3

Signals Carried over the Network

Services that the telecommunications networks provide have different characteristics. Required characteristics depend on the applications we use. To meet these different requirements, many different network technologies that are optimized for each type of service are in use. To understand the present structure of the telecommunications network, we have to understand what types of signals are transmitted through the telecommunications network and their requirements. In this chapter we look at the requirements of various applications, characteristics of analog voice channels, fundamental differences between analog and digital signals, analog-to-digital conversion, and a logarithmic measure of signal level, the decibel.

3.1 Types of Information and Their Requirements

Modern digital networks transmit digital information transparently; that is, the network does not necessarily need to know what kind of information the data contain. This information that is transmitted through the network may be any one of the following:

- Speech (telephony, fixed, or cellular);
- Moving images (television or video);
- Printed pages or still picture (facsimile or multimedia messaging);
- Text (electronic mail or short text messaging);
- Music;
- All types of computer information such as program files.

For digital transmission, analog signals such as speech are encoded into digital form and transmitted through the network as a sequence of bits in the same way computer files are transmitted. However, although all information is coded into digital form, the transmission requirements are highly dependent on the application; because of these different requirements, different networks and technologies are in use. Video and e-mail applications, for example, require different architectures. Network technologies have taken two main development paths: one for speech services and another for data services. The telephone network and ISDN have been developed for constant-bit-rate voice communication that is well suited to speech transmission. Data networks such as LANs and the Internet have been developed for bursty data transmission.

The constant-bit-rate requirement for speech follows from the principle that digitized voice signals have traditionally been transmitted in digital form as samples at regular intervals, as we will see in Section 3.6. Data transmission is bursty by nature. Sometimes we may copy a file across the network, whereas at other times we may work locally with our workstation.

When many different applications are integrated into multimedia communications, both basic types of service requirements of constant-bit-rate voice and bursty data have to be fulfilled and we need a concept that is able to meet both types of requirements.

In Table 3.1 different applications are compared from the communication requirements point of view. The applications are ordinary speech, *computer-aided design* (CAD) (a service in which high-resolution graphical information is transmitted), moving images (video), file transfer, and multimedia with integrated video, voice, and data. The importance of the transmission requirements for each application is explained next.

Data Rate or Bandwidth Requirement

Voice communication usually requires a constant data rate of 64 Kbps or less and high-resolution video a constant data rate of 2 Mbps or higher over the network. Characteristics of data communication are very different, for example, file transfer requires high-bit-rate transmission only during download, and high-resolution graphics on a Web page require high-data-rate transmission only when we download a new page. When we are reading a Web page we do not need transmission capacity at all. To define data transmission

Table 3.1
Communication Requirements of Different Applications

Transmission Characteristics	Voice	Video	File Transfer	Interactive Media
Bandwith requirement	Low, fixed	Very high, fixed	High, variable	High, variable
Data loss tolerance	Tolerant	Tolerant	Nontolerant	Tolerant or nontolerant
Fixed delay tolerance	Low delay	Tolerant	Tolerant	Low delay
Variable delay tolerance	No	No	Tolerant	No
Peak information rate	Fixed	Fixed	High	Very high

capacity, we sometimes use the term *bandwidth* instead of *data rate* because these terms are closely related to each other, as we will see in Chapter 4.

Data Loss Tolerance

Noise and other disturbances in the network may cause errors in the transmitted data. If errors occur, some amount of data may be lost. Voice and video transmission services are used by human beings, and they can tolerate accidental short disturbances. In computer communications a single erroneous bit usually destroys a whole data frame, which may contain a large amount of data. The loss of one frame destroys the transmission of a large file that is transferred in multiple frames. Most of the data communication systems are able to retransmit data frames in error. Systems designed for voice or video transmission do not use retransmission schemes because temporary retransmission delay is even more disturbing for human users than the loss of some data.

Fixed Delay Tolerance

When communication is interactive, as voice communication usually is, the two-way transmission delay should be very short for good quality. In the case of voice it should be of the order of some tens of milliseconds. Otherwise, we feel that quality is degraded because the response from the other party is delayed. We tolerate much longer delays in the case of ordinary data applications when we are waiting for a response to our "click" command.

Variable Delay Tolerance

Voice and video information is traditionally transmitted as samples at regular periods of time. The reconstruction of images and voice requires that all sample values be received sequentially and suffer the same delay. Conventional data networks recover from errors with the help of retransmission of the frames in error. This is a very efficient error recovery scheme, but it introduces some additional and variable delay. For voice applications this variable delay is often a worse solution than that of losing some data.

Peak Information Rate

Encoding of analog voice and video often produces a constant information data rate. Values of the samples with constant length contain voice or video information and they are transmitted at a constant rate. In data communication applications we usually work locally and every now and then a high data rate is needed to load graphical information or files. A peak load is typically of the order of 1,000 times higher than the average transmission capacity we use.

The different requirements just explained have supported development of the circuit-switched networks, such as PSTN and ISDN, for voice communications and packet-switched networks, such as LANs and the Internet, for data communications. *Asynchronous transfer mode* (ATM) technology was developed by ITU-T to be suitable and efficient for transferring all types of information. However, the expansion of the Internet has reduced its importance and Internet technology will be developed further to provide a platform for all kinds of communications.

3.2 Simplex, Half-Duplex, and Full-Duplex Communication

In telecommunications systems the transmission of information may be unidirectional or bidirectional. The unidirectional systems that transmit in one direction only are called *simplex,* and the bidirectional systems that are able to transmit in both directions are called *duplex* systems. We can implement bidirectional information transfer with *half-* or *full-duplex* transmission as shown in Figure 3.1.

In simplex operation the signal is transmitted in one direction only. An example of this principle is broadcast television, where TV signals are sent

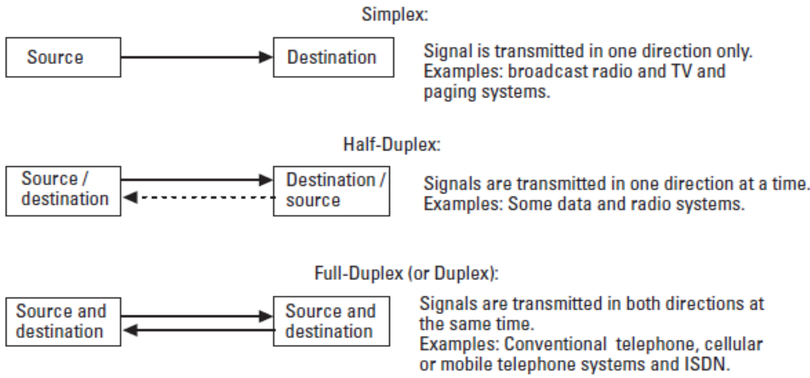

Figure 3.1 Simplex, half-duplex, and full-duplex transmission.

from a transmitter to TV sets only and not in the other direction. Another example is a paging system that allows a user to receive only alphanumerical messages.

In half-duplex operation the signal is transmitted in both directions but only in one direction at a time. An example of this is a mobile radio system where the person speaking must indicate by saying the word *over* that she is done transmitting and the other person is allowed to transmit. LANs use a high-speed, half-duplex transmission over the cable even though users may feel that the communication is continuously bidirectional, that is, full duplex.

In full-duplex operation signals are transmitted in both directions at the same time. An example of this is an ordinary telephone conversation where it is possible for both people to speak simultaneously. Most modern telecommunications systems use the full-duplex principle, which we call *duplex operation* for short.

3.3 Frequency and Bandwidth

To understand the requirements of different applications for a telecommunications network, we must understand the fundamental concepts of frequency and bandwidth. The information that we transmit through a telecommunications network, whether it is analog or digital, is in the form of electrical voltage or current. The value of this voltage or current changes through time, and this alteration contains information.

The transmitted signal (the alteration of voltage or current) consists of multiple frequencies. The range of frequencies is called the *bandwidth* of the signal. The bandwidth is one of the most important characteristics of analog information and it is also the most important limiting factor for the data rate of digital information transfer.

3.3.1 Frequency

We can see the telecommunications signal as a combination of many cosine or sine waves with different strengths and frequencies. The frequency refers to the number of cycles through with the wave oscillates in a second. As an example of the concept of frequency, we hear the oscillation of air pressure as sound. We are able to hear frequencies in the range of approximately 20 Hz to 15 kHz, where Hz (hertz) represents the number of cycles in a second. An example of the different frequencies is heard in the keys of a piano. The right-hand keys generate basic frequencies of the order of 1,000 Hz and the left-hand keys of the order of 100 Hz.

In electrical terms, an *alternating current* (ac) changes its direction of flow several times per second. This variation in direction is known as a *cycle*, and the term *frequency* refers to the number of cycles in a second that is measured in hertz. If a signal has 1,000 complete cycles in a second, then its frequency is 1,000 Hz or 1 kHz. A pure sine wave, like that shown in Figure 3.2, is generated with a loop of wire rotated in a magnetic field at a constant rate. This fundamental waveform can be seen as a cosine of the

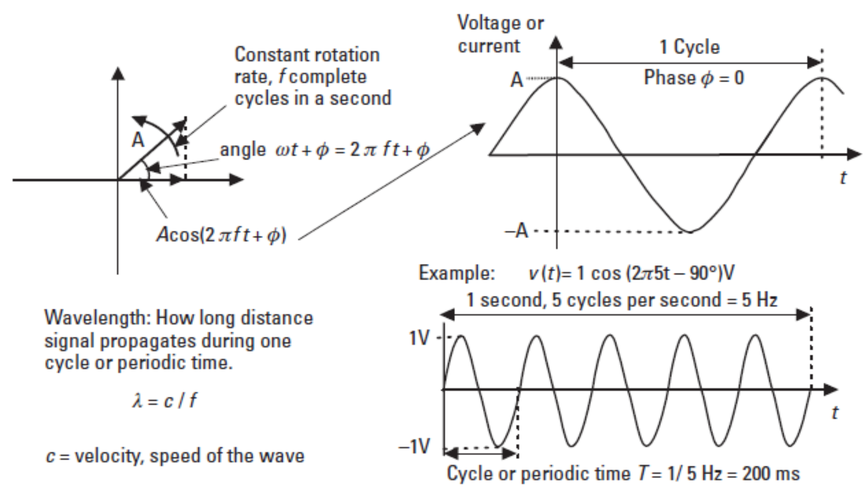

Figure 3.2 Cosine wave and frequency.

angle of the phasor rotating at a constant rate. The strength of the voltage or current alters according to the cosine curve when time increases. The length of the phasor corresponds to the maximum value of the signal and it is called *amplitude,* shown as A in Figure 3.2.

We can see any telecommunications signal as a sum of these fundamental waveform cosine waves that are expressed as

$$v(t) = A\cos(\omega t + \phi) = A\cos(2\pi ft + \phi) \qquad (3.1)$$

where f is frequency, the number of complete cycles in a second expressed in hertz, 1 Hz = 1/sec; t is time in seconds, and ϕ is the phase shift (phase of the cosine wave at time instant $t = 0$). The angular frequency ω in radians per second is $\omega = 2\pi f$, which comes from the fact that one complete cycle of a phasor makes up an angle of 2π radians.

The *periodic time* or *period* T in seconds represents the time of one complete cycle:

$$T = 1/f \text{ and } f = 1/T \qquad (3.2)$$

Wavelength λ represents the propagation distance in one cycle time, thus,

$$\lambda = c/f = cT \qquad (3.3)$$

where c is the velocity of the signal. For a sound wave, the velocity in the air is approximately 346 m/s; for light or radio waves, approximately, c = 300,000 km/sec.

The example in Figure 3.2 shows a waveform with a frequency of 5 Hz and amplitude of 1V. It corresponds to a phasor with length of A = 1V making five complete cycles in a second. At time instant $t = 0$, the waveform has a value of 0 and the phase or angle of the phasor is –90°. As the time increases and the phasor rotates, its projection at the horizontal axis of the phasor diagram increases, corresponding to an increase in the value of the wave with time. The equation for this example waveform is then $v(t) = A\cos(\omega t + \phi) = 1\cos(2\pi 5t - 90°)\,V$.

3.3.2 Bandwidth

The voice signal, which is the most common message in telecommunications network, does not look similar to a pure cosine wave in Figure 3.2. It contains many cosine waves with different frequencies, amplitudes, and phases combined together. The range of frequencies that is needed for a good enough quality of voice, so that the speaker can be recognized, was defined to be the range from 300 to 3,400 Hz. This means that the bandwidth of the telephone channel through the network is 3,400 – 300 Hz = 3.1 kHz, as shown in Figure 3.3. A human voice contains much higher frequencies, but this bandwidth was defined as a compromise between quality and cost. It is wide enough to recognize the speaker, which was one requirement for telephone channel.

Bandwidth is not strictly limited in practice, but signal attenuation increases heavily at the lower and upper cutoff frequencies. For speech, channel cutoff frequencies are 300 and 3.4 Hz, as shown in Figure 3.3. The bandwidth is normally measured from the points where the signal power drops to half from its maximum power. Attenuation or loss of channel is given as a logarithmic measure called a decibel (dB), and half power points correspond to a 3-dB loss. Decibels are discussed later in this chapter.

Bandwidth, together with noise, is the major factor that determines the information-carrying capacity of a telecommunications channel. The term *bandwidth* is often used instead of *data rate* because they are closely related, as we will see in Chapter 4.

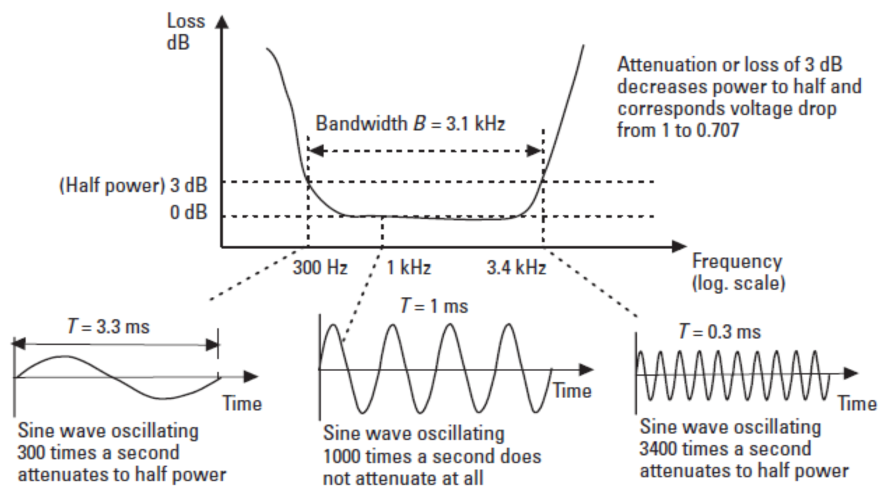

Figure 3.3 Bandwidth of the telephone speech channel.

3.4 Analog and Digital Signals and Systems

Most of the systems in the modern telecommunications network are digital instead of analog. In this section we look at the fundamental characteristics of analog and digital signals and how they influence the performance and operation of telecommunications systems.

3.4.1 Analog and Digital Signals

The difference between analog and digital form is easily understood by looking at the two watches in Figure 3.4. A true analog watch has hands that are constantly moving and always show the exact time. A digital watch displays "digits" and the display jumps from second to second and shows only discrete values of time.

Another example could be the slope of analog voltage where all values of voltage can be measured as shown in Figure 3.4. In "digital slope," only discrete values may be measured. In the example of the figure, we have eight discrete values, 0 to 7, in the digital slope. This does not mean that the digital systems perform worse than analog systems. If we want to improve the accuracy of the digital system, we just increase the number of steps and, in principle, any voltage level can be represented with the digital system as well.

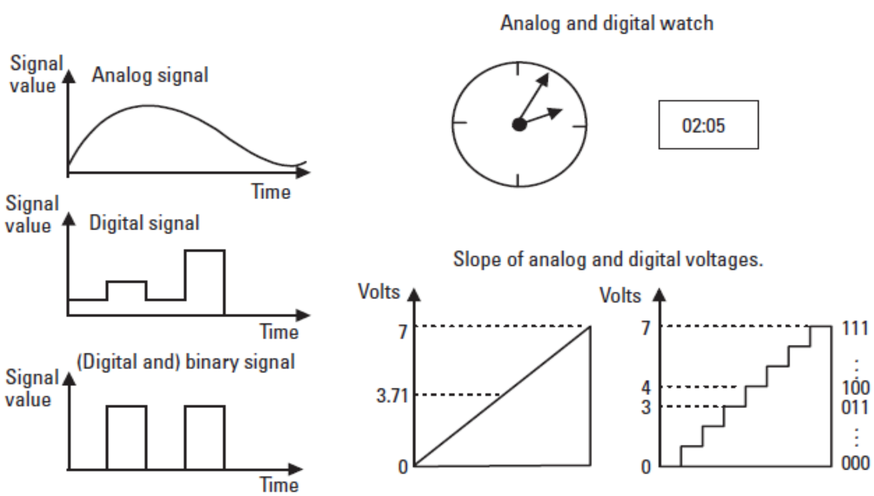

Figure 3.4 Analog and digital signals.

A special and very important case of digital signals is a binary signal where only two values, binary digits 0 and 1, are present as illustrated in Figure 3.4. Examples of binary signals are light on and off, voltage versus no voltage, and low current versus high current.

Binary signals are used internally in computers and other digital systems to represent any digital signal. For example, we can encode eight voltage levels of the slope in Figure 3.4 into three binary bits and each of these three bit words then represents one of the $2^3 = 8$ (0 (000) to 7 (111)) different values. As another example, a digital signal with eight-bit words or bytes (often called *octets* in digital telecommunications systems) can represent $2^8 = 256$ discrete values of a signal. These kinds of digital numbers are used to represent analog voice, in which each sample of a voice signal is encoded into eight-bit words, as we will explain in Section 3.6.

3.4.2 Advantages of Digital Technology

Analog systems in a telecommunications network have gradually been replaced with digital systems. Development of digital circuits and software technologies has made digital systems more and more attractive. The most important advantages of digital technology over analog technology are as follows:

- Digital functions make a high scale of integration possible.
- Digital technology results in lower cost, better reliability, less floor space, and lower power consumption.
- Digital technology makes communication quality independent of distance.
- Digital technology provides better noise tolerance.
- Digital networks are ideal for growing data communication
- Digital technology makes new services available.
- Digital system provides high transmission capacity.
- Digital networks offer flexibility.

An analog system requires the accurate detection of signal values inside its dynamic range, that is, between the maximum and minimum values of the signal. Digital systems use binary signals internally. A binary signal has only two values, and the only problem is to distinguish these two values from each other. The dynamic range is well defined and linearity is not required.

This makes the elements of digital circuits simple, and the utilization of compact technology for very complicated functions, such as integrated circuits, is feasible.

As a consequence, circuit integration leads to a smaller number of electronic components, smaller equipment, lower manufacturing costs, lower maintenance costs because of better reliability, and less power consumption. More and more complex integrated circuits are replacing many lower scale integrated circuits. This decreases system costs, because the increased complexity of components does not cost much in volume. When integrated circuits are manufactured in volume, complex ones do not cost much more than less complex circuits. In addition, the smaller number of separate components gives better reliability.

In long-distance connections, we have to amplify or regenerate the signal on the line many times. When we amplify an analog signal on the line, we amplify noise at the same time. This added noise decreases the quality of an analog signal, that is, decreases the *signal-to-noise* (S/N) ratio.

In the case of a digital system we use regenerators or repeaters instead of amplifiers. Repeaters regenerate the signal symbol by symbol, that is, transmit further the value that is closest to the received value. The regenerated signal is a sequence of digital symbols with nominal values and thus it contains no noise. If the noise is low in the input of each regenerator, symbols of the digital signal are regenerated without errors and we receive exactly the same digital message on the other side of the world as it was at the transmitting end. The operation of a digital repeater or regenerator is described in Chapter 4.

Modern switches digitize speech in the subscriber interface. If the path through the network is fully digital, conversion back to analog is done only at the far end. There is only one analog-to-digital and one digital-to-analog conversion regardless of the communication distance, that is, whether we make a call to our neighbor or to other side of the world.

The digital systems have to identify only signals from a set of discrete values. If symbols are not mixed because of too high a noise level, noise does not have any impact on the operation. Analog communication usually requires a much better S/N than low error rate digital communication. As a consequence, digital systems can utilize channels with much higher noise levels and they can tolerate higher interference than analog systems.

If the network is analog, a digital message has to be modulated into the frequency band of the analog telecommunications channel. This reduces the capacity available for the user. For example, a voice channel in the digital telephone network has a data capacity of 64 Kbps. If we use it via an analog subscriber loop with a voice-band modem, the data rate is restricted in practice to approximately 30 Kbps. With a *digital subscriber line* (DSL) (e.g., ISDN), the user data are exactly the same 64 Kbps used inside the network.

Digital systems are ideal for control via software because digital circuits operate in a numerical way. Integrated software makes systems flexible and new functions needed for new services are easier to implement. Intelligent network services, reviewed in Section 2.10, are good examples of these new services. As another example, we would not have cellular telephone service if we did not have digital software-controlled systems in the network.

The digital processing of information makes better utilization of channels possible; for example, several digital broadcast television channels fit into the band of one analog broadcast channel. In Chapter 4 we will see that digital signals tolerate higher disturbances than analog signals and this is one reason behind the better frequency efficiency. Low-cost multiplexing (no analog filtering and modulation circuitry required) and efficient use of optical transmission media make high-capacity digital systems feasible. Optical systems transmit digital signals as a series of short light pulses. The distortion of these digital pulses does not influence the quality of the message because distorted pulses are regenerated, which eliminates distortion.

All types of analog signals can be converted into digital signals. When this is done, the digital network is able to carry any information. Bits are handled in the same way whether they represent voice, video, or data.

Analog systems are different for each application because of different performance requirements. For example, a telephone connection requires channels with approximately 4-kHz bandwidth, but television signals require 5-MHz bandwidth with a much better S/N. In digital systems the corresponding characteristic is the data rate. For example, an analog telephone signal requires 64 Kbps and video with a much wider bandwidth requires 2 to 140 Mbps depending on the coding scheme in use. We can use one high-data-rate system for a single video channel or a large number of speech channels.

Digital technology provides efficient multiplexing for sharing capacity in high-data-rate connections. This makes high-capacity digital networks and systems flexible. The same system, if it provides a high enough data rate, can be used for any application.

3.4.3 Examples of Messages

In the previous sections, we described the characteristics of the digital and analog signals and systems. Now we look at some simple examples of information sources that produce messages that are transmitted through the network. There are many different information sources, including machines as well as people, and messages or signals appear in various forms. As for signals we can identify the two main distinct message categories: analog and digital.

3.4.3.1 Information, Messages, and Signals

The concept of information is central to communication. However, *information* is a loaded word, implying schematic and philosophical notions and, therefore, we prefer to use the word *message* instead. *Message* means the physical manifestation of information produced by a source. Systems handling messages convert them into electrical signals suitable, for example, for a certain transmission media.

3.4.3.2 Analog Message

An analog message is a physical quantity that varies through time, usually in a smooth and continuous fashion. Examples of analog messages are acoustic pressure produced when you speak or light intensity at one point in an analog television image. One example of an analog message is the voice current on a conventional subscriber telephone line as illustrated in Figure 3.5. In Section 2.2 we explained how the current is produced.

Because the information resides in a time-varying waveform, an analog communication system should deliver this waveform with a specific degree of fidelity. Because the strength of signals may vary in a range from 30 to 100 dB, depending on the application, the analog systems should have good linearity from the weakest signal to 1,000 to 10,000 million times stronger signal values.

3.4.3.3 Digital Message

A digital message is an ordered sequence of symbols selected from a finite set of discrete elements. Examples of digital messages are the letters printed on this page or the keys you press at a computer keyboard. When you press a key at your computer keyboard, each key stroke represents a digital message that is then encoded into a set of bits for binary transmission.

Because the information resides in discrete symbols, a digital communication system should deliver these symbols with a specified degree of accuracy in a specified amount of time. The main concern in the system design is that symbols remain unchanged, which is the final requirement for transmission accuracy.

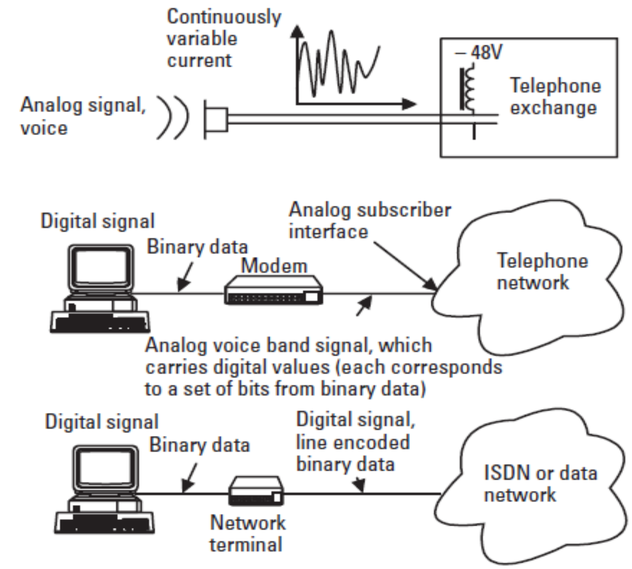

Figure 3.5 Examples of messages.

We need modems for the transmission of digital messages over analog channels. The modems receive a message from the terminal in the form of binary data and send it as an analog waveform to the speech channel as shown in Figure 3.5. Current modems do not modulate or change the analog waveform at the rate of the binary data they receive from the terminal. Instead they encode a set of bits into a digital symbol that may get many more values than just two. Each multilevel symbol corresponds to a set of bits and it is sent as one analog waveform to the line. When receiving a certain analog signal on the other end, the receiver detects a set of bits defined to correspond to that signal. Use of more than two signals increases the data rate through the speech channel compared with the binary principle, in which only two different signals are used. Speech channels have quite a narrow bandwidth, but a good S/N, which allows use of many different signals, as we will explain in Chapter 4.

When a digital network is used to transmit digital messages, signals are in digital form from end to end. Instead of a modem, a network terminal is needed at the subscriber's premises to encode binary signals into digital pulses suitable for cable transmission to an exchange site; see the ISDN example in Figure 3.5.

3.5 Analog Signals over Digital Networks

In this section we look at how analog signals are handled before transmission through a digital network. In the next section we concentrate on the pulse code modulation, which is performed in the network on our voice during a telephone call, and in Section 3.7 we present a brief review of other voice-coding schemes.

If a digital signal is to be transmitted through an analog network, it has to be converted into an analog signal suitable for the frequency band of the channel, as we saw in Figure 3.5. Digital networks provide communication only with a set of discrete symbols (in the binary case these symbols are called bits) at a certain data rate and the analog signal has to be converted into a series of these symbols for digital communication. The data rate of a digital network corresponds to the channel bandwidth of an analog network. The higher the data rate, the wider the required bandwidth and vice versa.

If the network is fully digital, analog voice is encoded into digital form at the transmitting end and decoded into analog form at the receiving end, as shown in Figure 3.6. This coding is performed in the subscriber interface of a digital telephone exchange and, in the case of ISDN service, in the subscriber's ISDN telephone or network terminal.

This process has two main phases, as shown in Figure 3.6:

1. *Analog-to-digital conversion (A/D):* An analog signal is sampled at the sampling frequency and the sample values are then represented as numerical values by the encoder. These values, presented as binary words, are then transmitted within regular time periods through the digital channel.
2. *Digital-to-analog conversion (D/A):* At the other end of the channel, the decoder receives numerical values of the samples that indicate the values of the analog signal at sampling instants. The sample pulses that have amplitudes corresponding to the values of the original signal at sampling instants are reconstructed and the series they form is filtered to produce an analog signal close to the original one.

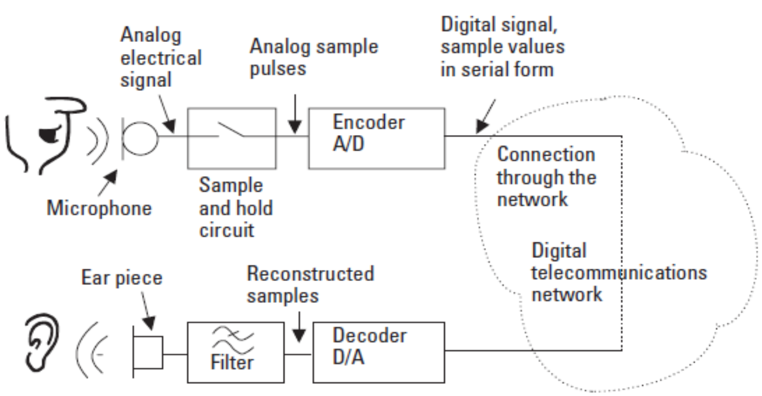

Figure 3.6 Analog voice signal through a digital network.

The methods for these A/D and D/A conversions have to be specified in detail so that the reproduction of the analog signal is compatible with the production of the digital signal that may have occurred on the other side of the world. In the next section we describe the method that is used in the telecommunications network and internationally standardized by the ITU.

3.6 PCM

PCM is a standardized method that is used in the telephone network to change an analog signal to a digital one for transmission through the digital telecommunications network. The analog signal is first sampled at a 8-kHz sampling rate; then each sample is quantized into 1 of 256 levels and then encoded into digital eight-bit words. This encoding process is illustrated in Figure 3.7. The overall data rate of one speech signal becomes $8{,}000 \times 8 = 64$ Kbps. This same data rate is available for data transmission through each speech channel in the network. In the United States one bit of eight in every sixth frame is "robbed" for in-band signaling and the available transparent data capacity of a single speech channel in the network is reduced to $8{,}000 \times 7 = 56$ Kbps.

Now we take a more detailed look at the three main processing phases of the PCM in the telecommunications network. Note that this principle is employed by all systems when there is a need to process analog signals with a digital system. Sampling rates and the number of quantizing levels vary from application to application, but the basic principle and phases of the process remain the same.

3.6.1 Sampling

The amplitude of an analog signal is sampled first. The more samples per second there are, the more representative of the analog signal the set of samples

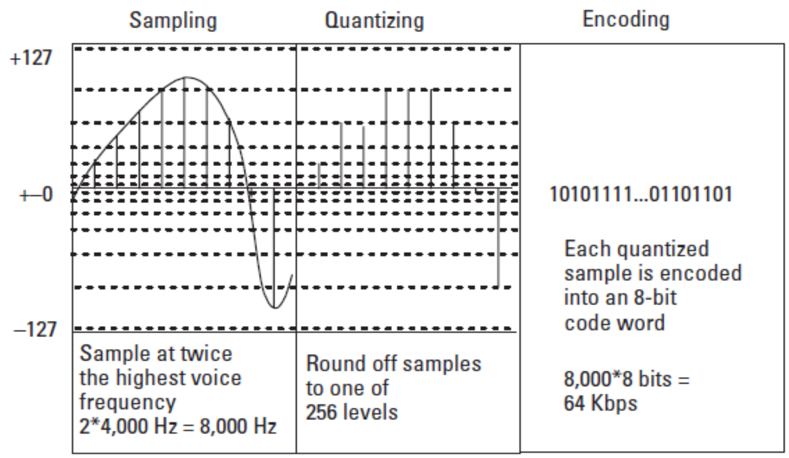

Figure 3.7 PCM.

will be. After sampling, the signal value is known only at discrete points in time, called sampling instants. If these points have a sufficiently close spacing, a smooth curve drawn through them allows us to interpolate intermediate values to any degree of accuracy. We can therefore say that a continuous curve can be adequately described by the sample values alone.

In a similar fashion, an electrical signal can be reproduced from an appropriate set of instantaneous samples. The number of samples per second is called the sampling *frequency* or sampling rate, and it depends on the highest frequency component present in the analog signal. The relation of sampling frequency and the highest frequency of the signal to be sampled is stated as follows:

> If the sampling frequency, f_s, is higher than two times the highest frequency component of the analog signal, W, the original analog signal is completely described by these instantaneous samples alone; that is, $f_s > 2W$.

This minimum sampling frequency is sometimes called the *Nyquist rate*. We can describe it in other words as an analog signal with the highest frequency component as W Hz. It is completely described by instantaneous sample values uniformly spaced in time within a period:

$$T_S = 1/f_S < 1/(2W) \tag{3.4}$$

Figure 3.8 represents the operating principle of a sampling circuit and an analog signal before and after sampling in both the time and frequency domains. The sampling circuit contains a generator, G, that produces short sampling pulses at the sampling frequency f_s. These sampling pulses close the switch of a relay at each sampling instant for a short period of time. The original analog signal $x(t)$ is sampled each time the switch is closed and a sampled signal $y(t)$ is produced. The sampled analog signal $y(t)$ contains short pulses that represent signal $x(t)$ values at discrete points in time. This sampling process that produces $y(t)$ is known as *pulse amplitude modulation* (PAM) because the amplitudes of the pulses contain the values of $x(t)$.

The time-domain curves in Figure 3.8 show the original continuous analog signal $x(t)$ and the sampled signal $y(t)$. The sampled signal $y(t)$ contains values of an analog signal at sampling instants. We can imagine that if the sampling frequency f_s is high, that is, the distance between sampling instants T_s is short, the sample pulses describe the original signal quite well. We could draw a line that connects the peak values of the pulses and the shape of this curve would be close to the original signal shape of $x(t)$.

The changes in $x(t)$ are related to the frequency content of $x(t)$. The more rapidly $x(t)$ changes, the higher frequency the components it contains. This explains why the sampling frequency is related to the highest frequency of the analog signal to be sampled. From the time-domain figure we understand that the sampling frequency must be much higher than the highest

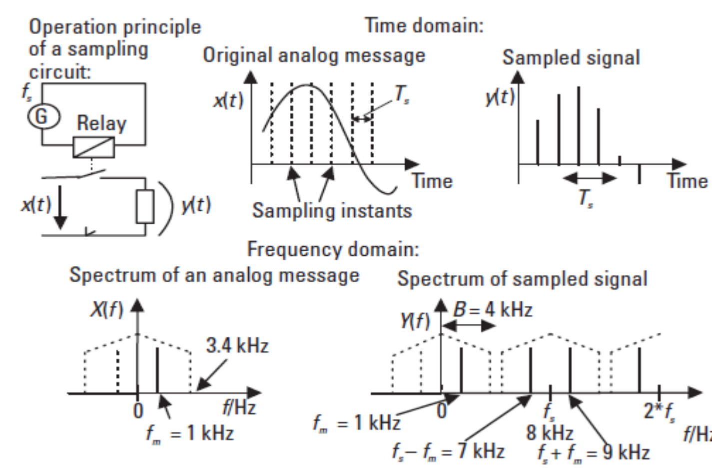

Figure 3.8 Sampling.

frequency of the analog message. Otherwise, rapid changes of signal $x(t)$ between sampling instants could not be described by sample values. The accurate answer to how much higher it should be can be understood more easily via the frequency domain.

The frequency-domain descriptions in Figure 3.8 show the spectrum of $x(t)$ and the sampled signal $y(t)$. Before sampling, the spectrum $X(f)$ of $x(t)$ contains speech frequencies up to 3.4 kHz, shown as a dashed line in the figure. As an example of the frequency components of speech we drew the spectrum of a 1-kHz cosine wave as a solid spectral line at the 1-kHz point on the frequency axis.

After sampling, the spectrum of the message also appears around the sampling frequency. If the message contains a single 1-kHz frequency component, after sampling we will have components at 1 kHz, 8 kHz – 1 kHz = 7 kHz, and at 8 kHz + 1 kHz = 9 kHz, as seen in the figure. In addition to these components, sampling also generates components around double sampling frequency, three times sampling frequency, and so forth.

The reproduction of an original signal from a sampled signal is performed by a lowpass filter and in the case of voice the bandwidth $B = 4$ kHz, that is, half the sampling frequency. We see from Figure 3.8 that this filter would let through only a 1-kHz component of the spectrum, that is, the actual original analog signal. With the help of the lowpass filter we have successfully reproduced the original analog message from the samples alone.

If we increase the frequency of an analog message $x(t)$ from 1 to 2 kHz we will have the lowest component of the sampled signal at 2 kHz, the solid spectral line at 1 kHz is moved to the right, the next spectral component at 8 kHz – 2 kHz = 6 kHz, and the solid line at 7 kHz is moved to the left. Lowpass filtering will still give the original 2-kHz message. Now if we increase the frequency beyond 4 kHz to, say, 5 kHz, we will get components at 5 kHz and 8 kHz – 5 kHz = 3 kHz, and lowpass filtering will give a 3-kHz signal instead of the original 5-kHz signal. Reproduction will not work anymore because the frequency of the analog signal has exceeded half of the sampling frequency.

We have seen that the sampling frequency must be more than twice the highest frequency component of the original signal to be encoded; otherwise, the message spectra around zero frequency and sampling frequency will overlap. This can be seen from the spectrum $Y(f)$ in Figure 3.8 if we imagine what happens if $W > f_s/2$. From the spectrum of the sampled signal $Y(f)$ in Figure 3.8, we also see that the message can be completely reconstructed from a PAM signal with a 4-kHz lowpass filter if $W < f_s/2$. This requirement is fundamental for all digital signal processing.

The highest frequency of voice that will be transmitted is chosen to be 3,400 Hz and the sampling frequency is standardized at 8,000 Hz, leaving enough guard band for filtering. Samples are then taken at intervals of $T_s =$ 125 μs.

In the sampling process a PAM signal $y(t)$ is created. The amplitudes of PAM pulses follow the original analog signal. Note that the samples are still analog, having any analog value between the minimum and maximum values of the original signal.

3.6.2 Quantizing

In the previous section we utilized sampling that produces a PAM signal that represents discrete but still analog values of the original analog message at the sampling instants. To transmit the sample values via a digital system, we have to represent each sample value in numerical form. This requires quantizing where each accurate sample value is rounded off to the closest numerical value in a set of digital words in use. Figure 3.9 represents the original and the quantized signal. The latter stays at the sample value until the next sampling instant.

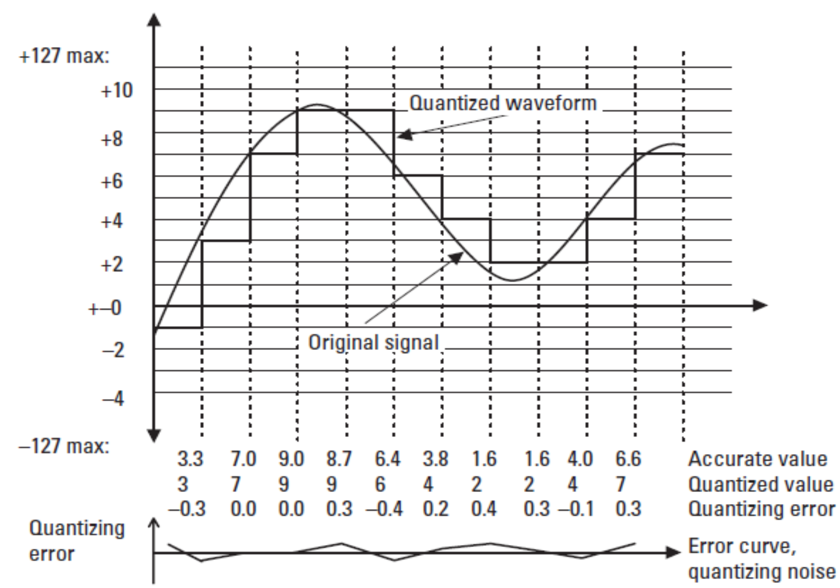

Figure 3.9 Quantizing and noise.

In this quantizing process the information in accurate signal values is lost because of rounding off and the original signal cannot be reproduced exactly any more. The quality of the coding depends on the number of quantum levels that is defined to provide the required performance. The more quantum levels we use, the better performance we get. For example, for a voice signal 256 levels (8-bit binary words) are adequate, but for music encoding (CD recording) 65,536 levels (16-bit binary word) are needed to give sufficient performance.

In the case of binary coding, the number of quantum levels is $q = 2^n$, where q denotes the number of quantum levels and n is the length in bits of the binary code words that describe the sample values.

The better quality we require, the more quantum levels we need and the longer sample words we have to use. This leads to the requirement of a higher bit rate for transmission of the data representing the original message. The data rate must be so high that the digital word of the previous sample will be transmitted before the next one is available for transmission. In each system, a certain compromise has to be made between quality and the data rate.

In uniform quantizing, the quantum levels are uniformly spaced between certain minimum and maximum values of the analog signal. In the next section we consider quantizing noise that the rounding off produces in the case of uniform quantizing.

3.6.3 Quantizing Noise

Quantizing causes signal distortion because the sample values no longer represent accurate values of the analog signal. Usually this distortion caused by rounding off in quantizing is small compared to the signal value. The maximum distortion, that is, maximum quantizing error, is half of the distance between quantum levels. This distortion is heard and theoretically modeled as noise; see the quantizing error curve in Figure 3.9. We can imagine that the decoder first receives accurate sample values and produces a perfect original signal. Then quantizing error is added on top of the perfect signal just as we hear, for example, background noise on top of an ideal voice or music signal.

The rounding off causes an error that is independent of the message because quantizing levels are close to each other and we can assume that the signal has the same probability to be anywhere between two levels at a certain sampling instant as shown in Figure 3.10. This error can be assumed to have a uniform probability density function and a zero mean. When we define the

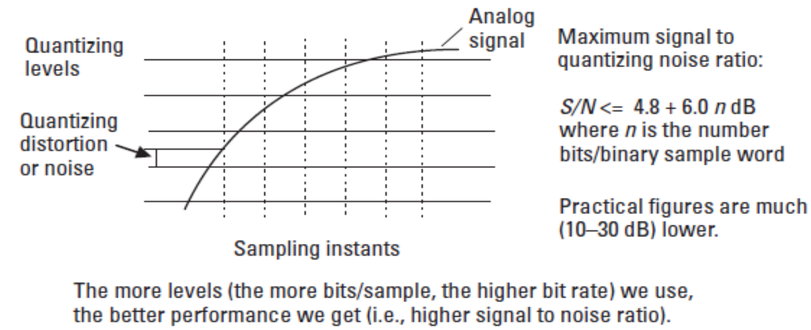

Figure 3.10 Quantizing noise and SQR.

signal to have values between –1 ... +1, it can be shown that the quantum noise power is equal to the variance of quantizing error and is given by

$$N = \sigma_q^{\ 2} = \frac{1}{3q^2} \tag{3.5}$$

where $N = \sigma_q^2$ = quantization noise power and q = the number of quantum levels. [Equation (3.5) gives the variance σ_q^2 of uniform distribution with a value of $q/2$ from $-1/q$ to $1/q$. The variance corresponds to the noise power N when the mean is zero.]

We see that if the number of quantum levels is increased, quantizing noise power decreases rapidly. We get the maximum *signal-to-quantizing noise ratio* (SQR) of linear quantizing when the maximum signal power is equal to one (power is a square of the signal value that was defined to be between –1 and +1):

$$SQR = S/N \le 3q^2 \tag{3.6}$$

where S = signal power, $N = \sigma_q^2$ = power of quantization noise, and q = number of quantum levels. The only noise we consider here is generated by quantizing and then SQR = S/N.

We can easily show further that in the case of linear quantizing and binary words, the absolute maximum S/N in decibels in the case of linear quantizing is

$$S/N \le 10\log_{10}\left(3q^2\right) = 10\log_{10}\left(3\cdot 2^{2n}\right) = 4.8 + 6.0n \text{ dB} \tag{3.7}$$

where n = the number of bits/word. The maximum S/N is achieved with the maximum signal power that is 1. The logarithmic measure decibel is described at the end of this chapter. The preceding formula gives the absolute maximum S/N of a system that uses uniform quantizing and codes sample values into n-bit binary words.

If we add one bit to the data word representing a linear sample value, we double the number of quantizing levels, which cuts the maximum quantizing error in half. On the other hand, from (3.7) we see that each bit increases the S/N by 6 dB. This means that the quantizing noise power is reduced by a factor of 4 corresponding to error voltage reduction by a factor of 2.

However, we assumed that the average power of the analog signal equals the maximum power, that is, all sample words have the maximum value. In practice, this cannot be the case and the average S/N is some tens of decibels lower than the maximum value given by (3.7). How much lower an average S/N we have in a practical system depends on the dynamic range that we reserve for the highest signal levels (the distance between the average signal power and the maximum signal power) to avoid the clipping of the signal and consequent severe distortion. As an example, if average signal power is 20 dB below maximum, the average S/N (or SQR) is 20 dB below its maximum value given by (3.7).

3.6.7 PCM Encoder and Decoder

The PCM coding schemes for digital voice communications were standardized by CCITT (now ITU-T) in the early 1970s. The standards were based on the technology of those days. The European standard was defined to be slightly different from the American standard, which is why conversion equipment is needed when communicating over the Atlantic or from Europe to Japan. Most countries in the world use the European A-law standard. As a conclusion to our discussion about PCM coding, we now look at the block diagrams of the PCM encoder and decoder that contain the processes that we have discussed in previous sections.

3.6.7.1 PCM Encoder

Figure 3.15 presents a block diagram of a PCM encoder based on the European standard. Before actual encoding, the analog signal is filtered into the frequency band from 300 to 3,400 Hz. This bandwidth was defined to be acceptable for sufficient quality human voice so that the speaker can be recognized at the other end. This filtering is mandatory to ensure that the

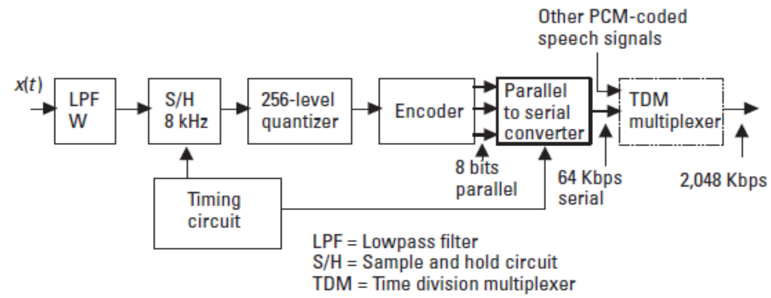

Figure 3.15 PCM encoder.

sampling theorem is satisfied, that is, that the analog signal does not contain frequencies higher than half of the sampling frequency. Then the analog signal is sampled at an 8-kHz sampling frequency and the samples are nonlinearly coded into 8-bit words by a quantizer and an encoder.

Words are then converted into serial form and multiplexed with other PCM-coded voice signals into a 2,048-Kbps primary rate signal that contains 30 voice channels according to the European standard. This 2-Mbps rate is a very common data rate in telecommunications networks. For example, digital exchanges build up 2-Mbps streams with 30 PCM-coded subscriber interfaces for internal transmission inside the equipment. The multiplexing process is described in Chapter 4.

In the United States the corresponding data rate is 1.544 Mbps instead of 2.048 Mbps. In this DS1 system, each frame contains 24 speech channels and a framing bit. The sampling rate is the same 8 kHz and we get

$$8{,}000 \cdot \{(8 \cdot 24) + 1\} = 1.544 \text{ Mbps} \tag{3.10}$$

3.6.7.2 PCM Decoder

At the receiver the demultiplexer separates 64-Kbps individual channels that are then converted into 8-bit parallel sample values, as shown in Figure 3.16. Sample pulses are reconstructed and the resulting series is filtered to create a voice signal that closely resembles the original.

3.7 Other Speech-Coding Methods

PCM was standardized during the 1970s and implementation of many more efficient coding methods has become feasible. By "more efficient" we mean

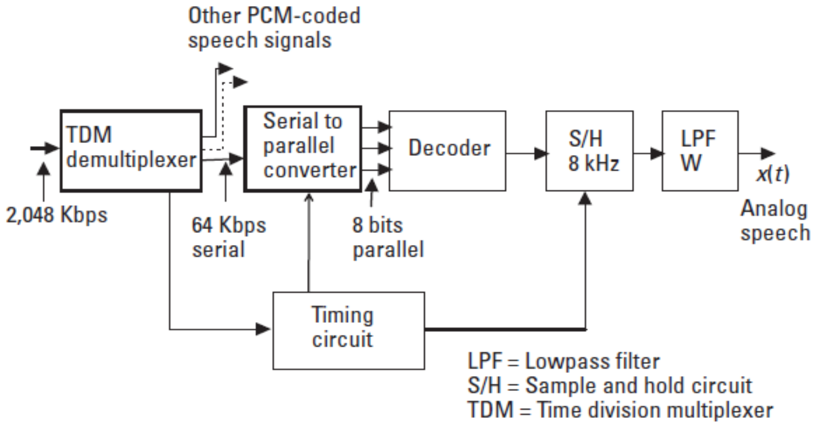

Figure 3.16 PCM decoder.

that we may get better quality at the same data rate or equal quality at a lower data rate. More sophisticated coding schemes are used, for example, in ISDN, where an ISDN telephone may transmit a better quality 7-kHz speech band at 64 Kbps than before. Another example that we will briefly review is GSM, where speech requires only 13 or 7 Kbps.

In the following sections, we review some methods that are used in telecommunications networks in addition to the PCM discussed in the previous section. We can divide voice coding methods into two categories: waveform and voice coding (vocoders) [1]. In waveform coding, such as PCM, we transmit information that describes a signal waveform in time domain.

In vocoders we use characteristics of human voice. To understand the basic principle of vocoders, imagine that we have a set of signal models each identified by a code. We divide speech into, for example, 50-ms segments and choose one of the models that is closest to the signal to be encoded and send its identification code to the other end. The decoder reproduces the signal corresponding to the received code. Vocoders may also split voice signals into several "components" in the frequency domain, each of them modeled separately for better quality. Vocoders introduce additional delay because each speech segment has to be analyzed before encoding. Waveform coding does not add delay and it usually give better quality but requires a higher data rate than vocoders. To achieve a suitable compromise between the quality and the data rate, the two basic principles are sometimes combined into hybrid coders.

In conventional PCM we encode all samples independently. We can improve encoding performance by assuming that the next sample value is not independent from the previous one, which is the case in practice.

3.7.1 Adaptive PCM (APCM)

APCM is a variation of conventional PCM in which signal strength information is transmitted periodically in addition to sample values. Now a smaller number of bits is needed for samples and they define the quantum level inside a given scale. If the signal level is high, the quantizing error is high because the same number of levels is used for all samples. On the other hand, for low signal levels the quantizing error is small and SQR can be kept high enough over a wide range of signal levels. This principle is used, for example, in original GSM as part of the voice coding process.

3.7.2 Differential PCM (DPCM)

In DPCM only the difference between a sample and the previous value is encoded as shown in Figure 3.17. Because the difference is typically much smaller than the overall value of the sample, we need fewer bits for the same accuracy as in ordinary PCM and the required bit rate is reduced [1]. In the example shown in Figure 3.17, PCM requires 5 bits (polarity and 4 bits for 16 quantum levels). DPCM, in which the only difference from the previous sample is encoded, 4 bits is clearly enough to describe the difference between subsequent samples.

For better quality or to further reduce the data rate, DPCM may use several of the preceding samples to predict the next sample. The example in Figure 3.17 shows that if two previous samples are used for prediction the encoder and decoder assume that the next sample follows the same slope. Now, 3 bits would be enough for the encoder to describe the difference between the prediction and the actual sample value. The decoder performs the same prediction and only the difference between predictions shown in Figure 3.17 need to be transmitted. Further improvement can be achieved if three previous sample values are used for prediction, but more than three samples do not add much advantage [2]. Actually, the first simple form of DPCM in Figure 3.17, which encodes the difference between preceding sample values, uses prediction as well but that prediction is based on only one sample and it equals the previous sample value.

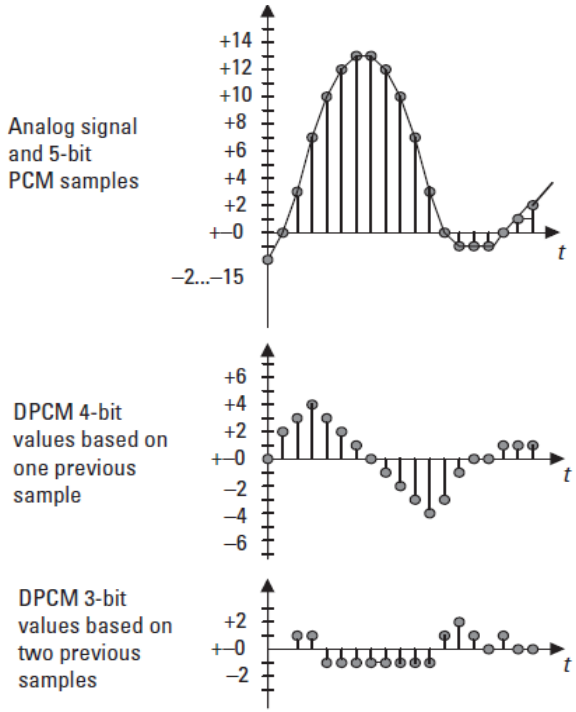

Figure 3.17 DPCM.

These waveform coding methods do not introduce much delay because prediction is based on previous sample values. DPCM methods require that absolute sample values be transmitted periodically to prevent propagation of errors. DPCM is sometimes used for digitized video transmission.

3.7.3 DM

DM is a very simple type of DPCM that transmits the binary value 1 if the sample is higher than the previous one. Binary value 0 is transmitted if the signal value has decreased. A variation of DM uses large quantizing steps when the signal contains steep slopes and small steps when the signal does not change much. This method is called *continuous variable slope delta* (CVSD) modulation and it is an alternative to ordinary PCM in Bluetooth speech transmission. CVSDM is also used in military voice applications [3].

3.7.4 Adaptive DPCM (ADPCM)

ADPCM combines two previously described methods, APCM and DPCM. Further compression is achieved by adapting the predictor and the quantizer to the characteristics of the signal. Both the encoder and the decoder use the same algorithm to estimate the values of the following samples with help of the preceding samples, and only the error to this estimate is transmitted as in DPCM in Figure 3.17. To further reduce the number of bits per sample, ADPCM adapts quantizing levels to the characteristics of the analog signal. Figure 3.18 shows a simplified example in which the prediction error is initially small and all bits can be used for half of the full quantizing error scale. Then the prediction error increases and the quantizing step size is doubled to describe higher values of prediction error. When the prediction error decreases, the quantizing step size is reduced again to describe properly small errors. Adaptation information is transmitted from encoder to decoder in addition to the prediction error.

In the original 32-Kbps ADPCM method, the difference between the predicted and actual sample value is coded with four bits, that is, into 15 quantum levels, and the data rate is half that of conventional PCM. If several subsequent samples vary widely, the quantizing steps are adapted to that change so that four bits are enough for prediction error. If prediction errors tend to increase, quantizing steps are increased and vice versa.

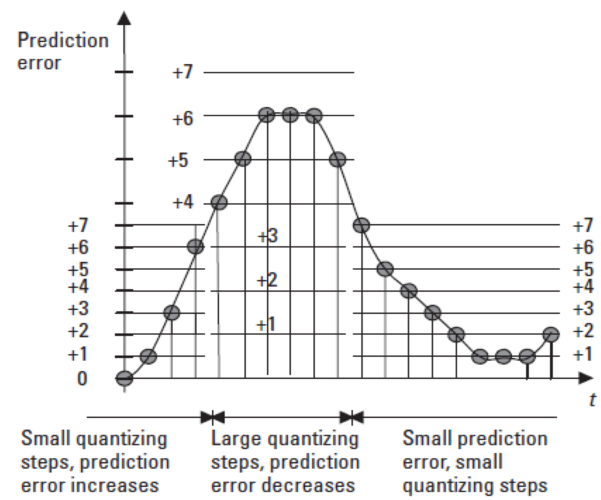

Figure 3.18 ADPCM principle.

According to the ADPCM standard, commercial voice quality is coded into 32 Kbps or even a lower (24 or 16 Kbps) bit rate. Samples are still taken at 8 kHz but transmitted with four bits (in the case of 32-Kbps ADPCM) and the quality is equal, or at least close, to the quality of ordinary PCM.

Recommendation G.728 for 16/24/32/46-Kbps ADPCM was approved by the ITU-T in 1990 and has been adopted worldwide for digital voice transmission between countries or within a country. It can partly resolve the current compatibility problem between North America's and Europe's PCM formats, due to their different companding schemes, by acting as a common language between the two PCM schemes.

There is also a recommendation for an ADPCM algorithm (G.722) that will code 7.1-kHz bandwidth audio signals into 64 Kbps. This coding scheme improves the quality of speech and it can be used for good quality voice over ISDN networks.

ADPCM systems are available on the market that convert two primary rate PCM streams into one data stream at the same rate by using ADPCM. Two 32-Kbps ADPCM channels occupy one ordinary PCM channel. Network operators use ADPCM to utilize long-distance transmission systems, for example, submarine systems, more efficiently. Another application example is in PABX networks where the offices of private enterprises are interconnected by leased-line 64-Kbps channels. ADPCM doubles the capacity of these expensive leased lines between PBX/PABXs. One application for ADPCM is also in cordless telephones such as *digital enhanced cordless telecommunications* (DECT).

The ADPCM coding scheme is based on the statistics of speech and it does not support modem or facsimile signals at higher data rates than 4,800 bps. Because of this, telecommunications network operators cannot use ADPCM instead of PCM coding for all calls. This is a problem if ADPCM systems are used inside a telecommunications network. One way to overcome this problem is to have the ADPCM encoder detect whether a data or facsimile connection is to be established and in that case disable the PCM/ADPCM transcoder for that channel.

Up to this point we have discussed primarily waveform coding methods. The phrase *waveform coding* refers to attempts to describe the shape or the waveform of the original analog signal, just as PCM, DPCM, and ADPCM do. In a more efficient coding scheme in terms of data rate, such as *voice coding*, which is implemented by vocoders, we divide speech into segments with lengths of some tens of milliseconds. Then we analyze each segment to find a model that describes it best and send parameters of the model instead of trying to imitate the shape of the signal. An example of the so-called hybrid methods that use both of the two main principles discussed earlier is the voice coding of a cellular network, as briefly reviewed next.

3.7.5 Speech Coding of GSM

In cellular networks an efficient coding scheme is needed in order to make maximum use of radio frequencies. The lower our data rate, the narrower the frequency band we need for each call and the more simultaneous calls a given frequency band supports, as we will see in Chapter 4. As an example of these efficient coding schemes we now briefly review the principle that is used in the GSM.

During efforts to standardize the speech-coding algorithm for GSM, the goal was to achieve a 16-Kbps data stream with the same speech quality as ordinary PCM. Waveform coding, such as PCM or ADPCM, did not give sufficient quality at this low data rate. Voice coding methods did give a low enough data rate but not good enough quality. In a voice coder or vocoder, the signal is modeled and the codes of the sound elements are sent. In the decoder the speech is reproduced.

A combination of these two basic principles was selected. The maximum processing delay was restricted to less than or equal to 65 ms, which requires the use of echo cancellers in the network. The original data rate became 13 Kbps, which was further reduced to 7 Kbps in 1995 with a more efficient coding algorithm.

The selected efficient speech coding is always used at the radio path where efficient utilization of transmission channels is more important than in the wireline network. We will see in Chapter 5 that to increase the radio interface capacity we need to make cells smaller and build more base stations, which is very expensive. For switching and interconnection to a fixed telecommunications network, GSM coding is changed into ordinary PCM.

GSM's operating principle is as follows. The voice signal is first divided into 20-ms slices. Each slice of the signal is analyzed and the periodicity is noticed. The periodical component is subtracted by an analysis filter from the original signal and the amplitude of the voice signal level is considerably reduced (Figure 3.19).

The periodical high-power component is transmitted as a set of parameters, and the low-level error or difference signal at the output of the analysis filter is waveform coded. This waveform coding does not require a high bit rate because the amplitude of the error signal is low.

At the receiving end, a synthesis filter is used and, with the help of the transmitted coefficients, it adds the periodical component to the error signal, which is reproduced from waveform-coded samples.

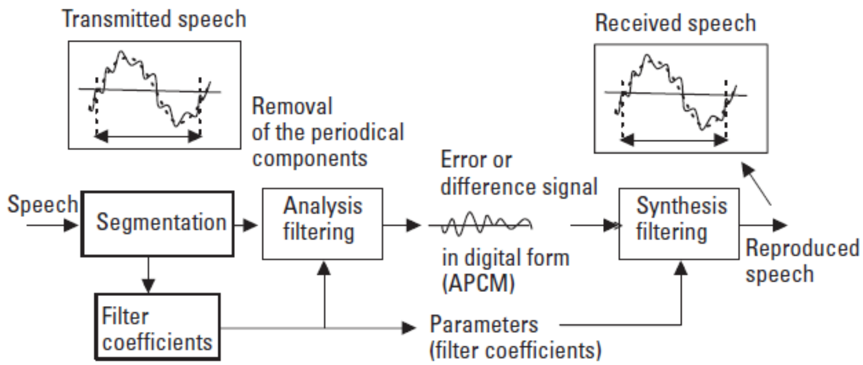

Figure 3.19 The principle of GSM speech coding.

3.9 Problems and Review Questions

Problem 3.1

Explain how the characteristics of digital data and voice communications differ.

Table 3.2
Data Sequence for Digital Milliwatt

Word	Bit Number							
	1	2	3	4	5	6	7	8
1	0	0	1	1	0	1	0	0
2	0	0	1	0	0	0	0	1
3	0	0	1	0	0	0	0	1
4	0	0	1	1	0	1	0	0
5	1	0	1	1	0	1	0	0
6	1	0	1	0	0	0	0	1
7	1	0	1	0	0	0	0	1
8	1	0	1	1	0	1	0	0
1	0	0	1	1	0	.	.	.
.	0	0	.	.	.	.	.	.

Problem 3.2

What is the wavelength λ of the radio signal for (a) a 100-MHz FM radio and (b) a 10-GHz microwave radio relay system?

Problem 3.3

A voltage waveform of a signal follows the equation $x(t) = 5\cos(1 \cdot 10^3 t)$ V, where t = time. What are the frequency, amplitude, radian frequency, and periodic time (period) of this signal?

Problem 3.4

Draw the signal $v(t) = 5\cos(1 \cdot 10^3 t + \pi/2)$ V. The vertical scale should be in volts and the horizontal scale in milliseconds.

Problem 3.5

Compare digital telecommunications technology with analog technology and list the most important advantages of digital technology.

Problem 3.6

What are the main three phases of PCM encoding (A/D conversion)? Explain how they are performed.

Problem 3.7

What is nonuniform quantizing and why is it used?

Problem 3.8

What is the minimum sampling rate of speech when the frequency band is 300 to 3,400 Hz and what is the minimum sampling frequency for high-fidelity music of 20 Hz to 20 kHz?

Problem 3.9

Draw the spectrum of an analog signal after sampling when the sampling frequency is 8 kHz and the signal that is sampled is a sine wave with a frequency of 1 kHz. Draw the spectrum for each case when the analog signal frequency is 2, 5, and 6 kHz. What happens in each case when we reconstruct the original signal from the sampled signal with a lowpass filter that has a bandwidth of 4 kHz?

Problem 3.10

The digital *compact disc* (CD) player is designed for a sound bandwidth of 20 kHz. Linear encoding with 16 bits per sample is used. Define (a) the minimum sampling rate, (b) the minimum binary data rate per channel (left or right), (c) the maximum SQR, and (d) the average SQR if the average signal level is 30 dB below the maximum value.

Problem 3.11

Estimate what bit rate would be needed for each voice channel in the digital telephone network if linear PCM coding is used. The same performance, with SQR at least 40 dB at signal levels higher than –40 dBm0 (sine wave), is required. (*Hint:* Estimate with the help of Figure 3.13 how much quantizing noise should be reduced at signal level –40 dBm0 and how much longer sample words would be required for this.)

Problem 3.12

How much PCM voice or stereo music (assume that CD-quality music requires 700 Kbps for both channels) can be stored in (a) a 1.44-MB (B = byte = 8 bits) disc and (b) a 20-GB memory space of a hard disk?

Problem 3.13

Explain how (a) DPCM and (b) ADPCM reduce data rates compared to ordinary PCM.

Problem 3.14

Explain the basic principle of GSM speech coding.

Problem 3.15

The input power of an amplifier is 2 mW and output power is 1W. What are the power levels (dBm) at the input and output and what is the gain of the amplifier in decibels?

Problem 3.16

The input and output powers of a circuit are in listed in Table 3.3. What is the absolute attenuation *L*, absolute gain *g*, attenuation in decibels, gain in decibels, and output power level for each case (a–e)?

Problem 3.17

What is the power in watts that corresponds to power levels of (a) 0 dBm, (b) 3 dBm, (c) –3 dBm, (d) 10 dBm, (e) 20 dBm, (f) 100 dBm, and (g) –100 dBm?

Problem 3.18

Figure 3.24 illustrates a telecommunications connection using a geostationary satellite. Calculate the input and output powers of the satellite amplifier and output power of the antenna at the receiving Earth station. Define both power levels in dBm and absolute power in watts. Use decibels and derive power levels, in dBm values, first.

Problem 3.19

The input power of a 40-km cable system is 2W (power at the beginning of the cable). An amplifier with a 64-dB gain is installed 24 km from the input. Define the signal power level, dBm, and absolute power at (a) the input of the amplifier and (b) the output of the system. The attenuation of the cable is 2.5 dB/km.

Table 3.3
Input and Output Powers of a Circuit

	P_{in}	P_{out}
(a)	1 mW	1 mW
(b)	1 mW	0.5 mW
(c)	1 mW	4 mW
(d)	10 mW	10 W
(e)	10 W	10 mW

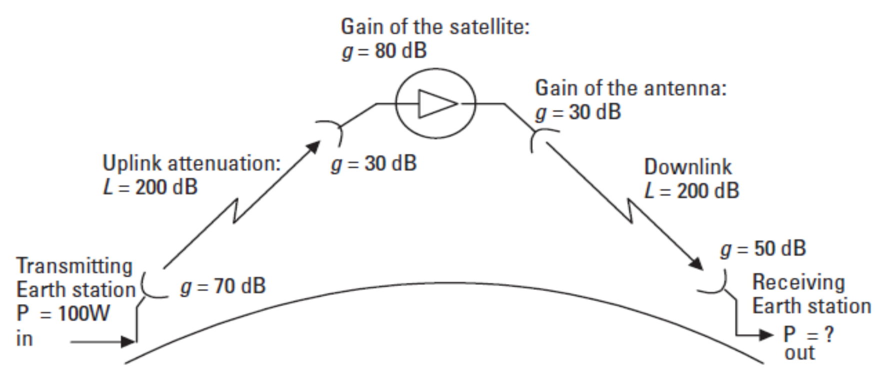

Figure 3.24 Satellite transmission link.