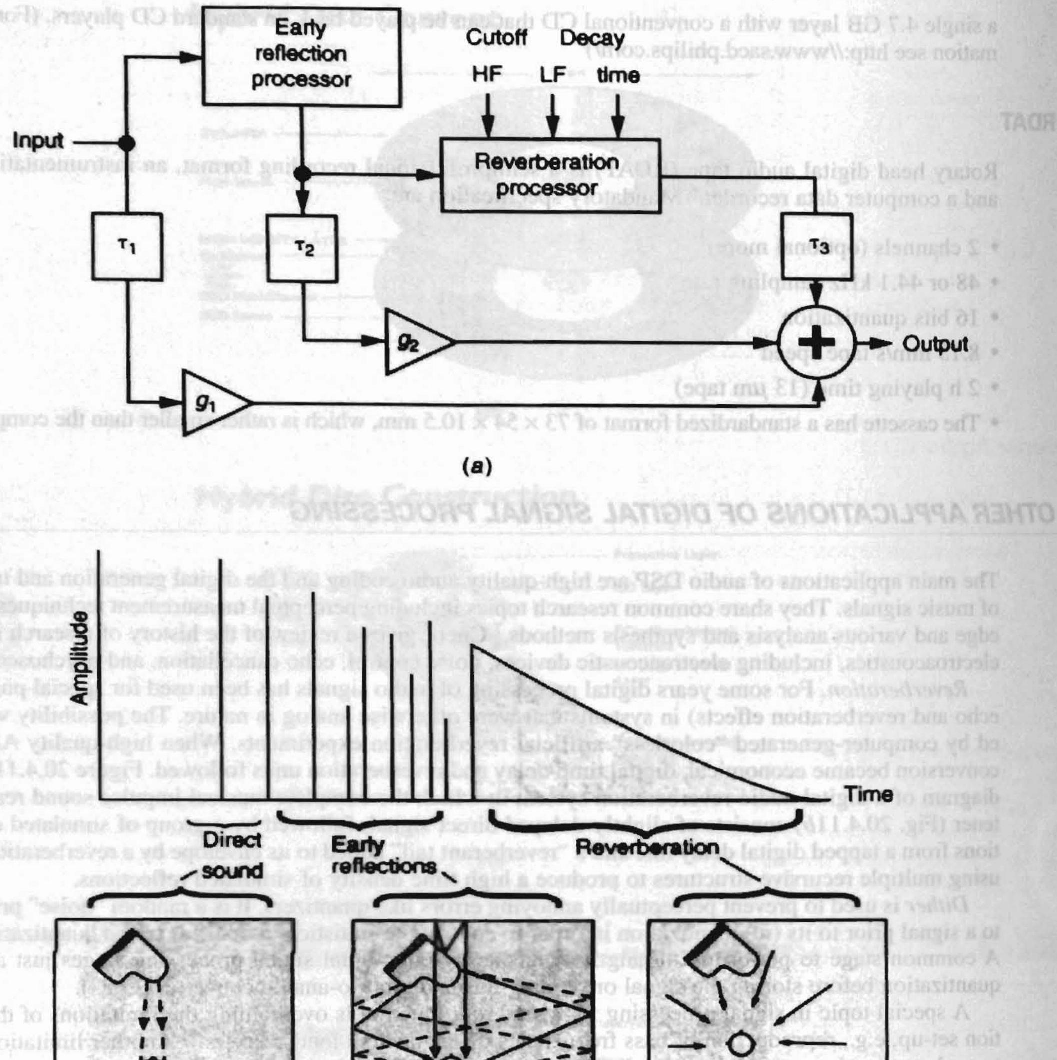

FIGURE 20.4.11 The basic operations in most lossless compression algorithms.

The third stage is an entropy coder, which removes further redundancy from the residual signal $e[n]$, and again in this process no information is lost. Most coding schemes use one of these three algorithms:

- Huffman, run length, and Rice coding, see Ref. 13 for more details
- Meridian Lossless Packing (MLP) for DVD-A
- Direct Stream Transfer for SACD

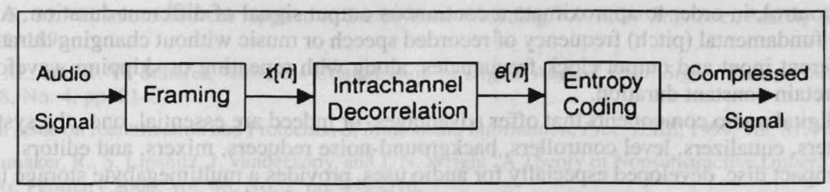

FIGURE 20.4.12 The basic operations in most lossless compression algorithms.

Watermarking. The advantages of digital processing and distribution of multimedia, such as noise-free transmission and the possibility of digital signal processing on these media, are obvious. The disadvantage, from the viewpoint of media producers and content providers, can be the possibility of unlimited coping of digital data without loss of quality. Digital copy protection is a way to overcome these problems. Another method is the embedding of digital watermarks into the multimedia.[14]

The watermark is an unremovable digital code, robustly and imperceptibly embedded in the host data and typically contains information about the origin, status, and/or destination of the data. While copyright protection is the most prominent application of watermarking techniques, other methods exist, including data authentication by means of fragile watermarks that are impaired or destroyed by manipulations, embedded transmission of value-added services, and embedded data labeling for other purposes than copyright protection such as monitoring and tracking.

Multimedia Content Analysis

Multimedia content analysis refers to the computerized understanding of semantic meanings of multimedia documents such as a video sequence with an accompanying audio track. There are many features that can be used to characterize audio signals. Usually audio features are extracted in two levels: short-term frame level and long-term clip level, where a frame is about 10 to 40 ms. To reveal the semantic meaning of an audio signal, analysis over a much longer period is necessary, usually from 1 to 10 s.[15]

Special Effects. If a single variably delayed echo signal ($\tau > 40$ ms) is added to direct signal at a low frequency (< 1 Hz), a sweeping comb filter sound effect is produced called *flanging*. When multiple channels of lesser delay (e.g., 10 to 25 ms) are used, a "chorus" effect is obtained from a single input voice or tone.

Time-Scale Modification. Minor adjustment of the duration of prerecorded programs to fit available program time can be accomplished digitally by loading a random-access memory with a sampled digital input signal and then outputting the signal with waveform sections of the memory repeated or skipped as needed under

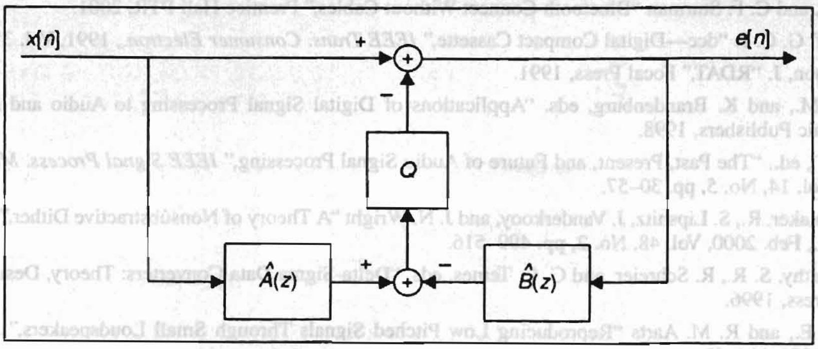

FIGURE 20.4.13 General structure for prediction.

computer control, in order to approximate a continuous output signal of different duration. A related need, to change the fundamental (pitch) frequency of recorded speech or music without changing duration, involves the use of different input and output clock frequencies, along with repeating or skipping waveform segments as needed to retain constant duration.

Other digital audio components that offer advantages, or indeed are essential, once the system goes digital, include filters, equalizers, level controllers, background-noise reducers, mixers, and editors.

The compact disc, developed especially for audio uses, provides a multimegabyte storage technique, which is very attractive in many other applications for read-only memories. Conversely, in the evolution of telecommunication networks, new techniques for signal decomposition and reconstruction, and for echo cancellations, suggest further audio improvements in conference pickup and transmission, for example. The interchange between digital audio and other branches of digital communication continues.

MPEG Audio Coding General

Moving Picture Experts Group (MPEG) is well known for its developments of a series of standards for the coding of audiovisual content [http://www.cselt.it/mpeg/]. Initially targeted at the storage of audiovisual content on compact disc media, the MPEG-1 standard was finalized in 1992 and included the first generic standard for low-bit-rate audio within the audio part. Then the MPEG-2 standard was completed and extended MPEG-1 technology toward the needs of digital video broadcast. On the audio side, these extensions enabled coder operation at lower sampling rates (for multimedia applications) and coding of multichannel audio. In 1997 the standard of an enhanced multichannel coding system (MPEG-2 Advanced Audio Coding, AAC) was defined. The so-called MP3 is the popular name for MPEG-1 Layer III. Then the MPEG-4 standard was developed, with new functionalities such as object-based representation, content-based interactivity, and scalability; the MPEG-4 standard was developed in several steps (called versions), adding extensions to the basic technology for audio. Reference 16 describes in some detail the key technologies and main features of MPEG-1 and MPEG-2 audio coders. In 1996 the effort behind MPEG-7 was started. MPEG-7 defines a universal standardized mechanism for exchanging descriptive data that are able to characterize many aspects of multimedia content with a worldwide interoperability,[17] or as the official name says, a "multimedia content description interface." Work on the new standard MPEG-21 "Multimedia Framework" was started in June 2000. The vision for MPEG-21 is to define a multimedia framework to enable transparent and augmented use of multimedia resources across a wide range of networks and devices used by different communities.

REFERENCES

1. Gravereaux, D. W., A. J. Gust, and B. B. Bauer *J. Audio Eng. Soc.*, 1970, Vol. 18, p. 530.
2. Bernhard, R. *IEEE Spectrum*, December, 1979, p. 28.
3. Rumsey, F., and J. Watkinson "The Digital Interface Handbook," Focal Press, 1995.
4. Bray, J., and C. F. Sturman "Bluetooth Connect Without Cables," Prentice Hall PTR, 2001.
5. Lokhoff, G. C. P. "dcc—Digital Compact Cassette," *IEEE Trans. Consumer Electron.*, 1991, Vol. 37, p. 702.
6. Watkinson, J. "RDAT," Focal Press, 1991.
7. Kahrs, M., and K. Brandenburg, eds. "Applications of Digital Signal Processing to Audio and Acoustics," Kluwer Academic Publishers, 1998.
8. Chen, T., ed., "The Past, Present, and Future of Audio Signal Processing," *IEEE Signal Process. Magazine*, September 1997, Vol. 14. No. 5. pp. 30–57.
9. Wannamaker, R., S. Lipshitz, J. Vanderkooy, and J. N. Wright "A Theory of Nonsubstractive Dither," *IEEE Trans. Signal Process.*, Feb. 2000, Vol. 48, No. 2, pp. 499–516.
10. Norsworthy, S. R., R. Schreier, and G. C. Temes, eds. "Delta-Sigma Data Converters: Theory, Design, and Simulation, IEEE Press, 1996.
11. Larsen, E., and R. M. Aarts "Reproducing Low Pitched Signals Through Small Loudspeakers," *J. Audio Eng. Soc.*, March 2002, Vol. 50, No. 3, pp. 147–164.

12. Aarts, R. M. "Phantom Sources Applied to Stereo-Base Widening," *J. Audio Eng. Soc.*, March 2000, Vol. 48, No. 3, pp. 181–189.

13. Hans, M., and R. W. Schafer, "Lossless Compression of Digital Audio," *IEEE Signal Process. Magazine*, July 2001, Vol. 18, No. 4, pp. 21–32.

14. Special Issue on Identification and Protection of *Multimedia Information, Proc. IEEE*, 1999, Vol. 87, No. 7, pp. 1059–1276.

15. Wannamaker, R., S. Lipshitz, J. Vanderkooy, and J. N. Wright "A Theory of Nonsubstractive Dither," *IEEE Trans. Signal Process*, February 2000, Vol. 48, No. 2, pp. 499–516.

16. Noll, P. "MPEG Digital Audio Coding," *IEEE Signal Process. Magazine*, Sept. 1997, Vol. 14, No. 5, pp. 59–81.

17. Lindsay, A. T., and J. Herre "MPEG-7 and MPEG-7 Audio—An Overview," *J. Audio Eng. Soc.*, July/August 2001, Vol. 49, Nos. 7/8, pp. 589–594.