

SIGNAL COMPRESSION

NIKIL JAYANT

*Lucent Technologies, Bell Laboratories
Murray Hill, New Jersey 07974, USA*

This article is an introduction to a special issue on signal coding and compression. We begin by defining the concepts of digital coding and audiovisual signal compression. We then describe the four dimensions of coding performance: *bit rate, signal quality, processing delay and complexity*. We illustrate the two basic principles of audiovisual coding, removal of signal redundancy and the matching of the quantizing system to the properties of the human perceptual system, with specific recent examples of coding algorithms. We then summarize standards for, and applications of audiovisual signal compression. A fast-emerging application is the internetworking of audiovisual information, a field that is too recent to be covered in the articles in this collection. We conclude our article by presenting our views about future research directions in the field.

1. Digital Signal Processing and Coding

The articles in this collection deal with digital coding for bit rate reduction. This discipline has evolved in at least two distinct schools in digital communications.

The compression of data signals is an integral part of Information Theory. In particular, the field of rate-distortion theory deals with the qualitatively different problems of lossless coding and lossy coding, using the concept of source entropy in the process.

Signal compression has also been a central topic in the fields of speech and image processing. Coding, synthesis and recognition are three key areas of interest in audiovisual signal processing. In *coding*, the purpose is to achieve a compact digital representation of the signal for economies in transmission or storage. A particular focus in *synthesis* is the creation of spoken or pictorial information starting from *text*, rather than from human speech or a real image, as in the case of coding. The customer for coded or synthesized audiovisual information is the *human*. In recognition the focus is on machine (computer) understanding of the information content of the signal, usually so as to aid in the completion of some task. Each of these areas for processing audiovisual signals is often synergistic with other areas. Thus, for example, coding and synthesis are used to build digital tape recorders for voice response systems; synthesis and recognition provide dialogue systems for voice control of machines; and recognition and coding provide complementary information in very low bit rate coding, as in the example of face location in a head-and-shoulders videophone scene, followed by more careful coding of facial features.

Focus in this special issue is entirely on the field of coding. The articles that follow this introductory paper¹⁻³ deal specifically with the coding of speech, wide-band audio and image signals. The article on image coding deals with still pictures as well as motion video.

The final paper⁴ deals with data compression. This subject is important in its own right. Lossless and lossy methods of data compression are also significant components of systems for signal compression. Examples are run-length coders for facsimile signals, and lossless coders that follow quantizing systems for audiovisual signals.

The remainder of this article focuses on audiovisual signals, in an attempt to overview the treatments in the three articles that follow it.

2. Audiovisual Signal Compression: Speech, Audio and Image Coding

The subject of signal compression has been described quite comprehensively in several texts⁵⁻⁷ and by relatively recent review articles,^{8,9} and the reader is referred to these sources for detailed discussions of coding principles as well as standards activities in the field. In this section, we provide a shorter, yet self-contained and updated account of the state-of-the-art in signal coding, in order to set the stage for the extensive updates in the signal-specific expositions of the three papers that follow this introductory article.

The four fundamental parameters of coding are *signal quality*, *bit rate*, *processing delay* and *complexity* of implementation. The primary purpose of coding is to decrease the bit rate while maintaining a specified level of signal quality. In general, it is also necessary to maintain specified levels of processing delay and implementation complexity.

Although the discussions of this article are inspired by audiovisual signals, the dimensions of coding performance that we are about to discuss apply to data signals as well. What is generally recognized however is that data transmission is often expected to be lossless, implying mathematically perfect signal quality. Further, the metric of delay is generally much less important for data than for audiovisual signals.

2.1. *Bit rate*

Table 1 defines typical bandwidths and sampling rates in audiovisual communications. The sampling rate is at least twice the bandwidth, as per the Nyquist theorem. In video signals, the bandwidth is generally only an implied quantity.

Bit rate is generally measured in *bits per second* or *bits per sample*. The number of bits per second is simply the product of the sampling rate (measured in Hertz or pixels per second) and the average number of bits per sample used in the quantizing system of the coder.

For example, based on the sampling rates in Table 1, 8 bits per sample in the coding of telephone speech would imply an overall bit rate of 64 kbps, and 16 bits per sample in the coding of each of two audio channels (left and right) in the CD-stereo format would imply an overall rate of 1410 kbps. An *average* rate of 0.25 bits per sample in the quantization of the transform coefficients in an HDTV coder would imply an overall rate of 15 Mbps. Most examples of low bit rate coding involve unequal bit allocations to the signal components resulting from time-frequency analysis, and average rates in the range of 0.2 to 2.0 bits per sample are quite common in the current art.

Table 1. Standard grades of audio and video signals.

Audio format	Sampling rate	Frequency range
Telephony	8 kHz	(200–3400 Hz)
Teleconferencing	16 kHz	(50–7000 Hz)
Compact disk (CD)	44.1 kHz	(20–20000 Hz)
Digital audio tape (DAT)	48 kHz	(20–20000 Hz)
Video format	Sampling rate	Spatio-temporal resolution
CIF	3 MHz	360 × 288 × 30
CCIR	12 MHz	720 × 576 × 30
HDTV	60 MHz	1280 × 720 × 60

CIF: Common intermediate format

CCIR: International consultative committee for radio

HDTV: One example of a high definition television format

Several coding systems also use variable-rate (ideally constant-quality) coding of different parts of the nonstationary audiovisual signal. Variable rate coding is particularly matched to packetized transmission.

2.2. Signal quality

Given that the ultimate judge of a signal compression system is the human observer, audiovisual signal quality is best described by a subjective criterion. Five-point scales of signal *quality* (or impairment) are widely accepted, and are sometimes supplemented by measurements of *intelligibility*. Measurement of subjective speech or image quality in a reliable and repeatable manner is a difficult and painstaking problem. Measurement of the *composite* quality of an audiovisual signal is at least as difficult, and data on this subject to date is quite scant.

In the case of image signals, subjective quality needs to be expressed also as a function of viewing distance. A distance of four times the picture height is a typical assumption.

2.3. *Processing delay*

The processing delay in coding is the sum of delays incurred in the *encoding* and *decoding* stages of the process. At the encoder, delay is introduced in the process of buffering blocks of data for purposes of efficient signal analysis for redundancy reduction. Examples are the block-processing methods of linear transforms and vector quantization. At the decoder, delay is introduced by block-based operations such as inverse transforms and by operations such as interpolation for increasing the displayed frame rate in video coding.

Another source of transmission delay in communications is that due to the networking of the digitized signals. Applications such as telephony and videoconferencing require the lowest possible values of delay, subject to the requirements of the compression algorithm in providing adequate levels of signal quality. In one-way communications such as broadcasting, delay is less important although it is still necessary to minimize processing delays to address requirements such as acceptable delays in station-switching. In storage applications, encoding delay is totally irrelevant, as long as the decoding delay is low enough for a good quality of service.

2.4. *Complexity*

Complexity is measured both by the arithmetic processing required by the algorithm (typically measured in mips, millions of instructions per second) and by its memory requirements (measured typically by kilobytes of ROM or RAM). The use of mips (or mflops) as a complexity measure is particularly appropriate for implementations on general purpose DSPs or computers. In application-specific integrated circuits (ASICs), other metrics of complexity (such as the number of transistors or gates) can be useful. Complexity is an important parameter of performance for at least two reasons—the need to minimize cost and the requirement (especially in applications with portable devices) to minimize power dissipation.

In applications such as broadcasting, it is particularly important to minimize complexity at the decoder. Complexity of the encoder is a relatively less important, if not insignificant, issue.

The available per-chip complexity (in arithmetic as well as memory) has been increasing exponentially over the last several years, with no saturation in sight at least until the end of the century. This will permit the practical use of increasingly sophisticated algorithms for signal coding (as well as other technologies for multimedia).

2.5. *Coding algorithms*

A variety of coding systems is needed to provide different tradeoffs of delay and complexity. However, there are only two basic principles of signal compression: (a) removal of the statistical (or deterministic) redundancies in the source signal,

and (b) the matching of the quantizing system to the properties of the human perceptual system. In the case of data signals, compression is based entirely on redundancy removal.

Speech signals have a well-understood universal model of production which permits very powerful techniques of redundancy-reduction: primarily, linear predictive coding in the time-domain. In principle, the signal structure can also be exploited in the unequal bit-allocation paradigm in frequency-domain coding, and this has been particularly appropriate in the coding of unrestricted audio.

Image and video signals have obvious inter-scan-line and interframe redundancies. This is very obviously true in the case of high-resolution facsimile where the structure appears in the form of significant 2-D clusters of black pixels (on a white background), as exploited by various forms of run-length coding. Signal structure is also very obvious in the case of videoconferencing scenes with head-and-shoulders inputs of relatively low spatiotemporal activity.

The very widely used MC-DCT coder for video uses the hybrid of a predictive operation (interframe predictive coding in the form of *motion compensation*) and a frequency-domain operation for the quantization of the prediction error (2-D *Discrete Cosine Transform*).

In speech as well as image and video coding, the quantizing system can take advantage of a perceptual phenomenon called noise-shaping to achieve even greater reductions of bit rate for a given level of signal quality. Masking is the phenomenon by which a strong stimulus (desired signal) completely covers up a weaker signal (quantizing noise) in its spectral or spatio-temporal vicinity. In other words, mathematically significant levels of quantizing distortion can be permitted with very low or sometimes zero loss of perceived signal quality. The most common form of perceptually-tuned coding is one where carefully selected components in a frequency-transform (short-time signal spectrum) are either coarsely quantized or even discarded. The greatest advances in perceptual coding have been in the compression of unrestricted audio. Unlike speech or videophone signals, unrestricted audio has no universal model or framework for redundancy reduction. In compressing these signals, the approach is to reduce redundancy as much as possible by classic techniques of prediction and transform coding, and to shift a great deal of the total burden to the perceptual (noise-shaping) side of the game. The gains of perceptual coding are the greatest when the bit rate is sufficient to provide critical, or at-the-threshold coding. In the audio coding illustration of Fig. 1, this threshold is described by the signal-specific JND (just noticeable distortion) as a function of frequency. The only signal components saved in the compression of a given audio block are those that extend above the JND for the given short-term power spectrum. The perceptual audio coder (PAC) provides extremely high levels of compression, approaching 20:1. A particular capability is the compression of the 1410 kbps CD-stereo signal to total bit rates on the order of 64 kbps with very little sacrifice of signal quality.⁹

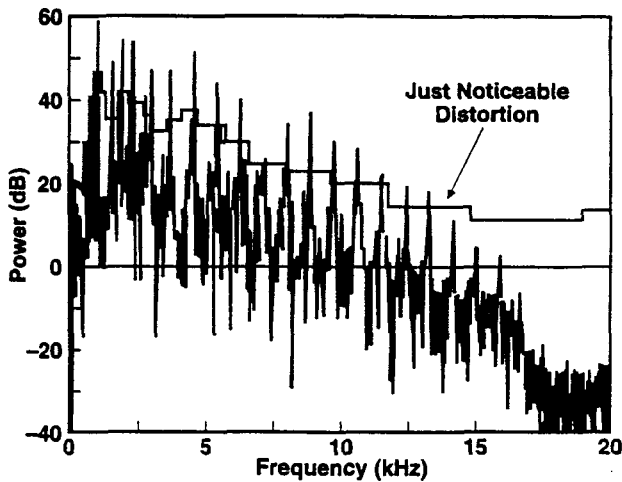


Fig. 1. Perceptual audio coding (after Ref. 9).

An important trend in very low bit rate coding is the use of object-modeling which utilizes signal structure as well as perceptual considerations. For example, in the coding of a head-and-shoulders scene, the face-object can be specially processed for better rendition, at the expense of a slight degradation of performance for the (perceptually less important) background scene. This can be done either by implicit face-processing such as a scene-adaptive vector quantizer which adapts its codebook to the dominant or sustained scene-object,¹¹ or by an explicit algorithm which models the face and head by convenient shapes such as an ellipse and a rectangle (Fig. 2), and diverts bit rate resources to the interior of the ellipse and/or rectangle.¹²

2.6. *State-of-the-art*

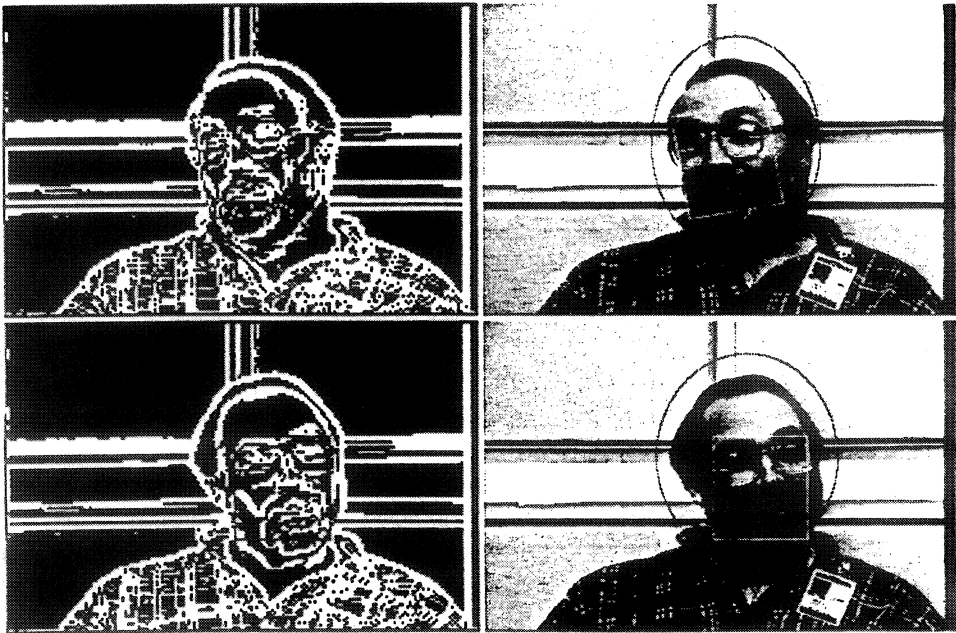
Figures 3 and 4 describe current capabilities in signal compression.

Figure 3 shows a range of applications that are supported by current capabilities in signal compression. At bit rates below 10 kbps applications such as secure voice, cellular radio, voice mail, and imagephone are practical. Between 10 and 20 kbps applications arise in network telephony and audio conferencing. Between 20 and 100 kbps several audiovisual applications emerge, including slide show graphics, internet video and voice, and music preview and broadcasting. Video conferencing becomes interesting at rates between 100 and 500 kbps, movies on demand at about 1 Mbps, and HDTV at about 20 Mbps.

Figure 3 has shown the bit rates (in kbps or Mbps) that support generic classes of current applications. Good communication quality (score of at least 3.5) is implied in many of these applications. Network-quality speech typically implies a score of at least 4.0 while broadcast-quality audio (and video) ideally requires scores in the



Automatically detected eyes-nose-mouth locations in sequences "jelena," and "roberto."



Automatically detected eyes-nose-mouth locations in sequence "jim."

Fig. 2. Face-tracking for object-based video coding (after Ref. 12).

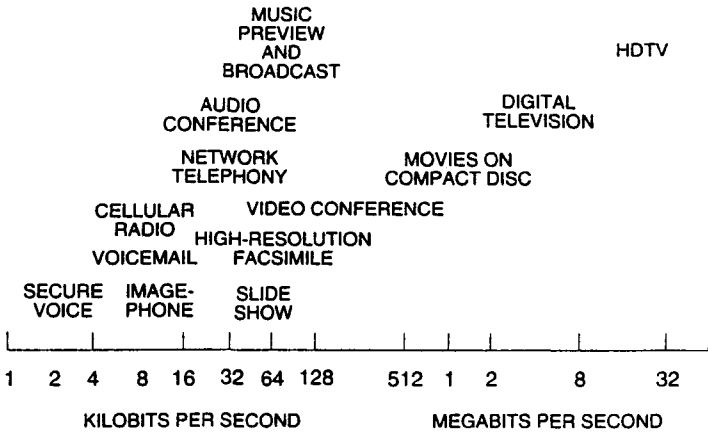


Fig. 3. Applications of signal compression (after Ref. 8).

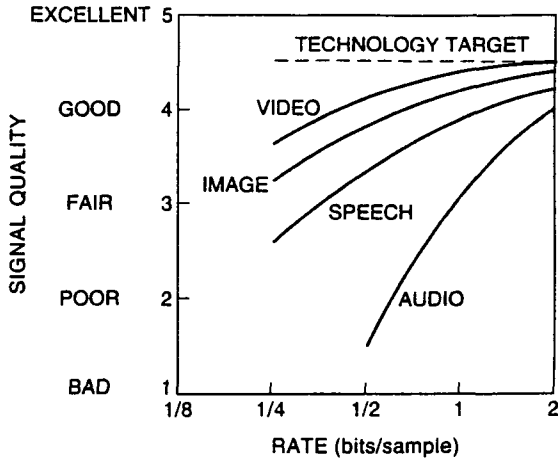


Fig. 4. Current capabilities in the coding of audiovisual signals (after Ref. 8).

neighborhood of 4.5. Figure 4 describes the normalized bit rates (in bits per sample) that support various levels of signal quality on a 5-point scale. These numbers can be converted to absolute rates (in kbps or Mbps) by multiplying the normalized bit rate by the sampling rates in Table 1. The results of Fig. 4 are also quite conservative, and the four characteristics in the figure are expected to rise toward the ideal technology-target as compression techniques become more sophisticated.

Figures 3 and 4 should be regarded as approximate, rather than rigorous, quantifications of coder performance. The horizontal line in Fig. 4 describes an obvious research target for signal coding. The rate of our progress in reaching this target is in general dependent on the input signal. It is also important to note that there are fundamental limits in signal coding, determined by the *perceptual entropy*, the

lowest possible bit rate for a given level of signal quality. The perceptual entropy is a function of input signal and the human perceptual mechanism.⁹

Figure 5 extends the technology perspective by including the complexity dimension. Included in the figure are several illustrative examples of multimedia compression. Also shown in the figure are the approximate domains of general-purpose and application-specific signal processing, as well as a few contrasting applications from the field of speech recognition.

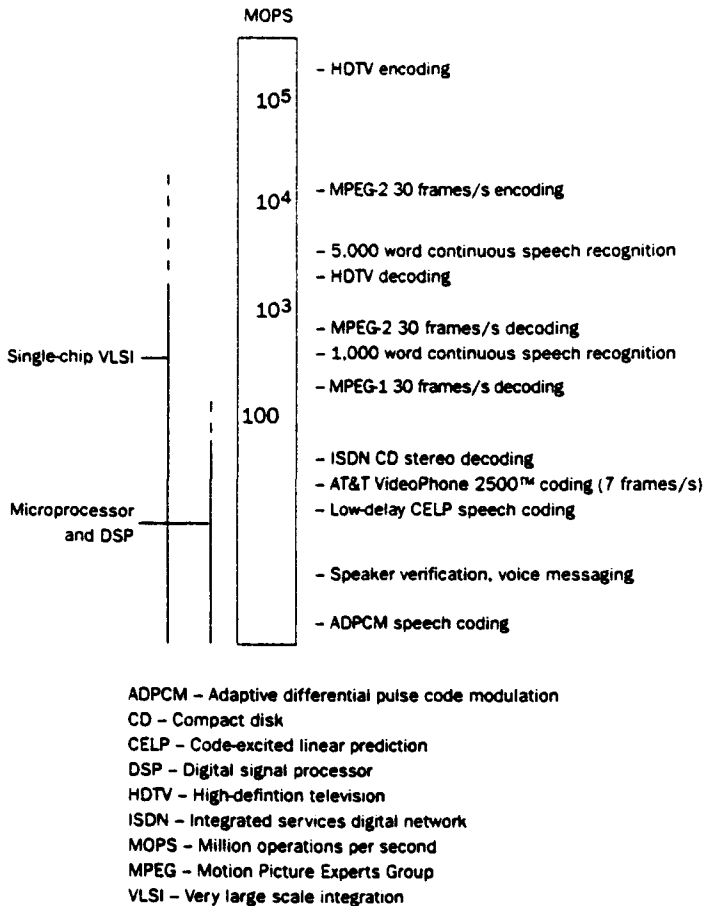


Fig. 5. The complexity of implementation needed for several examples of signal coding (after Ref. 10).

2.7. Coding standards

Table 2 summarizes a range of standards for the coding of audiovisual signals. The first group of standards is for telephone bandwidth speech with applications to network telephony, cellular communications, and secure voice. The second group

Table 2. Standards for speech, audio and image coding (after Refs. 1 and 10).

Standards body	Standard	Year	Algorithm	Bit rate	Application
			Mu-Law and A-Law PCM		
CCITT	G.711	1972		64 kbps	Network telephony
CCITT	G.721	1984	ADPCM	32 kbps	Network telephony
CCITT	G.723	1988	ADPCM	24, 40 kbps	Undersea cable
CCITT	G.726, 7	1990	ADPCM	16, 24, 32, 40 kbps	Undersea cable
ITU-T	G.728	1992	LD-CELP	16 kbps	Network telephony
ETSI	GSM	1988, 1994	RPLPC, VSELP	13.2, 5.6 kbps	Cellular telephony
TIA	IS-54	1989	VSELP	8 kbps	Cellular telephony
TIA	IS-96	1993	QCELP	0.8, 2, 4, 8.5 kbps	Cellular telephony
JDC (Japan)		1989, 1992	VSELP	8, 4 kbps	Cellular telephony
ETSI	GSM	1994	VSELP	5.6 kbps	Cellular telephony
NSA	FS1016	1989	CELP	4.8 kbps	Secure voice
NSA	FS1015	1975	LPC 10E	2.4 kbps	Secure voice
CCITT	G.722	1984	Subband-ADPCM	32–64 kbps	Teleconferencing
ISO	MPEG-1	1992	Musicam/ASPEC	128–384 kbps	Audio storage (Stereo)
ISO	MPEG-2	1996	—	320, 384 kbps	Audio storage (Five channel)
			Run		
ISO	JBIG	1991	length coding	0.05–0.10 bpp	Binary images
ISO	JPEG	1991	DCT	0.25–8.0 bpp	Still images
ISO	MPEG1, 2	1991, 1994	MC-DCT	1–8 Mbps	Addressable video
CCITT	Px64	1991	MC-DCT	64–1536 kbps	Videoconferencing
FCC	HDTV	1995	MC-DCT	20 Mbps	Advanced TV

includes a standard for wideband telephony and an audio standard. The last group includes standards for both still images and video at several rates.

Standardized algorithms lag behind current research capabilities by definition and due to the very nature of the standardization process. New, and sometimes proprietary advances, are often used to enhance a standard algorithm by incorporating the advancement in the form of a preprocessing or encoding process, while inter-operating with the standard decoder and bit stream syntax. Examples are the use of PAC-like psychoacoustics in an MPEG audio standard, and the use of face-tracking as a preprocessor in a CCITT or MPEG video standard.

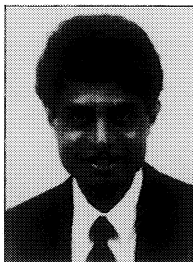
3. Trends in Signal Compression

As the articles that follow will attest, the field of signal compression is both intensely active and considerably mature. There is significant room however for future work and we mention below some of the more important directions with particular reference to audiovisual signals:

- Continued understanding of the characteristics of signals and of human perception, in an attempt to quantify and reach the fundamental limits in coding
- Applied research that identifies interesting points in the 4-D space of quality, bit rate, delay and complexity
- Advanced co-designs of source coding algorithms with techniques for channel coding, and finally,
- Collaborative designs of audiovisual coding with synthesis and recognition technologies for multimedia products and services.

References

1. R. V. Cox, "Current methods of speech coding", this volume.
2. P. Noll and D. Pan, "ISO-MPEG audio coding", this volume.
3. C. Podilchuk and R. J. Safranek, "Image and video compression: A review", this volume.
4. A. Moffat, T. C. Bell, and I. H. Witten, "Lossless compression for text and images", this volume.
5. N. S. Jayant and P. Noll, *Digital Coding of Waveforms*, Prentice Hall, 1984.
6. A. N. Netravali and B. Haskell, *Digital Images*, Plenum Press, 1987.
7. P. E. Papamichalis, *Practical Approaches to Speech Coding*, Prentice Hall, 1987.
8. N. S. Jayant, "Signal compression: Technology targets and research directions", *IEEE J. Selected Areas in Communications* (1992) 796–818.
9. N. S. Jayant, J. D. Johnston, and R. S. Safranek, "Signal compression based on models of human perception", *Proc. IEEE*, Oct. 1993, pp. 1385–1422.
10. N. S. Jayant, V. B. Lawrence, B. Ackland, and L. R. Rabiner, "Technological dimensions of multimedia", *AT&T J.*, Sept. 1995, pp. 14–33.
11. D. Wang and J. Hartung, "Codebook adaptation algorithm for a scene-adaptive video coder", *Proc. ICASSP '95*, Detroit, May 1995.
12. A. Eleftheriadis and A. Jacquin, "Automatic face location detection for model-assisted rate control in H.261-compatible coding of video", in *Image Communication* 7 (1995) 435–455.



Dr. Nikil Jayant is Director of the Multimedia Communications Research Laboratory at Bell Labs., the R&D arm of Lucent Technologies (formerly the Systems and Technology part of AT&T). In this position, Dr. Jayant is responsible for the creation and commercialization of technologies for audiovisual communication and multimedia information systems.

Earlier at AT&T Bell Laboratories, Dr. Jayant created and managed the Signal Processing Research Department and the Advanced Audio Technology Department. Dr. Jayant's personal research has been in the field of digital coding and transmission of information signals. He has published over a hundred papers and several books, and has been granted

twenty patents.

Dr. Jayant received his Ph.D. in electrical communication engineering from the Indian Institute of Science, Bangalore, India. He was a research associate at Stanford University for one year prior to joining Bell Labs, Murray Hill, in 1968. Dr. Jayant has received several honors, including the IEEE Browder J. Thompson Memorial Prize Award and the IEEE Donald G. Fink Prize Paper Award. Dr. Jayant is a Fellow of the IEEE and a member of the National Academy of Engineering.