DigiPoints Volume 1

Leader Guide

Module 8 – Digital Compression

Summary

This module describes the techniques by which digital signals are compressed in order to make it possible to carry one or more of the digitized TV signals within the 6 MHz channels available on CATV systems.

Outcomes

Students who complete this lesson will know the basics of compression technology for both audio and video digital signals as well as become familiar with the basic standards of GIF, JPEG, and MPEG.

Objectives

Upon successful completion of the module the student should be able to:

- Describe the difference between Lossless and Lossy compression.
- Describe how video compression is different from audio compression.
- Compare the different compression methods used for pictures.
- Describe, at a high level, the process used by MPEG-2 to digitize and compress a signal.

Prerequisites

Students should have read Chapter 8 of *DigiPoints, Volume 1*. In particular review Chapter 3 on Analog-to Digital-Conversion.

Length

90 minutes

©SCTE

Materials/Preparation

- One workbook per student
- Visual Aids
- Instructor should read Chapter 2 of *DigiPoints*, Volume 1.

Supplies/Equipment

- Dry Erase Board/Markers, chalkboard or flipchart
- Masking Tape
- Pointing Tool
- Tent Cards

Audience

The intended audience will be mid- to senior-level technicians or other associates who are seeking an understanding of digital basics.

©SCTE

Module Outline

This is an introductory level module that will provide a review of the following topics:

Objectives 1
Introduction 2
Basics of Compression Technology 4
Audio Compression 6
Video Compression 8
Video Compression Standards 9
Moving Pictures Experts Group MPEG 11
MPEG-2 Three Primary Concepts 12
Summary 16
Appendix 17

Objectives

EXPLAIN

REFER TO WB 8.1

Tell students that when they have completed this module, they will be able to accomplish these objectives.

- Describe the difference between Lossless and Lossy compression.
- Describe how video compression is different from audio compression.
- Compare the different compression methods used for pictures.
- Describe, at a high level, the process used by MPEG-2 to digitize and compress a signal.

©SCTE

Introduction

ASK & DISCUSS

Where is compression found in the analog TV world?

What are the positives of digital technology?

The negatives?

What would digital compression allow?

What might be a way of compressing a video signal?

Use as intro and as a way to evaluate skill base of students

Why Digital Compression?

Answer:

◊ *Digital technology is a bandwidth hog and the use of compression allows the total amount of 1s and 0s to be reduced to a manageable quantity without degrading the quality of the video or audio.*

DISPLAY VA 8.1
DISCUSS WB 8.2

Analog Compression on NTSC

NTSC color signal is a 26 MHz bandwidth consisting of three components:

- 13.5 MHz for Luminance or brightness, also called the grayscale
- 6.75 MHz for Chrominance - twice
- Information for Red, Green, Blue
- 13.5+6.75+6.75 MHz >> 6 MHz
- Three components are defined as:
  - ◊ Y for Luminance
  - ◊ Cb for the color difference R - Y
  - ◊ Cr for the color difference B - Y

NTSC analog systems:

- Luminance varies from 0-5 MHz
- Cb & Cr each varies from 0-1.5 MHz
- All three in 5 MHz bandwidth.

EXPLAIN

If three colors, Red-Blue-Green, transmitted in full resolution, each would need 5 MHz or 15 MHz total.

- NTSC does 3:1 compression
- Possible because of actual sensitivity of the receiver, which is the human eye
- This is a type of analog compression
  - ◊ Possible because of redundancies in video signal
- Digital compression is much better
  - ◊ More robust
  - ◊ Better and more refined handling techniques
  - ◊ Negative WAS the *computer power* needed; no longer a limitation

Basics of Compression Technology

ASK

How much storage is on a computer's hard drive?

Use board.

WRITE ON BOARD

Answer:

◊ *10, 30, 120 Gigabytes*

Make point about the need for compression

Without compression technology, 39 minutes of video requires 48 GBytes of storage.

WRITE ON BOARD

- Two broad categories of compression technology
  - ◊ Lossless Compression
  - ◊ Lossy Compression

DISPLAY VA 8.2
DISCUSS WB 8.3

Lossless Compression

- Guarantees digital output is identical to digital input to the bit level.
- Needed in data communications.
  - ◊ Numerical data
  - ◊ Monies and other critical items
- Variable Length Coding (VLC) is one of several types of Lossless methods.
  - ◊ Common patterns replaced with shorter code
  - ◊ PKZIP utility is one type
  - ◊ Huffman coding is a specific type of VLC
- Run length coding is another method.
  - ◊ Repeating pattern replaced with character that identifies it.
  - ◊ Another character indicating how many repeats.

DISPLAY VA 8.3
DISCUSS WB 8.4

Lossy Compression

- Does not guarantee identical reproduction to the bit level.
- Used for applications not requiring bit level reconstruction of input.
  - ◊ Audio and Video
  - ◊ CATV
- Human sensory system not that sensitive.
  - ◊ Exactness not required
  - ◊ Lossy encoding requires about 25% or 1/4 of the bandwidth of Lossless encoding
  - ◊ Designers must balance the economic tradeoffs in distortion
  - ◊ Distortion should not be detected by viewers
- Audio Considerations
  - ◊ Human hearing range is ~300 to 20,000 Hertz.
  - ◊ Telephony needs 300-4000 Hertz.
  - ◊ Surround sound uses full bandwidth to 20,000 Hertz, but not more than needed.
- Video Considerations
  - ◊ Transmit only what CHANGES between pictures.
  - ◊ Transmit only the amount of information that can be DETECTED by the viewer.
  - ◊ Color differences and brightness levels are key.

©SCTE

Audio Compression

Two basic methods to be discussed:

- Adaptive Pulse Code Modulation (ADPCM)
- Linear Predictive Coding (LPC)

DISPLAY VA 8.4
DISCUSS WB 8.5

Adaptive Pulse Code Modulation

- Audio frequencies do not change amplitude rapidly.
  - ◊ At 8000 samples/second, adjacent samples not very different.
  - ◊ Send the difference between samples using 4 bits versus 8 bits.
- CCITT in 1985 recommended this for 32 kbps voice digitization.

DISPLAY VA 8.5
DISCUSS WB 8.6

Line Predictive Coding

- The concept this is based on is very different.
  - ◊ Imagine speech as a buzzer at the end of a tube.
  - ◊ Speech is characterized by amplitude and frequency.
  - ◊ Throat and mouth form resonances called formants.
  - ◊ LPC estimates formants and removes them from the speech signal.
  - ◊ What remains is the residue of the amplitude and frequency.
  - ◊ Removing formants is called inverse filtering.
- Formants
  - ◊ Formants are represented as difference equations.
  - ◊ The difference equations express each sample as a linear combination of all previous samples.
  - ◊ The equation is called a “linear predictor”
- Prediction Coefficients
  - ◊ Coefficients of the difference equations.
  - ◊ Estimates computed by minimizing mean square errors between the predicted and the actual signal.
- Decoder
  - ◊ Use the residue to create a source signal.
  - ◊ The formants are used to create a filter.
  - ◊ Done on frames of speech, at a rate of 30 to 50 per second.
- Federal standard 1015 describes LPC-10e
  - ◊ Intelligible speech at 2400 bps.
  - ◊ Independently can alter pitch and speed.

Use board to discuss and demonstrate formants, i.e., resonances

Video Compression

ASK STUDENTS WHAT THE FOUR STANDARDS ARE

- Four commonly used standards
  - ◊ Ask the students what the technology is behind each of the four commonly used standards.

Write on board: GIF, JPEG, MPEG, H.261

Analog-to-Digital Conversion

See Module 3; if necessary review on board

- Basic three-step process
  - ◊ Sample - Nyquist Frequency requirement.
  - ◊ Quantize - Assign infinite values to finite values.
  - ◊ Encode - Represent analog signal as a binary number, 1s and 0s.

DISPLAY VA 8.6
DISCUSS WB 8.7

Video Compared to Audio

- Audio in narrow range of 300 to 20,000 Hertz.
- Video has wider bandwidth and more variables.
  - ◊ Luminance or brightness
  - ◊ Colors
- Video images are sets of dots called pixels which have:
  - ◊ Luminance - Brightness
  - ◊ Chrominance - Color
- Density of pixels determines image quality
  - ◊ Different standards

Human Vision

- Eye's cells detecting a "picture" are called "cones" and "rods".
  - ◊ Cones detect color
  - ◊ The more sensitive Rods detect differences in gray tones
  - ◊ Important to encode and transmit more luminance information, then color information because of how the eye sees things

Video Compression Standards

DISPLAY VA 8.7
DISCUSS WB 8.8

GIF - Graphics Interchange File

- Used for still images
  - ◊ Works best on line drawings with sharp, defined edges
- Limited colors stored
  - ◊ 8 bits per pixel
  - ◊ 256 colors maximum (2^7)
- Uses compression algorithm, LWZ
  - ◊ Can compress images to 5:1 ratio
  - ◊ Depends on type of image

DISPLAY VA 8.8
DISCUSS WB 8.9

JPEG - Joint Photographic Experts Group

- Compresses full-color, natural/real-world scenes.
- Seven-step process
  - ◊ Transform image into suitable color "space."
  - ◊ Process pixels in groups.
  - ◊ Group pixels in 8x8 blocks (64 frequency components) and transform these from time to frequency domain.
  - ◊ Divide each frequency component by a Quantization coefficient and round to an integer. Treat luminance and chrominance separately.
  - ◊ Convert these integers to strings of binary digits AND use a zig-zag pattern to maximize strings of 0s.
  - ◊ Compress these strings of 0s using Huffman coding.
  - ◊ Attach a header to the resultant binary digits; the header has the compression coefficients and Huffman coding tables.
- The above is design specific.
- Because of quantizing errors, the decompressed signal will not be EXACTLY like the original signal.
- JPEG standard has extensions that provide additional compression steps:
  - ◊ Progressive Mode
  - ◊ Hierarchical Mode

DISPLAY & DISCUSS VA 8.9

DISPLAY VA 8.10

DISPLAY VA 8.11

DISPLAY VA 8.12

Use this sequence of slides to show the complexity of the process. Point out where digitization and compression take place in the process.

DISPLAY VA 8.13
DISCUSS
WB 8.10

Huffman Coding

- Look for long strings of repeated sequences.
- Represent the repeated strings by simpler strings.
- The shortest string to represent the most repeated long sequence.

Moving Pictures Experts Group MPEG

DISCUSS

- Needed because JPEG was not designed for moving pictures.
- Motion pictures are many still pictures per second.
  - ◊ Movies are like photographs
  - ◊ Video/TV are built a line at a time
  - ◊ Computer monitors look similar, but operate differently than TV pictures

To represent this, remind the students of the "cartoons" they may have made via flipping paper, or demonstrate same.

ASK ABOUT THE RECEIVER

- The sense of motion on the screen depends on how the human eye detects and interprets what is occurring.
- Digital TV as used in cable must be able to handle all, both NTSC TV and computer generated, images.

The MPEG standard does this

DISPLAY VA 8.14
DISCUSS
WB 8.11

Flavors of MPEG

- Four (4) MPEG Versions
- MPEG-1
  - ◊ 1.5 Mbps bandwidth for video
  - ◊ 250 kbps audio
  - ◊ Non-interlaced video specified
  - ◊ Optimized for CD-ROMs, not TV
- MPEG-2
  - ◊ 100 Mbps bandwidth
  - ◊ Up to five audio channels for surround sound
  - ◊ Works with interlaced video
  - ◊ It is the standard used in CATV
- MPEG-3
  - ◊ Merged into MPEG-2
- MPEG-4
  - ◊ Low bit rates first proposed for videophone, at 64 kbps
  - ◊ 10 frames per second
  - ◊ Evolving standard that may include computer-generated graphics and audio

MPEG-2 - Three Primary Concepts

DISPLAY VA 8.15
DISCUSS
WB 8.12

See Appendix: Demonstrating MPEG-2

Interframe Motion Compression

- Processes frames in groups of frames.
  - ◊ Motion simulated by presenting series of frames
  - ◊ A frame can be one television frame or one television field
- Successive frames contain redundancies.
- Three-types of frames:
  - ◊ I-frame (Intra-frame)
  - ◊ P-frame (forward Predicted frame
  - ◊ B-frame (Bi-directional predicted frame
- I-frame
  - ◊ Constructed in similar manner as JPEG files
  - ◊ Building blocks for P & B frames
- P-frame
  - ◊ This is a predicted frame
  - ◊ Coded using motion vectors and error blocks
  - ◊ Uses either former P or I frame as a reference frame
- B-frame
  - ◊ Coded in same manner as P frame
  - ◊ Motion vector uses the previous, predicted, or both frames for a reference

DISPLAY VA 8.16
DISCUSS
WB 8.13

How I, P, and B frames are used

- Consider two frames, I & P
- Each is subdivided into blocks of 16 x 16 pixels called Macroblocks.
- Consider a straight line in motion.
  - ◊ This could be a line or the edge of an object
- In the I frame, the macroblock to be used is (2,3). It is the actual picture.
- The corresponding macroblock in the P frame is (2,3)

DISPLAY VA 8.17
DISCUSS
WB 8.14

- The following I frame, which is the actual picture, is used to compare the accuracy of the P-Frame.
- The difference is the error that is sent.
  - ◊ This error frame is encoded using DCT, Run Length and VL techniques.
  - ◊ The error block, which is 16 x 16, is broken into four 8x8 blocks of pixels for encoding.
  - ◊ Minimal errors, meaning minimal differences, provides many zeros to transmit.

DISPLAY VA 8.18
DISCUSS
WB 8.15

- B frames constructed like P Frames.
  - ◊ Encoder uses BEST of matching macroblock from previous OR future frame, or both.
  - ◊ Again, difference is transmitted.
- GOP – Group of Pictures
  - ◊ New I frame is stored every 15 frames.
  - ◊ Keeps errors with limits.
  - ◊ Each set of 15 frames called a Group of Pictures (GOP).

DISPLAY VA 8.19
DISCUSS VA 8.16

The (+2,-1) move has the following I frame overlaying the P Frame so that the error can be predicted. Point this out to the students.

DISPLAY VA 8.20
DISCUSS 8.17

Audio Encoding and Compression

- Multichannel audio
  - ◊ Five channels – Right, left, center, two-surround sound channels
  - ◊ Low frequency enhancement channel
  - ◊ Maximum of seven commentary and/or bi-lingual channels

Transport Layer

- Specifies how to multiplex compressed video and audio.
- Developed for error prone environments
  - ◊ 188-byte packet, fixed length
  - ◊ Four-byte header, 184-byte payload
  - ◊ Suitable for coaxial environments

DISCUSS

H.261 – It is not MPEG-2

- Video compression standard for videoconferencing
- *N* times 64 kbps
- Telephony network, such as ISDN
- Two resolutions supported
- Similar to MPEG-2
- Specialized subset of telephony application

DISPLAY VA 8.21

Discrete Cosine Transform

Do not get into the details of the formula

Defined as a mathematical function that transforms a temporal (function of time) relationship into an expression in the frequency domain.

DISPLAY VA 8.22
DISCUSS WB 8.18

- Used with JPEG and MPEG-2
- Easy to implement
- Aids the compression process
- Moving to frequency domain has the benefit of having more 0s in the expression than the time domain.
  - ◊ Makes compression more practical
  - ◊ Data is not lost
- DC component is key
  - ◊ Samples are quantized.
  - ◊ Quantized coefficients are “large” for DC component, but small for higher orders.
  - ◊ Quantization produces a large number of zero-value coefficients.
  - ◊ Huffman encoding reduces overall number of bits to be sent.
  - ◊ Zig-zag pattern of processing also increases efficiency of transmission.

DISPLAY VA 8.23
DISCUSS WB 8.19

SUMMARY

DISPLAY VA 8.24
DISCUSS WB 8.20

- Describe the difference between Lossless and Lossy compression.
- Describe how video compression is different from audio compression.
- Compare the different compression methods used for pictures.
- Describe, at a high level, the process used by MPEG-2 to digitize and compress a signal.

Review these items

REFER TO WB 8.21 AND WB 8.22

Have students answer the study questions at the end of Module 8. Review the answers.

Answers in Appendix

Appendix

Demonstrating MPEG-2 with a simple picture.

Materials

- A background, such as a seascape.
- A cutout picture of an object such as a boat.
- A square cutout, e.g., a frame.

Technique: The background holds steady, the "object" moves through the "frame". Check with the students about what changes and what remains the same. How might this be predicted? Discuss the effects of waves and gusts of wind (predicted versus actual).

Objective: To demonstrate to the student how easy? and how hard? it is to make a prediction. The prediction can be good, but unexpected motions can render it an error. If the error is zero, there are few bits to transmit. If the error is not zero, there are fewer bits to transmit since only the error is transmitted. This will also explain why the I frame is transmitted every 15th frame in the GOP.

Answers to the questions at the end of the Student Workbook

1. Why is compression needed for digital television?

   Digital technology is a bandwidth hog. Without compression, digital signals would not fit into the current channel bandwidth allocations. Also, the storage requirements for non-compressed video, which will be delivered from a video server, would be prohibitive. With no compression, 39 minutes of video requires 48 Gigabytes of digital storage. Placing this in perspective, a typical hard drive on a personal computer contains 20 Gigabytes, which must be allocated to many software applications as well as data storage.

2. What are the two general categories of compression, and how do they differ?

   The two categories are lossless and lossy. Lossless compression guarantees that the digital output of the compression system will be identical to the input, down to the bit level. Lossy compression does not guarantee identical reproduction down to the bit level, but ensures that the output of the system will meet certain preset criteria that make the output appear identical to the intended receiver.

3. What is Recommendation 601, and what are the associated sampling rates? What is the sampling structure called, and what other sampling structures are there?

   Recommendation 601 is the most widely used digital component video format. It uses a 4:2:2 sampling structure with co-sited samples. Co-sited means that each color pixel is described by the three samples, which are coincident in time. The 4 in 4:2:2 comes from the four times sub-carrier sampling frequency used in composite digital systems. In actuality, the sampling rate for luminance is 13.5 MHz. The number 4 is based on the 14.3 MHz NTSC sampling frequency that was once under consideration. The 2 in 4:2:2 represents the 6.75 MHz sampling rate for each of the color difference signals, Cb and Cr.

   Some early digital systems sampled the color components at one quarter of the luminance sample rate, creating a 4:1:1 system. A reduction in the chrominance data density equivalent to 4:1:1 can be achieved using vertical sub-sampling. This scheme is called 4:2:0 sampling. It interpolates chroma samples from two adjacent lines in a field to produce a single chroma sample that is spatially located half-way between one of the original samples and the location of the same line, but the opposite field. 4:2:0 sampling is generally inferior to both 4:1:1 and 4:2:2 sampling due to the poorer vertical resolution and resulting loss of color quality.

4. What are the three rules that drive compression algorithms?

- For audio:

  If you can't hear it, don't transmit it. The human ear can detect sound in the range of 300 to 20,000 Hertz. Anything beyond those levels is unnecessary for a human audience. Even within this range, there are different needs. 300 to 4000 Hertz adequately bounds human speech within recognizable limits for voice telephony. Digital audio with surround sound, on the other hand, requires the full 20,000 Hertz range for accurate sound reproduction.

- For video:

  Transmit only what changes between pictures. If the receiver already has the information, reuse what is at the receiver location from the previous transmission, and replace only the changed items. Most video is a composite of still and moving parts. What needs to be transmitted is determined by how much of the picture is actually moving, and how fast the motion is occurring.

  Transmit only the amount of information that can be detected by the end recipient. In most cases, this is a human being. This rule is similar to the one given for audio. In this case, it applies to color differences and brightness levels.

5. Why is the sample rate different for chrominance than for luminance?

   In human beings, the cells of the eye detect all video signals. The human eye has one set of cells called cones, which detect color, and another called rods, which detect black and white. Rods are more sensitive than cones. They can distinguish differences in levels of black (called greytones) better than the cones can detect differences in color. Essentially, the greytones correspond to luminance levels. Because of this sensitivity, when color video is converted from analog to digital, more luminance information needs to be coded and transmitted to retain the original visual quality of an image.

6. What happens in each of the steps in the general JPEG process?

   Steps 1-3 are sampling steps. In these steps, the luminance and chrominance information contained in each pixel of the video picture is sampled and the discrete cosine transform is applied to the samples to convert the spatial representation of the sample to a frequency representation.

   Step 4 is a quantizing step, where the sample values are divided by quantizing coefficients. The result is the rounding to zero of the higher frequency components of the sample.

   Steps 5-7 are encoding steps. In these steps, the quantized values from step 4 are turned into compressed binary strings, ready for transmission.

7. What is the chief difference between JPEG and MPEG?

   JPEG was developed for still images, while MPEG is a set of standards for motion pictures.

8. What are the first three parts of the MPEG-2 standard, and what do they specify?

   The first three parts of the MPEG-2 standard are Video, Audio, and Systems. The Video part specifies the techniques for interframe motion compression. The Audio part provides a low bit rate coding for multichannel audio. The Systems part specifies how to multiplex compressed video and audio into a single compressed bit stream.

9. Describe how prediction makes compression possible.

   Prediction allows reconstruction of a signal from a combination of a reference sample and another sample containing only data that has changed from the reference sample. Since the second sample only contains information that has changed from the first sample, less information needs to be coded.

©SCTE

10. What is the function of the Discrete Cosine Transform? How does it work?

    The Discrete Cosine Transform (DCT) is a mathematical function that transforms a temporal (function of time) relationship into an expression in the frequency domain. In more user friendly terms, this means that a "normal" equation, where the variable is time, becomes an equivalent expression with several frequency variables. The coefficients of the DCT can be placed in a matrix that is the same dimension as the original time domain description of pixel component magnitudes. Unlike the time domain expression, however, the frequency domain expression contains several near zero or zero coefficients, which can either be dropped or further compressed.

11. What are the two types of coding used to compress binary data?

    Variable length coding and run length coding can be used to compress binary data, either separately or together. Variable length coding replaces common patterns of digits with some shorter code. Run length coding replaces the repeating pattern with a character that describes what repeats, followed by the number of times the repeat has occurred.