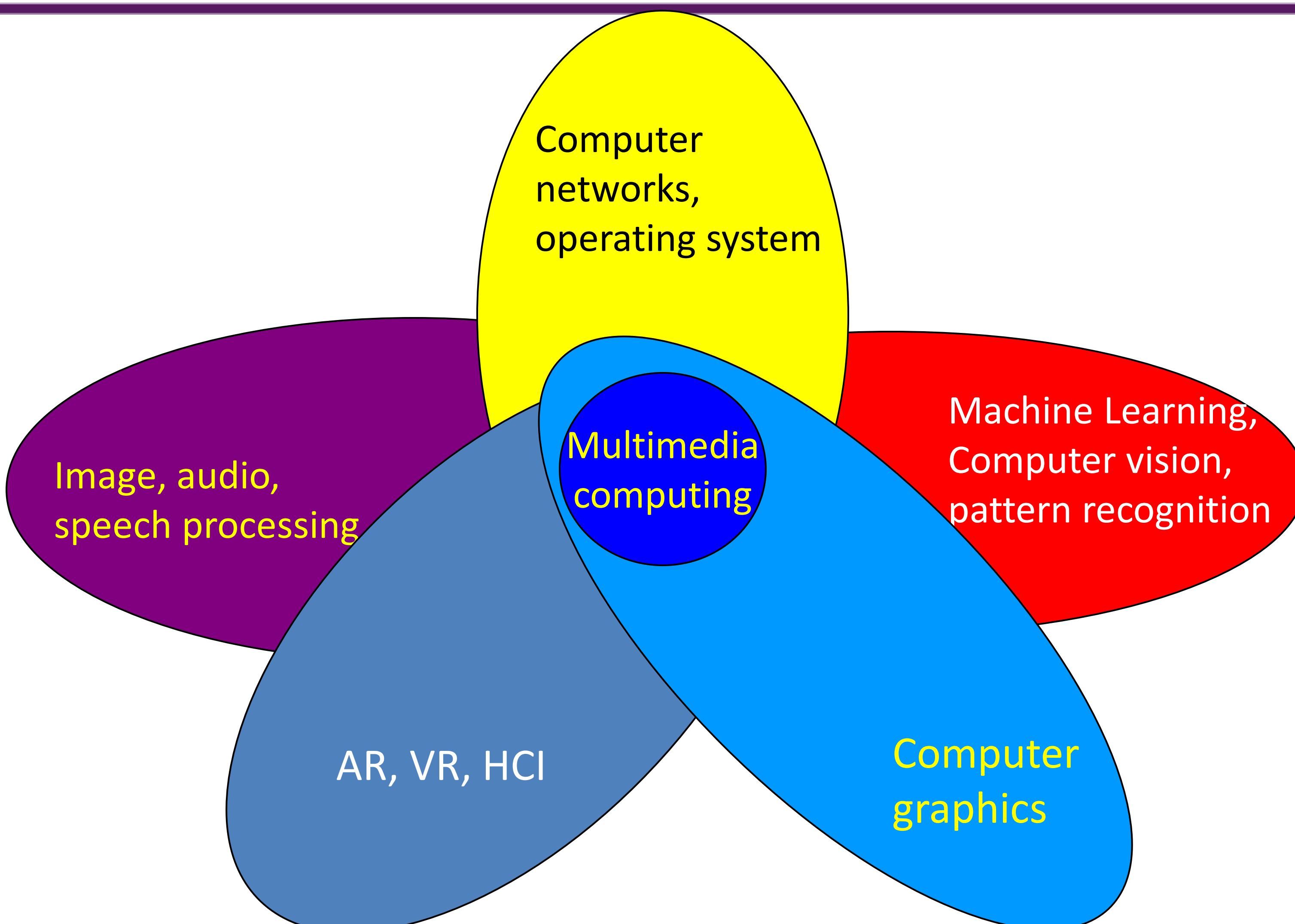
CS 3570
Introduction to Multimedia

Shang-Hong Lai
CS Dept., NTHU

What is Multimedia?

- "Multimedia" has no strict definition.
- In our context, multimedia indicates the computer technology (multimedia computing) for more efficient communication by using different media types:
  - Text
  - Audio and speech
  - Images
  - Graphics
  - Video

Multimedia is Multidisciplinary



Multimedia Systems

- Digital media is multimedia driven by computers. You can see it, hear it, maybe even touch it, and certainly interact with it.
- A Multimedia System is a system capable of processing multimedia data and applications.
- A multimedia system is characterized by the processing, storage, generation, manipulation and rendering of multimedia information.

Characteristics of a Multimedia System

- A multimedia system has the following basic characteristics:
  - Multimedia systems must be computer controlled.
  - Multimedia systems are integrated.
  - The information they handle must be represented digitally.
  - The interface to the final presentation of media is usually interactive.

Multimedia: Past and Present

Early History of Multimedia

- **Newspaper**: perhaps the *first* mass communication medium, uses text, graphics, and images.
- **Motion pictures**: conceived of in 1830's in order to observe motion too rapid for perception by the human eye. Silent feature films appeared from 1910 to 1927.
- **Wireless radio transmission**: Guglielmo Marconi conducted the first wireless radio transmission in Italy, in 1895.
- **Television**: the new medium for the 20th century, established video as a commonly available medium and has since changed the world of mass communications.

Multimedia: Past and Present

- 1985 – Negroponte and Wiesner co-founded the **MIT Media Lab**.
- 1989 – Tim Berners-Lee proposed the **World Wide Web**
- 1991 – **MPEG-1** was approved as an international standard for digital video — led to the newer standards, **MPEG-2**, **MPEG-4**, and further **MPEGs** in the 1990s.
- 1992 – **JPEG** was accepted as the international standard for digital image compression — led to the new JPEG2000 standard.
- 1993 – The University of Illinois National Center for Supercomputing Applications produced **NCSA Mosaic**—the first full-fledged browser.
- 1996 – **DVD video** was introduced; high quality full-length movies were distributed on a single disk.
- 1998 – **XML** 1.0 was announced as a W3C Recommendation. **Hand-held MP3 devices** first made inroads into consumerist tastes in the fall of 1998, with the introduction of devices holding 32MB of flash memory.

Multimedia in the New Millennium

- **2003** Skype: free peer-to-peer voice over the Internet.
- **2004** Web 2.0 promotes user collaboration and interaction. Examples include social networking, blogs, wikis.

  - Facebook founded. Flickr founded.
- **2005** YouTube created.

  Google launched online maps
- **2006** Twitter created: 500 million users in 2012, 340 million tweets per day.

  - Amazon launched its cloud computing platform.

  - Nintendo introduced the Wii home video game console -- can detect movement in three dimensions.
- **2007** - Apple launched iPhone, running the iOS mobile operating system.

  - Google launched Android mobile operating system.

Multimedia in the New Millennium

- **2009** - The first LTE (Long Term Evolution) network was set, an important step toward 4G wireless networking.
  - James Cameron's film, Avatar, created a surge on the interest in 3D video.
- **2010** - Netflix migrated its infrastructure to the Amazon's cloud computing platform.
  - Microsoft introduced Kinect, a horizontal bar with full-body 3D motion capture, facial recognition and voice recognition capabilities, for its game console Xbox 360.
- **2012** - HTML5 subsumes the previous version, HTML4. HTML5 is a W3C "Candidate Recommendation"; it is able to run on low powered devices such as smartphones and tablets.
- **2013** - Sony released its PlayStation 4 a video game console, which is to be integrated with Gaikai, a cloud-based gaming service that offers streaming video game content.
  - 4K resolution TV started to be available in the consumer market.

Multimedia in the New Millennium

- **2015** YouTube launched support for publishing and viewing 360-degree videos, with playback on its website and its Android mobile apps.

  - AlphaGo became the first program to beat a human professional player. Its core technology Deep Learning attracted significant attention and have seen success in multimedia content understanding and generation.
- **2016** HoloLens, a pair of mixed reality smartglasses developed and manufactured by Microsoft, started to be available in the market.

  Pokémon Go, an augmented reality (AR) mobile game, was released and credited with popularizing location-based and AR technologies.

- **2017** TikTok, a video-sharing social networking service for creating and sharing short lip-sync, comedy, and talent videos, was launched for the global market (it's Chinese version, Douyin, was launched in 2016).

Multimedia in the New Millennium

- **2018** The world's first 16K Ultra High Definition (UHD) short video film, Prairie Wind, was created.

  5G cellular systems started deployment, providing enhanced mobile broadband and ultra low latency access.

- **2020** Due to the outbreak of corona virus (COVID-19) around the world, work/study from home became a norm in early 2020. Multimedia-empowered online meeting and teaching tools were booming use.

- **2022** OpenAI announced ChatGPT, which is a conversational AI model built on the GPT (Generative Pre-trained Transformer) architecture with major milestones including GPT-3 (2020), GPT-3.5 (2022), and GPT-4 (2023).

- **2024**: OpenAI announced SORA for generating photorealistic videos from text

Multimedia Applications

- Video conferencing
- Interactive TV
- Computer games
- Virtual reality
- Augmented reality
- Searching large video and image databases for target visual objects, using semantics of objects
- Digital video editing and production systems
- Compositing of artificial and natural videos into hybrid scenes

Example Multimedia System

Chronoscope: Musée d'Orsay

The *Chronoscope* application spreads out an artist's works on a timeline. Paintings by different artists of the same period can be studied side-by-side to explore the cross-fertilization of ideas. Based on impressionist paintings from 1848 to 1914, in the collection of the *Musée d'Orsay* in Paris, France. Interface: Matthew Hodges. Content: Musée d'Orsay, Paris, France.

Chronoscope in MIT's Project Athena

Example Multimedia System

Training in VR: flight simulation

A flight simulator used by the US Air Force (photo by Javier Garcia). The user sits in a physical cockpit while being surrounded by displays that show the environment.

Audio

- Audio signals are continuous analog signals.
- Could contain music, speech, or other sounds
- Input: microphones and then digitized and stored CD Quality Audio requires 16-bit sampling at 44.1 KHz Even higher audiophile rates (e.g. 24-bit, 96 KHz)
- 1 Minute of Stereo CD quality (uncompressed) audio requires 10 MB.
- Usually saved in compressed format

Example-Audio and Speech

- Speech recognition has been widely used commercially
  - Apple Siri
  - Microsoft Azure Cognitive Services for Speech
  - Google Speech-to-Text API
  - Amazon Transcribe (AWS)
  - IBM Watson Speech to Text
  - OpenAI Whisper

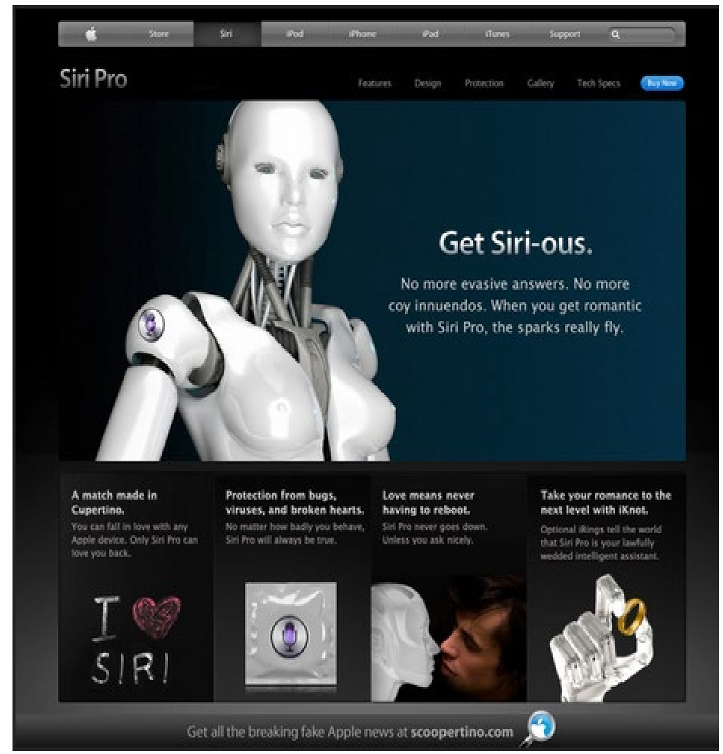


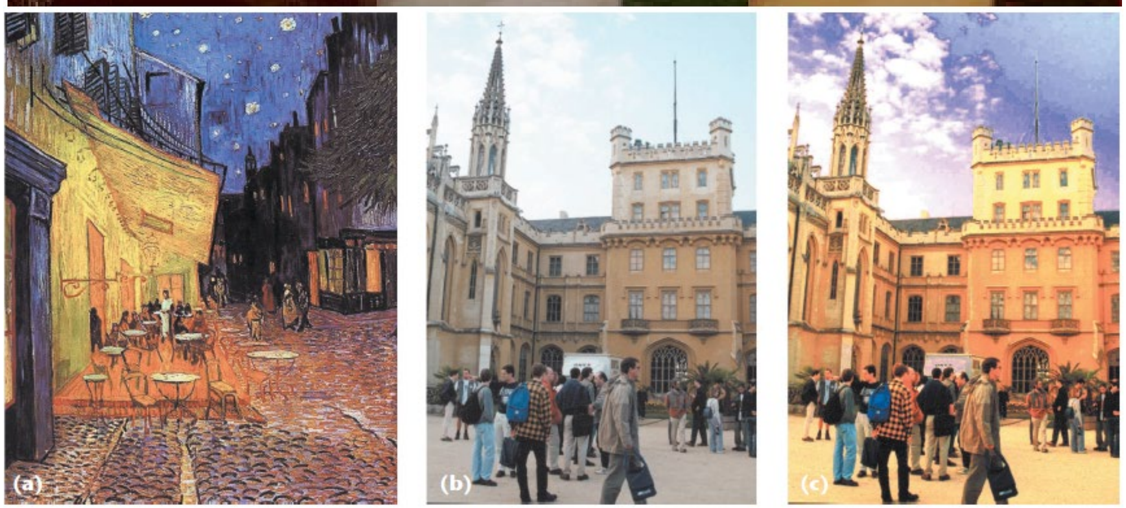
Image from google.

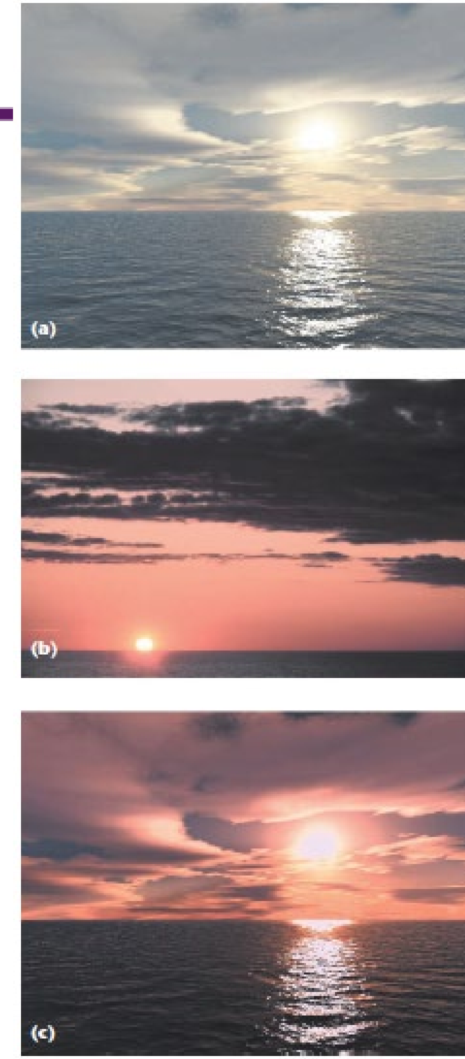
Images

- Still pictures are represented as a bitmap (a grid of pixels).
- Input: digitally scanned pictures or directly acquired from a digital camera.
- Input: May also be generated by programs.
- Stored at 1 bit per pixel (Black and White), 8 Bits per pixel (Grey Scale, Colour Map) or 24 Bits per pixel (True Colour)
- 10 Megapixels digital camera takes 29MB uncompressed!
- Compression is commonly applied.

Example-Image

- Color transformation.
  - Change color information from images to images.





Images from papers "**Color Transfer between Images**".

Example-Image

- Foreground & background segmentation

(a) Girl (4/2/12) (b) Ballet (4/7/14) (c) Boy (6/2/13)

(c) Grandpa (4/2/11) (d) Twins (4/4/12)

Images from papers "**Lazy Snapping**".

Graphics

- Format: constructed by the composition of primitive objects such as lines, polygons, circles, curves and arcs.
- Input: Graphics are usually generated by a graphics editor program (e.g. Illustrator) or automatically by a program (e.g. Postscript).
- Graphics are usually editable or revisable (unlike Images).
- Graphics usually store the primitive assembly

Example-Graphics

Images from NVIDIA

- Special effects in movies

孟加拉虎 Richard Parker 的影像合成

為創造電影中的貓鼬島所採用的合成技術

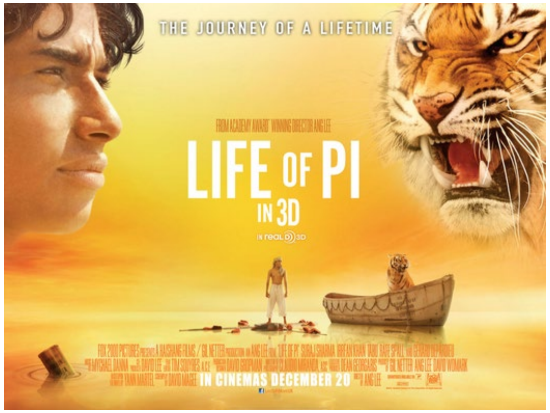
利用 Rampage 繪製《少年 Pi 的奇幻漂流》中的天景

Video

- Input: Analog video is usually captured by a video camera and then digitized.
- Raw video can be regarded as being a series of single images.
- There are typically 25, 30 or 50 frames per second.
- HD video on Blu-ray (up to 1920 x1080 = 2 Megapixels per frame) ~ 6.2 x 25 = 155MB/second to store uncompressed.
- Digital video clearly needs to be compressed for most times.

Example-Video

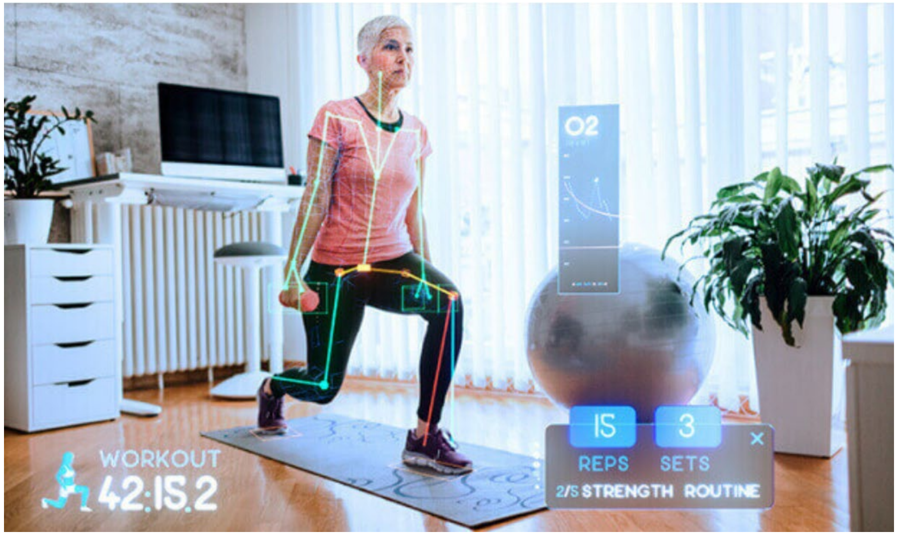
- Stereo videos and movies



Images from google.

Example- Human Interaction

- 3D Human Body Tracking

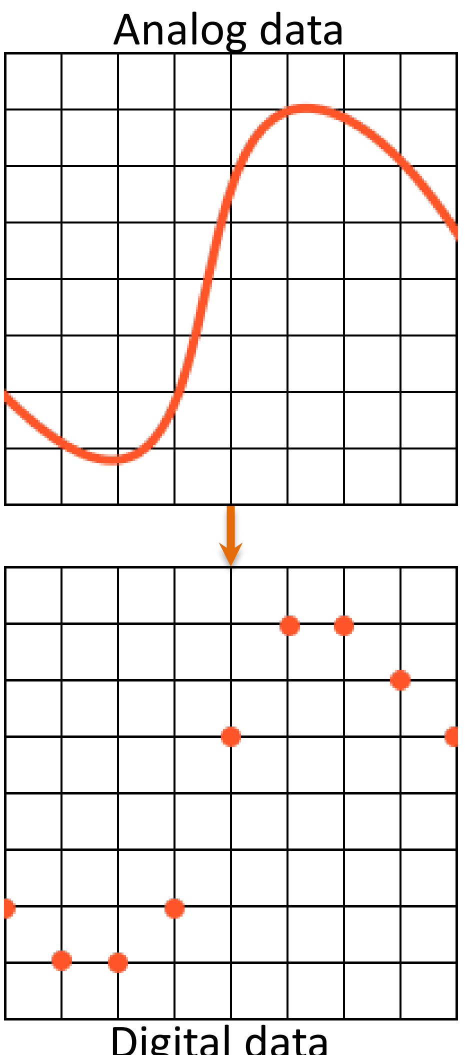


Control games with body motion/gesture and create immersive experiences

3D body tracking for exercise evaluation

Analog and digital data

- Analog data
  - Takes on continuous values
  - E.g. voice, video
- Digital data
  - Takes on discrete values
  - E.g. text, integer
- Converting the continuous phenomena of images, sound, and motion into a discrete representation that can be handled by a computer is called ***analog-to-digital conversion***.



Analog data compared with digital data

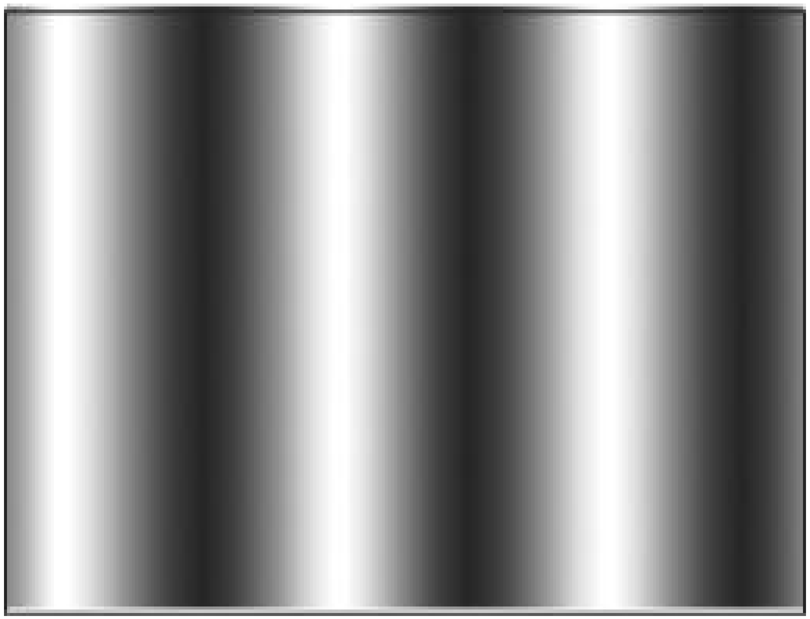
- Analog data > digital data
  - More information -> more precise and better quality
- Digital data > Analog data
  - Size -> communicated more compactly
  - Reliably -> less affected by noise when transmitted

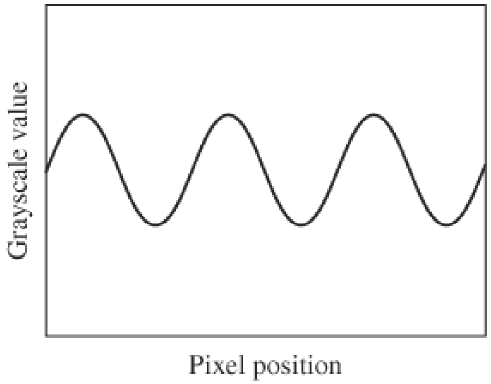

Analog **<->** *Digital*

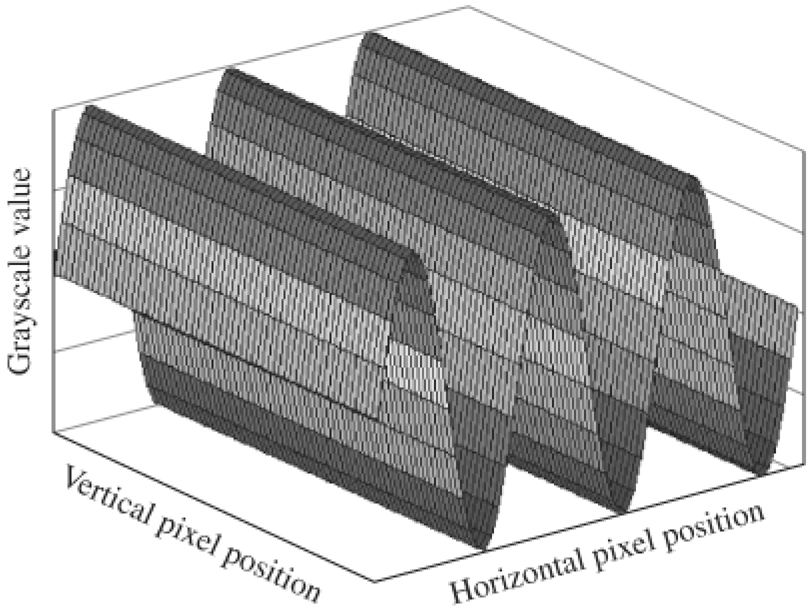
- **Computers** can **only** work with **digital data.**

Images from google.

An example of grayscale image





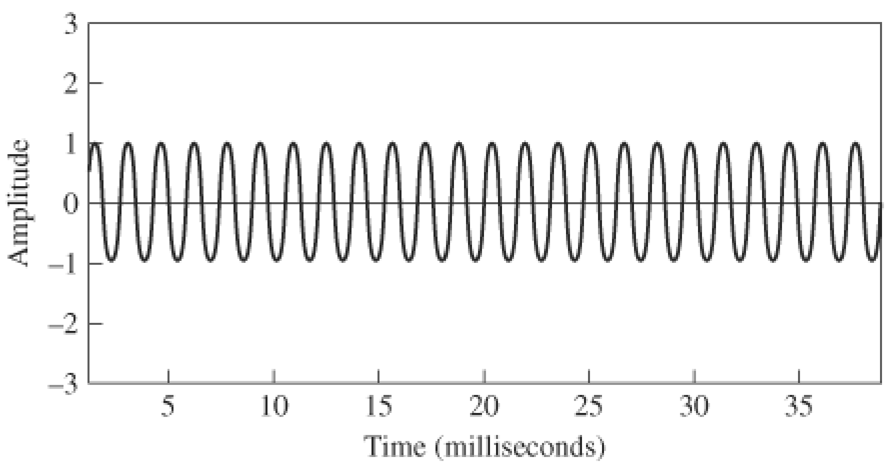


A/D Conversion

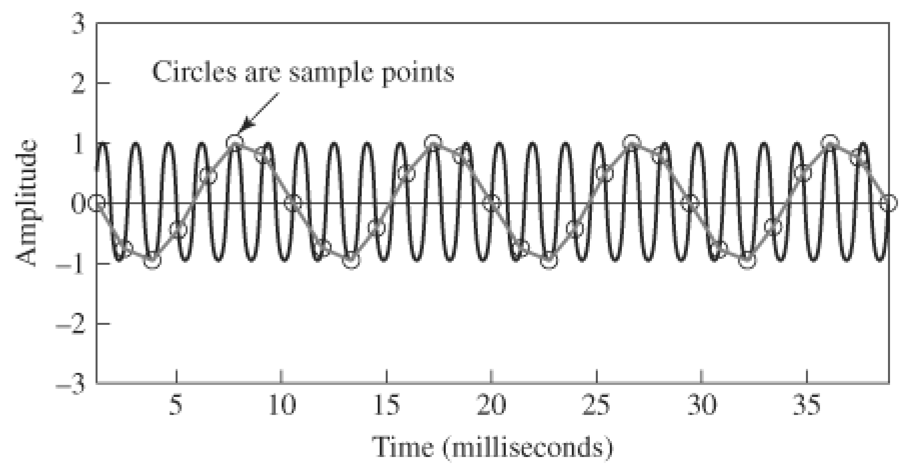
- Analog-to-digital conversion requires two steps: ***sampling*** and ***quantization***.
- The first step, sampling, chooses discrete points at which to measure a continuous phenomenon (***signal***).
- For images, the sample points are evenly separated in space. For sound, the sample points are evenly separated in time.
- The number of samples taken per unit time or unit space is called the ***sampling rate*** or, alternatively, the ***resolution***.
- The second step, quantization, requires that each sample be represented in a fixed number of bits, called the ***bit depth***. The bit depth limits the precision with which each sample can be represented.

Sampling and Aliasing

- ***Undersampling:*** sampling rate does not keep up with the rate of change in the signal.
- ***Aliasing*** in a digital signal arises from undersampling and results in the sampled discrete signal cannot reconstruct the original source signal.



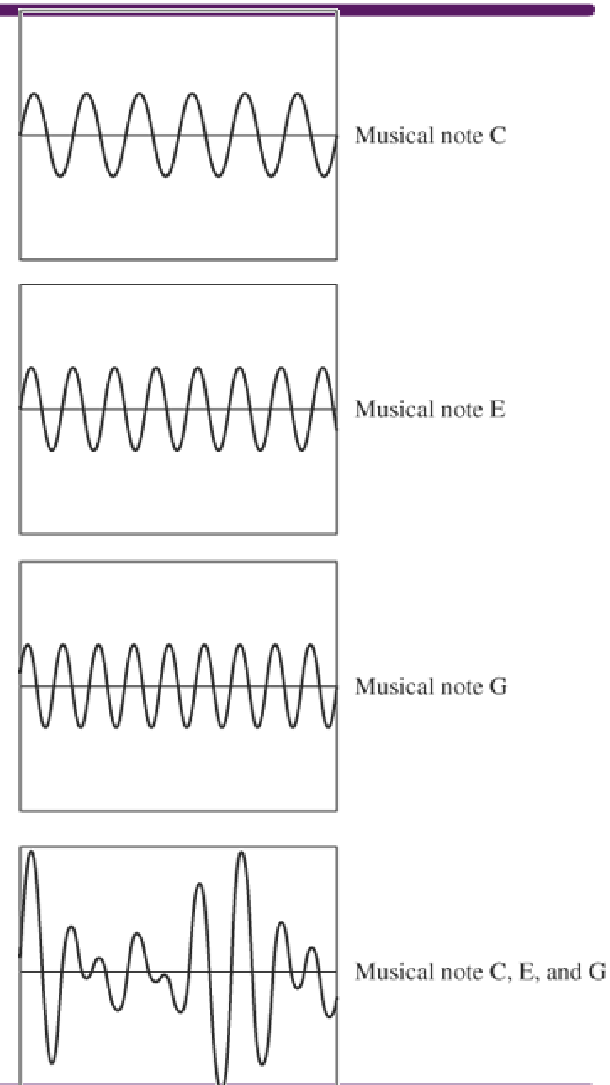
Audio wave at 637 Hz



637 Hz audio wave sampled at 770 Hz

Fourier transform

- Fourier analysis shows that any periodic signal can be decomposed into an infinite sum of sinusoidal waveforms.
- Fourier transform makes it possible to store a complex sound wave in digital form, determine the wave's frequency components, and filter out components that are not wanted.



The image of Barbara

Aliasing due to sampling

Nyquist theorem

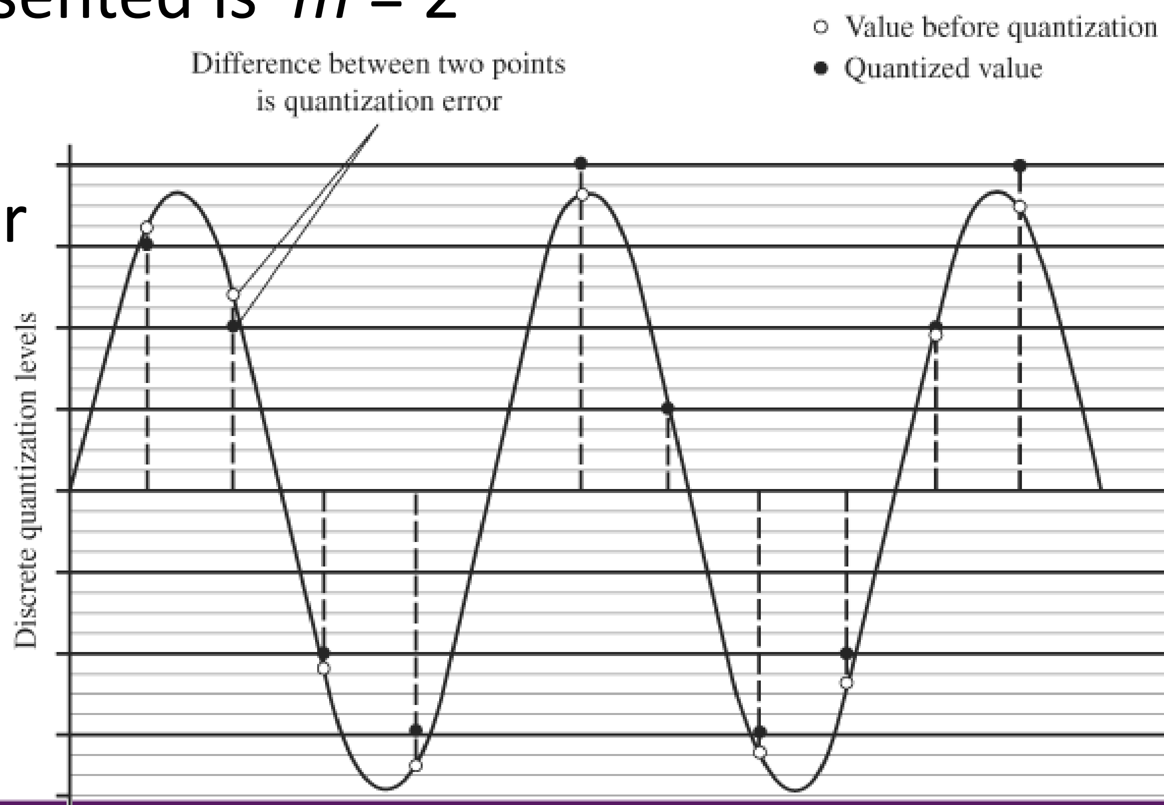
- The Nyquist theorem specifies the sampling rate needed for a given spatial or temporal frequency.
- To guarantee that no aliasing will occur, you must use a sampling rate that is greater than twice the maximal frequency in the signal being sampled.
- Let f be the frequency of a sine wave. Let r be the minimum sampling rate that can be used in the digitization process such that the resulting digitized wave is not aliased. Then the ***Nyquist frequency*** r is

$$r = 2f$$

https://en.wikipedia.org/wiki/Nyquist%E2%80%93Shannon_sampling_theorem

Quantization

- Let n be the number of bits used to quantize a digital sample. Then the maximum number of different values that can be represented is $m = 2^n$
- Quantization error



SNR and SQNR

- Signal-to-noise ratio (SNR) can generally be defined as the ratio of the meaningful content of a signal versus the associated noise.
- In analog data communication, SNR is defined as the ratio of the average power in the signal versus the power in the noise level.
- For a digitized signal, the signal-to-noise ratio is defined as the ratio of the maximum sample value versus the maximum quantization error. This can also be called ***signal-to-quantization-noise ratio*** (***SQNR***)

SQNR and Dynamic Range

- Let n be the bit depth of a digitized media, Then the ***SQNR*** is

$$SQNR = 20 \log_{10}(2^n) \quad \text{(in dB)}$$

- Signal-to-quantization-noise ratio is directly related to ***dynamic range***.
- Dynamic range, informally defined, is the ratio of the largest-amplitude sound (or color, for digital images) and the smallest that can be represented with a given bit depth.

Multimedia in the Future

- *Innovations now or in the near future:*
  - Accurate 2D/3D object tracking from video
  - Intelligent video understanding with Multimodal LLM
  - 3D capture technology for acquiring dynamic facial expression, and synthesizing realistic facial animation
  - Multimedia applications for handicapped persons
  - Deployment of "Digital fashion" + Wearable computing
  - Generative AI is fast evolving to interactively generate high-quality multimedia content very easily

Multimedia in the Future (cont'd)

"Grand challenge" problems, which act as a type of state-of-the-art for multimedia interests:

- Social Event Detection for Social Multimedia: discovering social events planned and attended by people.
- Sports Video Annotation: using video classification to label video segments with certain actions such as strokes in table tennis, penalty kicks in soccer games, etc.
- GameStory: a video game analytics challenge in which e-sport games often involving millions of players and viewers are analyzed.
- Convenient multimedia content generation: LMM

Multimedia in the Future (cont'd)

- Live Video Streaming: requiring ultra low end-to-end latency. The main challenge is the QoE (Quality of Experience), due to the latency constraint.
- Violent Scenes Detection in Film: automatic detection.
- Preserving Privacy in Surveillance Videos: methods obscuring private information (such as faces on Google Earth).
- Deep Video Understanding: understanding the relationships between different entities from a long duration movie. The relations can be family, work, social and other types.

Multimedia in the Future (cont'd)

- Large-scale Human-centric Video Analysis: analyzing various crowd and complex events such as getting off a train, dining in a busy restaurant, earthquake escape, etc.
- Searching and Question Answering for the SpokenWeb: searching for audio content within audio content by using an audio query, matching spoken questions with a collection of spoken answers.
- Multimedia Recommender Systems: improving the quality of recommender systems to produce items more relevant to users' interests. Applications include movie/news recommendation, etc.

Summary

- Multimedia systems & applications
- History of multimedia
- Samples of Multimedia Applications
- Digitization of Multimedia Signals
- Multimedia in the future