Module
1

Multimedia Systems

Lesson 1

Introduction to Multimedia Systems and Processing

Instructional Objectives

At the end of this lesson, the students should be able to:

1. Define multimedia signal.
2. Name at least five different sources of multimedia signal.
3. State the motivation behind the growth of multimedia technology.
4. State the elements of multimedia communication system.
5. State at least five challenges involved with multimedia signal processing and communication.

Introduction

In the era of information technology, we are dealing with free flow of information with no barriers of distance. Take the case of internet as one of the simplest examples. You can view and download information across the globe within a reasonable time, if you have a good speed of connection.

Let us spend a bit of time to think in which form we access the information. The simplest and the most common of these is the printed text. In every web page, some text materials are always present, although the volume of textual contents may vary from page to page. The text materials are supported with graphics, still pictures, animations, video clippings, audio commentaries and so on. All or at least more than one of these media, which we can collectively call as "multimedia", are inevitably present to convey the information which the web site developers want to do, for the benefit of the world community at large. All these media are therefore utilized to present the information in a meaningful way, in an attractive style.

Internet is not the only kind of information dissemination involving multiple media. Let us have a look at some other examples as well. In television, we have involvement of two media – audio and video, which should be presented together in a synchronized manner. If we present the audio ahead of video or video ahead of audio in time, the results are far from being pleasant. Loss of lip synchronization is noticeable, even if the audio and the video presentations differ by just 150 milliseconds or more. If the time lead or the lag is in order of seconds, one may totally lose the purpose of the presentation. Say, in some distance learning program, the teacher is explaining something which is written down on a blackboard. If the audio and the video differ in time significantly, a student will not be able to follow the lecture at all.

So, television is also a multimedia and now, we understand one more requirement of multimedia signals. The multimedia signals must be synchronized and if it is not possible to make them absolutely synchronized, they should at

least follow a stringent specification by which lack of synchronization can be tolerated.

Television is an example, where there is only unidirectional flow of multimedia information – from the transmitter to the receiver. In standard broadcast television, there is no flow of information in the reverse direction, unless you use a different device and a channel –say, by talking to the television presenter over telephone. In internet of course, you have interactivity in the sense that you can navigate around the information and make selections through hyperlinks, but bulk of the flow of information is happening from the web server to the users.

In some applications, we require free flow of multimedia signals between two or more nodes, as is the case with video conferencing, or what we should more appropriately call a multimedia teleconferencing. In case of multimedia conferencing, the nodes physically located in different parts of a globe will be equipped with microphones, cameras, a computer supporting texts, graphics and animations and may be other supporting devices if required. For example, suppose five eminent doctors across the continents are having a live medical conference to discuss about a patient's condition. The doctors should not only see and talk to each other, all of them should observe the patient at the same time, have access to the patient's medical history, live readings and graphs of the monitoring instruments, visual rendering of the data etc. In a multimedia teleconferencing application of this nature, one must ensure that the end-to-end delays and the turnaround time is minimum. Moreover, the end-to-end delays between different nodes should not be different very significantly with respect to each other, that is, the delay jitters must be small.

1.1 Multimedia Signals

So, are we now in a position to formally define a multimedia signal? At first, let us digress from multimedia and only try to define "signal". Signal is something that varies with respect to time and/or space, or with respect to some other quantity and represents something that is meaningful. It is very difficult, if not impossible to arrive at a definition that is complete, but we can accept the above definition to ease our understanding. If we can accept the above definition of "signal", we can attempt to define "multimedia signal" in the following way:

A multimedia signal is one that integrates signals from several media sources, such as video, audio, graphics, animation, text in a meaningful way to convey some information.

In the above definition, the stress is on the word "meaningful". For example, you may pick up different words at random from a dictionary one after the other and try to compose a sentence, but we can't call it a sentence unless it conveys a meaning. Same is the case with "multimedia signal". We can't just create a random mixture of audio, video, graphics, animation etc. in any manner we like

and claim it to be a multimedia presentation. The integration of different media must be done with some definite objective. And then, just like definite grammatical rules must be followed in a sentence, integration of different media must necessarily follow some rules and regulations, which are specified in the standards.

1.2 Motivation behind the growth of multimedia technology

So, what motivated the growth of multimedia technology? First and foremost, multimedia is an effective tool of communication. To express an idea, we need to use as many media as we may have at our disposal. A teacher cannot be called a good teacher if he/she only talks to the students without using a blackboard. Again another teacher, who only writes his lecture materials on the blackboard without verbally communicating with the students, is also not a good teacher, but he/she is definitely a better teacher than the earlier examples, if he/she uses the blackboard and at the same time, verbally interacts with the students. For an even better teaching, the teacher can take the help of overhead projectors, video recordings etc. The need for multimedia presentations have always been there, but few years back, say only 15-20 years earlier than today, we did not have sophisticated multimedia technology at our fingertips. Computers were costly; they were much less powerful than what it is today. The communication technology did not go through the recent revolutions and suffered from the problems of bandwidth limitations, unreliability, and moreover, very costly. It was then a very tough task, if not impossible to combine multiple media into an integrated data stream, given the limitations of processing speed, processing tools and the communication channels.

It is not a single factor alone that prompted the growth of multimedia technology. In recent years, advancements of processing speed, availability of very large-scale integrated circuits for real-time processing, effective data compression tools and algorithms to eliminate the redundancy and achieve bandwidth reduction, advancement of computer networking etc. have made the tasks of utilizing and integrating multiple media a reality and made "multimedia" a popular buzzword of today.

1.3 Elements of multimedia communication systems

For multimedia communication, we have to make judicious use of all the media at our disposal. We have audio, video, graphics, texts - all these media as the sources, but first and foremost, we need a system which can acquire the separate media streams, process them together to make it an integrated multimedia stream.

Now, click at the link given below to view the elements involved in a multimedia transmitter, as shown in Fig.1.1
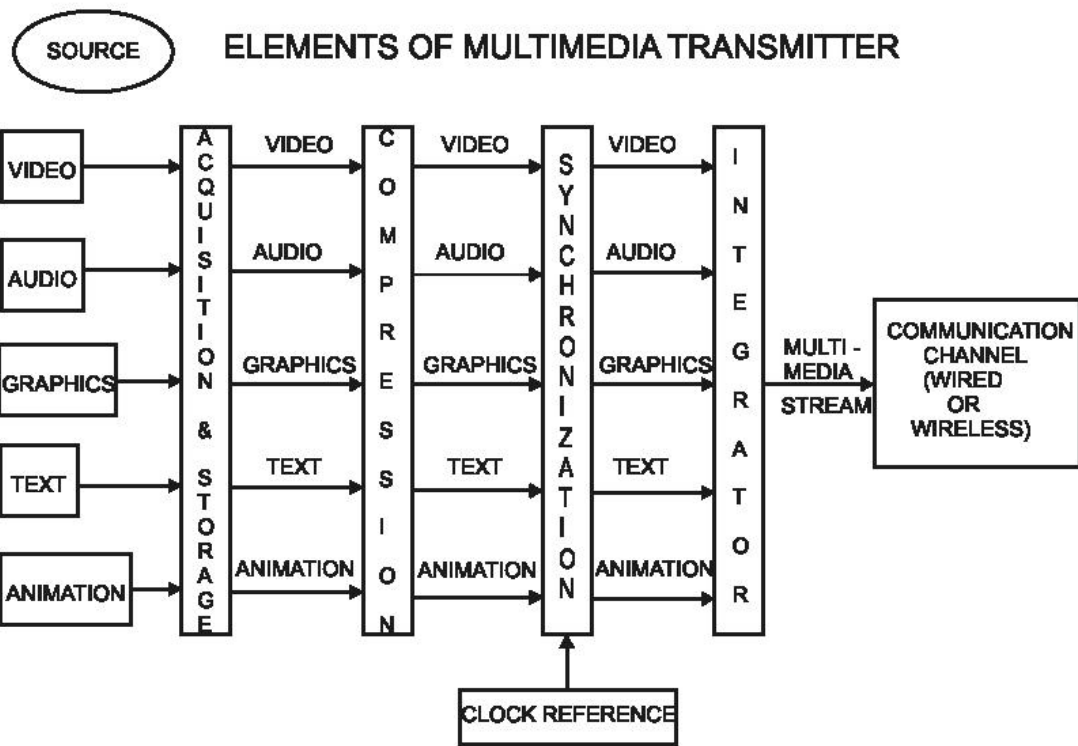
Fig 1.1: *Elements of Multimedia Transmitter*

Devices like cameras, microphones, keyboards, mouse, touch-screens, storage medium etc. are required to feed inputs from different sources. All further processing till the transmission is done by the computer. The data acquisition from multiple media is followed by data compression to eliminate inherent redundancies present in the media streams. This is followed by inter-media synchronization by insertion of time-stamps, integration of individual media streams and finally the transmission of integrated multimedia stream through a communication channel, which can be a wired or a wireless medium.

The destination end should have a corresponding interface to receive the integrated multimedia stream through the communication channel. At the receiver, a reversal of the processes involved during transmission is required. Now, click at the link below to view the elements involved in a multimedia receiver, as shown in Fig.1.2.
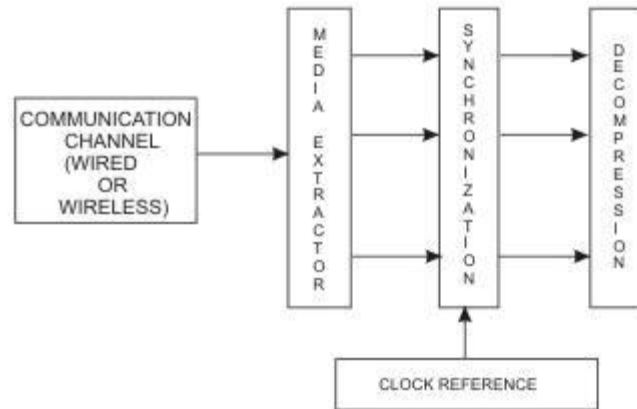
Fig 1.2 Elements of multimedia Receiver

The media extractor separates the integrated media stream into individual media streams, which undergoes de-compression and then presented in a synchronized manner according to their time-stamps in different playback units, such as monitors, loudspeakers, printers/plotters, recording devices etc.

1.4 Challenges involved with multimedia communication

Today, multimedia communication is no longer a dream, but a reality, although the technology is yet to reach its maturity. It has become possible to overcome some challenges, but still many challenges remains to be solved satisfactorily. The challenges involved with multimedia communication are listed below:

(i)   Bandwidth limitations of communication channels.

(ii)   Real-time processing requirements.

(iii)   Inter-media synchronization.

(iv)   Intra-media continuity.

(v)   End-to-end delays and delay jitters.

(vi)   Multimedia indexing and retrieval.

We now discuss each of these points separately and understand the technical aspects of these challenges.

Bandwidth limitations

Limited bandwidth of communication channels poses the most serious challenge in multimedia communication. Some of the media streams, such as video

sequences, large-sized still images, even stereo quality audio require large volumes of data to be transmitted in a short-time, considering real-time requirements. A few typical examples will make things clear.

Suppose, we want to transmit a color image of 1024 x 768 picture   elements (called pixels, or pels) through a telephone channel, supported with modem, having a speed of 14.4 Kbits/second. Assuming 24-bits for each pixel, i.e., 8-bits for each color components (Red, Green and Blue), the total number of bits to be transmitted for the entire image is given by

Total number of bits, $B = 1024 \times 768 \times 24 = 18.8 \times 10^6$

Therefore, the total time required to transmit this image is $(18.8 \times 10^3 /14.4)$, i.e., approximately 22 minutes, which is excessively high. The total time will be halved, if we double the modem speed, but 11 minutes of transmission time is still too high.

Let us now consider the example of a video sequence to be used for video conferencing. We consider a fairly small frame size of 352 x 288 pixels and a colored video sequence having 24-bits / pixel, acquired at a frame rate of 30 frames per second, which is to be transmitted through a leased line of bandwidth 384 Kbits/second. From the given data, the raw video bit rate is (352 x 288 x 24 x 30), i.e., 72.9 Mbits/second. Hence, we cannot transmit the video through the given channel bandwidth, unless the data is significantly compressed.

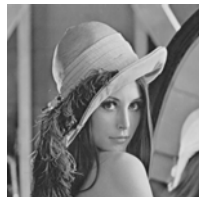The actual situation is not as bad as we may conclude from the above discussions.

Fig 3

 You can observe that most of the pixels have almost the same intensity as those of the neighbors. Only at the boundaries of the objects, the intensities change significantly. Thus, there is considerable redundancy present in the images, since the pixels are spatially correlated to a great extent. If this redundancy could be exploited, significant data compression can be achieved. We will look into this aspect in the coming lessons.

In case of video sequence, the amount of data to be handled is much more than those involved with still images, since the video data is continuously captured at the frame rate. Now, click at a video sequence in the link below to carefully observe the redundancies.

Foreman.mov

Fig 4

We can observe that the redundancies are not only present within a frame (which we can call as the spatial redundancy), but also between successive frames (which can call as the temporal redundancy). If you click the following, you will see a snapshot of few successive frames.

Fig 5

Can you now see that the successive frames are very similar to each other? It is because of the fact that between successive frame time, only a limited movement of the moving objects is possible and if the background is stationary, then most of the pixels do not exhibit any change at all between successive frames. In this example, you can only see lip movements and eye movements between successive frames. Thus, although the data to be handled is much more in video, there is a scope to exploit both spatial and temporal redundancy.

We are definitely convinced that data compression is an essential requirement for images and video. In case of CD-quality stereo audio sampled at 44.1 KHz, the raw data rate is 192 Kbps, which is also somewhat high, if we consider leased line bandwidth or wireless channel bandwidth. Hence, audio compression is also a requirement. Other media streams (text, graphics, animation etc) have relatively less data content in most of the applications and may not require compression.

However, we should not think that we can perform data compression to whatever extent possible. Most of the data compression techniques that achieve significant compression are irreversible. These therefore lead to loss of data and quality degradation. There is always a trade-off that is present between compression, i.e, bandwidth reduction and quality.

1.4.1 Real-time processing requirements:

Whichever techniques we may adopt to exploit the redundancies and achieve data compression, significant amount of processing will be involved. If the processing time is high, the advantage of data compression may be lost. In our still image example, if we are able to achieve a compression of 20:1, the entire image can be transmitted in one minute. If however, the processing time to achieve this compression had been in the order of minutes (imagine that a very old computer of the `60s is used to do the processing!), the advantage of compression would have been lost. Challenges are much more in the case of video. Video frames are captured at a rate of 30 frames per second and this leaves a time of 33 milliseconds between successive frames. Hence, video compression and whatever additional processing are required, needs to be completed within one frame time. This is very often the problem, as we shall later see, when we discuss the complexities of the compression algorithms. Although today's processors, operating at GHz rate are extraordinarily fast, quite often even such high speed processors are unable to perform real-time processing and high-speed, dedicated parallel processing hardware may be required. Development of multimedia processing hardware with real-time capability is a highly challenging research topic of today.

1.4.2 Inter-media synchronization:

As already outlined briefly in Section-1.3, inter-media synchronization is another challenging requirement in multimedia communication. The media streams are available from different and independent sources and are asynchronous with respect to each other. Lack of lip-synchronization is a commonly observed problem in multimedia systems involving audio and video. Lack of synchronization between the different media may even defeat the purpose of multimedia presentation. Multimedia standards have addressed this problem and adopted "time-stamping" to ensure synchronization. Time-stamps, with respect to a system clock reference (SCR) are appended with the different media (audio, video etc) packets before integrating the individual streams into a multimedia one. At the receiver end, the individual media streams are decoded and presented to the playback units in accordance with the time-stamps obtained from the pack and packet headers.

1.4.3 Intra-media continuity:

The extent of data compression with acceptable reconstruction quality is highly data-dependent. Wherever redundancy is more, high compression ratios are achievable, but redundancy may vary. Take the example of a video sequence. Each frame will undergo different extent of compression and this in turn will vary the bit rate (i.e., the number of bits to be transmitted per second) from one frame to the other. If we use a channel that supports constant bit rate, accommodating a variable bit rate source will be a challenging task. This is achieved by providing a buffer before the bit-stream generation. The buffer may be filled up at variable rate, but emptied during transmission at a constant rate. One must therefore ensure that at no instant should the buffer be completely emptied and the

channel starves for data (this is known as *buffer underflow problem).* On the other hand, at no instant should the buffer be completely filled up and further incoming data is lost (this is known as *buffer overflow problem*). In both these extreme situations of buffer overflow and underflow, the continuity is lost during presentation. The question is, to what extent can we tolerate such discontinuities? This depends upon the medium under consideration. If it is a video and you occasionally lose a frame, your eyes may not even notice that, whereas any discontinuity in the audio stream will be immediately detected by your ears. It is because our ears are more sensitive than our eyes as the ultimate detector. We can say that it is easy to fool our eyes, but it is not easy to fool our ears.

However, it must not be misunderstood that intra-media discontinuity can happen only due to the buffer overflow / underflow problems. It may happen because of inadequate processing speed, packet loss, channel errors and many other conditions.

1.4.4   End-to-end delays and delay jitters:
This is yet another important consideration, which we had mentioned in Section-1.0. In a multimedia broadcast or multimedia conferencing, if the users receive the multimedia contents after considerable delays or different users receive the same contents at different times, the interactivity is lost. The multimedia standards available till date have addressed this problem and specified what is acceptable.

1.4.5   Multimedia Indexing and retrieval:
As the price of digital storage media decreases, the storage capacity increases and multimedia applications gain more popularity, there is a requirement to store large number of multimedia files. Unless these files are properly indexed, retrieval of desired multimedia file becomes a tough task, in view of the search complexities. If the multimedia files are based on their contents and then a content-based query system is developed, efficient retrieval can be obtained. Often, for quick browsing of multimedia files, video summaries are needed and automated generation of video summaries is a very challenging task. All these issues are being addressed in the upcoming MPEG-7 multimedia standards.

Scope of this course

The subject of multimedia is studied from different perspectives in different universities and institutes. In some of the multimedia courses, the emphasis is on how to generate multimedia production, the authoring tools and the software associated with these etc. In this course, we don't cover any of the multimedia production aspects at all. Rather, this course focuses on the multimedia systems and the technology associated with multimedia signal processing and communication. We have already posed the technical challenges. In the coming lessons, we shall see in details how these challenges, such as compression,

synchronization etc can be overcome, how the multimedia standards have been designed to ensure effective multimedia communication, how to integrate the associated media and how to index and retrieve multimedia sequences. Other than the introduction, this course has been divided into the following modules:

(i)   Basics of Image Compression and Coding.

(ii)   Orthogonal Transforms for Image Compression.

(iii)   Temporal Redundancies in Video sequences.

(iv)   Real-time Video Coding.

(v)   Multimedia standards.

(vi)   Continuity and synchronization.

(vii)   Audio Coding.

(viii)   Indexing, Classification and Retrieval.

(ix)   Multimedia applications.

The course has been designed for senior undergraduate engineering students, graduate students, researchers and practicing engineers in related fields. Basic background of high-school and engineering mathematics has been assumed. Those having some basic knowledge of Digital Signal Processing and Digital Image Processing will find it easier to follow the lesson materials. Other readers may have to consult supplementary reference materials, as and when difficulty arises.

Questions

NOTE: The students are advised to thoroughly read this lesson first and then answer the following questions. Only after attempting all the solutions they should click to the solution button and verify their answers.

PART-A

A.1. Define a multimedia signal.

A.2. Name at least five different sources of multimedia signal.

A.3. Examine whether the following systems have multimedia capability:

      (a) Radio,   (b) Television,

      (c) Conventional wire-line telephone,   (d) A modern computer,

      (e) Cinema presentation system,   (f) Electronic diary.

A.4. List at least five factors which contributed to the growth of multimedia technology.

A.5. List the elements of a multimedia transmitter.

A.6. Why is it necessary to introduce time stamps with different media stream.

A.7. List the elements of a multimedia receiver.

A.8. State the role of media extractor in a multimedia receiver.

A.9. State at least five challenges involved with multimedia communication.

PART-B: Multiple-choice

In the following questions, click the best out of the four choices.

B.1. "Spatial redundancy" refers to the fact that

      (A) Neighboring pixels in a frame have varying intensities.

      (B) Neighboring pixels in a frame have similar intensities.

      (C) Pixels in the same location across successive frames have varying intensities.

      (D) Pixels in the same location across successive frames have similar intensities.

B.2. "Temporal redundancy" refers to the fact that

      (A) Neighboring pixels in a frame have varying intensities.

      (B) Neighboring pixels in a frame have similar intensities.

      (C) Pixels in the same location across successive frames have varying

intensities.

(D) Pixels in the same location across successive frames have similar intensities.

B.3. Lack of inter-media synchronization leads to

(A) Discontinuities in audio or video presentations.

(B) Incoherency in multimedia presentation.

(C) More end-to-end delays.

(D) Less utilization of channel bandwidth.

B.4. High data compression in audio/video is associated with:

(A) Poor reconstruction quality.

(B) Increased processing time.

(C) Increased transmission time.

(D) Increased channel error rate.

B.5. Buffer overflow/underflow results when:

(A) The channel is of variable bit rate.

(B) The channel is of constant bit rate.

(C) The rate of filling up the buffer and that of emptying are significantly different.

(D) There is no intra-media continuity.

B.6. End-to-end delays and delay jitters can play a nasty role in

(A) Broadcast television.

(B) Two-way voice-telephone communication.

(C) Internet.

(D) Video conferencing.

B.7. Real-time processing is a tough challenge for

      (A) Video

      (B) Audio

      (C) Text

      (D) Animation

B.8. Video indexing and retrieval problem is addressed by the following standard

      (A) JPEG

      (B) MPEG-2

      (C) MPEG-4

      (D) MPEG-7

PART-C : Problems

C.1 A monochrome video sequence uses a frame-size of 176 x 144 pixels and is having 8-bits/pixel. It is captured at a frame rate of 10 frames/ sec. The video is transmitted through a leased line of 64 Kbits/ sec bandwidth.

      (a) Calculate the compression ratio (Ratio of raw bit rate to the compressed bit rate) which will be necessary.

      (b) What will happen if the compression ratio is higher than that obtained in (a)?

      (c) What will happen if the compression ratio is lower than that obtained in (a)?

SOLUTIONS

A.1.
A multimedia signal is one that integrates signals from several media sources, such as video, audio, graphics, animation, text in a meaningful way to convey some information.

A.2
Different sources of multimedia signals are audio, video, still images, text, graphics, animation etc.

A.3

(a) Radio – No multimedia capability, since it supports only audio.

(b) Television – Has multimedia capability, since it supports both audio and video.

(c) Conventional telephone – No multimedia capability, supports only speech (audio).

(d) A modern computer – Has full multimedia capability, since it supports audio, video, still images, graphics, text, animation.

(e) Cinema presentation system – Has multimedia capability, since it supports both audio and video.

(f) Electronic diary – No multimedia capability, since it supports only texts.

A.4

There is no single factor that contributed to the growth of multimedia technology. In recent years, advancements of processing speed, availability of very large-scale integrated circuits for real-time processing, effective data compression tools and algorithms to eliminate the redundancy and achieve bandwidth reduction, advancement of computer networking etc. have all contributed to its growth.

A.5

The elements of a multimedia transmission system are:

- (i) Source devices (cameras, microphones etc.)
- (ii) Data acquisition
- (iii) Data compression
- (iv) Inter-media synchronization.
- (v) Integration of multiple streams into one (multiplexer)

A.6

Time-stamps introduced with different media streams are used to achieve inter-media synchronization.

A.7

The elements of a multimedia receiver are:

- (i) Media extractor.
- (ii) De-compression and reconstruction.
- (iii) Playback devices (monitors, displays, loudspeakers etc.)

A.8
The task of media extractor in multimedia receivers is to receive the integrated multimedia bit stream from the communication channel and extract the different media streams, such as audio, video, text etc. at its output. It acts as a de-multiplexer and performs a reverse role of integration in multimedia transmitters.

A.9
The challenges involved with multimedia communication system are data compression (especially for high data intensive media like video, still images and audio), real-time processing capability (for time critical media like video), inter-media synchronization, intra-media continuity, end-to-end delays and delay jitters, indexing and retrieval etc.

B-1: (B)   **B-2:** (D)   **B-3:** (B)   **B-4:** (A)

B-5: (C)   **B-6:** (D)   **B-7:** (A)   **B-8**: (D)

C-1:
(a) Raw bit-rate = (Pixel size/Frame) x (Bits/pixel) x (Frames/sec)

         = 176 x 144 x 10 bits/second = 253440 bits/second
    Channel bit-rate requirement is 64 Kbits/sec

    Required compression ratio = (Raw bit rate) / (Channel bit-rate)
               = 253440/64000 = 3.96

(b) and (c) : The readers are advised to think about these and if they are unable to solve, they should wait till the end of module-2. The answers will be provided at the end of module-2.