CHAPTER 2

Media and Data Streams

This chapter provides an introduction to the terminology used in the entire book. We begin with our definition of the term multimedia as a basis for a discussion of media and key properties of multimedia systems. Next, we will explain data streams and information units used in such systems.

2.1 The Term "Multimedia"

The word multimedia is composed of two parts: the prefix *multi* and the root *media.* The prefix *multi* does not pose any difficulty; it comes from the latin word *multus,* which means "numerous." The use of *multi* as a prefix is not recent and many Latin words employ it.

The root *media* has a more complicated story. *Media* is the plural form of the Latin word *medium. Medium* is a noun and means "middle, center."

Today, the term multimedia is often used as an attribute for many systems, components, products, and concepts that do not meet the key properties we will introduce later (see Section 2.3)**.** This means that the definition introduced in this book is (intentionally) restrictive in several aspects.

2.2 The Term "Media"

As with most generic words, the meaning of the word media varies with the context in which it is used. Our definition of medium is "a means to distribute and represent information." Media are, for example, text, graphics, pictures, voice, sound, and music. In this sense, we could just as well add water and the atmosphere to this definition.

[MHE93] provides a subtle differentiation of various aspects of this term by use of various criteria to distinguish between perception, representation, presentation, storage, transmission, and information exchange media. The following sections describe these attributes.

2.2.1 Perception Media

Perception media refers to the nature of information perceived by humans, which is not strictly identical to the sense that is stimulated. For example, a still image and a movie convey information of a different nature, though stimulating the same sense. The question to ask here is: *How do humans perceive information?*

In this context, we distinguish primarily between what we see and what we hear. Auditory media include music, sound, and voice. Visual media include text, graphics, and still and moving pictures. This differentiation can be further refined. For example, a visual medium can consist of moving pictures, animation, and text. In turn, moving pictures normally consist of a series of scenes that, in turn, are composed of single pictures.

2.2.2 Representation Media

The term representation media refers to how information is represented internally to the computer. The encoding used is of essential importance. The question to ask here is: *How is information encoded in the computer?* There are several options:

- Each character of a piece of text is encoded in ASCII.
- A picture is encoded by the CEPT or CAPTAIN standard, or the GKS graphics standard can serve as a basis.
- An audio data stream is available in simple PCM encoding and a linear quantization of 16 bits per sampling value.
- A single image is encoded as Group-3 facsimile or in JPEG format.
- A combined audio-video sequence is stored in the computer in various TV standards (e.g., PAL, SECAM, or NTSC), in the CCIR-601 standard, or in MPEG format.

2.2.3 Presentation Media

The term presentation media refers to the physical means used by systems to reproduce information for humans. For example, a TV set uses a cathode-ray tube and loudspeaker. The question to ask here is: *Which medium is used to output information from the computer or input in the computer?*

We distinguish primarily between output and input. Media such as paper, computer monitors, and loudspeakers are output media, while keyboards, cameras, and microphones are input media.

2.2.4 Storage Media

The term storage media is often used in computing to refer to various physical means for storing computer data, such as magnetic tapes, magnetic disks, or digital optical disks. However, data storage is not limited to the components available in a computer, which means that paper is also considered a storage medium. The question to ask here is: *Where is information stored?*

2.2.5 Transmission Media

The term transmission media refers to the physical means-cables of various types, radio tower, satellite, or ether (the medium that transmit radio waves)—that allow the transmission of telecommunication signals. The question to ask here is: *Which medium is used to transmit data?*

2.2.6 Information Exchange Media

Information exchange media include all data media used to transport information, e.g., all storage and transmission media. The question to ask here is: *Which data medium is used to exchange information between different locations?*

For example, information can be exchanged by storing it on a removable medium and transporting the medium from one location to another. These storage media include microfilms, paper, and floppy disks. Information can also be exchanged directly, if transmission media such as coaxial cables, optical fibers, or radio waves are used.

2.2.7 Presentation Spaces and Presentation Values

The terms described above serve as a basis to characterize the term medium in the information processing context. The description of perception media is closest to our definition of media: those media concerned mainly with the human senses. Each medium defines presentation values in presentation spaces [HD90, SH91], which address our five senses.

Paper or computer monitors are examples of visual presentation spaces. A computer-controlled slide show that projects a screen's content over the entire projection screen is a visual presentation space. Stereophony and quadrophony define acoustic presentation spaces. Presentation spaces are part of the above-described presentation media used to output information.

Presentation values determine how information from various media is represented. While text is a medium that represents a sentence visually as a sequence of characters, voice is a medium that represents information acoustically in the form of pressure waves. In some media, the presentation values cannot be interpreted correctly by humans. Examples include temperature, taste, and smell. Other media require a

predefined set of symbols we have to learn to be able to understand this information. This class includes text, voice, and gestures.

Presentation values can be available as a continuous sequence or as a sequence of single values. Fluctuations in pressure waves do not occur as single values; they define acoustic signals. Electromagnetic waves in the range perceived by the human eye are not scanned with regard to time, which means that they form a continuum. The characters of a piece of text and the sampling values of an audio signal are sequences composed of single values.

2.2.8 Presentation Dimensions

Each presentation space has one or more presentation dimensions. A computer monitor has two space dimensions, while holography and stereophony need a third one. Time can occur as an additional dimension within each presentation space, which is critical for multimedia systems. Media are classified in two categories with regard to the time dimensions of their presentation space:

1. Text, graphics, and pictures are called discrete media, as they are composed of time-independent information items. Indeed, they may be displayed according to a wide variety of timing or even sequencing, and still remain meaningful. We say that time is not part of the semantics of discrete media. The term discrete tends to blur, as modem computer-based text and graphics presentations are often value-discrete and time-continuous. For example, the text of a book is a discrete

   Each method used to process discrete media should be as fast as possible. On the other hand, time is not the critical factor, because the validity (and thus the correctness) of data does not depend on a time condition (at least not within a time frame of seconds or less). We could also speak about longer of shorter time conditions.
2. Continuous media refers to sound or motion video, where the presentation requires a continuous playout as time passes. In other words, time, or more exactly time-dependency between information items, is part of the information itself. If the timing is changed, or the sequencing of the items modified, the meaning is altered. We say that time is part of the semantics of continuous media. Continuous media are also called time-dependent media. Another technical consequence when dealing with continuous media is that they also require the networks that carry them to respect this time-dependency.

   How these media are processed is time-critical, because the validity (correctness) of data depends on a time condition. If an audio sampling value is transmitted too late it may become invalid or wrong, for example, since the audio data that follow this value have already been played out over the loudspeaker. In audio and video, the representation values form a continuous sequence, where video means pure

moving images. A combination of audio and moving images, like in television or movies, is not synonymous with the term video. For this reason, they are called continuous media. When time-dependent representation values that occur aperiodically are distinguished, they are often not put under the continuous media category. For a multimedia system, we also have to consider such non-continuous sequences of representation values. This type of representation-value sequence occurs when information is captured by use of a pointer (e.g., a mouse) and transmitted within cooperative applications using a common screen window. Here, the continuous medium and time-dependent medium are synonymous. By this definition, continuous media are video (moving images) of natural or artificial origin, audio, which is normally stored as a sequence of digitized pressure-wave samples, and signals from various sensors, such as air pressure, temperature, humidity, pressure, or radioactivity sensors.

The terms that describe a temporally discrete or continuous medium do not refer to the internal data representation, for example, in the way the term representation medium has been introduced. They refer to the impression that the viewer or auditor gets. The example of a movie shows that continuous-media data often consist of a sequence of discrete values, which follow one another within the representation space as a function of time. In this example, a sequence of at least 16 single images per second gives the impression of continuity, which is due to the perceptual mechanisms of the human eye.

Based on word components, we could call any system a multimedia system that supports more than one medium. However, this characterization falls short as it provides only a quantitative evaluation. Each system could be classified as a multimedia system that processes both text and graphics media. Such systems have been available for quite some time, so that they would not justify the newly coined term. The term multimedia is more of a qualitative than a quantitative nature.

As defined in [SRR90, SH91], the number of supported media is less decisive than the type of supported media for a multimedia system to live up to its name. Note that there is controversy about this definition. Even standardization bodies normally use a coarser interpretation.

2.3 Key Properties of a Multimedia System

Multimedia systems involve several fundamental notions. They must be computer-controlled. Thus, a computer must be involved at least in the presentation of the information to the user. They are integrated, that is, they use a minimal number of different devices. An example is the use of a single computer screen to display all types of visual information. They must support media independence. And lastly, they need to handle discrete and continuous media. The following sections describe these key properties.

2.3.1 Discrete and Continuous Media

Not just any arbitrary combination of media deserves the name multimedia. Many people call a simple word processor that handles embedded graphics a multimedia application because it uses two media. By our definition, we talk about multimedia if the application uses both discrete and continuous media. This means that a multimedia application should process at least one discrete and one continuous medium. A word processor with embedded graphics is not a multimedia application by our definition.

2.3.2 Independent Media

An important aspect is that the media used in a multimedia system should be independent. Although a computer-controlled video recorder handles audio and moving image information, there is a temporal dependence between the audio part and the video part. In contrast, a system that combines signals recorded on a DAT (Digital Audio Tape) recorder with some text stored in a computer to create a presentation meets the independence criterion. Other examples are combined text and graphics blocks, which can be in an arbitrary space arrangement in relation to one another.

2.3.3 Computer-Controlled Systems

The independence of media creates a way to combine media in an arbitrary form for presentation. For this purpose, the computer is the ideal tool. That is, we need a system capable of processing media in a computer-controlled way. The system can be optionally programmed by a system programmer and/or by a user (within certain limits). The simple recording or playout of various media in a system, such as a video recorder, is not sufficient to meet the computer-control criterion.

2.3.4 Integration

Computer-controlled independent media streams can be integrated to form a global system so that, together, they provide a certain function. To this end, synchronic relationships of time, space, and content are created between them. A word processor that supports text, spreadsheets, and graphics does not meet the integration criterion unless it allows program-supported references between the data. We achieve a high degree of integration only if the application is capable of, for example, updating graphics and text elements automatically as soon as the contents of the related spreadsheet cell changes.

This kind of flexible media handling is not a matter to be taken for granted-even in many products sold under the multimedia system label. This aspect is important when talking of integrated multimedia systems. Simply speaking, such systems should allow us to do with moving images and sound what we can do with text and graphics [AGH90]. While conventional systems can send a text message to another user, a highly

integrated multimedia system provides this function ***and*** support for voice messages or a voice-text combination.

2.3.5 Summary

Several properties that help define the term multimedia have been described, where the media are of central significance. This book describes networked multimedia systems. This is important as almost all modern computers are connected to communication networks. If we study multimedia functions from a local computer's perspective, we take a step backwards. Also, distributed environments offer the most interesting multimedia applications as they enable us not only to create, process, represent, and store multimedia information, but to exchange them beyond the limits of our computers.

Finally, continuous media require a changing set of data in terms of time, that is, a data stream. The following section discusses data streams.

2.4 Characterizing Data Streams

Distributed networked multimedia systems transmit both discrete and continuous media streams, i.e., they exchange information. In a digital system, information is split into units (packets) before it is transmitted. These packets are sent by one system component (the source) and received by another one (the sink). Source and sink can reside on different computers. A data stream consists of a (temporal) sequence of packets. This means that it has a time component and a lifetime.

Packets can carry information from continuous and discrete media. The transmission of voice in a telephone system is an example of a continuous medium. When we transmit a text file, we create a data stream that represents a discrete medium.

When we transmit information originating from various media, we obtain data streams that have very different characteristics. The attributes asynchronous, synchronous, and isochronous are traditionally used in the field of telecommunications to describe the characteristics of a data transmission. For example, they are used in FDDI to describe the set of options available for an end-to-end delay in the transmission of single packets.

2.4.1 Asynchronous Transmission Mode

In the broadest sense of the term, a communication is called asynchronous if a sender and receiver do not need to coordinate before data can be transmitted. In asynchronous transmission, the transmission may start at any given instant. The bit synchronization that determines the start of each bit is provided by two independent clocks, one at the sender, the other at the receiver. An example of asynchronous transmission is the way in which simple ASCII terminals are usually attached to host computers. Each time

the character "A' is pressed, a sequence of bits is generated at a preset speed. To inform the computer interface that a character is arriving, a special signal called the start signal—which is not necessarily a bit—precedes the information bits. Likewise, another special signal, the stop signal, follows the last information bit.

2.4.2 Synchronous Transmission Mode

The term synchronous refers to the relationship of two or more repetitive signals that have simultaneous occurrences of significant instants. In synchronous transmission, the beginning of the transmission may only take place at well-defined times, matching a clocking signal that runs the synchronism with that of the receiver. To see why clocked transmission is important, consider what might happen to a digitized voice signal when it is transferred across a nonsynchronized network. As more traffic enters the network, the transmission of a given signal may experience increased delay. Thus, a data stream moving across a network might slow down temporarily when other traffic enters the network and then speed up again when the traffic subsides. If audio from a digitized phone call is delayed, however, the human listening to the call will hear the delay as annoying interference or noise. Once a receiver starts to play digitized samples that arrive late, the receiver cannot speed up the playback to catch up with the rest of the stream.

2.4.3 Isochronous Transmission Mode

The term isochronous refers to a periodic signal, pertaining to transmission in which the time interval separating any two corresponding transitions is equal to the unit interval or to a multiple of the unit interval. Secondly, it refers to data transmission in which corresponding significant instants of two or more sequential signals have a constant phase relationship. This mode is a form of data transmission in which individual characters are only separated by a whole number of bit-length intervals, in contrast to asynchronous transmission, in which the characters may be separated by random-length intervals. For example, an end-to-end network connection is said to be isochronous if the bit rate over the connection is guaranteed and if the value of the delay jitter is also guaranteed and small. The notion of isochronism serves to describe what the performance of a network should be in order to satisfactorily transport continuous media streams, such as real-time audio or motion video. What is required to transport audio and video in real time? If the source transmits bits at a certain rate, the network should be able to meet that rate in a sustained way.

These three attributes are a simplified classification of different types of data streams. The following sections describe other key properties.

2.5 Characterizing Continuous Media Data Streams

This section provides a summary of the characteristics for data streams that occur in multimedia systems in relation to audio and video transmissions. The description includes effects of compression methods applied to the data streams before they are transmitted. This classification applies to distributed and local environments.

2.5.1 Strongly and Weakly Periodic Data Streams

The first property of data streams relates to the time intervals between fully completed transmissions of consecutive information units or packets. Based on the moment in which the packets become ready, we distinguish between the following variants:

- When the time interval between neighboring packets is constant, then this data stream is called a strongly periodic data stream. This also means that there is minimal jitter—ideally zero. Figure 2-1 shows such a data stream. An example for this type is PCM-encoded (Pulse Code Modulation) voice in telephone systems.

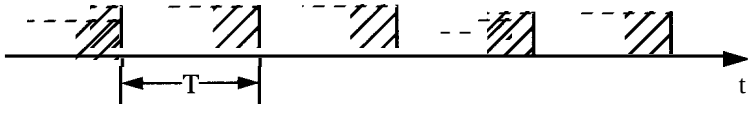
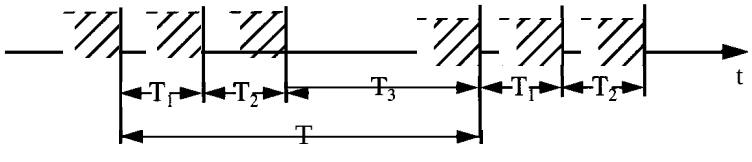


Figure 2-1 Strongly periodic data stream; time intervals have the same duration between consecutive packets.

- The duration of the time intervals between neighboring packets is often described as a function with finite period duration. However, this time interval is not constant between neighboring packets (or it would be a strongly periodic data stream). In the case shown in Figure **2-2,** we speak of a weakly periodic data stream.



- All other transmission options are called aperiodic data streams, which relates to the sequence of time interval duration, as shown in Figure **2-3.**

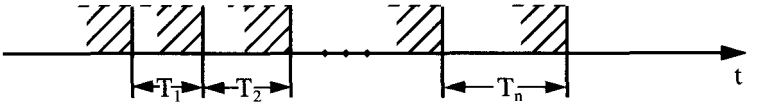


Figure 2-3 Aperiodic data stream; the time interval sequence is neither constant nor weakly periodic.

An example of an aperiodic data stream is a multimedia conference application with a common screen window. Often, the status (left button pressed) and the current coordinates of the mouse moved by another user have to be transmitted to other participants. If this information were transmitted periodically, it would cause a high data rate and an extremely high redundancy. The ideal system should transmit only data within the active session that reflect a change in either position or status.

2.5.2 Variation of the Data Volume of Consecutive Information Units

A second characteristic to qualify data streams concerns how the data quantity of consecutive information units or packets varies.

- If the quantity of data remains constant during the entire lifetime of a data stream, then we speak of a strongly regular data stream. Figure 2-4 shows such a data stream. This characteristic is typical for an uncompressed digital audio-video stream. Practical examples are a full-image encoded data stream delivered by camera or an audio sequence originating from an audio CD.

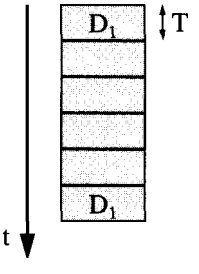


Figure 2-4 Strongly regular data stream; the data quantity is constant in all packets.

- If the quantity of data varies periodically (over time), then this is **a** weakly regular data stream. Figure 2-5 shows an example.

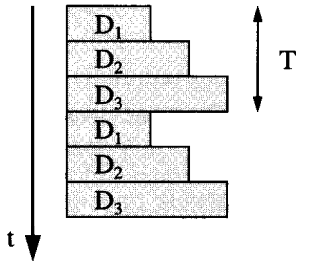


Figure 2-5 Weakly regular data stream; the packets' data stream varies periodically.

Some video compression methods use a concept that encodes and compresses full images individually. The resulting information unit is a relatively large data packet in the data stream. For reasons of simplicity, we will not consider the packet length, which is limited during the transmission, depending on the communication layer. These packets are transmitted periodically, e.g., every two seconds. For all images between two single images of the video stream, the differences between two consecutive images each form the information that is actually transmitted.

An example is MPEG (see Section 7.7), where *I* images are compressed single images, while the compression of *P* images and *B* images uses only image differences, so that the data volume is much smaller. No constant bit rate is defined for compressed *I*, *P*, and *B* packets. However, a typical average *I:B:P* ratio of the resulting data quantities is 10:1:2, which results in a weakly regular data stream over the long-term average.

Data streams are called irregular when the data quantity is neither constant, nor changing by a periodic function (see Figure 2-6). This data stream is more difficult to transmit and process compared to the variants described earlier.

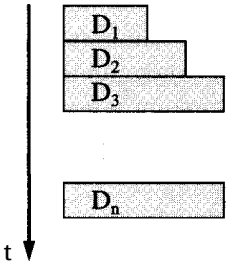


Figure 2-6 Irregular data stream; the packets' data quantity is not constant and does not vary periodically.

When applying a compression method that creates a data stream with a variable bit rate, the size of the single information units (each derived from a single image) is

determined from the image content that has changed in respect to the previous image. The size of the resulting information units normally depends on the video sequence and the data stream is irregular.

2.5.3 Interrelationship of Consecutive Packets

The third qualification characteristic concerns the continuity or the relationship between consecutive packets. Are packets transmitted progressively, or is there a gap between packets? We can describe this characteristic by looking at how the corresponding resource is utilized. One such resource is the network.

- Figure 2-7 shows an interrelated information transfer. All packets are transmitted one after the other without gaps in between. Additional or layer-independent information to identify user data is included, e.g., error detection codes. This means that a specific resource is utilized at 100 percent. An interrelated data stream allows maximum throughput and achieves optimum utilization of a resource. An ISDN B channel that transmits audio data at 64Kbit/s is an example.

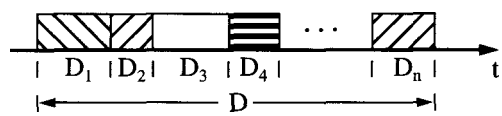


Figure 2-7 Interrelated data stream; packets are transmitted without gaps in between.

- The transmission of an interrelated data stream over a higher-capacity channel causes gaps between packets. Each data stream that includes gaps between its information units is called a non-interrelated data stream. Figure 2-8 shows an example. In this case, it is not important whether or not there are gaps between all packets or whether the duration of the gaps varies. An example of a non-interrelated data stream is the transmission of a data stream encoded by the DVI-PLV method over an FDDI network. An average bit rate of 1.2Mbit/s leads inherently to gaps between some packets in transit.

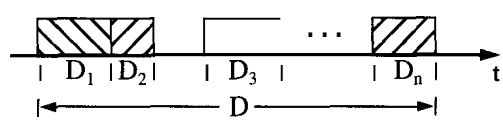


Figure 2-8 Non-interrelated data stream; there are gaps between packets.

To better understand the characteristics described above, consider the following example:

A PAL video signal is sampled by a camera and digitized in a computer. No compression is applied. The resulting data stream is strongly periodic, strongly regular, and interrelated, as shown in Figure 2-4. There are no gaps between packets. If we use the MPEG method for compression, combined with the digitizing process, we obtain a weakly periodic and weakly regular data stream (referring to its longer duration). And if we assume we use a 16-Mbit/s token-ring network for transmission, our data stream will also be noninterrelated.

2.6 Information Units

Continuous (time-dependent) media consist of a (temporal) sequence of information units. Based on Protocol Data Units (PDUs), this section describes such an information unit, called a Logical Data Unit (LDU). An LDU's information quantity and data quantities can have different meanings:

1. Let's use Joseph Haydn's symphony, *The Bear,* as our first example. It consists of the four musical movements, *vivace assai, allegretto, menuet,* and *finale vivace.* Each movement is an independent, self-sufficient part of this composition. It contains a sequence of scores for the musical instruments used. In a digital system, these scores are a sequence of sampling values. We will not use any compression in this example, but apply PCM encoding with a linear characteristic curve. For CD-DA quality, this means 44,100 sampling values per second, which are encoded at 16bits per channel. On a CD, these sampling values are grouped into units with a duration of 1/75 second. We could now look at the entire composition and define single movements, single scores, the grouped 1/75-s sampling values, or even single sampling values as LDUs. Some operations can be applied to the playback of the entire composition—as one single LDU. Other functions refer to the smallest meaningful unit (in this case the scores). In digital signal processing, sampling values are LDUs.

2. In Figure 2-9, we see that the uncompressed video sequence consists of single chips, each representing a scene. Each of these scenes consists of a sequence of single images. Each single image can be separated into various regions, for example regions with a size of 16x16 pixels. In turn, each pixel contains a luminance value and a chrominance value.

   This means that a single image is not the only possible LDU in a motion video sequence. Each scene and each pixel are also LDUs. The redundancies in single image sequences of an MPEG-encoded video stream can be used to reduce the data quantity by applying an inter-frame compression method. In this case, the smallest self-sufficient meaningful units are single-image sequences.

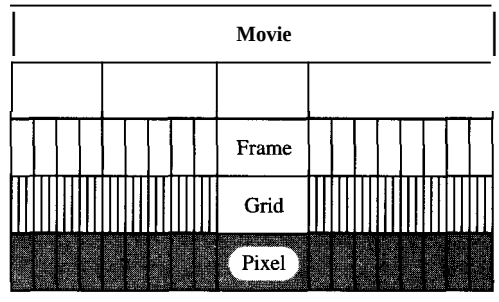


Figure 2-9 Granularity of a motion video sequence showing its logical data units **(LDUs).**

A phenomenon called granularity characterizes the hierarchical decomposition of an audio or video stream in its components. This example uses a symphony and a motion video to generally describe extensive information units. We distinguish between closed and open LDUs. Closed LDUs have a well-defined duration. They are normally stored sequences. In open LDUs, the data stream's duration is not known in advance. Such a data stream is delivered to the computer by a camera, a microphone, or a similar device.

The following chapter builds on these fundamental characteristics of a multimedia system to describe audio data in more detail, primarily concentrating on voice processing.