Audio Compression: Technology and Applications

Nikil S. Jayant
Edward Y. Chen

Recent algorithms for perceptual audio coding have provided the capability of compressing compact disk (CD)-stereo audio material (at 1,406 kbits/s) into rates approximating 128 kbits/s. This capability has enabled the development of several new classes of applications in audio transmission, broadcasting, and storage. In exploring the opportunities provided by these applications, rapid progress is occurring in three related technologies: signal processing and computing, for the creation of inexpensive audio decoders; communication techniques, for reliable transmission of compressed audio over radio channels; and, advanced memory, to supplement or replace the magnetic and optical media currently used to store high-quality audio. Perceptual audio coding will also enable the delivery of AM to FM-grade audio signals over voice-band modems at rates of about 32 kbits/s.

Introduction

The involvement of AT&T Bell Laboratories in audio processing has been a long and variegated one. It encompasses the domains of audio storage media, audio transducers, and various forms of audio signal processing. This paper discusses both recent research-and-development efforts in the specific field of high-quality digital audio and the implications of these efforts in potential new products and services for entertainment and communications.

A key enabler of these emerging technologies is the function of *audio compression*. This technique is the art and science of representing an audio signal in a compact, digital format for economies in transmission and storage. An audio signal is meant to be unrestricted in its type, although two very important subclasses are *music* and *speech*.

In this paper, the term *audio* refers to both music and speech if it is not otherwise specified as being only speech. The compression algorithms discussed are not specifically tuned to speech, however, as they are in a typical speech coder.

The Audio Processing section reviews the factors affecting speech and audio coding. The Compression Technology section discusses audio compression, focusing on perceptual audio coding (PAC) and the digital-signal-processing (DSP) technology needed to implement PAC coders and decoders (codecs). The Digital Audio Radio section explains the application of PAC technology to digital audio broadcasting (DAB), while the Audio Storage and Recording section clarifies the PAC application as it relates to next-generation technology for audio storage. The Network Service Opportunities section explores the implications of compression on emerging network services in communications and entertainment. The paper concludes with a summary of challenges in research, technology, and business.

Audio Processing

At least three dimensions of audio quality exist: *signal bandwidth*, *fidelity* of the reproduced (digitized) signal (for a given bandwidth), and *spatial integrity* in the case of multichannel audio (of which stereo is an

Table I. Signal bandwidths in audio

Audio grade	Bandwidth (Hz)
Telephone speech	200 - 3,200
AM audio	50 - 7,000
FM audio	50 - 15,000
CD audio	20 - 20,000

Table II. Current capabilities in speech and audio coding

Capability	Signal bandwidth (kHz)
Communications-quality speech at 4-8 kbits/s	3.2
Network-quality telephony at 16 kbits/s	3.2
Wideband speech coding at 16 to 32 kbits/s	7 - 10
Commentary audio at 32 kbits/s	10
CD-like stereo at 64 kbits/s	15
CD stereo at 128 kbits/s	20
Five-channel audio at 320 kbits/s	20

important example).

Table I provides a summary of various, well-recognized grades of audio bandwidth. Nowadays, fidelity in low-bit-rate digitized audio is appropriately measured by an internationally standardized five-point impairment scale.[1,2] Multichannel audio is typified either by a left-right stereo pair or the so-called 5.1 format (left, right, center, rear left [left surround], rear right [right surround] and a low-bit-rate subwoofer for very low frequencies).

Signal compression is a key element in contemporary and emerging technology for digital audio. Table II summarizes current capabilities in audio compression—in terms of the bit rate in the (compressed) digital representation—for various grades of audio bandwidth and spatial content. These capabilities are particularly reflective of current AT&T technology for speech and audio compression. In terms of needed bit rate and complexity of implementation, there is very little overhead in making the transition from FM-grade audio to compact disk (CD)-grade audio.

Speech Coding. Most speech-coding technology has addressed the telephone-bandwidth signal. During the last two decades, the bit rate needed for so-called *toll-quality* or *network-quality* digital telephony has decreased from 64 to 32 to 16 kbits/s. Recently, coding of 7-kHz speech has received more attention, and the bit rate needed for high-quality speech in this bandwidth has decreased from 64 to 32 kbits/s, with 16 kbits/s as a nearly realized current target.

On the other hand, the technology of unrestricted audio has focused on signals in the 15- to 20-kHz bandwidth, as in FM and CD formats. Recently, techniques for digital audio coding have been extended to include lower signal bandwidths, such as 7 to 10 kHz, using correspondingly lower bit rates in the digital representation.

Conventional methods of speech coding rely on a universal model of speech production to remove *statistical redundancy* for signal compression. Audio coding relies on typically weaker models for redundancy reduction. It gains significant efficiencies, however, by eliminating *perceptual irrelevancy* in the signal to be compressed.

As in speech coding, the principal criteria for measuring audio-coder performance are *quality, bit rate, delay*, and *complexity*. In the one-way application of broadcasting, delay is not critical in the same sense as in the two-way speech-communication example. It is still important to maintain processing delay within reasonable limits, however, so that such parameters as the latency in station-switching are not objectionable. Many applications of audio coding (including broadcasting) are decoder intensive. As such, the focus will be on the *decoder* in addressing complexity.

The paper by Cox et al. in this issue discusses speech coding in greater detail.[3]

Audio Coding. High-quality, digital stereo media include the CD, digital audio tape (DAT), mini-disk (MD), and digital compact cassette (DCC). All these media assume a 20-kHz bandwidth for the audio signal. The CD and DAT use a 16-bit pulse-code modulation format and sampling rates of 44.1 and 48 kHz. Thus, (uncompressed) bit rates of 1.406 and 1.536 Mbits/s are used, respectively, for two-channel stereo. Audio quality provided by both the CD and DAT formats is considered to be nearly identical. The MD and DCC use efficient signal compression by a certain factor, which falls within the range of four to five. This technique provides near-CD quality. The compression capability of ten to one or greater, implied in Table II, is even more powerful. It results in very little loss of perceived audio quality.

The capability of compressing audio by a factor of ten to one (or better) leads to many new and important application classes, which are listed in Panel 2. Due to the emerging generation of quality-sensitive users and the capacity-limited characteristics of transmission and storage media, these applications demand both the highest levels of audio compression and the very best modem and storage technologies. Great sophistication in software and hardware is also required for DSP and computing.

As with speech coding, several current and emerging standards for audio compression currently exist. These standards include the International Telegraph and Telephone Consultative Committee (CCITT) G.722 standard for commentary (7-kHz bandwidth) audio at 64 kbits/s, and the Motion Picture Experts Group 1 standard for compression of CD stereo to rates of 128, 192, and 256 kbits/s.[2] Most low-bit-rate telephone speech coding is based on linear predictive algorithms. Low-bit-rate audio coding, however, depends on the explicit frequency-domain algorithms of transform coding and sub-band coding. Low-bit-rate audio coding has a high degree of tuning to the characteristics of the human perceptual (auditory) system.

Panel 1. Abbreviations, Acronyms, and Terms

AM—amplitude modulation
CCITT—International Telegraph and Telephone Consultative Committee
CD—compact disk
codec—coder/decoder
DAB—digital audio broadcasting
DAR—digital audio radio
DAT—digital audio tape
DBS—direct-broadcast satellite
DCC—digital compact cassette
DSP—digital signal processor
EIA—Electronic Industries Association
FCC—Federal Communications Commission
FFT—Fast Fourier Transform
FM—frequency modulation
IBAC—in-band adjacent channel
IBOC—in-band on channel
IBRC—in-band reversed channel
ISDN—integrated services digital network
JND—just noticeable distortion
MD—mini-disk
mflops—millions of floating-point operations per second
NB—new band
NRSC—National Radio Systems Committee
PAC—perceptual audio coding
POF—point of failure
PSK—phase shift keying
RAM—random-access memory
RF—radio frequency
RISC—reduced-instruction-set computing
ROM—read-only memory
TOA—threshold of audibility

Compression Technology

The algorithm for PAC, as well as the current implementations of the PAC codec, are discussed in this section. The elastic compression capabilities of PAC are summarized in the last four rows of Table II.

Perceptual Audio Coding. Figure 1 is an illustration of the power spectrum of an audio signal and the just-noticeable-distortion (JND) profile, a function of the audio spectrum and the characteristics of the human auditory system. The "treads" in the JND "staircase" correspond to the so-called *critical bands* in hearing. The meaning of a critical band is explained later in this subsection. At this point, however, note that there are 25 critical bands shown, and each band's bandwidth increases with frequency. As the JND profile in Figure 1 is re-evaluated as a function of input audio spectrum (say, once every 10 ms), the band edges of the staircase treads remain invariant while the heights of the treads adapt to the new spectrum.

The JND represents a critical distortion level as a function of frequency. If the distortion (introduced in the compression process) is at or below the staircase function at all frequencies, such distortion is simply not heard. A consequence of this result is that the only frequency components needing preservation in transmission are those jutting above the JND; all other components are merely discarded. This leads to extremely high levels of compression with no loss of audio quality. The mathematically computed signal-to-distortion ratio can be very low (as low as, say, 20 dB) while the perceived signal-to-distortion ratio is infinite, in principle.

This paradigm results in a constant-quality, variable-bit-rate coder. Typically, one can realize such a coder in which the signal is slightly overcoded most of the time and slightly undercoded some of the time. This effect is accomplished by using a bit-rate buffer and by entering feedback of buffer status into the coder.

The intelligence in the PAC algorithm resides in the computation of the JND function, which is an empirical derivation based on an understanding of the human auditory system. In the current version of PAC, the JND threshold is calculated as an interpolation between fairly well understood thresholds for *noise-masking-tone* and *tone-masking-noise*.[4] The interpolation parameters depend on how tone-like or noise-like the input signal is estimated to be in each critical band.

The JND technique utilizes the phenomenon of *frequency-domain masking*, which is the capability of a signal to mask a (somewhat) weaker signal (or distortion) in its frequency vicinity. Such masking is greatest in the critical band in which the input signal is located, and lower degrees of masking are realized for signals (or distortions) residing in farther critical bands. The phenomenon of *time-domain masking*, the capability of a signal to mask a (somewhat) weaker signal (or distortion) in its temporal vicinity, is not as well understood. Temporal masking, however, is also implicitly used in PAC. For example, by adapting the audio block length used for frequency-analysis to the characteristics of the audio signal, the possibility of temporal masking can be enhanced. Specifically, the distortions in the coding of sudden onsets of audio activity are best masked by using a short block length. Longer block lengths are more efficient for compressing steady-state sounds. The resulting system is a *window-switching* perceptual coder. The overall compression efficiency of PAC is maximized by following the perceptual coding algorithm with no-loss compression of the digital signals at the coder output.

Composite Coding of Multichannel Audio. In the coding of both two-channel and five-channel signals, the dependencies that typically exist among the various channels can be used to achieve greater compression than that provided by independent coding of the component channels. PAC uses perceptually meaningful criteria for composite coding of multichannel audio. For example, a stereo pair of left (L) and right (R) signals is compressed either by coding the (L,R) pair or an (L+R, L-R) pair. The switching between the (L,R) and (L+R, L-R) analyses, as well as bit rates used to code these components, are based on advanced psycho-acoustic criteria.

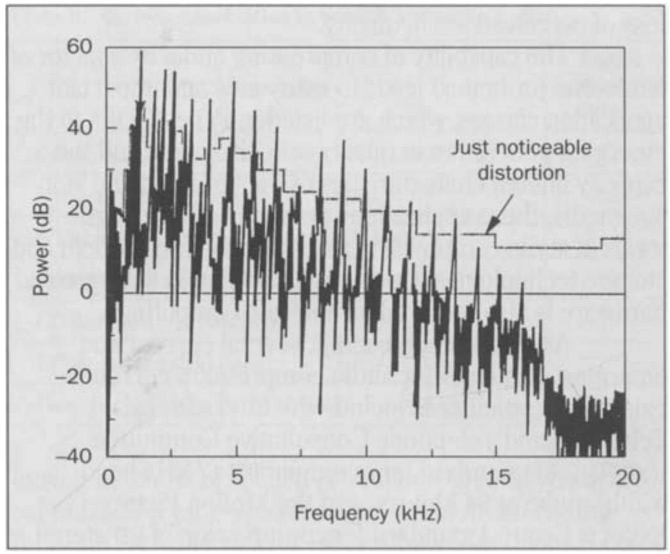

Figure 1. This illustration traces the power spectrum of an audio signal and the just-noticeable-distortion (JND) criterion for perceptual audio coding. The "treads" in the JND "staircase" correspond to the so-called *critical bands* in hearing. Note that there are 25 critical bands and that each band's bandwidth increases with frequency. As the JND profile is reevaluated based on a function of input audio spectrum, the band edges of the staircase treads remain invariant while the heights of the treads adapt to the new spectrum.

Variable-Rate PAC. By making the bit-rate buffer very long (say, seconds or even minutes long), PAC can operate in the *constant-quality mode*, which is fundamentally more appropriate than the *constant-rate* mode because of the non-stationary nature of the audio signal. The variable-rate method, which has a long buffer, is impractical for a transmission or broadcast application. It is ideal, however, for a storage application. For a given level of quality, such as CD-quality coding, the variable-rate method results in an *average* bit rate that could be 25 to 40 percent lower than the rate of a constant-rate coder.

Implementation of the PAC Codec Prototype. The PAC *stereo decoder* is implemented on a single, general-purpose digital signal processor: the AT&T DSP3210, which operates at a clock speed of 55 MHz. The instruction cycle time is 72 ns. There are 8 Kbytes (2K $*$ 32 bits) of on-chip RAM, which

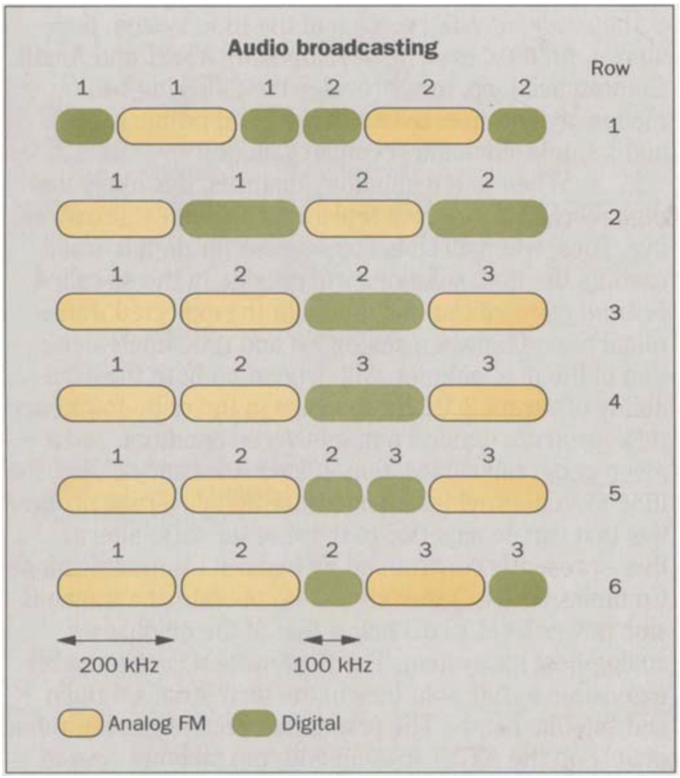

Figure 2. This depiction illustrates the bandwidth allocation schemes in audio broadcasting. Several paradigms for simulcasting analog and digital versions of the same audio program in the FM terrestrial band are possible. The same paradigms also apply to digital broadcasting of a separate audio program. Numbers on the figure refer to station (channel) number. Rows 2, 3, and 4 typify the in-band adjacent channel (IBAC) and in-band reversed channel (IBRC) scenarios. Row 1 typifies the in-band on-channel (IBOC) double-sideband mode. Rows 5 and 6 represent the IBOC single-sideband mode.

can be used for the critical, speed-sensitive code. The total memory requirement for real-time decoding with the current algorithm design is about 64 Kbytes.

On the other hand, the PAC *stereo encoder* is implemented on a dual Intel i860 processor system. (Intel is a registered trademark of Intel Corp.) This system is a RISC-based, symmetric, multiprocessor platform having addressable memory shared equally between the two i860 processors. The on-board processors run at a clock rate of 50 MHz. The dual i860 processor system is capable of executing a 1,024-point, complex Fast Fourier Transform (FFT) in fewer than 300 µs. The memory requirement for the real-time PAC encoder is about 1.5 Mbytes. In the current implementations described earlier, the processing requirements for PAC decoding and encoding are approximately 28 and 140 mflops (millions of floating-point operations per second).

The PAC *five-channel decoder* uses the same platform as the stereo PAC encoder. The difference is that the five-channel decoder uses only one i860 processor instead of two. The processor's clock rate is the same. The memory requirement for the five-channel decoder is much lower than the 1.5 Mbytes used in the current implementation of the two-channel encoder.

The input/output interface for the encoder and decoder boards is the PC/AT Industrial Standard Architecture bus. (PC/AT is a registered trademark of International Business Machines Inc.) In each case, the host platform is a basic, everyday PC. This arrangement has facilitated prototyping efforts, simplified system development, and greatly reduced costs.

Clearly, the foregoing descriptions emphasize that the complexities of both the encoder and decoder are quite different. In such applications as broadcasting or audio playback, this asymmetry is an important issue. The decoder market and decoder needs would outweigh those of the encoder. In either teleconferencing or audio recording-and-playback applications, a more symmetrical design is desirable. The reason is that every decoder must coexist with an encoder in such applications. At present, less-complicated PAC encoders are being investigated. The hope is that their design will parallel that of the simple decoders. Thus, more efficient and pervasive use of PAC technology will be facilitated in various applications.

Digital Audio Radio (Broadcasting)

The following three topics are discussed in this section: the application of DAB, a DAB system recently developed at AT&T Bell Laboratories, and an ongoing process for creating a DAB standard in the United States. For purposes of this discussion, digital audio radio (DAR) is used as a synonym for DAB.

DAR Technology. In recent years, DAR—having transmission formats ranging from terrestrial, cable, and satellite—has proven to be a powerful and timely technology through various tests. Many view DAR as the logical progression in sound transmission for the next century. DAR's fundamental promises include CD-quality sound and superior immunity to interference. A DAR system also offers the potential of many new broadcast services, including both information transfer and messaging, which can easily be implemented into digital technology.

Figure 2 depicts several possible paradigms for simulcasting analog and digital versions of the same audio program in the FM terrestrial band. The same paradigms also apply to digital broadcasting of a separate audio program. Within the 88- to 108-MHz band, the individual, analog, FM-broadcasting license is currently based on 200-kHz spacing of carrier frequencies. *In-band* DAR systems propose to operate in this same 88- to 108-MHz band, and they can make use of vacant spacings in any one of the following three ways:

- Two 100-kHz spacings on either side of the analog channel (row 1);
- A single 200-kHz spacing (rows 2, 3, and 4); or
- A single 100-kHz spacing (rows 5 and 6).

The frequency-allocation arrangements in rows 2, 3, and 4 of Figure 2 are said to be the *in-band adjacent channel* (IBAC) systems. In the ever-so-crowded FM-broadcasting markets, broadcasters and station owners welcome the efficient use of the current broadcasting-band allocation. By using only 100 kHz of bandwidth spacing per station transmission instead of 200 kHz, the number of potential FM-broadcasting stations would increase greatly. This increase would certainly provide better frequency granularity. For example, in Figure 2, simulcasting (in general, broadcasting) is provided for stations 2 and 3 in rows 5 and 6 but only for station 2 in row 3 and station 3 in row 4. The capability of compressing CD stereo into bit rates approximating 128 kbits/s is crucial to the robust transmission of a high-quality stereo signal over a 200-kHz facility. In the current state of audio-compression and radio-modem technologies, however, it may not be possible to provide rigorous CD-quality stereo, combined with transmission robustness, in a bandwidth spacing of 100 kHz.

Figure 2 also shows other in-band solutions. The *in-band on-channel* (IBOC) system operates simultaneously with an analog host FM transmitter and the same carrier frequency. In row 1 of Figure 2, simulcasting for station 1 or station 2 is accomplished by means of a so-called *double-sideband* IBOC system. In rows 5 and 6, the *single-sideband* version of the IBOC system is shown. An IBOC system, developed by AT&T and Amati Communications, Inc., provides the following two modes of operation: a double-sideband primary mode and a single-sideband secondary mode.

When DAR technology matures, it is likely that digital broadcasting will replace all analog FM broadcasting. Thus, when all slots are *reserved* for digital broadcasting, the IBAC solutions will operate in the so-called *in-band reserved channel* mode. In the expected transitional period between analog FM and DAR, implementation of the IBAC solution will depend on both the availability of vacant 200-kHz spacings in the radio-frequency (RF) spectrum under a non-interfered condition and a given geographical location. If such a vacancy exists, the IBAC system provides interference and coverage properties that can be superior to those of the IBOC alternative—presently constrained by Federal Communications Commission (FCC) guidelines—to operate at a transmission power level 25 dB below that of the on-channel, analog-host FM system. The IBAC system is also readily extensible to DAR solutions in the terrestrial AM-radio and satellite bands. The rest of this section concentrates mainly on the AT&T in-band adjacent channel system.

In an over-the-air transmission environment, broadcasting signals suffer through many types of channel impairments, such as:

- Fading due to multipath reception;
- Doppler shifting due to vehicles traveling at high speeds;
- Deep fades;
- Nulls; and
- Stoplight fades due to slow traffic, narrow streets, and tall buildings in urban areas.

When such impairments occur, one familiar characteristic of an analog FM system is that reception quality degrades continuously as the signal-to-noise ratio of the received signal decreases. In a DAR system, one of the primary goals is to maintain a high level of CD-quality audio reception over a significant majority of transmission impairments, allowing almost perfect reception until the point of total failure.

The 200-kHz slots in Figure 2 are well matched to the audio coding rate of 128 kbits/s. With anticipated improvements in coding schemes and modeling techniques for the stereo perceptual audio coder, a 64-kbits/s

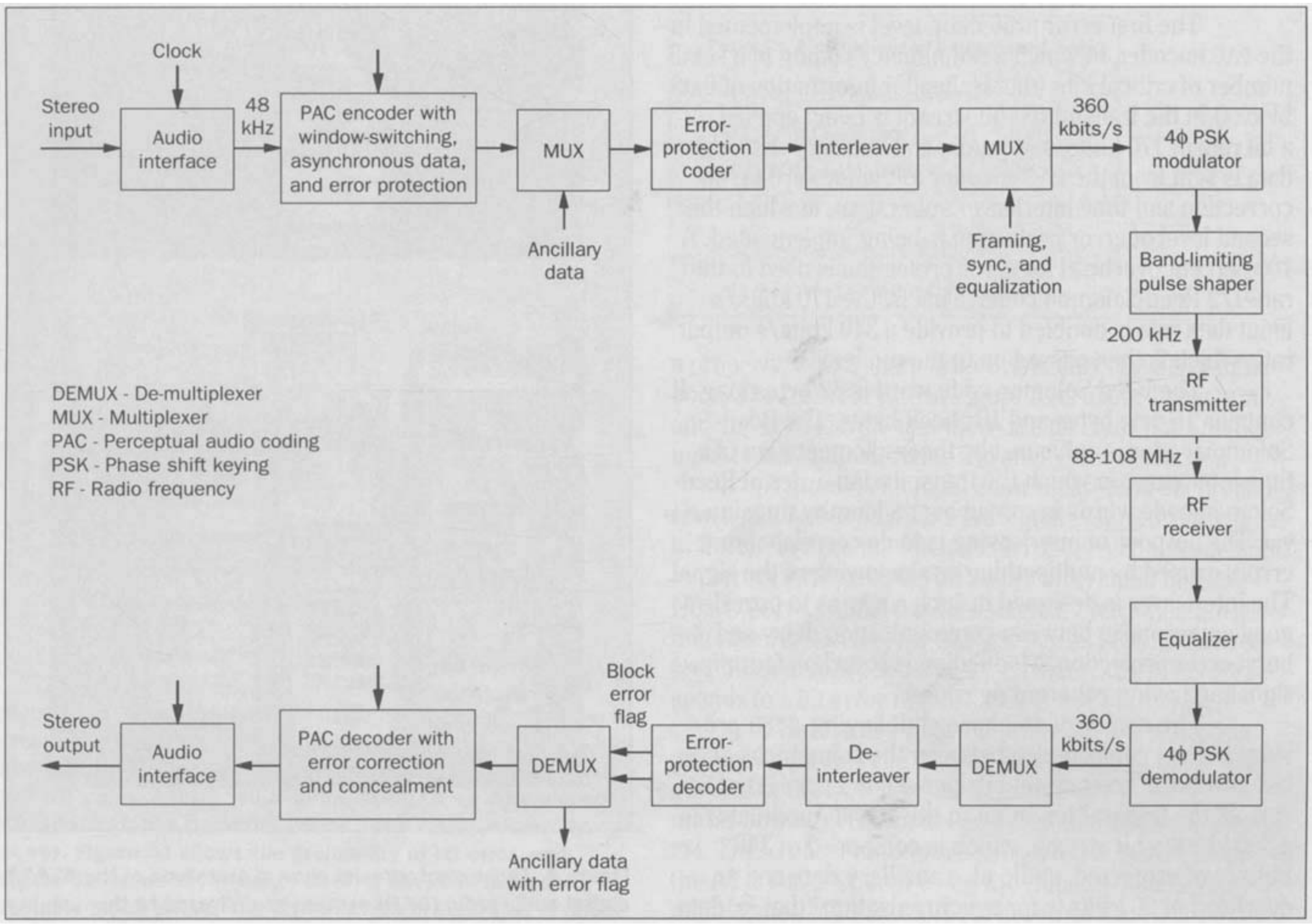

Figure 3. This is a drawing of a high-level block diagram of the AT&T in-band adjacent channel (IBAC) and in-band reversed channel (IBRC) system for digital audio radio (DAR). The AT&T DAR system provides an adaptive equalization design to alleviate frequency-selective channel fading, time-interleaving to randomize burst errors, and three levels of protection against residual bit-error effects.

CD-quality codec could eventually be realized. The implementation of this new compression technology might increase the potential of available channels in the demanding RF spectrum. Thus, such implementation presents a better granularity (that is, 100 kHz per channel for the FM band instead of the currently defined 200 kHz). Depending on the nature of the transmission and broadcasting applications, the low bit rate and elastic characteristics of the PAC algorithm could help allocate some of the given spectrum bandwidth to channel-coding improvements (that is, synchronization, equalization, and timing-recovery schemes) and different types of forward-error-correction schemes. This allocation would significantly strengthen transmission robustness without sacrificing audio quality.

The AT&T IBAC/IBRC System. Figure 3 shows a high-level block diagram of the AT&T IBAC/IBRC system for DAR. The AT&T DAR system provides adaptive equalization design to alleviate frequency-selective channel fading, time-interleaving to randomize burst errors, and three levels of protection against residual bit-error effects.

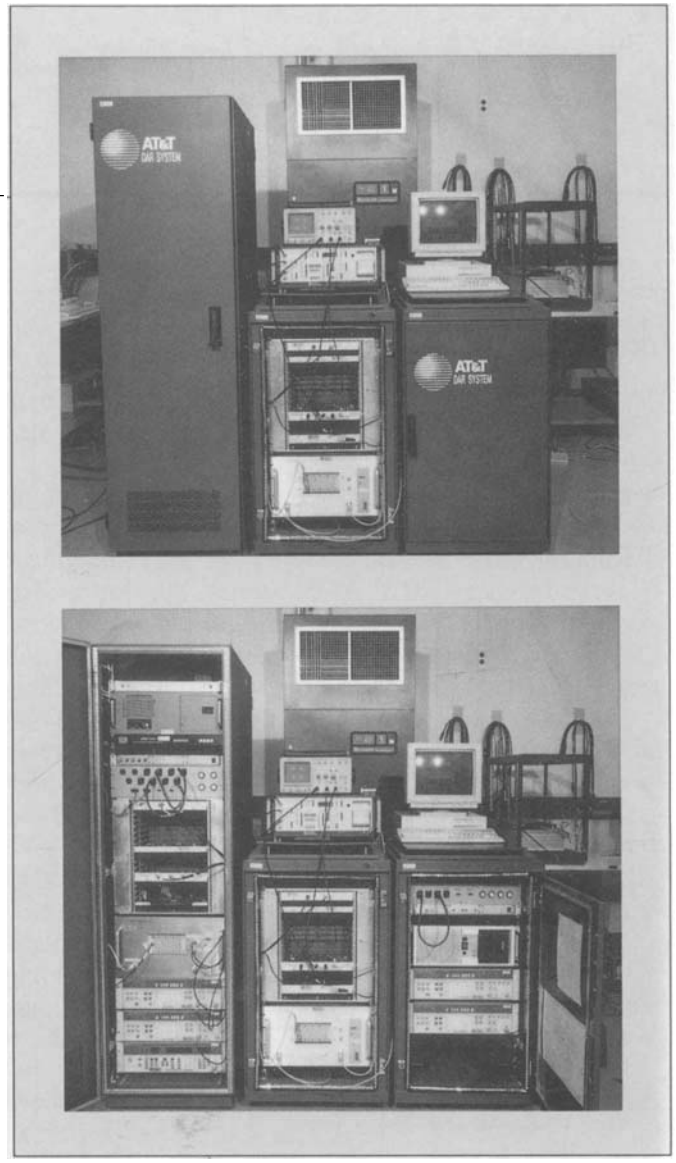

Figure 4. These photographs show a prototype of the AT&T digital audio radio (DAR) system as delivered to the Electronic Industries Association (EIA) National Radio Systems Committee (NRSC) testing process.

The first error-protection level is implemented in the PAC encoder, in which a redundancy coding of a small number of critical bits (that is, header information of data blocks) in the transmitted bit stream is being applied. At a bit rate of 170 kbits/s for audio and ancillary data, the data is sent from the PAC encoder to the forward-error-correction and time-interleaver subsystem, in which the second level of error protection is being implemented. A 100-percent overhead for error protection is used in the rate-1/2 Reed-Solomon coder. That is, the 170-kbits/s input data rate is doubled to provide a 340-kbits/s output rate, which is then passed on to the modem.

The Reed-Solomon code word is 32 bytes long. It contains 16 data bytes and 16 check bytes. The Reed-Solomon coder is enhanced by the implementation of a time interleaver in which the transmission order of Reed-Solomon code words is spread over a lengthy time interval. The purpose of interleaving is to de-correlate burst errors caused by multipathing or shadowing of the signal. The interleaver is designed in such a way as to provide a good compromise between communication delay and burst-error protection. Modulation is based on four-phase signaling having coherent detection.

The choice of 4ϕ phase shift keying (PSK) provides a good compromise between the robustness of 2ϕ PSK and the efficiency (in bits-per-second per Hz) of 8ϕ PSK. At the transmitter, input to the 4ϕ PSK modulator is a 360-kbits/s bit stream, which is composed of 340 kbits/s of protected, audio-plus-ancillary data and an overhead of 20 kbits/s for synchronization (that is, data block, interleaver block, and symbol clock).

At the receiver, the 20-kbits/s overhead is used for synchronization, channel equalization, and timing recovery. The channel equalizer is a separable, non-cross-coupled passband equalizer using a fractional-spacing algorithm having a 180 kilosymbols-per-second symbol rate. Equalization is periodically adaptive, occurring once for every block of 3,400 bits (or 1,700 symbols, with 2-bit symbols).

If the number of received errors exceeds the correction capability of the Reed-Solomon decoder, a block-error flag signal at the hardware level is asserted to the PAC decoder, indicating that the data in the current block is not valid. This is the third level of error protection. It is called *error concealment*, an algorithm for the concealment of residual errors at the PAC decoder. The concealment algorithm uses audio-signal redundancies, including left-right correlations in stereo, to provide a reconstruction quality that is significantly better than that of muting. Muting is used only when there is a burst of several consecutive block-error flags.

The communication delay in the DAR system is approximately 650 ms. This figure is the sum of an encoder delay of 50 ms, an interleaver delay of about 400 ms, and a decoder delay approximating 200 ms. Latency in station-switching is about 600 ms. Figure 4 shows photographs of a prototype AT&T DAR system, as delivered to the Electronic Industries Association (EIA) National Radio Systems Committee (NRSC) testing process.

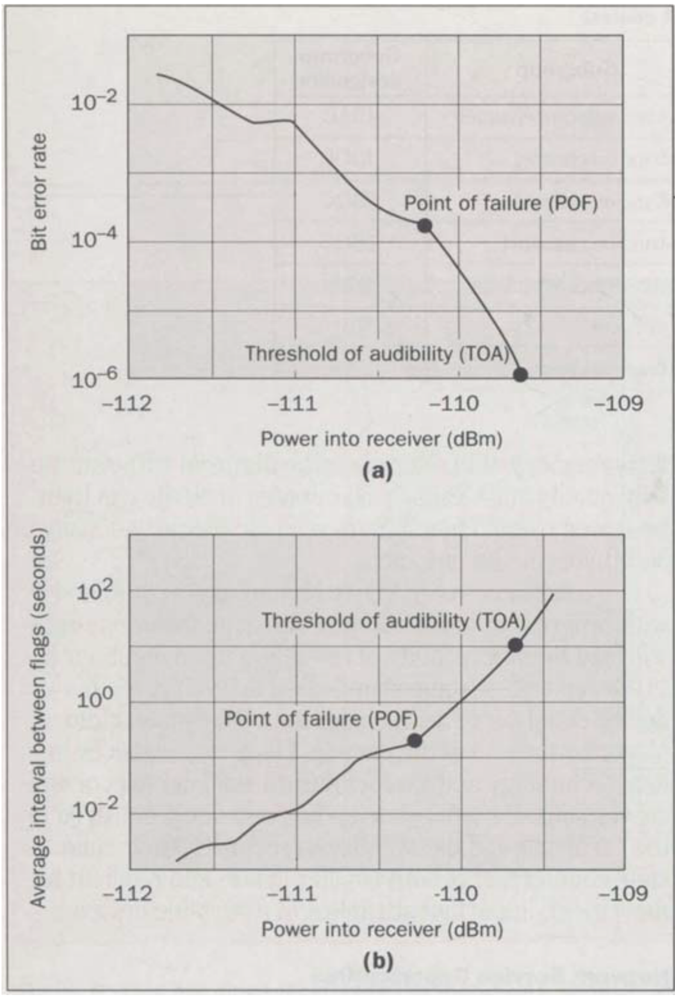

Figure 5. Implementation of both digital-channel-coding techniques and associated, multiple levels of error protection has shown that the difference between perfect audio reception and the point of total audio failure could be as little as one dB. These graphs trace data error rates versus received power. Figure 5a shows the probability of bit error, and Figure 5b shows the average interval between non-correctable block errors. The bit error rate in Figure 5a is about 10^{-6} at the threshold of audibility (TOA) and 10^{-4} at the point of failure (POF).

The following two metrics have been defined in DAR testing (as discussed in the next paragraphs) to describe the perceived output quality of received audio:

- *Threshold of audibility* (TOA) (of distortion), defined as an occurrence of first-noticeable audible distortion; and
- *Point of failure* (POF), defined as the first-noticeable occurrence of very annoying audio quality.

Implementation of both digital channel-coding techniques and associated, multiple levels of error protection has shown that the difference between perfect audio reception (one-quarter dB before TOA) and the point of total audio failure (one-quarter dB after POF) could be as little as one dB. The error-data plot in Figure 5 shows such a property, with a sharp drop within only one dB of difference. The bit error rate in Figure 5a is about 10^{-6} at TOA and 10^{-4} at POF. Note that these are bit error rates at the input to the audio decoder and after error correction.

Panel 2. Applications of compressed audio
–Digital audio broadcasting
–Advanced television
–CD-ROM multimedia
–Solid-state audio album
–Music preview and distribution
–Personalized or just-in-time audio

A more meaningful error measure is the probability of a *block-error flag*. This is a signal asserted at the input to the audio decoder when the error-correcting code fails to correct the (multiple) bit errors in an audio block. Nearly perfect audio can be obtained, even when the assertion rate of a block-error flag goes up as high as three to four block errors per second (see Figure 5b). This corresponds to a bit error rate of about 10^{-4} (Figure 5a).

The characteristic of nearly perfect reception during deep fades can also be interpreted as being a bigger and better coverage area for DAR (for a given transmitter power level) when compared to traditional, analog FM. These observations are supported in recent testing of the AT&T DAR system using the transmission facility of radio station WPRB, 103.3 FM in Princeton, New Jersey. The Electronic Industries Association – National Radio Systems Committee tests mentioned in the following subsection include measurements of TOA and POF for various, typical characterizations of the mobile DAR channel.

The EIA-NRSC Contest for a DAR Standard. In early spring 1992, the EIA was chartered by the FCC to establish a DAR standard subcommittee. This group, together with a subcommittee of the NRSC, is responsible for testing, documenting, and recommending DAR systems to the FCC. In April 1994, AT&T, along with four other proponents, submitted proposals to the EIA. Seven DAR systems (Table III) were submitted to the EIA for the DAR standard competition in the United States. Test procedures include more than 40 weeks of laboratory testing for audio quality, transmission impairment and receiving impairment, and 12 weeks of mobile field testing.

Table III. Candidate systems in the U.S. DAR contest

Proponent	Band (MHz)	Subgroup	Subgroup designator
AT&T	88-108	In-band adjacent channel	IBAC
AT&T/Amati*	88-108	In-band on channel	IBOC
USADR-FM#1	88-108	In-band on channel	IBOC
USADR-FM#2	88-108	In-band on channnel	IBOC
VOA/NASA/JPL	2,310-2,360	Direct-broadcast satellite	DBS
Eureka 147*	1,452-1,492	New band	NB
USADR-AM	0.54-1.7	In-band on channel	IBOC

*Systems include a second mode

The following four DAR systems were submitted to the EIA by their proponents:
- In-band adjacent channel (IBAC);
- In-band on-channel (IBOC);
- Direct-broadcast satellite (DBS); and
- New band (NB).

AT&T submitted two systems: an internally designed IBAC DAR system, and an externally designed IBOC DAR system (the latter was submitted in partnership with Amati Communications, Inc.) Both are in-band systems, and they are designed to be tested in the 88- to 108-MHz terrestrial FM radio band. Of the seven systems submitted to the EIA, AT&T PAC digital audio-coding technology is being used in three: the AT&T IBAC System, AT&T/Amati IBOC System, and the Voice of America/NASA/JPL DBS System. The PAC codec provides CD-quality stereo for most tested signals at bit rates ranging between 128 and 160 kbits/s. The single-sideband mode of the AT&T/Amati System uses PAC at 128 kbits/s. All other PAC codecs entered into the EIA-NRSC contest for a DAR standard use a rate of 160 kbits/s.

Audio Storage and Recording

Compression factors approximating ten to one point to a revolution in the technology for audio playback and recording. For example, a CD using compression technology can contain about ten hours of high-quality music, rather than only about one hour without compression. Alternatively, one hour of audio recordings can be stored on a considerably smaller medium, perhaps even a postage-stamp-size silicon chip.

Figure 6 plots the number of minutes of CD-stereo that can be stored on a ROM chip as a function of year (1994 through 1998), assuming the continued evolution of memory technology. A stereo encoding rate of 128 kbits/s is also assumed, leading to approximately *one minute per megabyte*. With today's technology, capacities of approximately 60 Mbytes per square inch are impractical for general-purpose ROMs. Such capacities may be possible, however, in the context of a robust audio coder at 128 kbits/s, which can work with a so-called slow-and-dirty memory. If CD-stereo can be digitized with satisfactory quality at 64 kbits/s, 60 minutes of music can then be stored using a much more realistic memory density of 30 Mbytes per square inch.

Advances in memory technology—combined with progress in audio compression and *decoder* design—will lead to the capability of recording up to one hour of CD-stereo on a postage-stamp-sized ROM chip. Such a device could serve as the basis for a solid-state audio player by the end of the decade. Likewise, advances in RAM technology and *encoder* design will lead to economical versions of a solid-state audio recorder. Compared to the CD player and the MD player/recorder, their solid-state counterpart is both smaller in size and resilient to movement, important attributes of a portable device.

Network Service Opportunities

In the future, audio signals are expected to constitute a rapidly increasing part of telecommunications traffic. Audio communications includes *business services*, such as audio transmissions on studio links, audio retrieval from remote databases for just-in-time creation of albums, and high-quality audio conferencing. Audio communications also includes such consumer services as music preview, downloading of music on a high-speed transmission link into the home, pay-per-listen and narrowcasting services, creation of personalized albums, and ultimately—high-quality music and voice transmitted over advanced voice-band and ISDN modems.

The size of the network-services opportunity can be gauged by means of the following numbers, which are based on 128-kbits/s stereo: a 1-terabyte database for 10,000 hours of music; 10^{17} bits for 100-million just-in-time tapes; and, 10^{17} bits, assuming 24-hour transmission into 10,000 studios. If 64 kbits/s is considered to be a basic call unit, such traffic represents tens of billions of call-minutes per year.

The growth in audio communications in all its many varieties and formats depends on suitable refinements of current practices in the domains of copyrighting, royalties, and content ownership.

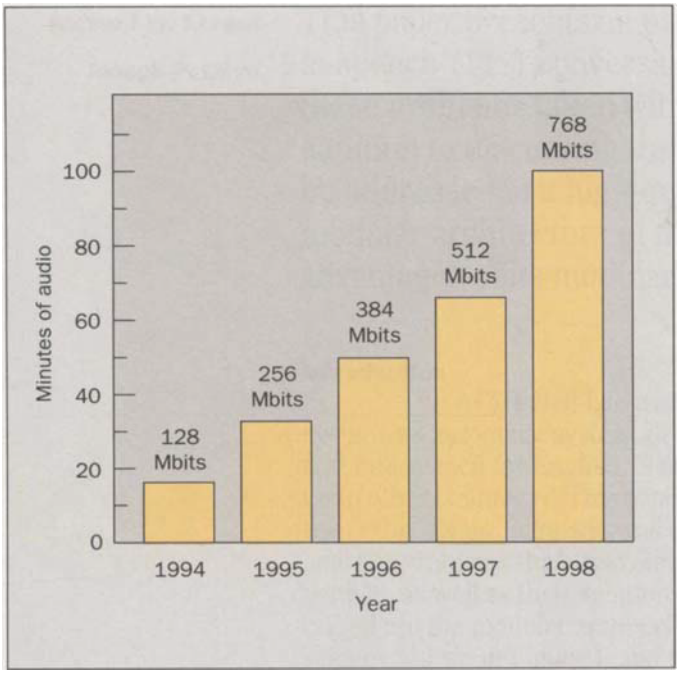

Figure 6. This bar chart illustrates the increasing capacity of the solid-state audio chip over a five-year time span. It plots the number of minutes of CD-stereo that can be stored on a ROM chip as a function of year, assuming the continued evolution of memory technology. A stereo encoding rate of 128 kbits/s is also assumed, leading to approximately one minute per megabyte.

Conclusion

Recent advances have been made in audio compression, as well as in various new classes of applications, resulting from compression factors in excess of ten to one. The CD-stereo entertainment signal has successfully replaced analog music as the standard of high-quality audio, and compression of the 1.4 Mbits/s CD-stereo signal into about 128 kbits/s has suggested the evolution of several classes of new applications.

In parallel, the compression of five-channel audio into 320 kbits/s provides the basis for high-quality, advanced-television sound. Single-channel audio compression—at speeds between 64 and 32 kbits/s—creates fundamental new opportunities in audio services, teleconferencing, and telephony applications.

Major business opportunities arising from such applications include royalties from algorithm licensing, sale of large quantities of audio decoders and consumer products, and revenues from a high volume of audio traffic transported over telecommunications networks. In short, the vision is that digital audio technology will achieve a pervasiveness, within the next ten years, approaching or even equaling that of digital speech.

Exciting challenges will appear along the way—in research, technology development, and the digital-audio business. Compression factors well in excess of ten to one require continued advances in coding, psychoacoustics, and signal processing. Broad technological advances will depend on continued progress in the allied disciplines of transducers, echo cancelers, modems, and memory devices. Finally, new business paradigms will be needed as proprietary technology merges with standards and as the disparate disciplines of the entertainment and communications industries come together.

Acknowledgments

The authors thank the following members of the AT&T Bell Laboratories Advanced Audio Technology Department for their contributions to PAC and DAR technologies: J. D. Johnston, S. R. Quackenbush, S. M. Dorward, K. Thomson, R. L. Cupo, J-D Wang, C-E W. Sundberg, and N. Seshadri. The authors also thank the following colleagues for their supporting contributions: J. Chang, J. S. Palwyk, L. V. Kanel, J. Garibay, R. O. Marabuto, R. Rogoszewicz, C. W. Schaible, K. L. Sherman, R. Rana, R. Jhoty, and J. Heinzmann. Additionally, the authors express appreciation for the contributions of J. Berndt, W. H. Elliott, J. Hanley, V. B. Lawrence, L. R. Rabiner, and I. Wolff to the DAR project. In particular, they are grateful to A. G. Fraser for his sustained support and encouragement of this work.

References

1. N. S. Jayant and P. Noll, *Digital Coding of Waveforms: Principles and Applications to Speech and Video*, Prentice-Hall, Englewood Cliffs, New Jersey, 1984.
2. P. Noll, "Wideband Speech and Audio Coding," *IEEE Communications*, Vol. 31, No. 11, November 1993, pp. 34-44.
3. R. Cox, P. Kroon, J-H Chen, R. Thorkildsen, K. O'Dell, and D. Isenberg, "Speech Coders: From Idea to Product," *AT&T Technical Journal*, Vol. 74, No. 2, March/April 1995, pp. 14-22.
4. N. S. Jayant, J. D. Johnson, and R. J. Safranek, "Signal Compression Based on Models of Human Perception," *Proceedings of the IEEE*, Vol. 81, No. 10, October 1993, pp. 1385-1422.

(Manuscript approved February 1995)

Nikil S. Jayant *is department head in the Signal Processing Research and Advanced Audio Technology group at AT&T Bell Laboratories, Murray Hill, New Jersey. He is responsible for research on audio-visual communications, and in particular, digital audio broadcasting and videotelephony over voice-band modems and wireless multimedia networks. Mr. Jayant holds a B.S. degree in physics and mathematics from Mysore University in India, as well as B.E. and Ph.D. degrees in electrical communications from the Indian Institute of Science in Bangalore. He joined AT&T in 1968.*

Edward Y. Chen *is a supervisor and technical manager in the Advanced Audio Technology department at AT&T Bell Laboratories, Murray Hill, New Jersey. He is responsible for research on digital audio radio and home networking as they apply to audio-coder and network application-specific integrated circuit (ASIC) design. Mr. Chen holds a B.S. degree in electrical engineering from the Worcester Polytechnic Institute in Massachusetts, and an M.S. degree in engineering and applied science from Yale University in New Haven, Connecticut. He joined AT&T in 1972.*