

## Chapter 2. Advanced Telecommunications and Signal Processing Program

### Academic and Research Staff

Professor Jae S. Lim

### Visiting Scientists and Research Affiliates

M. Carlos Kennedy

### Graduate Students

John G. Apostolopoulos, David M. Baylon, Shiufun Cheung, Raynard O. Hinds, Eric C. Reed, Lon E. Sunshine, Chang Dong Yoo

### Technical and Support Staff

Cindy LeBlanc, Christopher R. O'Donnell, Denise M. Rossetti

## 2.1 Introduction

The present television system was designed nearly 50 years ago. Since then, there have been significant developments in technology which are highly relevant to the television industries. For example, advances in very-large-scale-integration (VLSI) technology and signal processing theories make it feasible to incorporate frame-store memory and sophisticated signal processing capabilities into 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 a consortium of U.S. companies established the Advanced Television Research Program (ATRP) at MIT.

The major objectives of ATRP are:

1. To develop the theoretical and empirical basis for the improvement of existing television systems, as well as the design of future television systems.
2. To educate students through television-related research and development, and to motivate them to undertake careers in television-related industries.

3. To facilitate the continuing education of scientists and engineers already working in the industry.
4. To establish a resource center to which problems and proposals can be brought for discussion and detailed study.
5. To transfer technology developed from this program to the industries.

In the past, the program's research areas have focused on a number of issues related to digital television design. As a result of this effort, significant advances have been made, and these advances have been included in the U.S. HDTV standard. Specifically, the ATSP group represented MIT in the Grand Alliance, which consisted of MIT, AT&T, Zenith Electronics Corporation, General Instrument Corporation, David Sarnoff Research Center, Philips Laboratories, and Thomson Consumer Electronics. The Grand Alliance digital television system served as the basis for the U.S. digital television (DTV) standard, which was formally adopted by the U.S. Federal Communications Commission.

In addition to research on issues related to the design of digital television systems, our program also includes research on signal processing for telecommunications applications and research on speech enhancement.

### 2.1.1 Patents

Lim, J.S., and P.A. Monta. "Digital Advanced Television Systems." Patent #5,485,210, January 1996.

Lim, J.S. "Advanced Television System." Patent #5,508,746. April 1996.

## 2.2 Signal Processing for Signals with Arbitrarily Shaped Regions of Support

### Sponsors

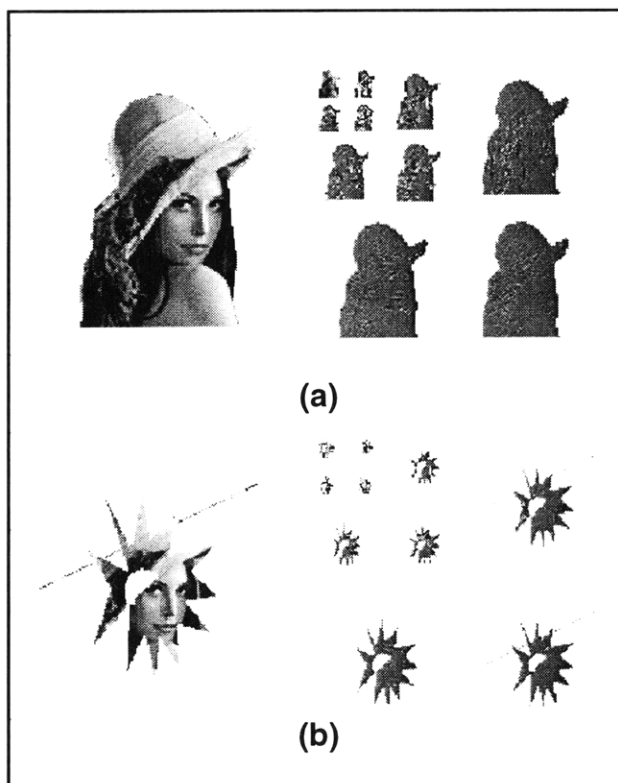
Advanced Telecommunications Research Program  
AT&T Fellowship

### Project Staff

John G. Apostolopoulos

Many applications today and in the near future will entail efficient processing of multidimensional ( $M$ -D) signals with arbitrarily shaped (non-rectangular) regions of support. For example, most high-level representations of images or video incorporate 2-D or 3-D models which decompose the scene into arbitrarily shaped objects or regions. Medical imaging often results in 2-D or 3-D imagery where the relevant information is localized over an arbitrarily shaped region. Furthermore, many areas of scientific research involve problems defined over arbitrarily shaped domains. High quality and computationally efficient processing of these "arbitrarily shaped signals" is important for the success of these applications.

Transform/subband representations form a basic building block for many signal processing algorithms and applications. Most of the effort has focused on developing representations for infinite-length signals, with simple extensions to finite-length 1-D and rectangular support 2-D signals. However, the important problem of representing 2-D or general  $M$ -D signals, with arbitrarily shaped regions of support, has received relatively little attention. We have developed a novel framework (see figure 1) for creating critically sampled perfect reconstruction transform/subband representations for discrete 1-D, 2-D, and general  $M$ -D signals defined over arbitrarily shaped regions of support. Our method selects an appropriate subset of vectors from a (easily obtained) basis for a larger (superset) signal space, in order to form a basis for the arbitrarily shaped signal. In particular, we have developed a number of promising wavelet representations for arbitrary-length 1-D signals and arbitrarily



**Figure 1.** Our group has developed a novel, critically sampled, perfect reconstruction wavelet representation for signals with arbitrary regions of support. Figure 1a illustrates a 2-D object from an image with an arbitrarily shaped region of support and its two-level wavelet representation using our novel approach. Figure 1b illustrates a 2-D signal with a much more complex region of support and its two-level wavelet representation. To aid in interpretation, the lowpass subbands have been scaled, and the higher-frequency subbands have been offset so that the zero amplitude corresponds to gray.

shaped 2-D/ $M$ -D signals that provide high performance with low complexity.

## 2.3 Multiplexing of Variable Bit Rate Video Sources

### Sponsors

Advanced Telecommunications Research Program  
INTEL Fellowship

### Project Staff

David M. Baylon

Traditional video compression algorithms for fixed bandwidth systems typically operate at fixed targeted compression ratios. A problem with fixed bit rate systems is that video quality can vary greatly within an image and across an image sequence. In

order to achieve more constant quality video, variable bit rate coding can be performed, where bit allocations are allowed to vary more widely as the video statistics change.

Although it may be inefficient to transmit a single variable rate encoded source over a constant rate channel, it becomes more appropriate if several variable rate sources are combined to share a given fixed bandwidth. In source multiplexing, bit rate fluctuations of each source are effectively averaged, making the aggregate bitstream more suitable for the constant rate channel. Applications for source multiplexing include video transmission over broadband networks or cable channels, where it may be desirable to transmit several video programs simultaneously over a fixed rate channel.

This research is aimed at finding intermediate representations for variable bit rate encoded video which are suitable for real-time source multiplexing. Specifically, the focus is on efficient MPEG-2 compatible representations that provide constant quality but that can also be easily modified if the given source needs to be recompressed in order to be properly multiplexed within the fixed bandwidth. Related issues being studied include (1) determining which statistics of the compressed bitstreams can be multiplexed within a given buffer size and (2) determining how best to process the bitstreams if they cannot be directly multiplexed. Since approaches which require passing back and forth between compressed and uncompressed representations may not be appropriate for real-time operation, approaches which operate directly in the compressed domain appear promising. We are also studying how knowledge of the sequence obtained during the intermediate representation stage can aid in the multiplexing stage.

### 2.3.1 Publication

Baylon, D.M., and J.S. Lim. "Optimal Transform Coefficient Selection for Images." *SPIE Applications of Digital Image Processing 19* 2847: 469-478 (1996).

## 2.4 Biorthogonality in Lapped Transforms: A Study in Audio Compression

### Sponsor

Advanced Telecommunications Research Program

## Project Staff

Shiufun Cheung

The demand for high-quality audio in transmission systems such as digital audio broadcast (DAB) and high-definition television (HDTV), as well as commercial products such as the minidisc (MD) and the digital compact cassette (DCC), has generated considerable interest in audio compression schemes. The common objective is to achieve high quality at a rate significantly smaller than the 16 bits/sample used in current compact disc (CD) and digital audio tape (DAT) systems. We have been considering applications to HDTV; an earlier implementation, the MIT audio coder (MIT-AC), is one of the systems that was considered for inclusion in the U.S. HDTV standard. In this research, we seek to build upon our previous efforts by studying one important aspect of audio coder design: the short-time spectral decomposition.

In conventional audio coders, the short-time spectral decomposition serves to recast the audio signal in a representation that is not only amenable to perceptual modeling but also conducive to deriving transform coding gain. This decomposition is commonly achieved by a multirate filter bank, or equivalently, a lapped transform.

To improve the performance of audio compression schemes, we have formulated a biorthogonal cosine-modulated filter bank which is a generalization of Malvar's extended lapped transform (ELT). The ELT, a popular implementation of cosine-modulated filter banks, is of particular interest because it forms the major building block of signal decomposition schemes in many audio coders.

Conventional lapped transforms are designed to be orthogonal filter banks in which the analysis and synthesis filters are identical. Allowing the analysis and synthesis filters to differ leads to a biorthogonal transform which has more degrees of design freedom than its orthogonal counterpart. We have proven that the incorporation of biorthogonality into an  $M$ -channel ELT yields an increase of  $M/2$  degrees of freedom. This additional flexibility allows the design of synthesis filter banks with improved sidelobe behavior which should be beneficial to audio coder performance.

### 2.4.1 Thesis

Cheung, S. *Biorthogonality in Lapped Transform: A Study in High-Quality Audio Compression*. Ph.D. diss., Dept. of Electr. Eng. and Comput. Sci., MIT, 1996.

## 2.5 Real-Time Video on the Internet

### Sponsors

Advanced Telecommunications Research Program  
AT&T Fellowship

### Project Staff

Raynard O. Hinds

The Internet has become efficient at transmitting data over packet-switched networks. This is because data has no inherent delay constraints and can handle the delay jitter that occurs due to variable queuing delays across the network as well as the excess delay that occurs from retransmission of lost packets. This research looks at transmitting real-time video over this same network. Real-time video cannot tolerate excessive delay. Packets arriving after their scheduled playback point at the receiver are discarded. Delay jitter can be compensated for only by queuing at the receiver which increases the overall delay in transmission. However, video sequences are capable of tolerating loss. This research seeks to understand the operation of the Internet as well as how real-time video transmission can best be accomplished over these networks.

## 2.6 Very-Low-Bit-Rate Video Coding

### Sponsors

Advanced Telecommunications Research Program  
U.S. Navy - Office of Naval Research  
NDSEG Graduate Fellowship

### Project Staff

Eric C. Reed

Video compression algorithms are used in many applications to achieve transmission bit rates over a wide range. This range can be divided into three broad categories: high, medium, and low bit rates. Roughly speaking, applications utilizing 1Mb/s or greater can be considered a high bit rate while bit rates between 10-30 kb/s can be placed into the (very) low bit rate category. The characteristics of the source and the objectives of the coder are quite different for high and low bit rate video. For example, HDTV video has spatial resolutions as large as 720 X 1280 pixels per frame (ppf) with frames up to 60 f/s. In order to transmit an HDTV signal at 20 Mb/s, we need a compression factor of about 75 (from 24 to 0.34 bpp). However, at low bit rates, we think of transmitting video with resolutions

less than 176 X 144 ppf with frame rates lower than 12 f/s. Even with this much reduction of the source, compression factors of 200-800 are needed to achieve bit rates between 10-30 kb/s. This allows a video signal to be represented with approximately 0.01-0.03 bpp. In order to obtain acceptable video quality, the bits need to be used in a very efficient manner.

With the increasing popularity of wireless video and the transmission of video through the Internet, very low bit rate coders need to be improved. The goal of this research is to develop fundamentally new video coding techniques which can produce acceptable video quality at 10 kb/s. There are many issues we must consider that are not relevant in high bit rate coders. For example, overhead, while not a major issue at high bit rates, can dominate the bitstream at low bit rates leaving little space for the actual data. We must also make sure that bits are placed in the more important regions of the video. This research is in its beginning stages. Currently, we are investigating techniques that increase the compressed bits-per-pixel rate by reducing the number of actual pixels to be coded along with the associated overhead.

### 2.6.1 Thesis

Reed, E.C. *Improvement of MPEG-2 Compression by Position-Dependent Encoding*. S.M. thesis, Dept. of Electr. Eng. and Comput. Sci., MIT, 1996.

## 2.7 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 is 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. This research considers topics relating to conversion among the six formats proposed for the U.S. HDTV standard.

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

## 2.8 Speech Enhancement

### Sponsor

Maryland Procurement Office  
Contract MDA904-93-C-4180

### 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 based solely 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, the assumption of stationarity must be assured. 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 are investigating some of these algorithms. The benefits from using a variable length window are 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 LP coefficients of the unvoiced component.

### 2.8.1 Publications

Yoo, C.D. "Selective All-Pole Modeling of Degraded Speech Using M-Band Decomposition." *Proceedings of ICASSP 96*, Atlanta, Georgia, May 7-10, 1996, pp. 641-644.

### Thesis

Yoo, C.D. *Speech Enhancement: Identification and Modeling of Stationary Time-Frequency Regions*. Ph.D. diss. Dept. of Electr. Eng. and Comput. Sci., MIT, 1996.

