Hardware for Multimedia

Input and Output Devices

- Most important components of a multimedia system
- Devices classified as per their use
- Key devices for multimedia output
  - Monitors for text and graphics (still and motion)
  - Speakers and MIDI interfaces for sound
  - Specialized helmets and immersive displays for virtual reality
- Key devices for multimedia input
  - Keyboard and OCR for text
  - Digital cameras, scanners, and CD-ROMs for graphics
  - MIDI keyboards, CD-ROMs and microphones for sound
  - Video cameras, CD-ROMs, and frame grabbers for video
  - Mice, trackballs, joy sticks, virtual reality gloves and wands, for spatial data
  - Modems and network interfaces for network data
- Monitors
  - Most important output device
  - Provides all the visual output to the user
  - Should be designed for the highest quality image, with least distortion
  - Large vacuum tube with electron gun at one end aimed at a large surface (viewing screen) on the other end
  - Viewing screen is coated with chemicals that glow with different colors; three different phosphors are used for color screens
  - Source of electron beam is electrically negative pole or cathode (hence the name Cathode Ray Tube, or CRT)
  - Two different sets of colors used in monitors – RGB and CMY, with either set capable of full color spectrum
  - Electron beam strikes the screen many times per second
    * Phosphors are re-excited at each electron strike for a brief instance
    * Refresh rate, measured in Hz
    * Preferred refresh rate is 75 Hz or more
  - Electron beam sweeps across the screen in a regular pattern
    * Required to refresh phosphors frequently and equally
    * Raster scan pattern
    * Always strikes when going from left to right (trace), and turned off to go from right to left (retrace)
  - Three separate electron beams for three colors, for better focus and higher refresh rates
  - Screen divided into individual picture elements, or *pixels*
    * Each pixel is made of its own phosphor elements to give the color
    * Memory chip contains a map of what colors to display on each pixel
  - Bit map
    * Mostly used in context of binary images (black or white)

    * One bit per pixel to indicate whether pixel is black or white
  - Color maps, or pixmap
    * One byte for each color for every pixel (24-bit color)
  - Image changed in the memory map associated with screen
    * For realistic motion images and for flicker-free screen, bit-map must be modified faster than the eye can perceive (30 frames/sec)
    * For a 640×480 screen, number of bits is: $640 \times 480 \times 24 = 7,372,800$
    * To refresh the screen at 30 times per second, the number of bits transferred in a second is: $640 \times 480 \times 24 \times 30 = 221,184,000$ or 221 Mb
    * Larger screen requires more data to be transferred
    * Transfer rate limitation can be overcome by using hardware accelerator board to perform certain graphic display functions in hardware
    * Full-screen 30 image per second performance may not be possible even with graphics accelerator board
  - Physical size of monitor
    * Important factor in the quality of multimedia presentation
    * Typically between 11 and 20 inches on diagonal
    * Another important factor is the number of pixels per inch
      · Too few pixels make the image look grainy
      · For best quality images, pixels should not be wider than 0.01 inches (28mm) in diameter
      · Latter quantity is used for marketing the monitors (25mm dot pitch)
  - Graphics display board
    * Used in addition to monitor to speed up graphics
    * Special hardware circuits for 2D and 3D graphics
    * Simple graphics boards just translate image data from RAM into one usable by monitor
    * Complex boards can even speed up the refresh rate of screen
  - Qualities of a good multimedia monitor
    * Size, refresh rate, dot pitch
  - Other concerns about monitor include weight and ambient light
  - Liquid crystal display monitors
    * Flat screen displays
    * Crystals allow more or less light to pass through them, depending upon the strength of an electric field
    * Not appropriate for multimedia presentation as the view angle is extremely important
  - 3D monitors in the future
  - Human factor concerns

• Speakers and MIDI interfaces
  - Production of sound
    1. Digitized representation of frequency and sound transmitted at appropriate time to the loudspeaker (`.WAV` files) – common method
    2. Commands for sound synthesis can be transmitted to a synthesizer at appropriate time (`midi` files) – used for the generation of music
  - Musical Instruments Digital Interface (MIDI)
    * Standard to permit interface for both hardware and control logic between computers and music synthesizers
    * Adopted in 1982

    * Consists of two parts
      1. Hardware standard
         · Specifies cables, circuits, connectors, and electrical signals to be used
      2. Message standard
         · Types and formats of messages to be transmitted to/from synthesizers, control units (keyboards), and computers
         · Messages consist of a device number, a control segment to tell the device the function to be performed (turn on/off a specified circuit), and a data segment to provide the information necessary for the action (volume of sound, or frequency of basic sound)
    * An entire piece of music can be described by a sequence of MIDI messages
  - MIDI interface
    * Required in the computer to communicate with MIDI instruments
    * Circuit board to translate the signals

- Alphanumeric keyboards and optical character recognition
  - Used for textual input
  - Pressing a key on a keyboard closes a circuit corresponding to the key to send a unique code to the CPU
  - Printed text can be input using OCR software
    * OCR software analyzes an image to translate symbols into character codes
    * Systematically checks the entire page, searching for patterns of dark and light recognizable as alphabetic, numeric, or punctuation characters
    * Choose the best match from a set of known patterns
  - Quality of scanned page as well as output

- Digital cameras and scanners
  - Real image – something present in nature
  - Digital image – Representation of real image in terms of pixels
  - Still image – Snapshot of an instance
  - Motion image – Sequence of images giving the impression of continuous motion
  - Graininess in real images
    * Individual dots observed when a photograph taken by conventional camera is enlarged sufficiently
  - Digital image capture
    * Light is focused on photosensitive cells to produce electric current in response to intensity and wavelength of light
    * Electric current is scanned for each point on the image and translated to binary codes
    * Codes correspond to pixel values and can be used to rebuild the original picture
  - Scanners scan an image from one end to the other
    * Scanning mechanism shines bright light on the image and codes and records the reflected light for each point
    * Scanner does not store data but sends it to the computer, possibly after compression of the same
  - Quality of images
    * Depends on the quality of optics and sharpness of focus
    * Perceived by sharpness of resulting image
    * Accuracy of encoding for each pixel depends on the precision of photosensitive cells
    * Resolution of scanner/camera (number of dots/inch)
    * Amount of storage available

  - Preferable to scan at the highest possible resolution under given hardware and storage space constraints to get the most detail in the original image

- Video camera and frame grabbers
  - Standard video camera contains photosensitive cells, scanning one frame after another
  - Output of the cells gets recorded as analog stream of colors, or sent to digiting circuitry to generate a stream of digital codes
  - Video input card
    * Required for use of video camera to input video stream into computer
    * Digitizes the analog signal from camera
    * Output can be sent to a file for storage, CPU for processing, or monitor for display (or all of them)
  - Frame grabber
    * Allows the capture of a single frame of data from video stream
    * Not as good resolution as a still camera
    * Typical frame grabbers process 30 frames per second for real time performance
- Microphones and MIDI keyboards
  - Used to input original sounds (analog)
  - Microphone has a diaphragm that vibrates in response to sound waves
  - Vibrations modulate a continuous electric current analogous to sound waves
  - Modulated current can be digitized and stored as standardized format for audio data, such as `.WAV` file
  - Microphone plugs into a sound input board
    * Developer can control the sampling rate for digitizing
    * Higher sampling rate gives better fidelity but requires more space
    * Sampling rate for music – 20,000 Hz
    * Sampling rate for speech – 10,000 Hz
  - Editing digital audio files (cut and paste)
- Mice, trackballs, joy sticks, drawing tablets, ...
  - Used to enter positional information as 2D or 3D data from a standard reference point
  - Latitude, longitude, altitude
  - Common to define a point on the computer screen
  - Mouse defines the movement in terms of two numbers – left/right and up/down on the screen, with respect to one corner
  - Movement of mouse is tracked by software, which can also set the tracking speed
  - Trackball works the same way as the mouse
  - A joystick is a trackball with a handle
  - Pressing the button associated with the mouse/trackball/joystick sends a signal to the computer asking it to perform some function using the cursor for context
  - Multimedia software should be able to determine the positional information as well as the signal context (mouse press)
- CD-ROMs and video disks
  - Popular media for storage and transport of data
  - Data written on disk by burning tiny holes, interpreted as binary 0 and 1 by software

  - Read-only devices; data can be written only once
  - CD-ROMs can typically store about 600MB of information
  - With time, the speed has improved (4X in 1995 to more than 50X now)
  - DVD-ROMs allow a few gigabytes of data on a single disk
  - Ideal media for distributing multimedia productions (low cost)

Virtual Reality Devices

- Provide artificial stimuli to the senses of the user
- Substitute for input from physical world surrounding the system
- Virtual reality output devices
  - *Immersion* of the VR system
    - Extent of user isolation from the world
    - Reception of artificially generated stimuli in lieu of the world
    - Greater immersion requires sophisticated output devices
    - Expensive in terms of hardware, programming, and computing power
  - Design requirements for a particular multimedia system and cost/benefit of using a particular piece of VR hardware
  - Primary stimuli are visual and aural
  - Motion may be possible using hydraulics that are programmed in conjunction with visual and audio data
  - Not much in terms of touch and smell
- Visual output
  - Presented on a screen or head-mounted projection device
  - Immersion environments
    - CAVE
      - CAVE Automatic Virtual Environment
      - Most immersive VR visual output environment
      - Developed at NCSA at UIUC
      - Room about 10 feet square formed by rear projection screens
      - Images controlled by a high-speed graphics computer
      - User needs to wear special headgear with 3D glasses and a head motion tracking device
      - 3D glasses make the image appear to be actual 3D objects within the room
      - Head tracking device is coupled to a controlling computer which varies the images so that they appear to move in response to head movements
      - Expensive to build and maintain
    - ImmersaDesk
      - An inexpensive version of CAVE for desktop systems
      - Has only one rear projection screen
      - Applications include versions of Quake and Doom
  - Head-mounted displays
    - Disables visual stimuli from outside world from reaching the user
    - A large helmet to go on top of user's head
    - Small screen suspended in front of eyes

        - Could be two small screens, one in front of each eye
        - Two screens can have two phases of the same image to give stereoscopic effect
        - Screens should have excellent focus, extremely high resolution, and realistic colors
    - HMD should be light in weight (human factors)
    - Should provide at least 120° vertical view and about 160° horizontal view
  - Limitations
    - Small flat screens are made using LCD
    - Problem with the resolution and brightness levels of LCD
    - The response time to change for LCD may not be acceptable
  - Parallax
    - Change in position of stationary object when viewed from slightly different position
    - Each eye views the objects at slightly different position
    - Amount of apparent motion of object is a function of distance from the eye
      - As the distance to object approaches infinity, apparent motion goes to zero
    - Problem in capturing parallax information with motion of camera
      - Parallax information may not be due to motion of user's head
    - Problem in capturing and storing views with 360° scope
      - Partially solved by panning camera
  - Retinal images
    - Project the image directly on the retina of viewer's eyes
    - Image projected by LEDs and reflected onto retina by a small mirror
    - Display limited to monochrome images with moderate resolution

- Aural output
  - Two primary factors related to perception of sound – localization and identification
  - Sound output must change subtly so that it appears to come from the same location no matter where the head is pointed
  - Current sound systems are not realistic with regard to controlling the precise location of the source
- Virtual reality input devices
  - Most input performed by using mechanical devices such as buttons of a joystick
  - Problem to employ unobtrusive virtual input devices that perform like the real devices
  - Position sensing
    - Accomplished by means of some form of radiated signal
    - Signal could be visible light, infrared, ultrasound, or laser
    - Signal emitted from a device mounted on subject, or reflected off the subject
    - Subject can be made to wear devices containing sensors/emitters to send signals
      - Wearable devices can transmit information about many points simultaneously
      - A glove can transmit information about all fingers
    - Position is given in terms of three mutually perpendicular axes
    - It may be required to get the orientation of the object as well
      - Orientation defined in terms of terminology used by pilots
      - *Yaw* – Rotation along the Y (vertical) axis
      - *Pitch* – Rotation along the Z (left-right) axis
      - *Roll* – Rotation along the X (front-back) axis
  - Motion

    * Specified in terms of change in position and orientation
    * Six degree of freedom corresponding to six parameters
    * Sensor output can be a continuous stream of data or sent only upon request
    * Polling reduces the amount of network traffic but may miss quick changes in position
    * Lag of latency
      · Delay from actual time of motion and when it is interpreted
      · Should not exceed 50 msec to avoid being perceived by user
    * Update rate
      · Rate at which measurements are made
      · Slow update rate makes the motion look jerky
    * Precision and accuracy of measurements
      · Accuracy varies with particular application but should be as high as possible
      · Accuracy depends on analog to digital converters
    * Range of sensors
      · Maximum range/distance over which motion can be sensed
      · Dimensions of a room, geocells in flight simulators, distance over which a hand can move
    * Degree to which sensor screens out interference from ambient sources
  – Voice input
    * Speech or voice recognition
    * Form of pattern recognition
    * Spoken sound patterns are matched against previously recorded patterns
    * Problems
      · Voice quality of different people – pitch, timbre, volume, rate of speech, accent
    * Computer can be trained by the subject by speaking certain words repeatedly
    * Limited vocabulary
    * Natural language processing
      · People use different words for same thing (can i use your pen?)
      · Some sentences make sense but cannot be properly parsed
      · Accentuating a word may be important
      · Tone of speaker's voice can alter the meaning of words
      · Cultural or language issues (In India, you always pass out from college)
      · Homonyms (see vs sea, know vs no)
      · Relative position of words (Only the son praised his sister.)
    * Limited vocabulary can still be used for commands to substitute point-and-click

Modems and Network Interfaces

- Network interface
  - Translate the signals from computer to network and the other way round
- Serial and parallel
  - Each character represented by a set of bytes (typically from 7 to 16)
  - Bits may be transmitted in parallel (within computer) or serial (over the network)
  - Parallel transmission is faster but requires extra wires (more expensive)
  - Interface can convert from serial to parallel and vice versa

- Character encoding
  - ASCII and EBCDIC
  - ASCII uses 7 bits per character, but extended ASCII uses 8 bits to represent special characters
  - Unicode
    * Fixed-width. uniform text and character encoding scheme
    * Includes characters from world's scripts, including technical symbols
    * Uses 16-bits
    * No escape sequences required for characters
  - ISO/IEC 10646-1:1993 standard
    * 32-bit character encoding
    * Includes Unicode as one 16-bit portion of the standard
- Start/Stop/Error-checking codes
  - Used to inform the device of beginning and end of serial transmission
  - Needed to identify a change of state on the transmission medium
    * Transmission medium with 0 shows no data being transmitted
    * Need to transmit data starting with 0
    * Achieved by sending a start bit that is opposite of idle state
    * Next eight bits contain data
  - Serial data needs to be converted to parallel as eight bits are needed together to signal a character
  - Stop bit ensures that the translation from serial to parallel has been achieved before more data is sent
  - Some bits may be used for error detection/correction
- Transmission rate
  - Internal transmission rate is much faster than transmission rate across machines over the network
  - Interface needs to account for the change in data transmission rate
  - Signal from interface to computer (interrupt) informs about when it has received a byte and is ready to transmit it forward
- Transmission form
  - Signal can be transformed from two voltage levels (binary) to something suitable for transmission as voice over phone lines
  - Translation achieved through a modem (modulator/demodulator)
  - No special communication lines are required, except phone lines
  - Limited in transmission speed
  - A speed of 56K still may not be fast enough for image downloading
  - Multimedia designer needs to be concerned about the number of images being transmitted, possibly over slow connections