Chapter 2. Advanced Telecommunications and Signal Processing Program

Academic and Research Staff

Professor Jae S. Lim

Visiting Scientists and Research Affiliates

Dr. Hae-Mook Jung

Graduate Students

John G. Apostolopoulos, Shiufun Cheung, Ibrahim A. Hajjahmad, Kyle K. Iwai, Peter A. Monta, Aradhana Narula, Julien J. Nicolas, Alexsander Pfajfer, Lon E. Sunshine, Chang Dong Yoo

Technical and Support Staff

Cindy LeBlanc, Debra L. Harring, Denise M. Rossetti

2.1 Introduction

The present television system was designed nearly 40 years ago. Since then, there have been significant developments in technology, which are highly relevant to the television industries. For example, advances in the very large scale integration (VLSI) technology and signal processing theories make it feasible to incorporate frame-store memory and sophisticated signal processing capabilities in a television receiver at a reasonable cost. To exploit this new technology in developing future television systems, Japan and Europe established large laboratories, funded by government or industry-wide consortia. The lack of this type of organization in the U.S. was considered detrimental to the broadcasting and equipment manufacturing industries, and in 1983 the Advanced Television Research Program (ATRP) was established at MIT by a consortium of U.S. Companies.

The major objectives of the ATRP are:

- To develop the theoretical and empirical basis for the improvement of existing television systems, as well as the design of future television systems.
- To educate students through television-related research and development and to motivate them to undertake careers in television-related industries.
- To facilitate continuing education of scientists and engineers already working in the industry.
- To establish a resource center to which problems and proposals can be brought for discussion and detailed study.
- To transfer the technology developed from this program to the industries.

The research areas of the program include the design of receiver-compatible ATV system and digital ATV system and development of transcoding methods. Significant advances have already been made in some of these research areas. A digital ATV system has been designed and tested in the fall of 1992 by the FCC for its possible adoption as the U.S. HDTV standard for terrestrial broadcasting. Some elements of this system are likely to be included in the U.S. HDTV standard.

In addition to research on advanced television systems, the research program also includes research on speech processing. Current research topics include development of a new speech model and development of algorithms to enhance speech degraded by background noise.

2.2 ATRP Facilities

The ATRP facilities are currently based on a network of eight Sun-4 and three DecStation 5000 workstations. There is approximately 14.0 GB of disk space distributed among the various machines. Attached to one of the Sun-4s is a VTE display system with 256 MB of RAM. This display system is capable of driving the Sun-4 monitors or a 29-inch Conrac monitor in the lab at rates up to 60 frames/sec. In addition to displaying high-resolution real-time sequences, the ATRP facilities include a Metheus frame buffer which drives a Sony 2k×2k monitor. For hard copy output, the lab uses a Kodak XL7700 thermal imaging printer which can

produce 2k×2k color or black and white images on 11"×11" photographic paper.

Other peripherals include an Exabyte 8 mm tape drive, a 16-bit digital audio interface with two channels and sampling rates up to 48 kHz per channel, and an "audio workstation" with power amplifier, speakers, CD player, tape deck, etc. Additionally, the lab has a 650 MB optical disk drive, a CD-ROM drive, and two laser printers. For preparing presentations, the ATRP facilities also include a Macintosh SE30 microcomputer, a Mac IIx, and an Apple LaserWriter.

Obtaining a fast network (FDDI) is under consideration to augment the current 10 Mbps Ethernet. The new network would enable much faster data transfer to display devices, and it would support large NFS transfers more easily.

2.3 Very-low-bit-rate Video Representations

Sponsor

Advanced Telecommunications Research Program

Project Staff

John G. Apostolopoulos

Video plays an important role in many of today's applications, and it is expected to gain even greater importance in the near future with the advent of multimedia and personal communication devices. The large raw data rate of a video signal, together with the limited available transmission capacity in many applications, necessitates compression of the video signal. A number of video compression algorithms have been developed for different applications from video-phone to high-definition television. These algorithms perform reasonably well at respective bit rates of 64 kb/s to tens of Mb/s. However, many applications in the near future, particularly those associated with portable or wireless devices, will most likely be required to operate at considerably lower bit rates, possibly as low as 10 kb/s. The video compression methodologies developed thus far are not applicable at such low bit rates. The goal of this research is to create a video compression approach which can produce acceptable video quality at approximately 10 kb/s.

Conventional video compression algorithms may be described as block-based coding schemes. They partition each frame into square blocks and then independently process each block. Examples of these redundancy-reduction techniques include block-based temporal motion-compensated prediction and spatial block discrete-cosine transform. Block-based processing is the basis for virtually all video compression systems today because it is a simple and practical approach to achieve acceptable video quality at the required bit rates. However, block-based coding schemes cannot effectively represent a video signal at very low bit rates. This is because the source model used is extremely limited. Block-based schemes inherently assume a source model of (translational) moving square blocks. However, a typical video scene is not composed of translated square blocks. In effect, block-based schemes impose an artificial structure on the video signal and then try to encode this structure, instead of recognizing the structure inherent to a particular video scene and attempting to exploit it.

In this research a conceptual framework is developed for identifying the structure that exists within a video scene. This is one of the most important perceptual aspects of the information within the scene. Therefore, by identifying and efficiently representing this structure, it may be possible to produce acceptable video quality at very low bit rates. Since real scenes contain objects, a promising source model includes two- or three-dimensional moving objects. A sizable amount of current research involves image segmentation, however, most of this research has focused primarily on image analysis and extremely little on the possibility of compression. Very importantly, significant statistical dependencies exist in the image regions belonging to each object, and this must be exploited. This point has not been addressed; only the previously mentioned block-based schemes have been discussed in the literature.

An efficient compression algorithm must therefore consist of a powerful and compact framework for accurately representing possible objects or structure in a video signal, as well as an efficient approach to represent the information that exists within each object.

The goal of this research is to create a video compression approach which can produce acceptable video quality at approximately 10 kb/s. This will allow video applications in the near future, particularly those associated with portable or wireless devices, to operate at performance levels not achievable by current video compression schemes. In order to attain this goal, we are developing a conceptual and mathematical framework for identifying and exploiting the structure that exists within a video signal.

2.3.1 Publications

Apostolopoulos, J.G., and J.S. Lim. "Video Compression for Digital Advanced Television Systems." In *Motion Analysis and Image Sequence Processing.* Chapter 15. Eds. M. Sezan, and R. Lagendijk. Dordrecht, the Netherlands: Kluwer Academic Publishers, 1993.

Apostolopoulos, J.G., P.A. Monta, J.J. Nicolas, and J.S. Lim. "Designing a Video Compression System for High Definition Television." Invited paper for special video session within ICASSP, 1993.

2.4 Audio Compression Using Hierarchical Nonuniform Filterbanks

Sponsor

Advanced Telecommunications Research Program

Project Staff

Shiufun Cheung, Peter A. Monta

Interest in coding high-fidelity audio has increased in recent years; audio source coding algorithms are finding use in transmission applications such as digital audio broadcast (DAB) and high-definition television (HDTV) and in commercial products such as MiniDisc and Digital Compact Cassette. The common objective is to achieve high quality at a rate significantly smaller than the 16 bits/sample used in CD and DAT systems. We have been considering applications to HDTV, and an earlier implementation, the MIT Audio Coder (MIT-AC) has successfully completed testing at the Advanced Television Test Center. In this research, we study an improved signal representation scheme which uses a hierarchical nonuniform filterbank.

In many waveform coders, the first step in the audio coding process is a short-time spectral decomposition of the signal. Until recently, the representations used by many audio coders have uniform filterbank bandwidths. This results in an undesirable tradeoff between time and frequency resolution. For example, if the analysis bandwidth is chosen to be narrow enough to resolve the critical bands in the low frequencies, the resulting poor temporal resolution can result in temporal artifacts, such as the "pre-echo" effect.

There have been various efforts to correct this deficiency. For example, some coders include an adaptive mechanism for improving temporal resolution by adjusting the block transform size when a transient is detected. Another more fundamental approach is the use of a nonuniform filterbank, which can simultaneous satisfy the requirements of good pre-echo control and frequency resolution consistent with critical-band analysis. It also allows a more uniform architecture without an explicit adaptation step.

Previous development along these lines has considered the use of wavelet representations. In our prototype implementation, we use a hierarchical filterbank structure based on M-band perfect-reconstruction cosine-modulated filterbanks (for various values of M). These filterbanks have been developed by Malvar[1] and in an equivalent formulation by Vaidyanathan.[2]

Testing of the new representation has shown improvement over an uniform filterbank scheme. While pre-distortion still exists in transient signals, listening sessions show that pre-echo is inaudible with the new filterbank.

One potential disadvantage of a fixed filterbank containing some wide subbands is that the system may not realize the theoretical coding gain if the signal spectrum is very peaked. Further research is now being performed to devise schemes which can adaptively subdivide the filterbank tree based on spectral flatness measurements.

2.5 Transform Coding for High-Definition Television

Sponsor

Advanced Telecommunications Research Program

Project Staff

Ibrahim A. Hajjahmad

The field of image coding is useful for many areas. For of these areas is the reduction of channel bandwidth needed for image transmission systems, such

---

1 H.S. Malvar, "Extended Lapped Transforms: Properties, Applications and Fast Algorithms," *IEEE Trans. Signal Proc.* 40: 2703-2714 (1992).

2 P.P. Vaidyanathan, *Multirate Systems and Filter Banks*, (Englewood Cliffs, New Jersey: Prentice Hall, 1993).

as HDTV, video conferencing, and facsimile. Another area is the reduction of storage requirements. One class of image coders is known as the transform image coder.[3] In transform image coding, an image is transformed to another domain, more suitable for coding than the spatial domain. The obtained transform coefficients are quantized and then coded. At the receiver, the coded coefficients are decoded and then inversely transformed to obtain the reconstructed image.

One transform which has shown promising results is the discrete cosine transform (DCT).[4] The DCT is a real transform with two important properties that make it very useful in image coding. One is the energy compaction property, where large amount of energy is concentrated in a small fraction of the transform coefficients (typically low frequency components). This property allows us to code a small fraction of the transform coefficients with a small sacrifice in quality and intelligibility of the coded images. Another is the correlation reduction property. In the spatial domain there is a high correlation among image pixel intensities. The DCT reduces this correlation and redundant information does not need to be coded.

Current research is investigating the use of the DCT for bandwidth compression. New adaptive techniques are also being studied for quantization and bit allocation that can further reduce the bit rate without reducing image quality and intelligibility.

2.6 Pre-Echo Detection and Reduction

Sponsor

Advanced Telecommunications Research Program

Project Staff

Kyle K. Iwai

In recent years, there has been an increasing interest in data compression for storage and data transmission. In the field of audio processing, various kinds of transform coders have successfully demonstrated reduced bit rates while maintaining high audio quality. However, there are certain coding artifacts which are associated with transform coding. The pre-echo is one such artifact. Pre-echos typically occur when a sharp attack is preceded by silence. Quantization noise added by the coding process is normally hidden within the signal. However, the coder assumes stationarity over the window length, an assumption which breaks down in a transient situation. The noise is unmasked in the silence preceding the attack, creating an audible artifact called a pre-echo.

If the length of the noise can be shortened to about 5 ms, psycho-acoustic experiments tell us that the noise will not be audible. Using a shorter window length shortens the length of the pre-echo. However, shorter windows also have lower quality frequency selectivity and lower quality coder efficiency. One solution is to use shorter windows only when there is a quiet region followed by a sharp attack.

In order to use adaptive window length selection, a detector had to be designed. A simple detector was implemented which compares the variance within two adjacent sections of audio. In a transient situation, the variance suddenly increases from one section to the next. The coder then uses short windows to reduce the length of the pre-echo, rendering the artifact inaudible.

2.7 Video Source Coding for High-Definition Television

Sponsor

Advanced Telecommunications Research Program

Project Staff

Peter A. Monta

Efficient source coding is the enabling technology for high-definition television over the relatively narrow channels envisioned for the new service (e.g., terrestrial broadcast and cable). Coding rates are on the order of 0.3 bits/sample, and high quality is a requirement. This work focuses on new source coding techniques for video relating to representation of motion-compensated prediction errors, quantization and entropy coding, and other system issues.

Conventional coders represent video with the use of block transforms with small support (typically 8x8 pixels). Such independent blocks result in a simple

---

[3] J.S. Lim, *Two-Dimensional Signal and Image Processing,* (Englewood Cliffs, New Jersey: Prentice Hall, 1990); R.J. Clarke, *Transform Coding of Images,* (London, England: Academic Press, 1985).

[4] N. Ahmed, T. Natarajan, and K.R. Rao, "Discrete Cosine Transform," *IEEE Trans. Comput.* C-23: 90-93 (1974).

scheme for switching a predictor from a motion-compensated block to a purely spatial block; this is necessary to prevent the coder from wasting capacity in some situations.

Subband coders of the multiresolution or wavelet type, with their more desirable localization properties, lack of "blocking" artifacts, and better match to motion-compensated prediction errors, complicate this process of switching predictors, since the blocks now overlap. A novel predictive coding scheme is proposed in which subband coders can combine the benefits of good representation and flexible adaptive prediction.

Source-adaptive coding is a way for HDTV systems to support a more general imaging model than conventional television. With a source coder that can adapt to different spatial resolutions, frame rates, and coding rates, the system may then make tradeoffs among the various imagery types, for example, 60 frames/s video, 24 frames/s film, highly detailed still images, etc. In general, this is an effort to make HDTV more of an image transport system rather than a least-common-denominator format to which all sources must either adhere or be hacked to fit. These techniques are also applicable to NTSC to some extent; one result is an algorithm for improved chrominance separation for the case of "3-2" NTSC, that is, NTSC upsampled from film.

2.8 Error Concealment for an All-Digital HDTV System

Sponsor

Advanced Telecommunications Research Program

Project Staff

Aradhana Narula

Broadcasting high-definition television (HDTV) requires the transmission of an enormous amount of information within a highly restricted bandwidth channel. Adhering to the channel constraints necessitates the use of an efficient coding scheme to compress the data. Compressing the data dramatically increases the effect of channel errors. In the uncompressed video representation, a single channel error affects only one pixel in the received image. In the compressed format, a channel error affects a block of pixels in the reconstructed image, perhaps even the entire frame.

One way to combat the effect of channel errors is to add well-structured redundancy to the data through channel coding. Error correction schemes generally, however, require transmitting a significant number of additional bits. For a visual product like HDTV, it may not be necessary to correct all errors. Instead, removing the subjective effects of channel errors using error concealment techniques may be sufficient and require fewer additional bits for implementation. Error concealment may also be used in conjunction with error correction coding. For example, it may be used to conceal errors which the error correction codes are able to detect but not correct.

Error concealment techniques take advantage of the inherent spatial and temporal redundancy within the transmitted data to remove the subjective effects of these errors once the location of the errors has been detected. In this research, error concealment techniques were developed and analyzed to help protect the system from errors occurring in several parameters transmitted for HDTV images. Specifically, error concealment for errors in the motion vectors and discrete cosine transform (DCT) coefficients were investigated.

2.9 Transmission of HDTV Signals in a Terrestrial Broadcast Environment

Sponsor

Advanced Telecommunications Research Program

Project Staff

Julien J. Nicolas

High-definition television systems currently being developed for broadcast applications require 15-20 Mbps to yield good quality images for roughly twice the horizontal and vertical resolutions of the current NTSC standard. Efficient transmission techniques must be found in order to deliver this signal to a maximum number of receivers while respecting the limitations stipulated by the FCC for over-the-air transmission. This research focuses on the principles that should guide the design of such transmission systems.

The major constraints related to the transmission of broadcast HDTV include (1) a bandwidth limitation (6 MHz, identical to NTSC), (2) a requirement for simultaneous transmission of both NTSC and HDTV signals on two different channels (Simulcast approach), and (3) a tight control of the interference effects between NTSC and HDTV, particularly when the signals are sharing the same frequency bands. Other considerations include complexity and cost

issues of the receivers, degradation of the signal as a function of range, etc.

A number of ideas are currently under study. Most systems proposed to date use some form of forward error-correction to combat channel noise and interference from other signals. The overhead data reserved for error-correction schemes represents up to 30 percent of the total data, and it is therefore well worth trying to optimize these schemes. Current work is focusing on the use of combined modulation/coding schemes capable of exploiting the specific features of the broadcast channel and the interference signals. Other areas of interest include use of combined source/channel coding schemes for HDTV applications and multi-resolution coded modulation schemes.

2.10 Position-Dependent Encoding

Sponsor

Advanced Telecommunications Research Program

Project Staff

Alexsander Pfajfer

In typical video compression algorithms, the DCT is applied to the video, and the resulting DCT coefficients are quantized and encoded for transmission and storage. Some of the DCT coefficients are set to zero. Efficient encoding of the DCT coefficients is usually achieved by encoding the location and amplitude of the non-zero coefficients. In typical MC-DCT compression algorithms up to 90 percent of the available bit rate is used to encode the location and amplitude of the non-zero quantized DCT coefficients. Therefore, efficient encoding of the location and amplitude information is extremely important for high quality compression.

Position-dependent encoding, a novel approach to encoding of the location and amplitude information, is being examined. Position-dependent runlength encoding and position-dependent encoding of the amplitudes attempts to exploit the inherent differences in statistical properties of the runlengths and amplitudes as a function of their position. This novel method is being compared to the classical, separate, single-codebook encoding of the runlength and amplitude, as well as to the joint runlength and amplitude encoding.

2.11 HDTV Transmission Format Conversion and the HDTV Migration Path

Sponsor

Advanced Telecommunications Research Program

Project Staff

Lon E.Sunshine

The current proposal for terrestrial HDTV broadcasting allows for several possible transmission formats. Because production and display formats may differ, it will be necessary to convert between formats in an effective way. A key to this process is the de-interlacing process. Since HDTV will presumably move toward progressive display systems, it will be necessary to de-interlace non-progressive source material. The research will consider topics relating to conversion among the six formats being proposed for the U.S. HDTV standard.

As HDTV evolves, it is probable that more transmission formats will be allowed. Furthermore, additional bandwidth may be allocated for some channels (terrestrial and/or cable). This research will consider the issues related to the migration of HDTV to higher resolutions. Backward compatibility and image compression and coding issues will be addressed.

2.12 Speech Enhancement

Project Staff

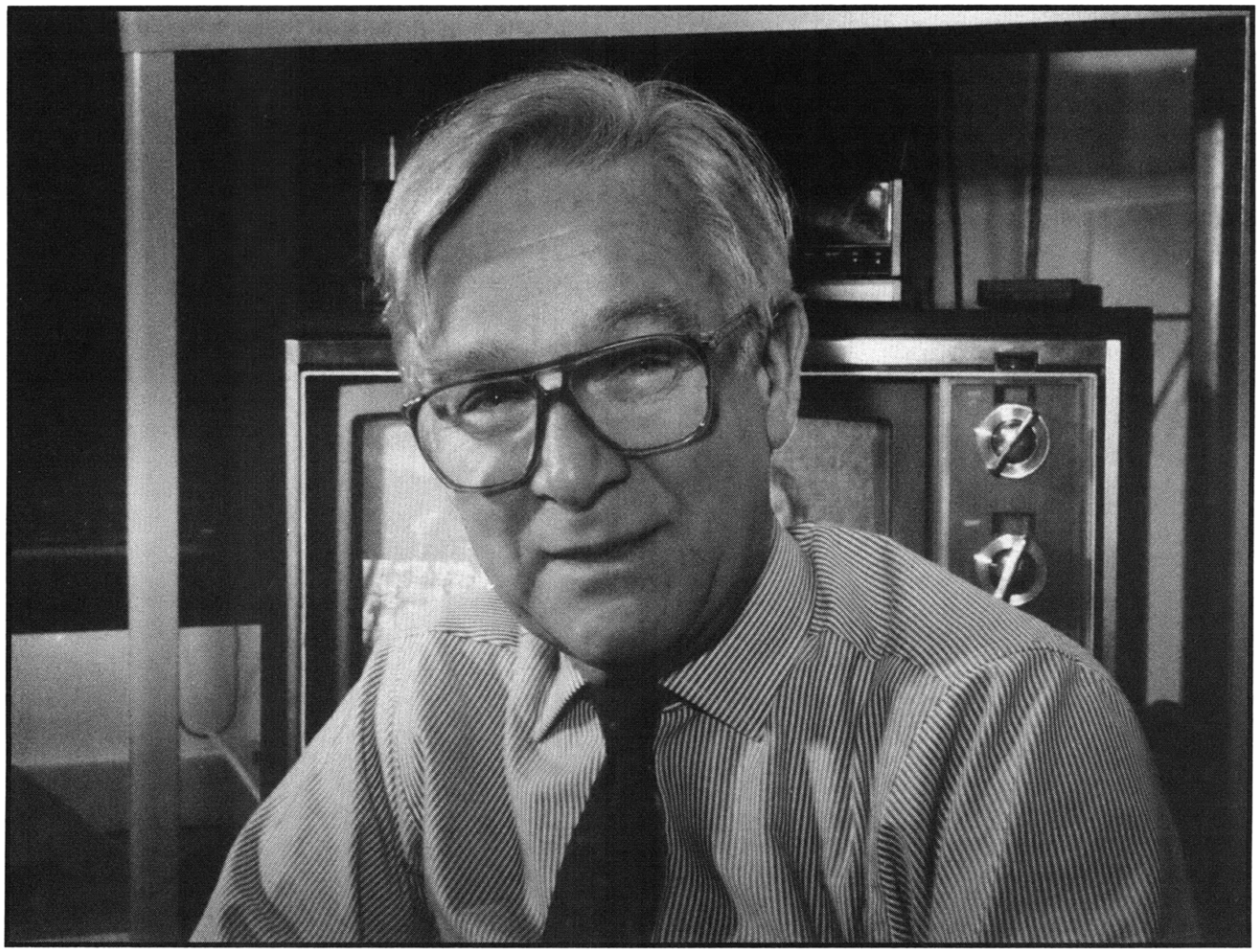
Chang Dong Yoo

The development of the dual excitation (DE) speech model has led to some interesting insights into the problem of speech enhancement, and based on the ideas of the DE model, a new speech model is being developed. The DE model provides more flexible representation of speech and possesses features which are particularly useful to the problem of speech enhancement. These features, along with a variable length window, are the backbone of the new speech model being developed.

Because the DE model does not place any restrictions on its characterization of speech, the enhancement system based on the DE model performs better than the one based on any of the previous speech models. While a model should be inclusive in its characterization, it should have some restrictions. Specifically, a speech model should pertain to speech. The DE model is somewhat

unrestrictive and simple in its characterization of speech. It is solely based on the separation of the voiced and unvoiced components. Whether it makes sense to represent a stop as a voiced and an unvoiced component is just one of many interesting issues which are being investigated. An extension of the DE model which deals with these issues better is currently being studied.

All model-based enhancement methods to date have been formulated on the premise that each segment of speech be stationary for a fixed window length. To improve the performance of the enhancement algorithm, this assumption of stationarity must be assured. To do this, a variable-length window should be used to capture varying durations of stationarity in the speech. There are several algorithms which adaptively detect changes in auto-regressive model parameters in quasi-stationary signals which have been successfully used in speech recognition. We propose to investigate some of these algorithms. The benefit from using a variable length window is two-fold: (1) It will allow better and "cleaner" separation of the voiced and unvoiced components; and (2) It will allow for a greater reduction in the number of characteristic parameters, such as the amplitudes of the voiced components and the LP coefficients of the unvoiced component.



Professor Emeritus William F. Schreiber