The Digital Mystique: A Review of Digital Technology and its Application to Television

A. MICHAEL NOLL

ABSTRACT *This article explores the divergences and convergences between the worlds of analog and of digital. It reviews some of the basic concepts of digital conversion and then compares and contrasts analog and digital signals, including their application to telephone and television signals. Issues arising from the use of digital and analog compression to save bandwidth are discussed. Lessons learned from the use of compression of cellular telephone signals are applied to the world of digital television.*

Key words: compression; digital; television; digital television; television.

Introduction

Today is the age of digital! But, the real world remains analog. What then is the place of analog in today's digital universe? What sense should we make of the digital revolution? What is the future of television technology?

This article explores the divergences and convergences between the worlds of analog and of digital. After reviewing some of the basic concepts of digital conversion, the article compares and contrasts analog and digital signals, including their application to telephone and television signals. Issues arising from the use of digital and analog compression to save bandwidth are discussed. Lessons learned from the use of compression of cellular telephone signals are applied to the world of digital television.

Digital has its advantages, but it also has some shortcomings: for one, it consumes large amounts of bandwidth. In their beginnings and in their endings, nearly all signals are analog. So too are the human beings who create and consume signals of various kinds.

Today's fascination with all things digital needs clarification and a more realistic perspective. This article is an attempt to demystify the digital mystique.

Digital: A Review of Basic Concepts

Digital is a means for representing the instantaneous amplitudes of a signal as a series of numbers. These numbers, or digits, are usually encoded in binary form as a series of binary 'ones' and 'zeros.'

A digital binary signal is itself an analog signal in the sense that it is a waveform that varies continuously with time and in amplitude, although, in theory, only two possible amplitudes. Digital occurs in the interpretation of the signal by deciding at regular intervals in time that it must be one of two possible amplitudes, corresponding to a

0810-9028/98/020145-09

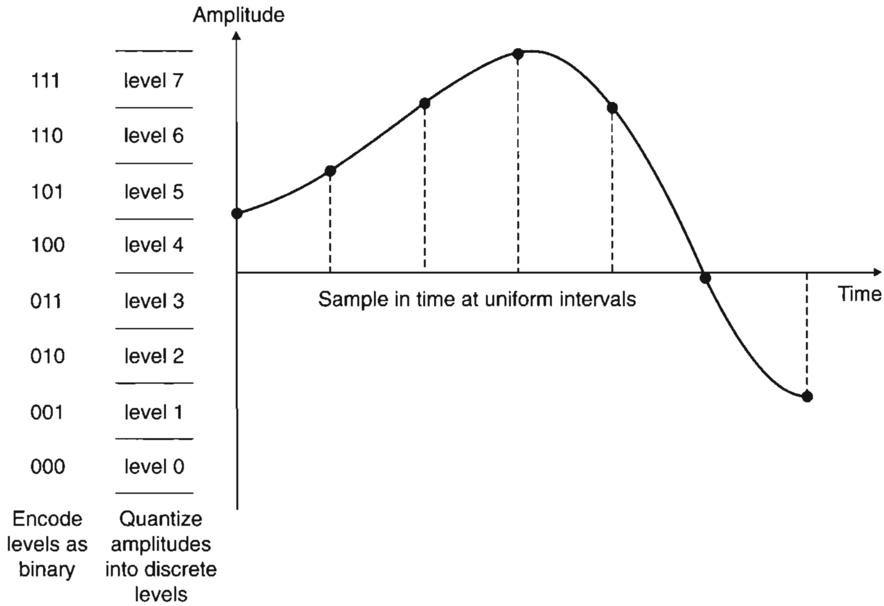

Figure 1. The process of converting an analog waveform into digital consists firstly of sampling it in time at uniform intervals. Next, the sample values are quantized into discrete levels. Lastly, the quantized levels are encoded as binary 'ones' and 'zeros' and expressed as a new signal varying between two amplitudes.

binary 'zero' or a binary 'one'. Hence, digital signals being analog occupy bandwidth. If all this seems confusing, perhaps an example (see Figure 1) might help clarify the concepts.

Consider a telephone speech signal with a maximum frequency of about 4000 Hz, or 4 kHz. To convert this signal to a digital format, it is first sampled in time at a rate of 8000 samples per second. This sampling rate is adequate to capture the time variations in the analog signal perfectly, since according to the Nyquist sampling theorem the sampling rate is at least twice the maximum frequency in the analog signal. The sample values are next quantized into a fixed number of levels, usually chosen to be a power of 2. In our example, 256 levels, which is 2 raised to the 8th power, are usually used to quantize a telephone speech signal.

Each quantization level can be encoded in binary form using 8 binary digits, or bits, for short. As an example, if the instantaneous amplitude of the sampled signal were level 175 it would be encoded in binary as 10101111. Since 8 bits are being used to encode each sample, the overall bit rate is 8 bits per sample times 8000 samples per second, which is 64,000 bits per second, or 64 kbps.

The last step is to represent the bits as a signal that can be sent over a transmission medium or recorded on some appropriate medium. This is usually done by assigning one voltage level to represent a binary 'one' and another voltage level to represent a binary 'zero'. In our example, we will use 0 volts to represent a zero and 3 volts to represent a one. The final 'digital' waveform would look like a square wave varying between 0 and 3 volts.

The digital signal looks like a square wave and is itself an analog waveform. As such, it occupies bandwidth—actually much more bandwidth than the original analog signal. One way to estimate the bandwidth of a digital signal is to realize that its fastest variation is repetitive alternations of a zero followed by a one. Such an alternation looks like a single cycle of a sine wave: thus two bits is equivalent to one Hertz. Therefore, the bandwidth of a digital signal can be approximated as the bit rate divided by 2. Or alternatively, a communication channel can carry a digital signal with a bit rate about twice the bandwidth of the channel.

For the preceding example of a digital telephone signal at 64 kbps, the bandwidth of this digital signal would be about half the bit rate, or 32 kbps. Thus, although the original analog signal had a bandwidth of 4 kHz, its digitally encoded version has a bandwidth that is eight times as much. In fact, the bandwidth of a digital signal increases in direct proportion to the number of bits used to encode each sample.

Actually, it is usually possible to fit more bits per second than only twice the bandwidth of a communication channel. The equation for the theoretical capacity of a communication channel of bandwidth W corrupted by Gaussian white noise with a signal-to-noise ratio of S/N is:

$$C = W \log_2 (1 + S/N).$$

This equation was derived by the Bell Labs' mathematician Claude E. Shannon.

In the real world, a digital signal would be corrupted by additive noise and by the finite bandwidth limitations of a communication channel. The effects of these corruptions would be to make the waveform appear noisy and also to smear the shape of the waveform. As long as the corruptions were not too severe, a simple threshold decision, in our example, set at the halfway voltage of 1.5 volts, would be sufficient to decide whether a zero or a one was transmitted. However, should the corruptions become too severe, reliable binary decisions would suddenly no longer be reliable, and the original signal could not be recovered.

In most cases, the amount of noise is well below the amount that would make recovery impossible. In these cases, the increase in bandwidth for a digital representation of an analog waveform is accompanied by a significant increase in noise immunity.

Noise Immunity

Trading bandwidth for noise immunity occurs frequently in communication systems. Wideband frequency modulation, invented by Major Edwin Armstrong at Columbia University in the early 1930s, is one example. Although the baseband monophonic hi-fi signal transmitted over Armstrong's wideband FM had a bandwidth of 15 kHz, the frequency-modulated radio signal had a much greater bandwidth of 200 kHz. This 13-fold increase in bandwidth was accompanied by a considerable increase in noise immunity. Armstrong's ingenuity was in realizing that wideband FM would increase immunity against noise.

Digital is most appropriate when a signal needs to be stored for posterity without the prospects for any degradation over time. Most electronic storage media—particularly those based on electromagnetism—deteriorate over time. An analog signal would thus be corrupted. However, a digital representation could be recovered in its entirety as long as the noise did not exceed the levels needed to make reliable decisions about ones and zeros.

Digital Telephony: The Invention of Pulse Code Modulation

Digital was first applied to telephone signals with the invention of pulse code modulation (PCM) in 1948 by Bell Labs' scientists Bernard M. Oliver, John R. Pierce, and Claude E. Shannon. Their invention was first applied to telephony in 1962 in the T1 time-division multiplexing system which combined together 24 digitized telephone signals for transmission over a single pair of copper telephone wires.

The T1 system was a success because it was far less costly to install the multiplexing equipment than to rip up city streets to install more cables between central offices when traffic exceeded capacity.

Long-distance telephony remained analog, however, for decades until only the latter half of the 1980s. The challenge was not a lack of knowledge about digital but rather a lack of bandwidth. It was not until a transmission medium was developed and installed that was capable of carrying the much larger bandwidths required for digital that it became feasible. That medium was optical fiber. Today, time-division multiplexing of digital telephone signals over optical fiber is widespread and has replaced other media.

Digital Bandwidth: Some Examples

We saw above that digital takes bandwidth—much more bandwidth than the original analog signal. The following examples will clarify this important observation.

As explained previously, a telephone signal requires an analog bandwidth of 4 kHz. This analog signal when converted to a digital format requires 64 kbps, which normally would require a bandwidth of about 32 kHz. The digital representation of a telephone signal could not be sent over a 4 kHz telephone channel. It is this considerable gain in bandwidth that delayed the use of digital techniques in long-distance telephony until a transmission medium with enough bandwidth was available, namely, optical fiber.

A high-fidelity audio signal has a maximum frequency of 20,000 Hz. Converting this analog signal to digital at a sampling rate of 44,100 samples per second and using 16 bits for quantization, as used for compact discs, results in a bit rate of 705,600 bps. The two signals of stereo would double this rate to about 1.4112 Mbps.

Although the now nearly extinct stereo phonograph record could accommodate the two separate signals required for stereophonic audio, it could not accommodate the large bandwidth needed for a digital audio signal. The solution was the use of a laser-read disc that had sufficient bandwidth. A similar problem occurred at the recording studio where the analog audio tape recorders did not have sufficient bandwidth to record a digital audio signal. The solution here was the use of suitably modified video tape recorders which already had the sufficient bandwidth to record the megahertze required by video and thus more than sufficient bandwidth for digital audio.

Lessons From Digital Cellular Telephony

The cellular telephone service originally used analog concepts. However, in order to increase the capacity of the system, digital technology has been introduced. Since compression was necessary to conserve bandwidth, a compromise with quality has occurred that is a surprises to many consumers who expected quality comparable to digital audio compact discs. This section reviews these lessons form digital cellular telephony.

In order to gain immunity to noise and other problems, cellular telephony uses wideband frequency modulation of the radio signals for each channel. A 4 kHz telephone signal frequency modulates its radio carrier to a channel bandwidth of 30 kHz. Clearly this is not bandwidth efficient, although the gain in immunity to noise and other problems justifies the inefficiency.

As cellular congestion increased, solutions were sought to increase the capacity of cellular telephony without adding to the overall band of radio spectrum allocated to the system. One way of achieving this increase in overall capacity is to pack more speech signals into each 30 kHz channel.

One way of packing together speech signals is the use of time-division multiplexing, or TDM. With TDM, speech signals are converted to digital, and samples are then interspersed together to create a single stream of bits that fits within the 30 kHz channel space.

One of today's approaches to digital cellular shares the 30 kHz channel with three digital speech signals, thereby allocating 10 kHz per digital signal. This means that the bit rate required for each speech signal must be reduced from the normal 64 kbps. Thus, digital cellular must utilize compression of the speech signal.

Most digital cellular compression schemes use a variant of a compression technique known a linear predictive coding (LPC), invented at Bell Labs by Bishnu S. Atal in the 1960s. The LPC technique in use today in the United States for digital cellular operates at a rate of 8 kbps. Additional bits are added for error correction, synchronization, and other purposes, which increases the final data rate for each channel to 16 kbps. The European GSM digital cellular system uses LPC at 13 kbps in order to improve the speech quality over LPC at 8 kbps.

The LPC compression used in digital cellular compresses the speech signal from an uncompressed 64 kbps to a compressed 8 kbps—a factor of 8. However, some speech quality is lost in the compression process. The compressed speech can sound a little buzzy and harsh. There also can be a small amount of delay required for the compression processing. However, when used in a noisy environment—such as an automobile—the deteriorations in speech quality would not be noticeable. For this application, the compromise with quality is acceptable.

There is, however, a more interesting problem that occurs with digital cellular. Since the signal portion of digital cellular is limited to 8 kbps, digital cellular when used for data communication is limited to the same 8 kbps.

Digital Television and Compression

An analog television signal has a maximum frequency of 4.5 MHz. When transmitted over the airwaves, an analog television signal occupies a bandwidth of 6.0 MHz, using conventional vestigial-sideband amplitude modulation. There is no noise immunity, and the transmitted signal is easily corrupted by additive noise and by multipath distortion, or ghosts.

An analog NTSC color television signal when converted to digital requires about 4.5 Mbps, which would require a bandwidth nearly 10 times that of the analog signal. Such bandwidths—and more—are today available in professional digital video tape recorders for use in television studios. But such bandwidths are not available for transmission over the airwaves. One digital television signal would require all the spectrum space used today by nearly a dozen conventional TV channels.

The digital bit rates required for professional quality video are three times that of the preceding, or about 167 Mbps (see Table 1). High-definition television would require tremendous bit rates, nearly 1 billion bits/sec.

The solution to the problem of the excessive bandwidth required by digital television

Table 1. Example of digital television bit rates

NTSC professional quality:
- 720 pixels/line
- 483 active lines/frame
- 30 frames/sec
- 16 bits/pixel

(8 luminance; 4 $C_r = Y - R$;
4$C_b = Y - B$)
- digital bit rate: 167 million bps

NTSC home quality:
- 3.6 MHz $\times$ 4 = 14 million samples/sec
- 6 bits/sample
- digital bit rate: 84 million bps

HDTV quality:
- 1920 pixels/line
- 1080 lines/frame
- 30 frames/sec
- 16 bits/pixel
- digital bit rate: 1 billion bps

TV channel:
- bandwidth: 6 MHz
- maximum useable digital capacity: about 18 Mbps

signals is compression. There is much redundancy, or excess information, in most signals, particularly those signals representing the moving images of television. One television field changes little from the next, unless there is considerable motion. Similarly, one scan line changes little from the next. Furthermore, there are many repetitive patterns in a single image. If appropriate signal processing were used, these redundancies could be eliminated and a compressed version of the signal created that required considerably less bandwidth for transmission.

The Motion Pictures Experts Group (MPEG) has developed standards for compressing video images. The MPEG standard in use today reduce the digital bit rate to about 3 Mbps, with reasonable image quality. The compression is accomplished by encoding changes from one television frame to another (inter-frame encoding) and also by capitalizing on similar portions of the image within a frame (intra-frame encoding).

MPEG encoding accomplishes an impressive reduction in bit rate. However, should large portions of the picture move, some blurring might be noticeable. Also, edges might appear overly crisp, which might be bothersome to some viewers.

Such artifacts are normal with compression. There are circumstances where any degradations of quality are not acceptable, such as the digital audio compact disc. In such cases, full bandwidth digital is then used. But consumers do not watch television because of the technical quality of the image, unlike much listening to recorded music where the technical quality of the audio is very important to many consumers.

Compression, by definition, is a compromise with quality—in today's world of digital, and in yesteryear's world of analog. The challenge is to make the compromise not noticeable to most people.

MPEG compression is performed uniformly across a television frame. However, most people are probably looking at only a small portion of the frame, perhaps that portion in which the action is occurring, such as the face of a person who is talking. This suggests a more powerful form of compression which takes into account that some portions of a frame might be more important than surrounding areas and concentrates on giving more resolution to these more important areas. This, however, is the type of decision best made by humans, which suggests that human observers somehow control the compression process. But human intervention can be both time consuming and costly. Is television really worth all this effort? Is picture quality really all that important to most people?

Digital Television: HDTV or More Programs?

The Federal Communications Commission is intent on foisting digital television on consumers, even if it means that analog NTSC television sets are all obsolete in 10 years. But the television broadcast industry is not yet clear on what digital television will become.

Since a compressed digital version of a conventional television signal requires only 3 Mbps, about four such digital television signals can be placed in the 6 MHz bandwidth today allocated for each television channel. This would allow broadcasters to send many more programs to each viewer and would thus increase program variety allowing broadcasters to compete with the many programs of direct broadcast satellite TV.

Alternatively, the full bandwidth of the 6 MHz channel could be used for one or two digital high-definition television (HDTV) signals that offer 1000 scan lines and widescreen images. But what is not clear is which, if any, of these differing definitions of digital television would be most meaningful to consumers. I have stated on many occasions that HDTV will only allow viewers to see more clearly how poor the program content really is! But then again, simply creating the opportunity for more violence, more sex, and more obscenity would also be a waste of the technology. But I am reminded that the inventors of television were horrified at the mass-market low-brow programming that their invention was used for.

Analog Compression

It is frequently forgotten today that the basic concepts of compression were invented decades ago for analog signals. Telephone speech signals in analog form were compressed using vocoders (for *voice coders*) by a factor of 10, or even more, so that a 4 kHz signal needed only about 400 Hz.

In fact, the NTSC system used for color television employs a number of different compromises which could be considered to be forms of compression. The frame rate of 30 frames per second is chosen based on the persistence of human vision that creates the illusion of motion when different images are shown faster than about 24 times per second. The field rate of 60 fields per second was chosen based on the flicker fusion rate for human vision. The scan rate of 525 scan lines was chosen based on the acuity of human vision, assuming a viewing distance of less than four times the picture height.

If a form of compression were not used, the addition of color to a monochrome television signal would triple the bandwidth from 4.5 MHz to 13.5 MHz. But again by taking into account properties of human color perception, the color information can be restricted to much less bandwidth than the monochrome information. In fact the color information is separated into two components: one requiring 1.5 MHz and the other 0.5 MHz. By sharing spectrum space through the technique of frequency interleaving, the color information is encoded along with the monochrome signal without any additional bandwidth. In theory, the quality of the color television image is degraded from what would be obtained with more bandwidth, but few, if any, TV viewers are able to notice the deterioration since the compromises are all based on the psychological properties of human vision.

Digital NTSC Television

The NTSC color television standard has lasted nearly a half century in the United States. Clearly for a standard to last this long, much about it must have been correct. However, many color TV sets still fail to extract the full capabilities inherent in the broadcast signal. A trip to a TV studio to witness the NTSC signal in its full glory shows its full quality before it has been corrupted by noise, multipath distortions, and the inadequacies of most TV set circuits.

Yet many of the inadequacies of NTSC could be resolved through digital processing in the TV set. The full 1.5 MHz bandwidth of the color information contained in the l-component of the chrominance signal could be recovered and displayed. Noise and multipath corruptions could be eliminated and removed through digital filtering. Sequential scanning could be replaced through superior progressive scanning. And extra scan lines could be interpolated between the existing lines to create 1000 lines.

But in the head-first rush to digital, the far simpler concepts of digital processing of the NTSC signal have been mostly forgotten. Also forgotten is the fact that a compressed digital signal does not have the inherent quality of an uncompressed NTSC signal. It would not be a surprise if the 'new' television of the next century is digital NTSC television.

Future Innovations

Much progress in technology is evolutionary. However, major revolutions do occur. The digital compact disc is one such example and holds some lessons for digital television.

Yesteryear's black-vinyl phonograph record stored an analog signal with a maximum frequency of about 15,000 Hz. Converting this analog stereo audio signal to digital created a digitally encoded signal with a bit rate of about 1.4 Mbps. Such a signal required a bandwidth of about 0.7 MHz—far more than could be recorded on a phonograph record. The digital audio signal could have been compressed to fit within the limited bandwidth of a phonograph record, but the compromise with quality would not have been acceptable.

The solution was the development of the laser video disk, which had sufficient bandwidth to store the megahertzes of an analog video signal. Rather then storing analog video, the laser disk was shrunk in diameter and its bandwidth was used to store an uncompressed digital audio signal. Thus, the digital audio compact disc was invented.

Imagine if a digital storage medium for the home market were invented with sufficient bandwidth to store an uncompressed video signal. All of today's emphasis on compression with its apparently unending progression of MPEG standards would be extinct. Video compression would only be needed for over-the-air broadcast in which bandwidth was still a problem. The capacity of today's digital disc is increasing to such an extent that the capacity needed to store uncompressed digital video would seem to becoming closer and closer for the consumer market. It is interesting to speculate what such an innovation would mean for the video industry and the physical storage of video.

Decades ago when the earliest experiments in television were being conducted, a major fork occurred in the development of television standards. This fork was the decision to use horizontal scanning to analyze the image to create the serial signal that was required for radio broadcast. The alternative method of image analysis was parallel transmission using a two-dimensional matrix of light sensors. With today's computer technology and newer methods of radio transmission, a re-examination into the most basic principles of television might result in totally new methods which might well make obsolete today's methods of compression and thoughts about digital television.

Conclusion

In our headlong rush toward all things digital, we have forgotten many of the analog techniques and other concepts of the past. However, innovation frequently occurs by avoiding the vogues of the day and instead revisiting the ideas of the past. Digital is a great way to process analog signals.

So, do not place all your bets on today's digital television. It clearly has its place, particularly when signals need to be stored for posterity. But its place in the world of the fleeting signals of broadcast television seems less clear.