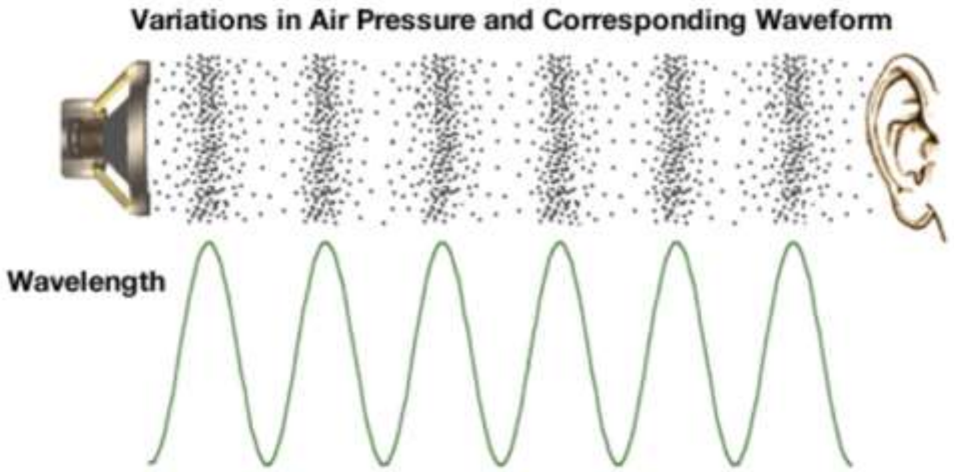
# CHAPTER 2: SOUND AND AUDIO

Course: Multimedia Systems

# SOUND

- Physical phenomemon produced by the vibration of matter. E.g. String, Instruments
- Sound is a form of energy that travels through mediums like air, water and solilds as vibration of waves but not in Vaccum
- As the matter vibrates, pressure variation occurs that triggers in the surrounding.
- This alteration of high and low pressure is propagated through different meduims and the sound is heard.



# SOUND..CONTD

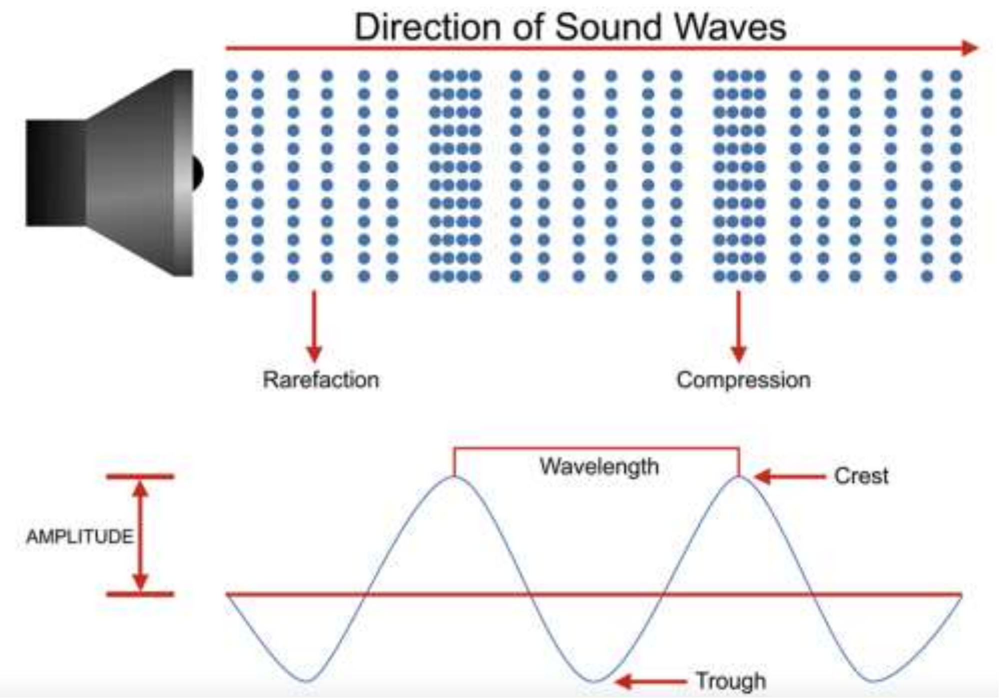
- The speed of sound depends upon which medium it travels.
- The mediums are Solid, Liquid and Gas(Air)
- Sound waves travel faster in those mediums where there are closely packed molecules.

Speed of Sound in Different Materials
(20°C and 1 atm)

| Material | Speed (m/s) |
|---|---|
| Air | 343 |
| Air (0 °) | 331 |
| Helium | 1005 |
| Hydrogen | 1300 |
| Water | 1400 |
| Sea Water | 1560 |
| Iron and Steel | ≈ 5000 |
| Glass | ≈ 4500 |
| Aluminum | ≈ 5100 |
| Hardwood | ≈ 4000 |
| Concrete | ≈ 3000 |

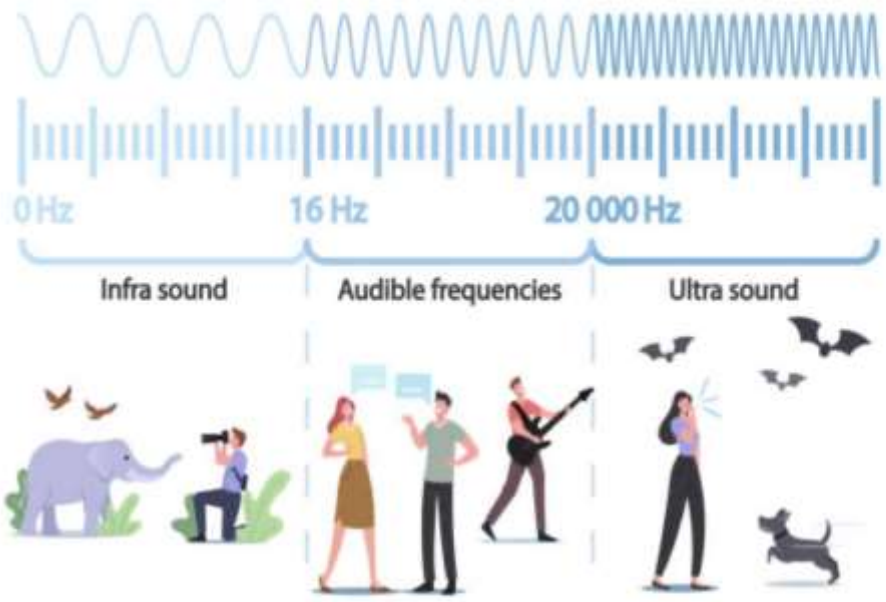
# CHARACTERISTICS OF SOUND

- Loudness: related to the intensity of sound waves (as the volume increases the amplitude of the waves increases)
- Pitch: Related to the frequency of the waves
- Audiable range : audiable to human(20 Hz – 20000Hz)
- UltraSound: High pitch sound(above 20 KHz)
- InfraSound: Very low pitch sound (Below 20 Hz)



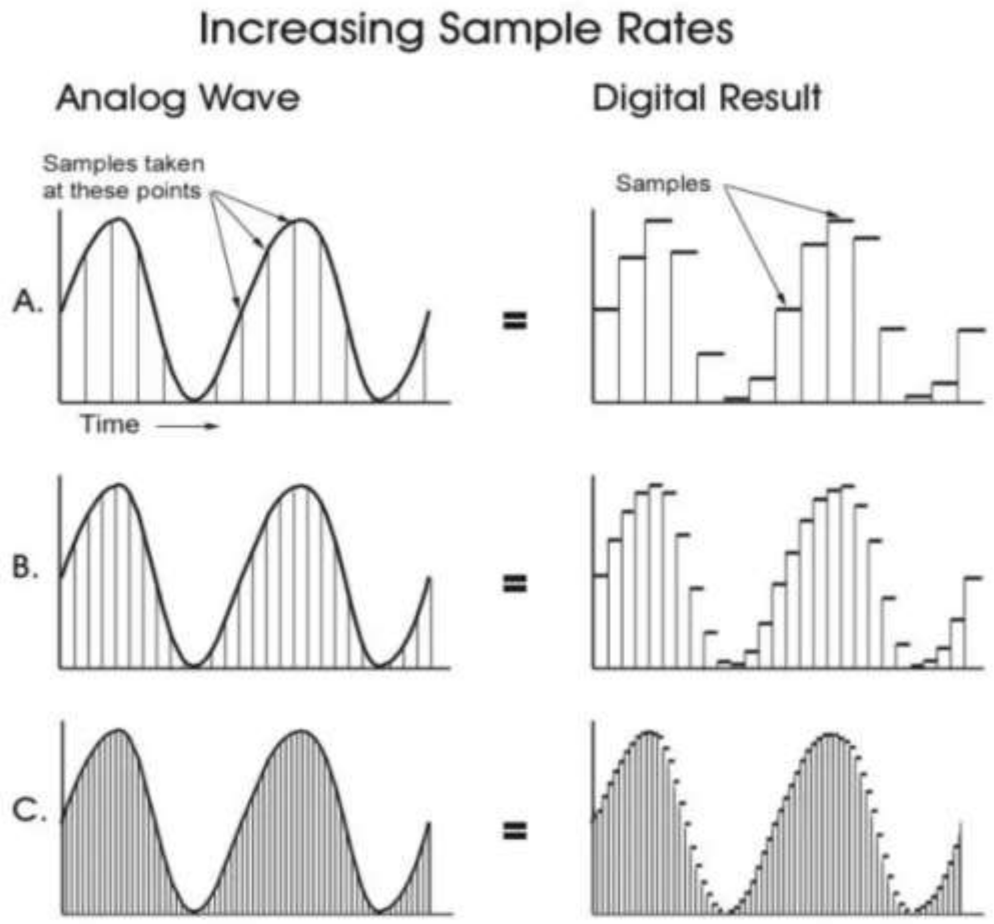
# CONTD..

- Frequency: Reciprocal of time period(f=1/t)
- Measered in Hz (Hertz): cycles per period
- 1KHz= 1000 Hz
- Periodic waveform: repeats after certain interval, easier to predict E.g. sinewave, squarewave
- Non-Periodic waveform: random sound, Harder to predict. E.g. noise, slushing of water



# COMPUTER REPRESENTATION OF SOUND

- Computer measures the amplitude of the waveform at regular time intervals to produce a series of numbers. These are called Samples.
- Sampling Rate: The rate at which continious waveform is sampled is called sampling rate.
- For standard CD Quality Audio(Uncompressed) sampling rate is 44100Hz (44.1KHz)



# NYGUIST SAMPLING THEORM

- " *For Lossless digitization, the sampling rate should be at least twice of the maximum frequency of the waveform*"
- Sampling theorem refers to the process of converting an analog signal into a sequence of uniformly spaced numbers, allowing the original signal to be accurately reconstructed from its samples taken at a specific rate
- Sampling theorem states that "continues form of a time-variant signal can be represented in the discrete form of a signal with help of samples and the sampled (discrete) signal can be recovered to original form when the sampling signal frequency Fs having the greater frequency value than or equal to the input signal frequency Fm. i.e. : **$F_s \geq 2F_m$**
- When sampling frequency equals twice the input signal frequency is known as "Nyquist rate".

# SAMPLING AND QUANTIZATION

**Sampling**

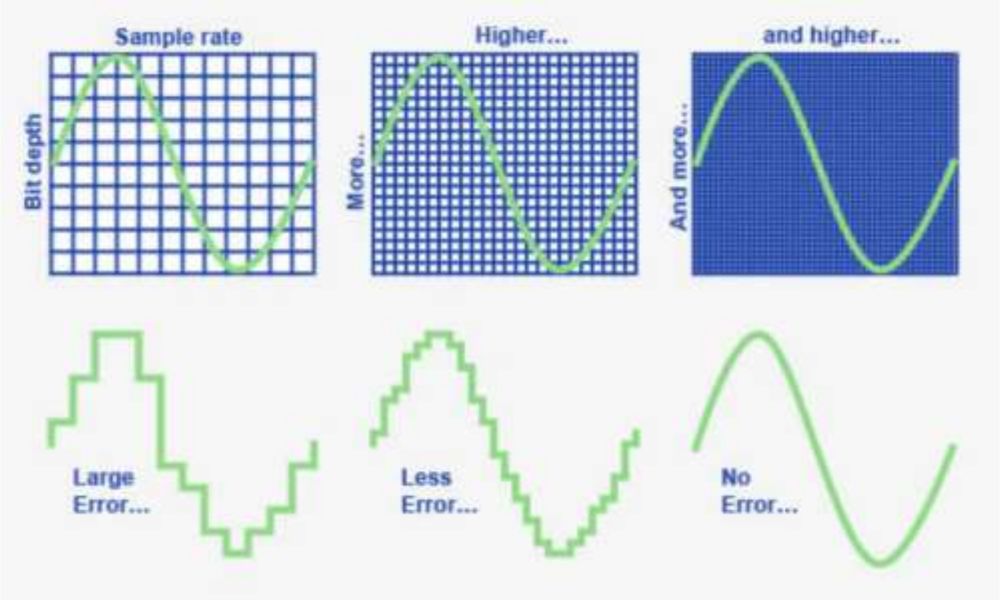
- Provides Resolution in X- axis or Time axis
- Its like taking the snapshot of the audio waveform.
- More number of samples, more accurate the analog signal
- Standard CD-quality audio is sampled at 44.1 kHz . That is 44100 samples per second.
- 48kHz is standard for audio-video

**Quantization**

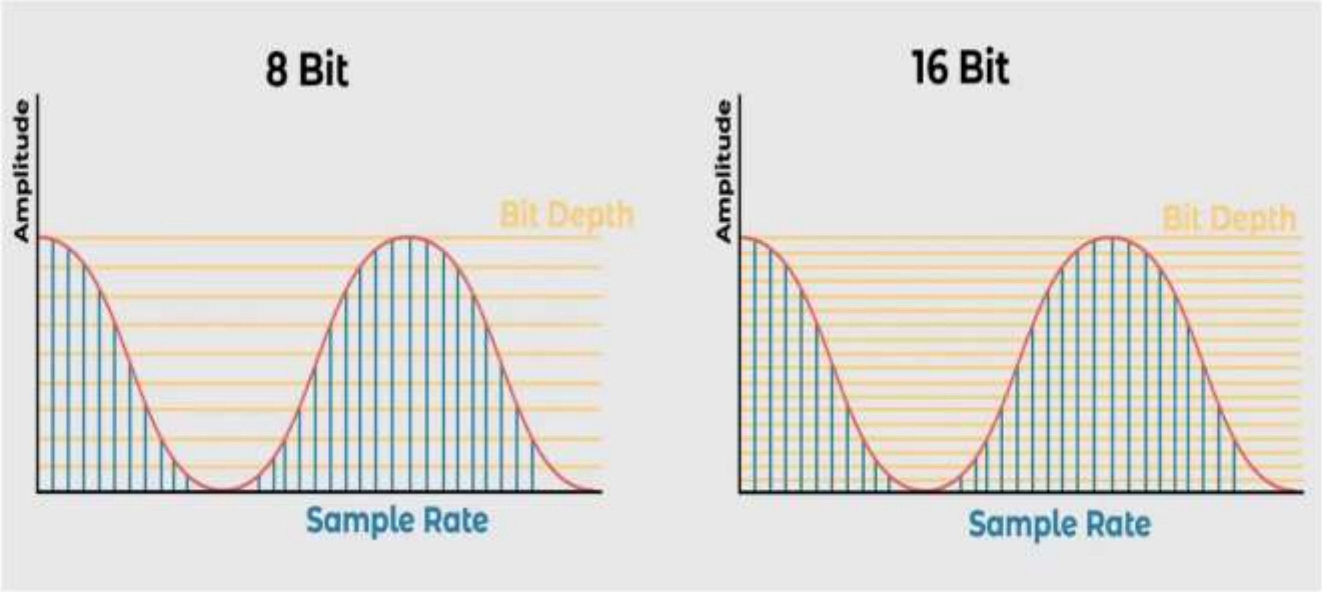
- Provides Resolution in Y- axis or Amplitude Axis
- Set at bit-depths
- The resolution or quantization of a sample value depends upon the number of bits used in measuring the height of the amplitude and rounding them off to the nearest values.
- i.e. 8 bit, 16bit (CD-Quality), 24-bit (Blue Ray). Therefore n-bits = $2^n$ possible values
- Lower the quantization= lower the quality of sound
- Higher the quantization= more dymanic range of the sound

# SAMPLING AND QUANTIZATION

**Sampling**




**Quantization**



# CAN YOU HEAR THE DIFFERENCE AT HIGHER SAMPLING RATES AND BIT DEPTHS?

- This is a very controversial topic in the audio world - the fact is that the human ear can only hear up to 20 kHz anyway, and 44.1 kHz is perfectly adequate.
- In some cases, higher sample rates can be advantageous in audio processing, but in the finished mix there is no audible difference to 44.1 kHz.
- In the past, with 16-bit recorders, you always had to be careful to record as loud as possible so that the background noise was not audible
- 24-bit is very convenient for recording because you don't have to worry about levels. The dynamic range is so wide and the noise floor is very low

# BIT DEPTH AND DYNAMIC RANGE

| Bit Depth | No. of Discrete Values | Dynamic Range |
| --- | --- | --- |
| 4-bit | 16 | 24dB |
| 8-bit | 256 | 48dB |
| 16-bit | 65,536 | 96dB |
| 24-bit | 16,777,216 | 144dB |
| 32-bit | 4,294,967,296 | 1,528dB |

# BIT DEPTH AND FILE SIZE

| Bit Depth | Sampling Rate | File Size Per Minute of Audio (in MB = megabytes) |
|---|---|---|
| 4-bit | 44.1kHz | 1.323MB |
| 8-bit | 44.1kHz | 2.646MB |
| 16-bit | 44.1kHz | 5.292MB |
| 24-bit | 44.1kHz | 7.938MB |
| 32-bit | 44.1kHz | 21.168MB |

# AUDIO FILE SIZE

- Calcualte the file size for CD-quality Audio(Stereo) with the play-back time of 5 minutes.
- For say a 5 min track first we convert time to seconds = 300sec.
- For CD- Quality Audio Sampling Rate = 44.1 KHz = 44100 Hz
- Bit Depth = 16, No. of channel = 2 (For Stereo)
- Formula = (Sample rate x Bit-depth x Time) x No. of Channel
- i.e. 44, 100 x 16 x 300) x 2 = (423, 360, 000) x 2 = 423,360,000 bits

# AUDIO FILE SIZE

- To convert to Bytes, we divide by 8 because a Byte contains 8bits.

i.e. 423,360,000/8 = 52,920,000 bytes.

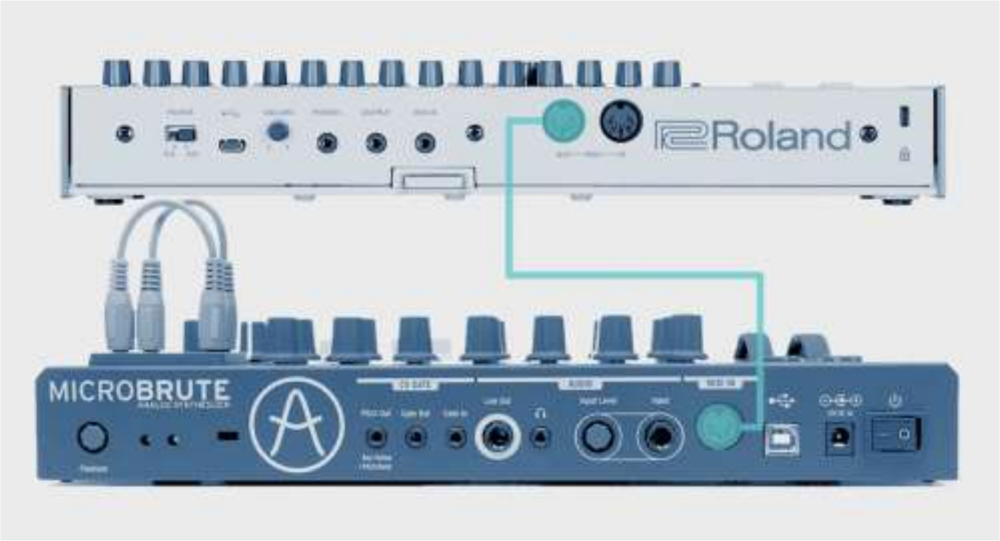
- Converting to Kilobytes by dividing by 1024 (the number of Bytes in a Kilobyte), then dividing by 1024 again (the number of Kilobytes in a Megabyte).
- i.e. 52,920,000/ (1024 x 1024) =52,920,000 / 1,048,576 = 50.468 MB

Question: Calculate the File size(in Megabyte) of an 10 second audio sampled at 48 KHz having bit depth 24 with Mono channel.

# MIDI

- Musical Instrument Digital Interface
- Relationship between Music and computers has become more interconnected considering the development of MIDI
- It is a small piece of equipment that plugs directly into the computer's serial port and allows the transmission of signals
- Now it has become a standard for every manufacturers of electronic musical instruments
- It has a set of specifications so that instruments of different manufacturers can, without difficulty, communicate musical information between one another

# MIDI

# MIDI

Out 1 + 2
MIDI Out
MIDI In
Audio Out
MIDI In
MIDI Out
MIDI Thru
USB
AUDIO
MIDI

# HOW DOES IT WORK?

- A musical instrument takes an input MIDI input signal, which can be as simple as an audio signal, and produces a MIDI output signal.
- MIDI-capable devices can send/receive data in several ways
- A MIDI compatible instrument like the Roland FT-817 produces a MIDI note and plays it on its own devices or can copy it and play it on another instrument.
- The notes are produced when a key is pressed, or a gate button is pushed, sending the resulting electrical signal to an amplifier to generate a pitch that is the audio amplitude of the note.
- MIDI also has considerable digital and analogue manipulation capabilities for the creation of and modification of sounds

# WHAT DOES MIDI DO?

- It is basically a score that is read by a compatible music synthesizer.
- If you play a note on your keyboard, the MIDI converts it into numbers and sends it to the software or hardware that will then make that sound.
- If you change the score, the sound changes.
- MIDI does not send the sound wave made by an instrument; instead, it sends information about the music notes, and the receiving device uses its own internal mechanisms to generate the sounds.
- Many video games use MIDI files for music rather than recorded music because the size of the files is so small.
- This is a popular way for amateur and independent musicians to distribute their music. It cuts out the record company and makes it easier for people to get the music to the audience.

# COMPONENTS OF MIDI INTERFACE

## A MIDI INTERFACE HAS TWO COMPONENTS

1. Hardware:
   - Consits of different physical hardware devies to built the system
   - Specifies the physical connection between musical instruments using MIDI cables

2. Data Format:
   - The information that travels through the hardware devices
   - The encoding includes the notion of the beginning and end of the note, frequency and volume information
   - Also includes information such as stroke intensity(hammering effect)

# MIDI MODES

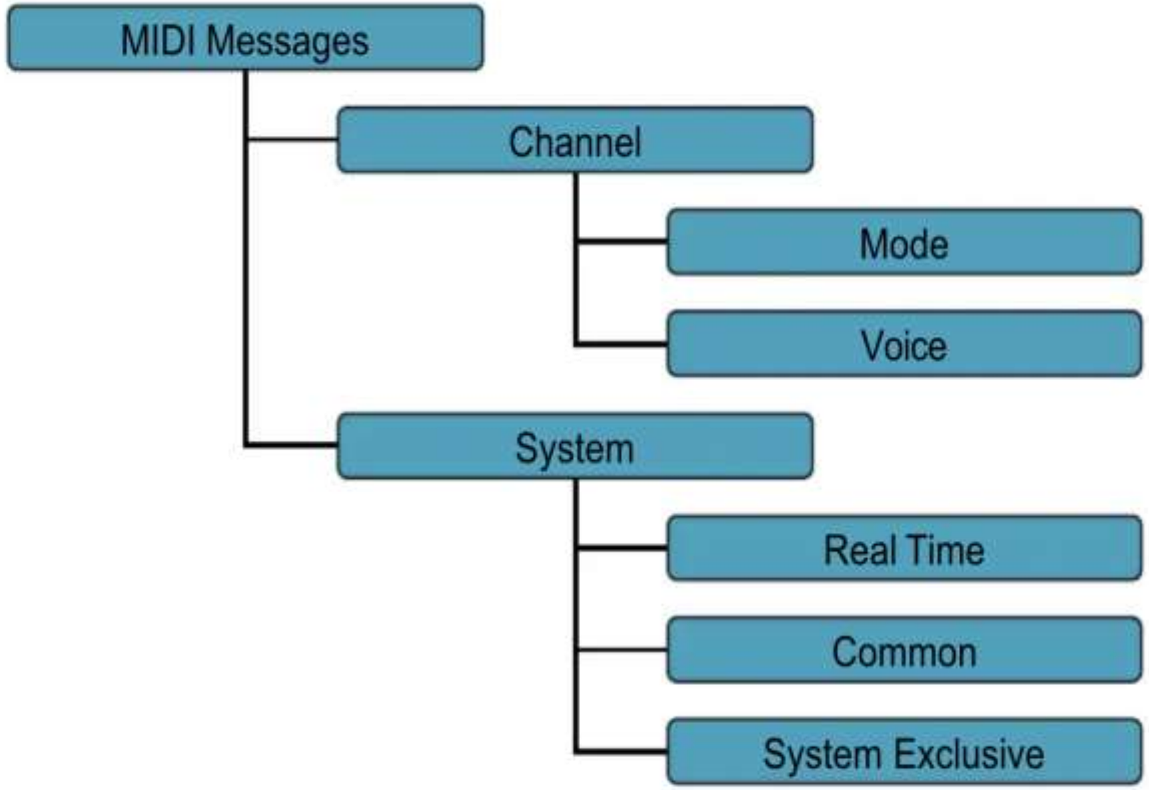
- To tune a MIDI device to one or more channels, the device must be set to one of the 4 reception modes
  - Mode 1: Omni ON/ Poly
  - Mode 2: Omni ON/ Mono
  - Mode 3: Omni OFF/Poly
  - Mode 4: Omni OFF/ Mono
  - If Omni is turned ON, the MIDI device monitors all the MIDI channels and responds to all channels.
  - If Omni is turned OFF, the MIDI device responds only to the channel messages sent on channel the device is set to receive.
  - If Poly is set, the device can play several notes at a time.
  - If Mono is set, the device plays one note at a time.

# MIDI DEVICES

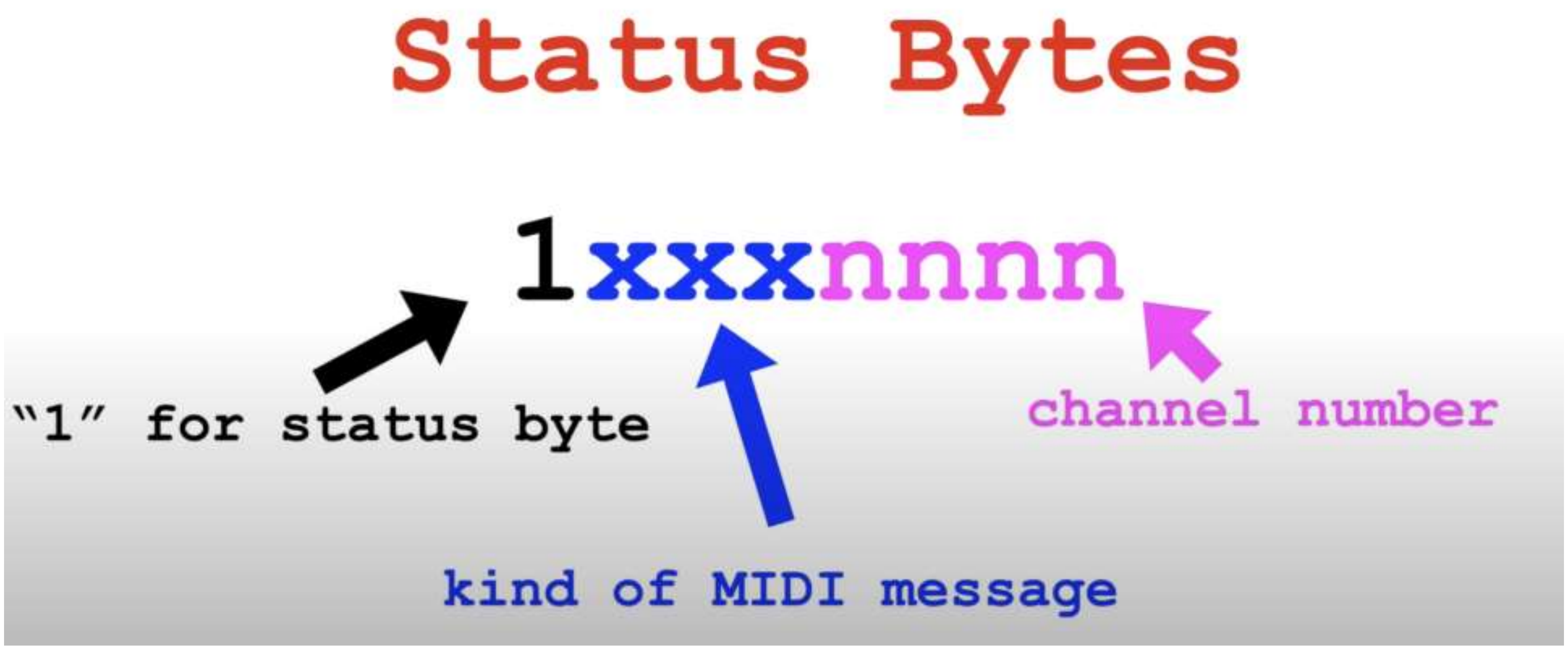
- The heart of any MIDI system is the MIDI Synthesizer device.
- A typical synthesizer looks like a simple keyboard with a pannel full of buttons.
- Most synthesizers have the components as:
  - Sound Generators
  - Microprocessor
  - Keyboard
  - Control Pannel
  - Auxillary Controllers
  - Memory

# MIDI MESSAGE

- MIDI message transmit the info. Between MIDI devices.(8-bits)
- Every MIDI message would consist of the *status byte*( First Byte-1)
- i.e. 1XXXXXXX
- It describes the kind of message and the following bytes would contain the *data bytes.* ( First Byte-0)
- i.e. 0XXXXXXX



# FORMAT


Status Bytes
1xxxnnnn
"1" for status byte
channel number
kind of MIDI message

# MIDI MESSAGES

- 1. Channel Messages
- Go to only specific device
- Two types :
  - 1. Channel Voice Message: It sends the actual performance data between MIDI devices that describes keyboard action, controller action and control pannel changes.
  - Defines pitch, amplitude, timbre, duration, note ON, note OFF, pressure
  - 2. Channel Mode Message: Determines the way the receiving MIDI device responds to the channel voice message
  - Eg. Omni OFF, Omni ON

- 2. System Messages
- Go to all devices in MIDI system.
- i.e. 1111XXXX
- Three Types:
  - 1. System Real Time Message: Very Short 1 byte message that is used to synchronize the timing of devices (Clock)
  - 2. System Commom Message: They are commands that prepare the synthesizers to play a song. Song Select, Tune Select
  - 3. System Exclusive Message: They are the message that depend upon different types of instuements. Yamaha, Roland etc

# MIDI CHANNEL MESSAGE

**Note On** messages send information about **which key is pressed, at what velocity**, and **on which channel.**

| 1001**nnnn** | 0**kkkkkkk** | 0**vvvvvvv** |
|---|---|---|
| *status byte* | *data byte* | *data byte* |
| **nnnn** is the channel | **kkkkkkk** is the key | **vvvvvvv** is the velocity |

# MIDI SYSTEM MESSAGE

System Exclusive Messages

11110000 & 11110111

System Common Messages

11110xxx (e.g. 11110001)

System Real-Time Messages

11111xxx (e.g. 11111000)

# MIDI MESSAGES_VIDEO

# SPEECH ANALYSIS

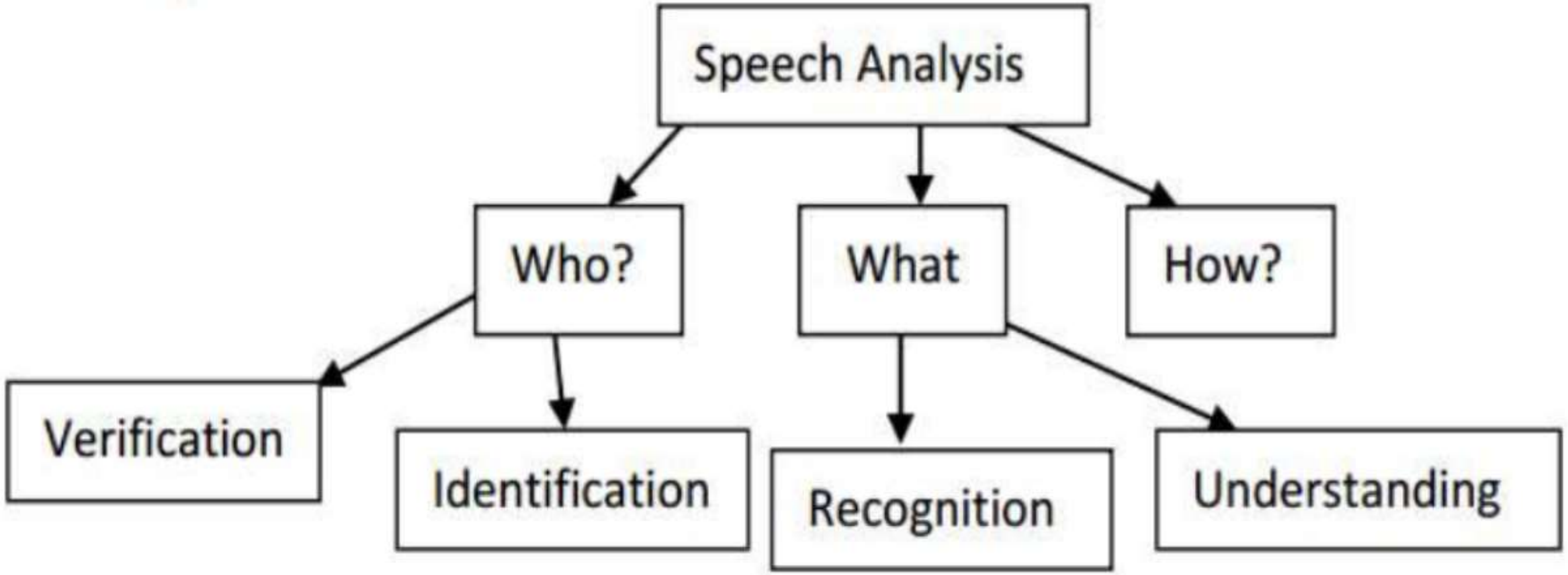
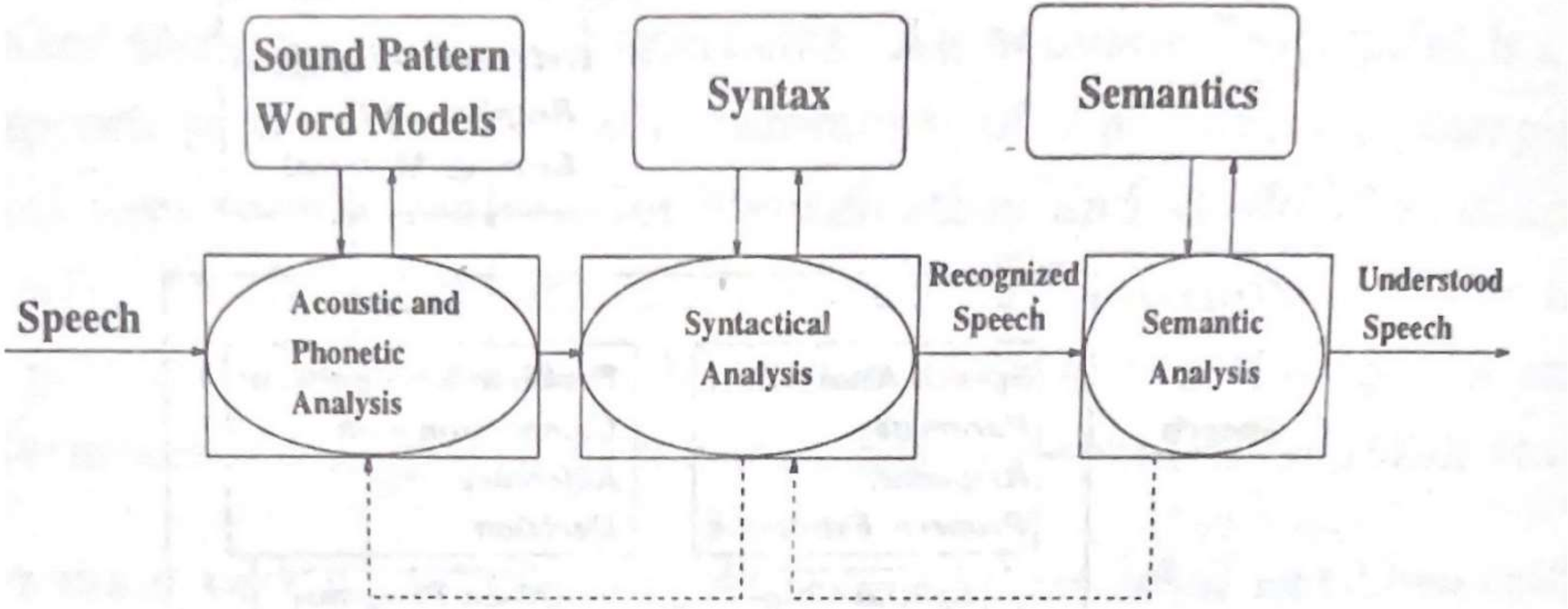



**Figure:** Research Areas of speech analysis

# SPEECH ANALYSIS

- Human Speech has certain characteristics determined by the speaker.
- This analysis can serve to analyze WHO is speaking i.e. recognize the speaker for his/her identificaiton as acoustic fingerprint.
- Another task of analysis is to analyze WHAT has been said i.e to recognize and understand the speech signal itself.
- Another area of analysis is to analyze how it has been said by the respective person. Focuses on the style of delivery and finding out whether the person said it anger or calm mood.
- For these analysis, environment noise, room acoustics and speakers physical and psychological conditions plays a vital role.

# COMPENENTS OF SPEECH RECOGNITION



# SPEECH RECOGNITION

- In the first step, acoustical and phonetical analysis is performed

  i.e. pronounciation of words

- In the second step, the speech unit is passed through a syntactical analysis(Syntax of the sentence)

  i.e. correctness of the sentence

- In the third step, the semantics of the preciously recognized speech if analyzed.

  i.e. Meaning of the speech or sentence

# PROBLEMS FACED DURING SPEECH RECOGNITION

- 1. Overlap due to echo of the aucostics from the wall of primary sound wave
- 2. Co-articulation: Very often the neighbouring words flow into one another. This would create a recognition issue or confusion.
- 3. Time taken by the individuals. (Speaker dependent)

# END OF CHAPTER 2