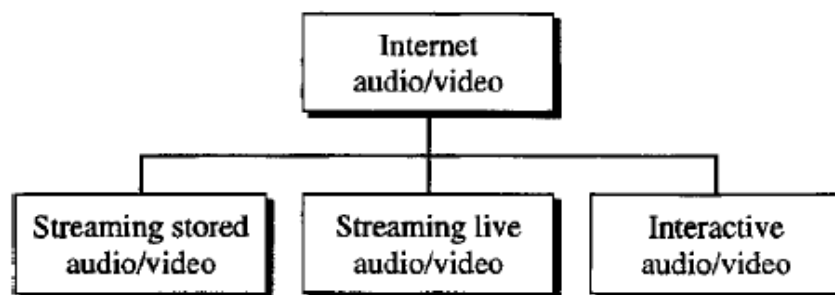


## ***Multimedia***

Recent advances in technology have changed our use of audio and video. In the past, we listened to an audio broadcast through a radio and watched a video program broadcast through a TV. We used the telephone network to interactively communicate with another party. But times have changed. People want to use the Internet not only for text and image communications, but also for audio and video services. In this chapter, we concentrate on applications that use the Internet for audio and video services. We can divide audio and video services into three broad categories: streaming stored audio/video, streaming live audio/video, and interactive audio/video, as shown in Figure 29.1. Streaming means a user can listen to (or watch) the file after the downloading has started.



---

Figure 1 *Internet audio/video*

In the first category, streaming stored audio/video, the files are compressed and stored on a server. A client downloads the files through the Internet. This is sometimes referred to as on-demand audio/video. Examples of stored audio files are songs, symphonies, books on tape, and famous lectures. Examples of stored video files are movies, TV shows, and music video clips. Streaming stored audio/video refers to on-demand

requests for compressed audio/video files. In the second category, streaming live audio/video, a user listens to broadcast audio and video through the Internet. A good example of this type of application is the Internet radio. Some radio stations broadcast their programs only on the Internet; many broadcast them both on the Internet and on the air. Internet TV is not popular yet, but many people believe that TV stations will broadcast their programs on the Internet in the future. Streaming live audio/video refers to the broadcasting of radio and TV programs through the Internet. In the third category, interactive audio/video, people use the Internet to interactively communicate with one another. A good example of this application is Internet telephony and Internet teleconferencing. Interactive audio/video refers to the use of the Internet for interactive audio/video applications. We will discuss these three applications in this chapter, but first we need to discuss some other issues related to audio/video: digitizing audio and video and compressing audio and video.

## **DIGITIZING AUDIO AND VIDEO**

Before audio or video signals can be sent on the Internet, they need to be digitized. We discuss audio and video separately. Digitizing Audio

When sound is fed into a microphone, an electronic analog signal is generated which represents the sound amplitude as a function of time. The signal is called an *analog audio signal*. An analog signal, such as audio, can be digitized to produce a digital signal. According to the Nyquist theorem, if the highest frequency of the signal is  $f$ , we need to sample the signal 21 times per second. There are other methods for digitizing an audio signal, but the principle is the same. Voice is sampled at 8000 samples per second with 8 bits per sample. This results in a digital signal of 64 kbps. Music is sampled at 44,100 samples per second

with 16 bits per sample. This results in a digital signal of 705.6 kbps for monaural and 1.411 Mbps for stereo.

## **Digitizing Video**

A video consists of a sequence of frames. If the frames are displayed on the screen fast enough, we get an impression of motion. The reason is that our eyes cannot distinguish the rapidly flashing frames as individual ones. There is no standard number of frames per second; in North America 25 frames per second is common. However, to avoid a *SECTION 29.2 AUDIO AND VIDEO COMPRESSION 903* condition known as flickering, a frame needs to be refreshed. The TV industry repaints each frame twice. This means 50 frames need to be sent, or if there is memory at the sender site, 25 frames with each frame repainted from the memory. Each frame is divided into small grids, called picture elements or pixels. For black-and-white TV, each 8-bit pixel represents one of 256 different gray levels. For a color TV, each pixel is 24 bits, with 8 bits for each primary color (red, green, and blue). We can calculate the number of bits in 1 s for a specific resolution. In the lowest resolution a color frame is made of 1024 x 768 pixels. This means that we need  $2 \times 25 \times 1024 \times 768 \times 24 = 944$  Mbps. This data rate needs a very high data rate technology such as SONET. To send video using lower-rate technologies, we need to compress the video. *Compresswn* is needed to send video over the Internet.

## **AUDIO AND VIDEO COMPRESSION**

To send audio or video over the Internet requires compression. In this section, we discuss audio compression first and then video compression.

**Audio Compression** Audio compression can be used for speech or music. For speech, we need to compress a 64-kHz digitized signal; for music, we need to compress a 1.41 I-MHz signal. Two categories of techniques are

used for audio compression: predictive encoding and perceptual encoding.

*Predictive Encoding* In predictive encoding, the differences between the samples are encoded instead of encoding all the sampled values. This type of compression is normally used for speech. Several standards have been defined such as GSM (13 kbps), G.729 (8 kbps), and G.723.3 (6.4 or 5.3 kbps). Detailed discussions of these techniques are beyond the scope of this book.

*Perceptual Encoding: MP3* The most common compression technique that is used to create CD-quality audio is based on the perceptual encoding technique. As we mentioned before, this type of audio needs at least 1.411 Mbps; this cannot be sent over the Internet without compression. MP3 (MPEG audio layer 3), a part of the MPEG standard (discussed in the video compression section), uses this technique. Perceptual encoding is based on the science of psychoacoustics, which is the study of how people perceive sound. The idea is based on flaws in our auditory system: Some sounds can mask other sounds. Masking can happen in frequency and time. In frequency masking, a loud sound in a frequency range can partially or totally mask a softer sound in another frequency range. For example, we cannot hear what our dance partner says in a room where a loud heavy metal band is performing. In temporal masking, a loud sound can numb our ears for a short time even after the sound has stopped. MP3 uses these two phenomena, frequency and temporal masking, to compress audio signals. The technique analyzes and divides the spectrum into several groups. Zero bits are allocated to the frequency ranges that are totally masked. A small number of bits are allocated to the frequency ranges that are partially masked. A larger number of bits are allocated to the frequency ranges that are not masked.

MP3 produces three data rates: 96 kbps, 128 kbps, and 160 kbps. The rate is based on the range of the frequencies in the original analog audio.

**Video Compression** As we mentioned before, video is composed of multiple frames. Each frame is one image. We can compress video by first compressing images. Two standards are prevalent in the market.

Joint Photographic Experts Group (JPEG) is used to compress images.

Moving Picture Experts Group (MPEG) is used to compress video. We briefly discuss JPEG and then MPEG.