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),

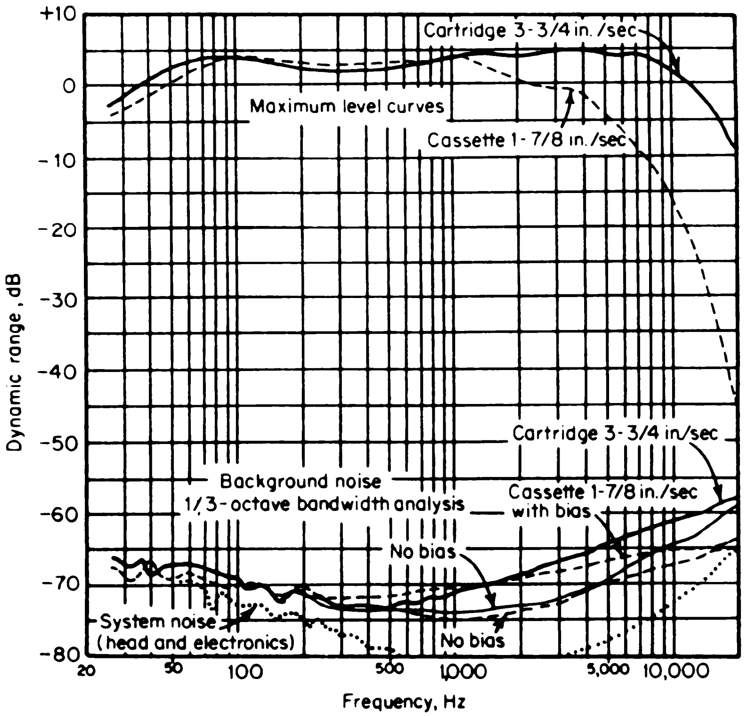


FIGURE 20.4.1 Dynamic range of analog tape cartridges and cassettes. (*After Ref.* 1)

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

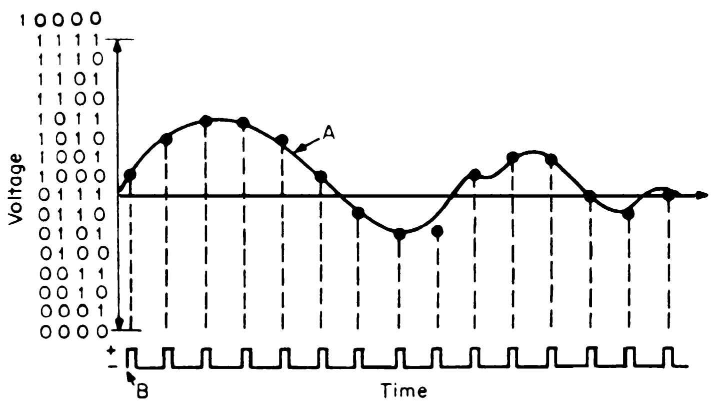


FIGURE 20.4.2 Digital encoding of an analog waveform: (*a*) continuous analog signal wavetrain; (*b*) clock-driven pulse train. At equal time intervals, sample values are encoded into nearest digital word.

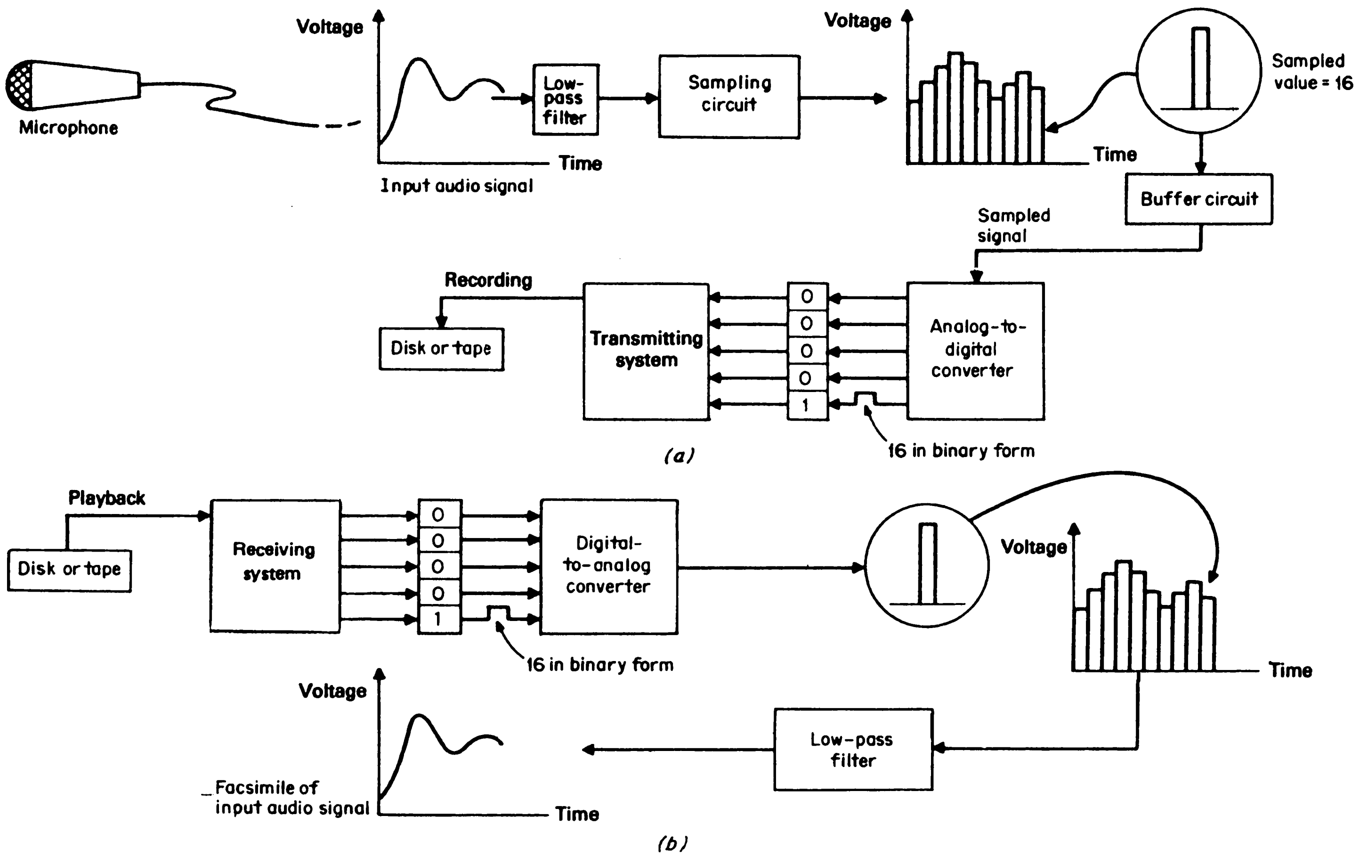


FIGURE 20.4.3 The basic electronic system components for encoding and decoding digital audio signals for (*a*) transmitting (or recording) and (*b*) receiving (or playback). (Ref. 2)

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

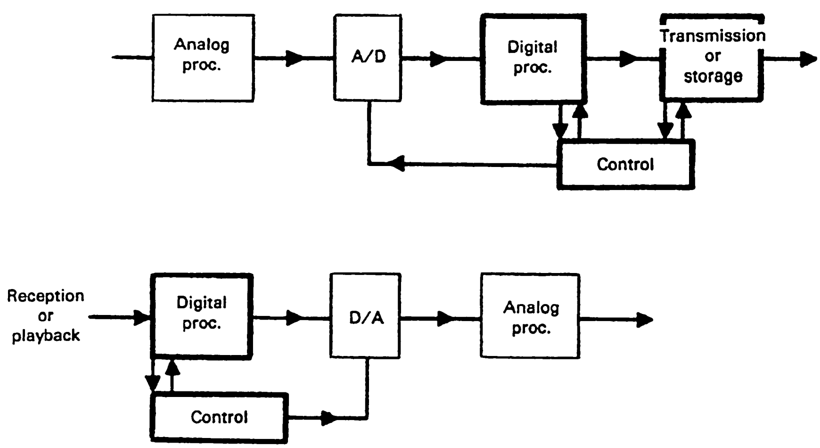


FIGURE 20.4.4 Block diagram of the basic functions in a digital audio system.

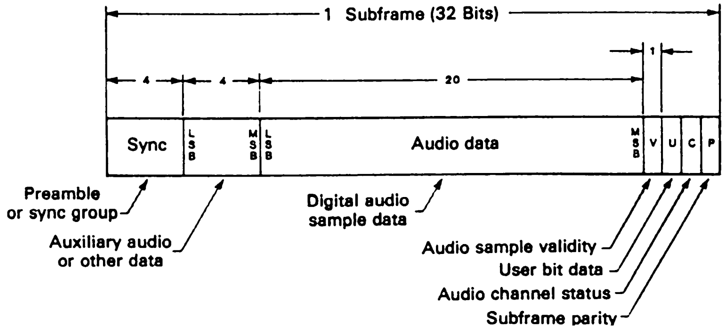


FIGURE 20.4.5 Subframe format recommended for serial transmission of linearly represented digital audio data.

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 } with ±12.5% vernier
(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 μms/15 μs (EIAJ format and compact disc compatible)
Format: DASH-F (fast)
Channel coding: HDM-1
Error control: cross-interleave code

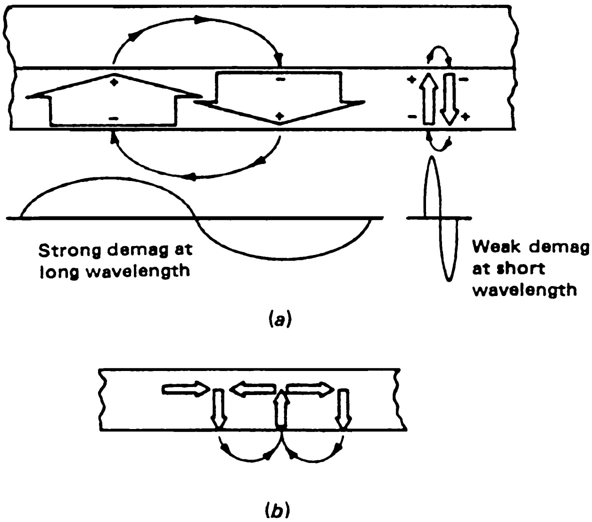


FIGURE 20.4.6 New recording modes: (*a*) perpendicular recording (CoCr); adjacent dipole fields aid; (*b*) vector field recording (isotropic medium): longitudinal and perpendicular fields aid at short wavelength.

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

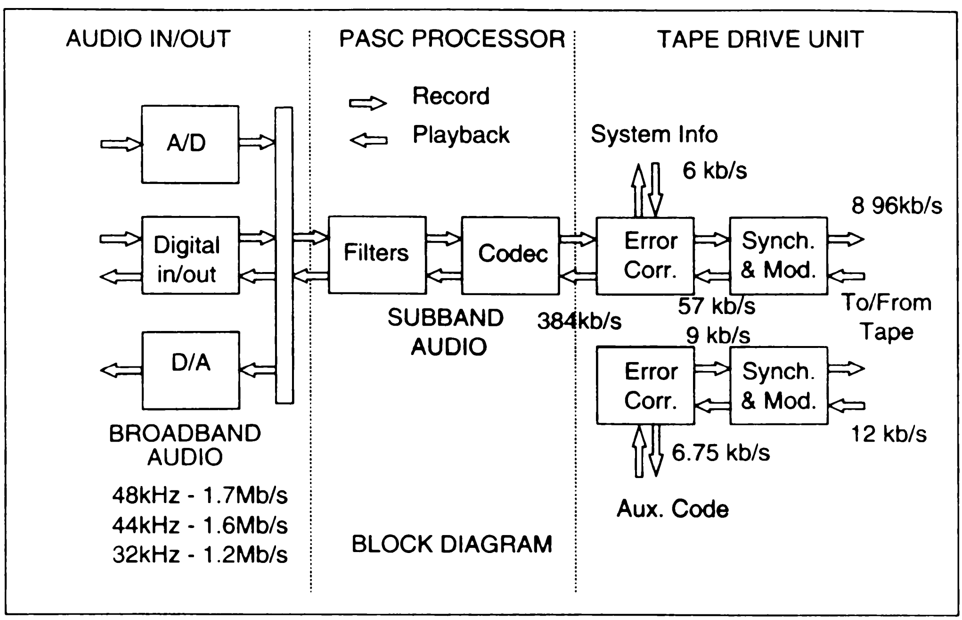


FIGURE 20.4.7 DCC recorder block diagram (Ref. 5).

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

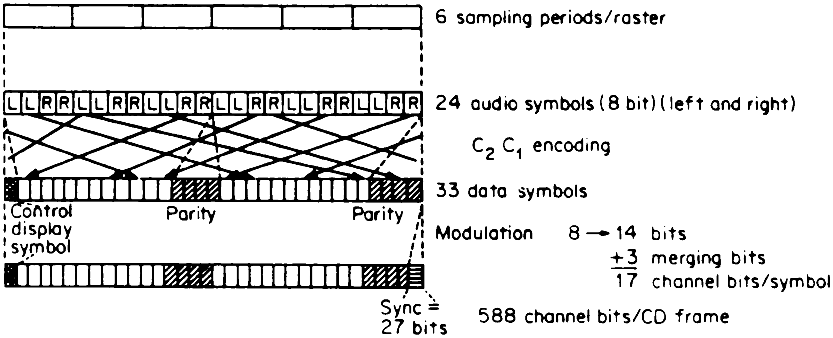


FIGURE 20.4.8 Formats in the compact disc encoding system.

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)

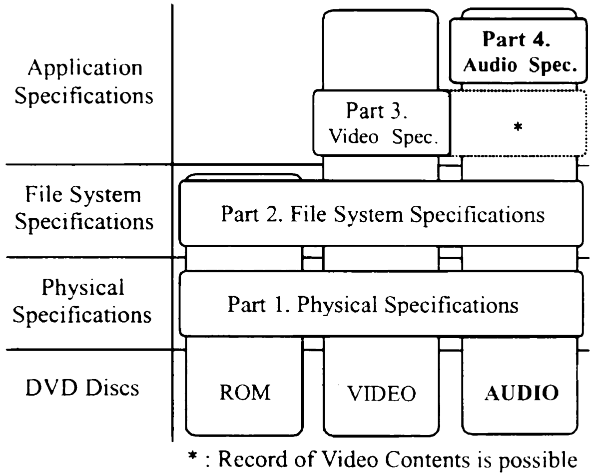


FIGURE 20.4.9 DVD specifications for DVD-ROM, DVD-video, and DVD-audio read only disks, parts 1 to 4.

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 ($2^1/_2$ 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.

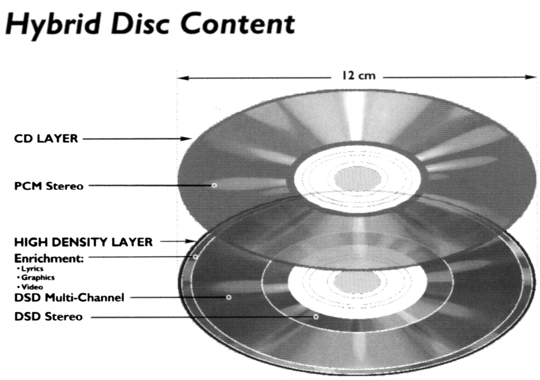
The main idea of the hybrid disc format (see Fig. 20.4.10) is to combine both well-known technologies, CD and DVD, respectively, to keep compatibility with the CD players in the market, and to use the existing DVD-video process tools to make a two-layer disc, i.e., to add a high-density layer to a CD 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

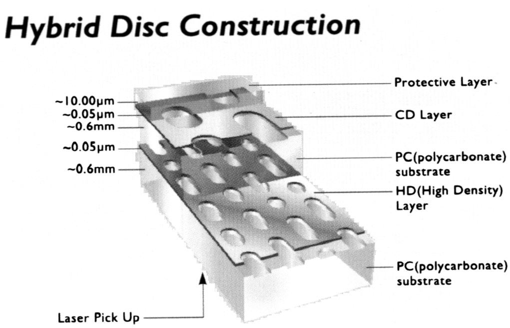
The solution came in the form of the DSD signal processing technique. Originally developed for the digital archiving of priceless analog master tapes, DSD is based on a 1-bit sigma-delta modulation together with a fifth-order noise-shaping filter and operates with a sampling frequency of 2.8224 MHz (i.e., 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

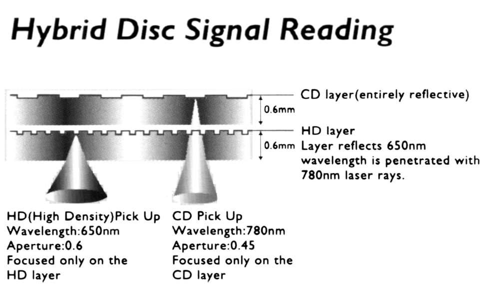
	Conventional compact disc	Super Audio CD
Diameter	120 mm (4–3/4 in.)	120 mm (4–3/4 in.)
Thickness	1.2 mm (1/20 in.)	2 × 0.69 mm = 1.2 mm (1/20 in.)
Max. substrate thickness error	+/–100 μm	+/–30 μm
Signal sides	1	1
Signal layers	1	2: CD-density reflective layer + high-density semitransmissive layer
Data 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



(a)



(b)



(c)

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 $73 \times 54 \times 10.5$ 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 $x[n-1]$, $x[n-2]$, …, 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.

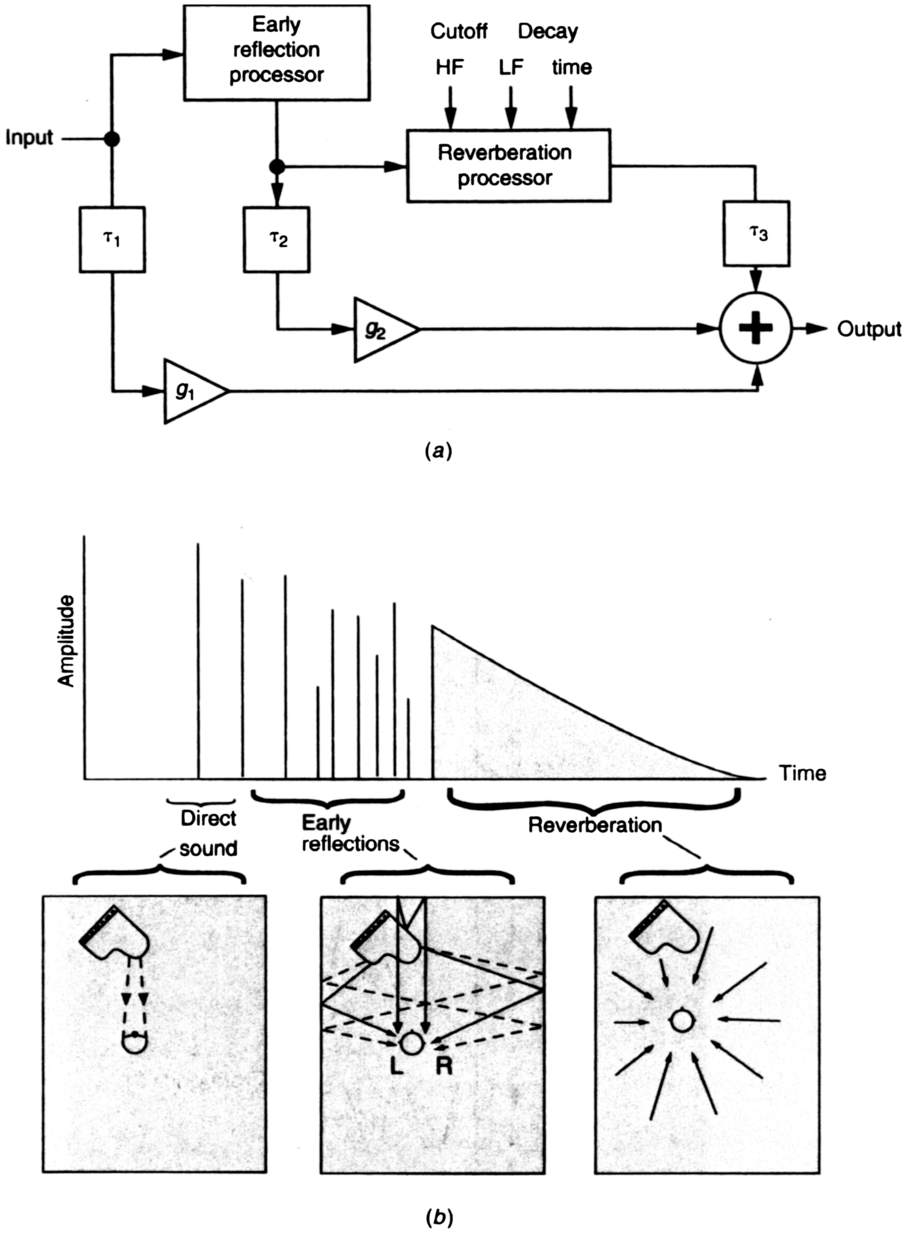


FIGURE 20.4.11 The basic operations in most lossless compression algorithms.

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

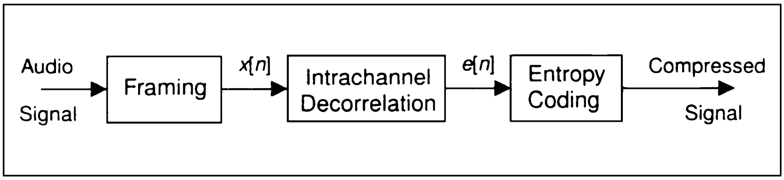


FIGURE 20.4.12 The basic operations in most lossless compression algorithms.

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

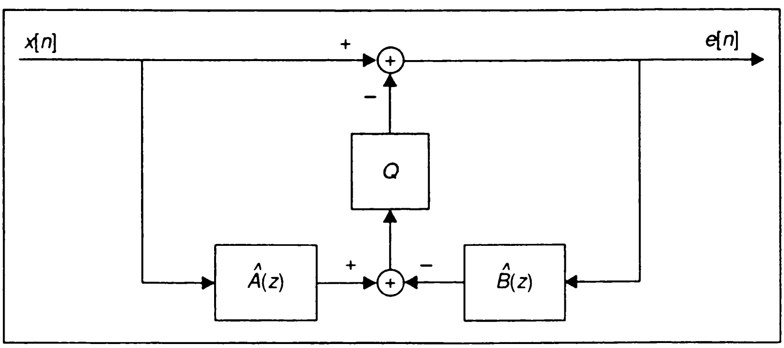


FIGURE 20.4.13 General structure for prediction.

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.