Chapter Two

Digital Data

Analog and Digital Signal Conversion

- The world we sense is full of analog signals:
  - Electrical sensors convert the medium they sense into electrical signals
    - E.g. transducers, thermocouples, microphones.(usually) **continuous** signals
  - **Analog**: continuous signals must be converted or digitized for computer processing.
  - **Digital**: discrete digital signals that computer can readily deal with.

Effective and Efficient Codes

- An effective code is one that can represent each desired data item with a unique combination of symbols.
- An efficient code is one that does not waste processing, storage, or transmission resources.
- A basic concern in all multimedia development is assuring that the digital encoding (file format) being used can effectively and efficiently represent the required range of media data.
- For example, **ASCII** used 7 bits to represent letters, numbers and other symbols. Then **extended ASCII** or **ASCII-8** added the 8th bit to double the number of available codes. **Unicode**, a later standard, uses 16-bit codes to effectively designate over 65,000 individual characters.

Digital Files

- A computer file is a container for binary code, which is the universal language of a computer.
- File sizes the size of a file is usually measured in numbers of bytes.
- Kilobyte(KB)=1024 bytes= 2 to power 10
- Megabyte(MB)=1,048,576 bytes= 2 to power 20
- Gigabyte(GB)=1,073,741,824=2 to power 30
- Terabyte(TB)=1,099,511,627,776= 2 to power 40

DIGITAL FILES

- A file format is the convention that specifies how instructions and data are encoded in a computer file. Without a specific file format, a binary code has no meaning (HTML format, text file format, ogg format for multimedia, png..)

- File Extension is a series of letters that designate a file type.

- Extensions are important for multimedia developers for two reasons:
  - First: They immediately identify a file type and ,often, even the program that created the file.
  - Second: File extensions are often used by a computer's operating system to identify and launch an appropriate program to open the file.

Digital Files

- Different computer systems, or platforms, use different hardware and software.
- File compatibility (a file that is work/compatible with more than one platform) a file that can be read and processed on one computer platform, such as a Windows PC, often can be used on another, such as a Macintosh.
- One of the most important concerns for multimedia developers is file compatibility, assuring that a computer can process the instructions or data that are encoded in a particular file format.

File compatibility

- Program files (OS or application programs) give the computer its instructions (OS (Windows, OS X), programming languages, and applications).
- Programs are developed for specific computer platforms and are generally not compatible with other platforms.
- Applications such as Word or Photoshop are usually available for PCs or Macs but the developer will have to own separate copies for each platform.

File compatibility

- **Multimedia data files-** text, images , sounds, video and animation- pose two principal compatibility challenges.

1. Whether or not the file format is *cross-platform compatible*. For example, Microsoft developed the BMP image format and Apple developed PICT. These two formats are not cross-platform compatible.
2. Whether or not different application programs on *a given platform* can process the format. For example, some image formats, such as TIFF, can be used by most image editing applications and are also widely compatible with authoring software.

File compatibility

- Native file formats are an important example of specialized files. These are coding conversions used by specific computer applications, such as Adobe Photoshop or Corel painter. Native formats *contain information specific to the application that created them.*

- For example, a file in Photoshope's native format (PSD) may contain information about filters and other effects applied to the image during editing.

- In word .doc is the native format but still the files can be saved as txt or rtf and opened by word.

Digital Files

- File Conversion the process of transforming one file type to another.
- File conversion can be carried out by specialized applications, such as Prism Video Converter Software, which convert AVI MP4 WMV MOV MPEG FLV and other video file formats, or using the "Save as" function in popular media-specific software.
- File Maintenance Effective file maintenance involves three major steps: *identification, categorization,* and *preservation.*
- **(follow these steps when you create your Assignments and Project)**

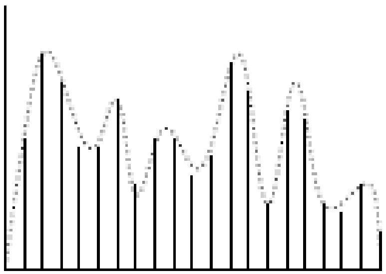
File Maintenance

- Files should be clearly *identified* in terms of their specific contents, their general type, and/or their originating program. This is done by file name and extension.
- *Categorization* is the process of meaningfully grouping related files. Electronic folders serve the same general functions as their paper counterparts. Files are more readily located when grouped as images, text documents, video clips, and so on.
- File *preservation* includes the preparation and storage of back-up copies as well as their distribution to individuals or departments that may need them for future work.

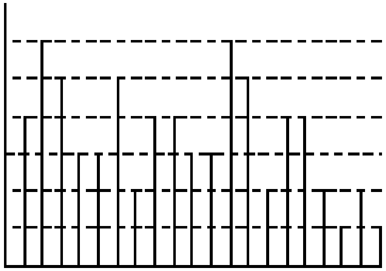
Digitization

- Letters and numbers are discrete units. Other data- still or moving images, and sounds- are often presented as continuous, or analog, phenomena.
- Digitization is the process of converting analog data to a digital format (Analog to digital format).
- The first step in the digitization process is *sampling* which takes samples of the continuous signal.
- The second step is known as *quantization* where we restrict the value of the samples to a fixed set of levels.

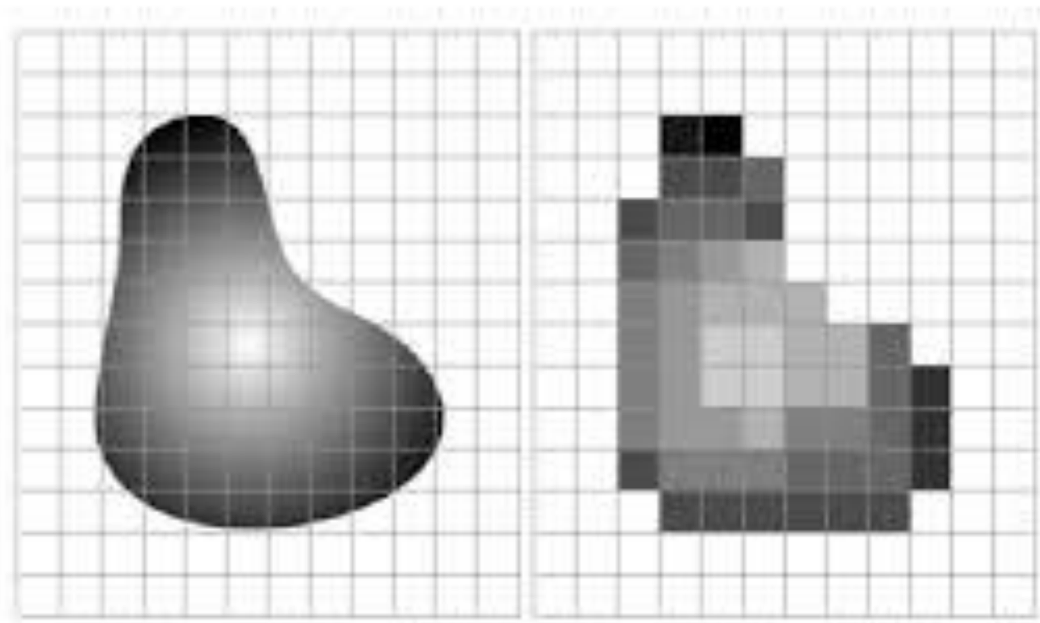
A Continuous Signal



Samples of the signal



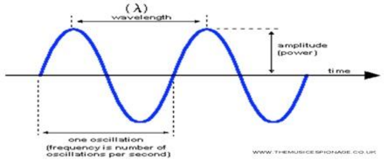
Quantised Samples



a b

Sampling

- Sampling is the process of analyzing a small element of an image or sound and representing that element in a digital code.



- To digitize ***analog sound***,

1. thousands of samples of the varying amplitudes of the sound are collected each second.
2. Each sample is assigned a binary code indicating its amplitude at that instant.
3. The computer processes these values to re-create the sound from the individual samples.

- Analog images, such as photographs, are digitized by sampling their color at many different points.

- The re-created image is a grid of *picture elements*, pixel, each having a particular color.

DIGITIZATION

- Sample rate is the number of samples taken in a given unit of time (sounds) or space (images).

- In the case of sound, the rate is given in kilohertz, thousands of samples per second.

- Sample rate for images are defined spatially and are referred to as ***spatial resolution***. Spatial resolution typically varies from seventy two to several thousand pixels per inch.

Sample rate

- Low sample rates reduce the quality of digital images and sounds because they omit original analog information.

Digital encoding

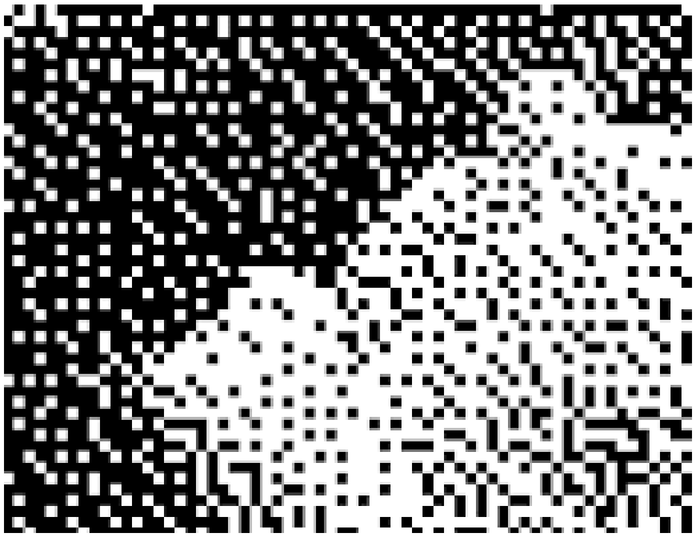
- For images, a lower number of samples will miss details such as fine gradations of lines, often producing an incomplete, "fuzzy" reproduction.

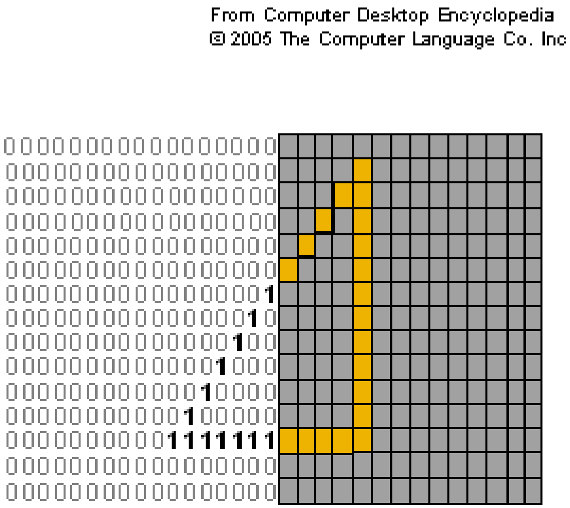
Spatial resolution. 300ppi (left) v. 50ppi (right).

Digital encoding of media

- There are two major approaches to digitally encoding media: description-based and command-based.
- In descriptive approach a digital media file contains a detailed representation of the many discrete elements that comprise the image or sound.
- An image is described by recording the colors of each of its many individual pixels thus creating a grid or "map" of their locations. This type of digital encoding is therefore often called a bitmapped image (raster) [JPEGs, TIF, GIFs and PNGs] .

RASTER IMAGES





Digital encoding of media

- Sound is described through thousands of individual amplitude samples and is known as sampled sound.
- Descriptive encoding is the preferred method for representing *natural scenes and sounds*. It also supports very detailed editing.
- Descriptive digital encoding takes advantage of the computer's ability to store and process very large numbers of discrete media elements.

COMMAND ENCODING

- Command digital encoding takes advantage of another property of computers- their ability to execute commands. The *command approach* stores a set of instructions (mathematical formulas) that the computer then follows to produce digital images and sounds.

- Sounds can be encoded as commands by specifying musical actions such as striking a particular organ key with a certain force and duration. This is used in the popular MIMD format.

- Digital images can be encoded as drawing commands. The image that is stored using command encoding is called vector-image [EPS, AI (by Adobe) and PDF].

- Macromedia Flash a popular authoring application for the Web, uses command-based graphics to lower the file sizes and speed the delivery of animation.

Command encoding

- *Command-based media have important advantages over their descriptive counterparts*. File sizes are usually much smaller, making them particularly useful in applications with limited bandwidth, as is often the case on the Web.
- They can also be scaled without introducing distortion, unlike most descriptive media.

File compression

- Compression is the process of re-encoding digital data to reduce file size.
- A specialized program called a codec, for COmpressor/DECompressor, changes the original file to the smaller version and then decompresses it to gain present the data in a usable form.
- A basic understanding of compression is particularly important in multimedia development for two reasons: (Why important for Multimedia)
  - File compression is one of the major strategies developers use to reduce file sizes that leads to more efficient multimedia applications and often essential for effective delivery on the Web.
  - **The choice of a particular form of compression can dramatically affect media quality.**

FILE COMPRESSION

- There are two major types of compression: **lossy** and **lossless**.
- In *lossy compression*, the number of bits in the original file is reduced and some data is lost. MP3 is an example of sound file that incorporates lossy compression.
- Lossy compression is not an option for files consisting of text and numbers , so called *alphanumeric* information.
- *Lossless compression* substitutes a more efficient encoding to reduce the file size while preserving all of the original data. When the file is decompressed it will be identical to the original.

Error detection and Correction

- Changing or eliminating even a few of bits from the digital data can produce serious distortions.
- For example, the pits and lands of a CD can be altered by a scratch or dust. Data may also be lost or altered in electronic transmission.
- **For digital data to be reliable, some means of detecting and correcting errors must be found.**

Error detection and Correction

- The first challenge in ensuring reliable digital information is to find a means of detecting an error. A simple strategy for doing this is the "parity bit.“
- A **parity bit** is an extra bit added to a data code to maintain either an even or an odd number of 1s in the code. In this approach, the number of 1 bits in each block of data is counted. An extra bit is then added to make the total number 1 bits either odd, or even. The added bit is called a parity bit. If the number of 1 bits is intended to be odd the coding scheme is called "odd parity"; even numbers of 1s are "even parity."

Error detection and Correction

- Even parity means that the number of 1 bits must be an even number. If it is, a 0 is added. If it isn't, a 1 bit is supplementary.

- For example, the ASCII code for the letter *A* is 01000001. The parity bit added would be a 0 and the code for the letter would become 010000010.
- On receipt of the data block, the bits are again counted. If the number of sits is not the expected odd or even value, an error must have occurred. This simple system can detect many errors, but it does have serious limitations. More advanced approaches are often used to detect errors. **Once the error is detected, the second challenge is to correct it with accurate data bits.**

ERROR DETECTION AND CORRECTION

- For some applications, such as data transmission, knowing an error occurred may be enough. The data error will generate a request from the receiver to retransmit the flawed data item. The process continues until an error-free message is received. (The receiver asks the sender to resend the data item)

- **This strategy will not work for all types of digital media,** however. For example, accurate playback of audio CDs requires *immediate* detection and correction of data errors. To accomplish this, error correction code is added to reconstruct the original bit sequence. This requires redundant data in the bit stream, which increases data storage requirements. It also dramatically improves reliability. CD audio discs devote approximately one third of their storage to error detection and correction.

ADVANTAGES OF DIGITAL INFORMATION

- Reproduction digital information cab be repeatedly copied with no loss of quality. While the video copy (analog format as tape) does not preserve all the information of the master, and copies of copies soon produce significant distortions, a process known as generation decay.

- Editing analog editing of photos tends to be slow and expensive, requiring specialized equipment and techniques. Once a photo is in digital form, programs such as Adobe Photoshop can be used to instantly change its size, crop it, change its brightness or contrast, add special effects and so on.

ADVANTAGES OF DIGITAL INFORMATION

- Integration
- All media are in a common digital format and can be stored and accessed by a single device, a digital computer (compatibility).

- Distribution
- The development of the Internet, particularly the World Wide Web, dramatically improved the distribution of digital information.

Digital Challenges

- File Sizes
- A minute of CD-quality sound produces a file of over 10MB. A minute of uncompressed full-screen, full-fidelity (24-bit)[depth of the pixel] video takes up approximately 17GB.

- Solutions to the challenge of large file sizes involve improvements in hardware and software.

  1. Many of these focus on compressing the original files.

2. Other solutions include hard drives with storage capacities in the hundreds of gigabytes and optical technologies such as DVD.

DIGITAL CHALLENGES

- Processor Demand (for rendering)
- Large digital files also burden a computer's processor. Some forms of digitization are particularly processing-intensive. In the final stages of producing a 3-D animation, computers must make many calculations. This process, called "rendering," may take many hours, or in the case of lengthy and complex presentations, days and weeks, to complete.

DIGITAL CHALLENGES

- Standardization
- The absence of a common standard fragments the multimedia market and discourage the development of new applications. These are often incompatible: data formats that work with one type of hardware or software do not work with others.

- Over time, a combination of market forces and the work of standards committees generally produce a dominant standard. This, in turn, affected the development of multimedia products and services.

DIGITAL CHALLENGES

- In the early stages of CD technology, different computer manufactures used different standards for encoding data. **The development of a common standard (ISO 9660) made it possible for most computers to access the information on any CD, thus encouraging the development of many more multimedia applications.**

Digital Challenges

- Bandwidth
- A challenge for network users is bandwidth. *Bandwidth* is the amount of digital data that can be transmitted over a communications medium, or band.

- We need to have high bandwith
- Improvements in bandwidth include cable modems, DSL (Digital Subscriber Line, 1.5Mbps), digital lines such as T1 (1.54Mbps), and satellite communications (2Mbps).

Digital Challenges

- Preservation
- the durability and reliability of long-term storage media such as digital tape (must be periodically refreshed), CDs, and DVDs (the durability of optical format is not completely known); and the availability of the hardware and software required to read the archived files.
- Migration (transferring from on format to another that can be manipulated by the current software or hardware).
- Emulation is the use of a new technology to reproduce the operability of an older one