Chapter 2

Audio System

Basic concepts

- Sound is produced by vibration of medium.Molecules move back and forth.
- It is continuous signal with some frequency, period and amplitude
- Period is time to move air molecule back and forth.
- Periodic Sound signal repeats itself after some time interval called period.

Basic concepts

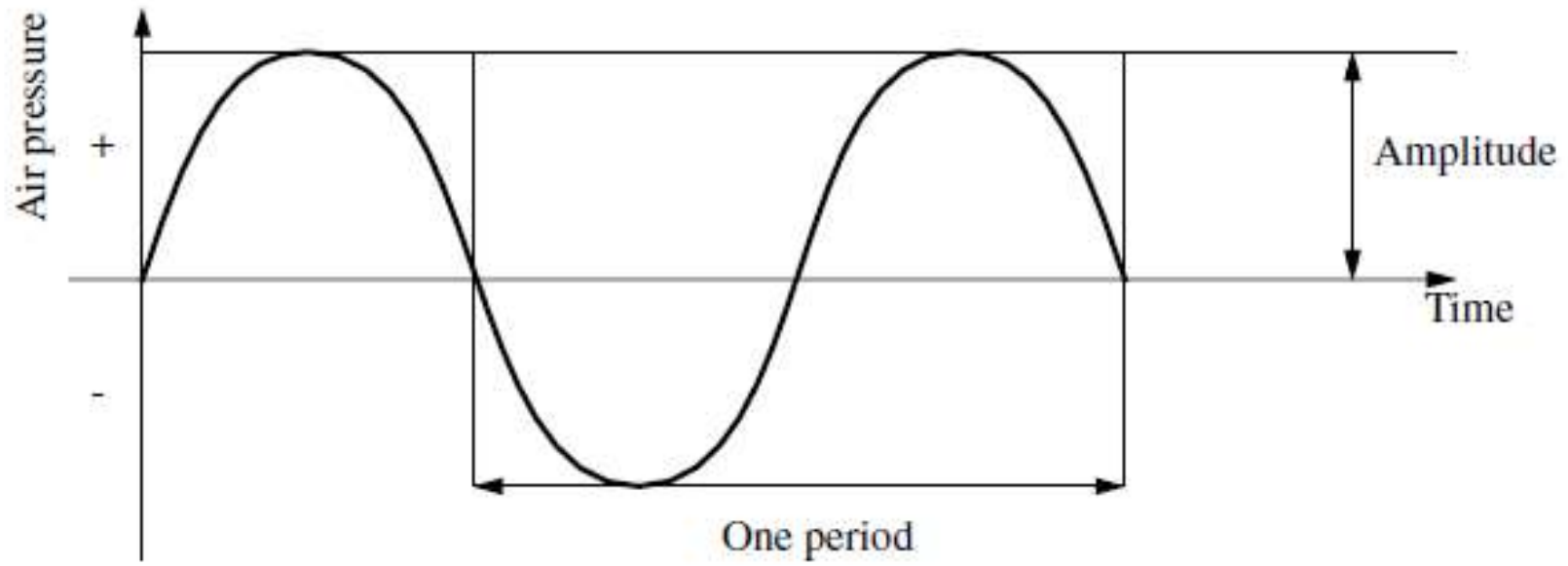
- Sound is pressure waves.
  i.e. travels on pressure difference
- Is longitudinal wave.
  ie. Oscillation is in the direction of propagation
- Females has higher frequency
- Female voice can travel larger distance

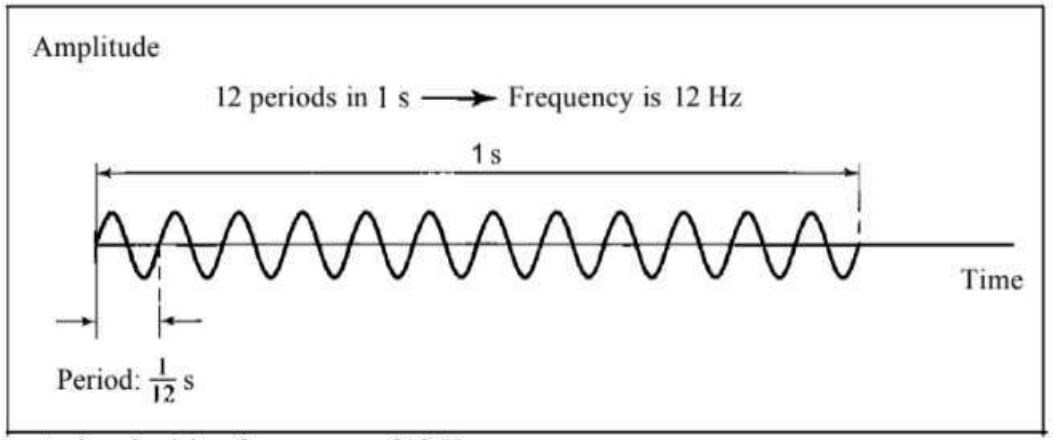
We hear when vibration is detected by ear

Basic concepts

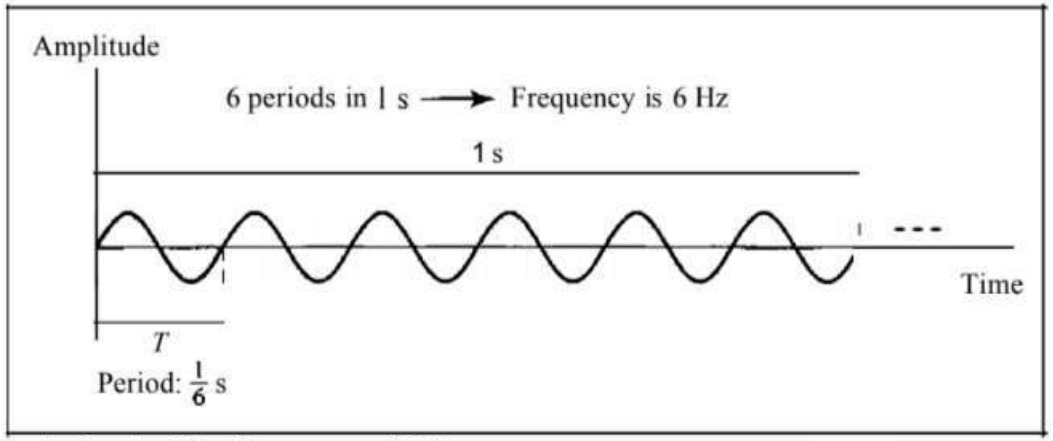
- Ear converts vibration into electrical impulses
- Impulses are transmitted to our brain

Amplitude and period



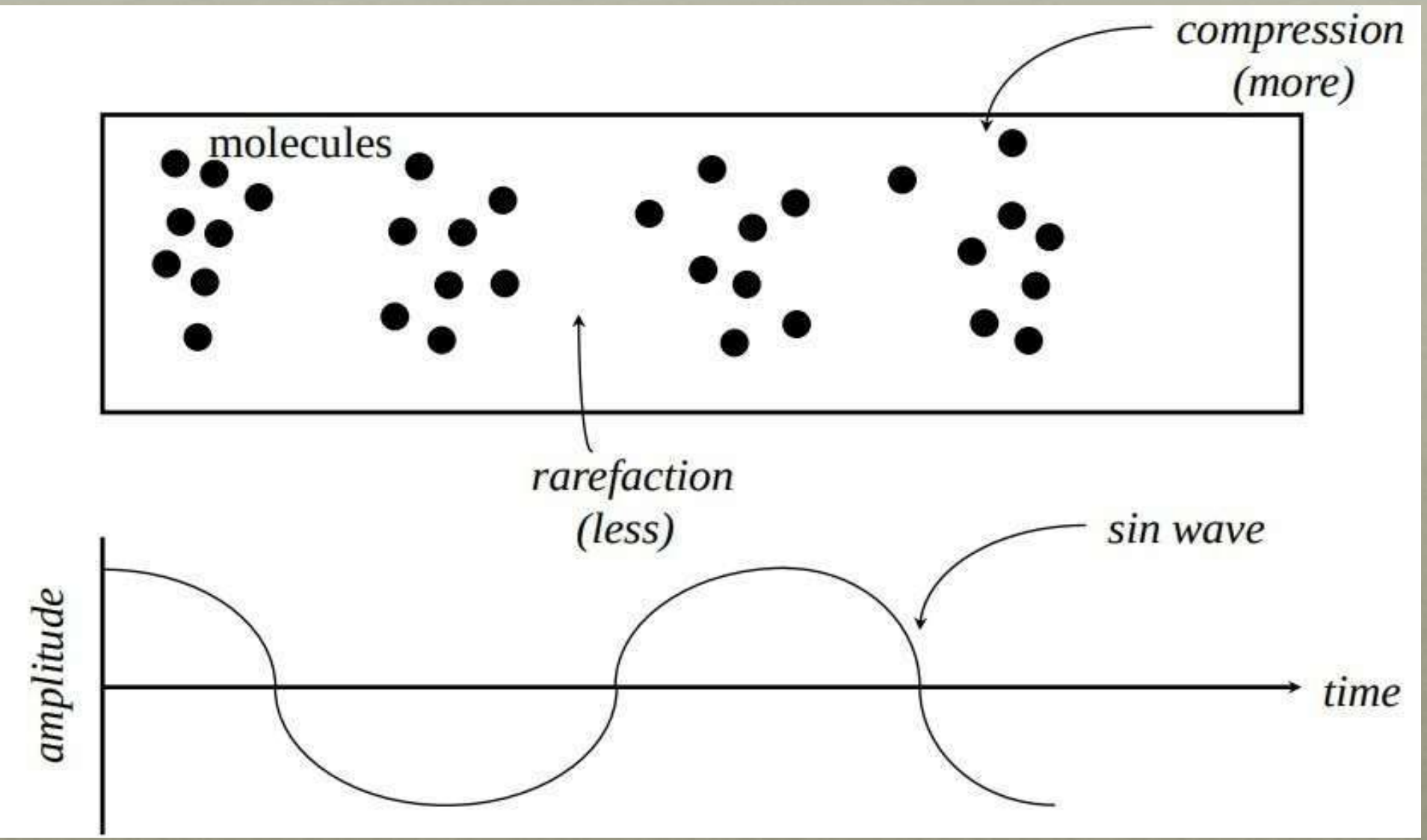


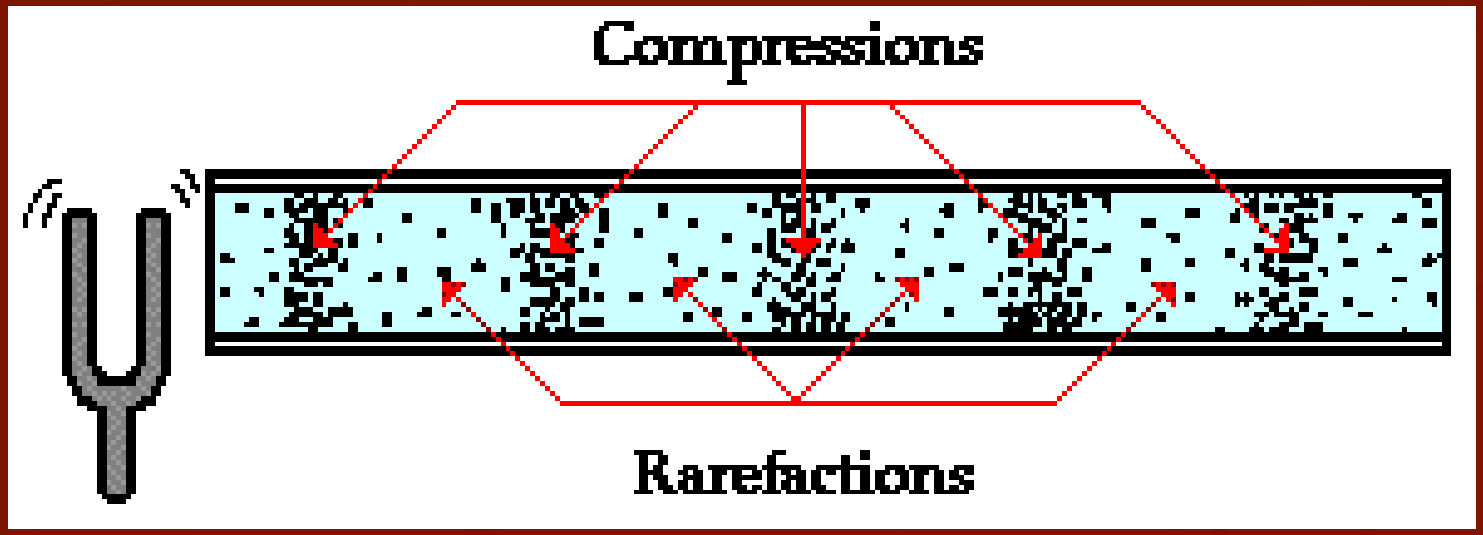
a. A signal with a frequency of 12 Hz



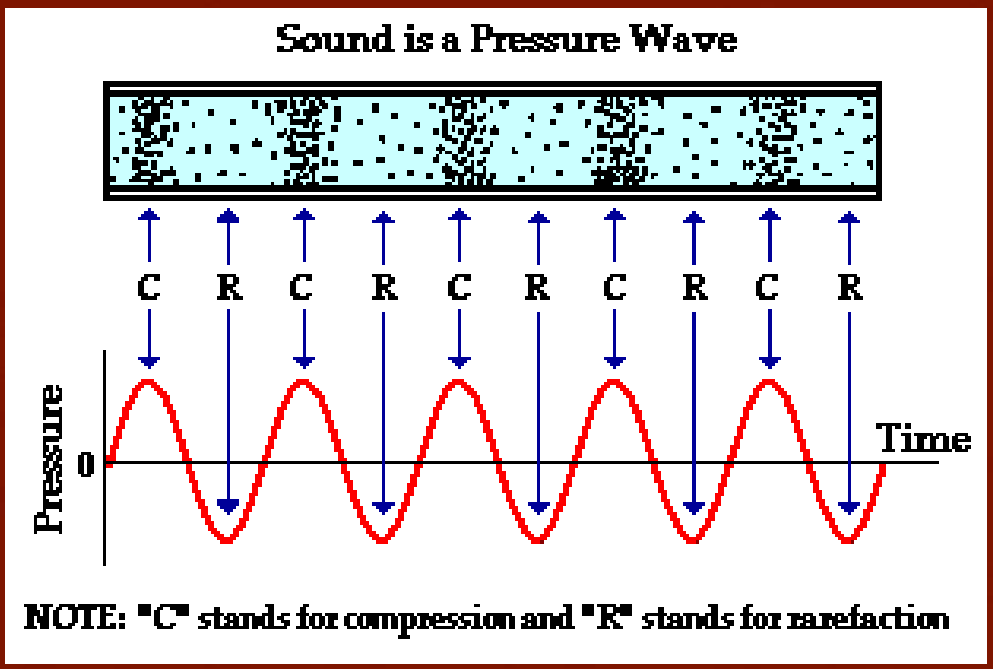
b. A signal with a frequency of 6 Hz

Basic concepts




Compressions
Rarefactions

Sound is a Pressure Wave



NOTE: "C" stands for compression and "R" stands for rarefaction

Basic concepts

- Source of periodic sound are musical instruments
- Non periodic sound sources include cough, sneeze.
- Some regions of molecule get compressed and some get far apart called rarefaction.
- Compression region has higher pressure
- Distance between two consecutive compressed region is wavelength

Frequency

- Number of periods in a second
- Measured in hz
- Audible range 20hz to 20khz
- Sound within human hearing range is audio
- Music signals in 20hz to 20khz

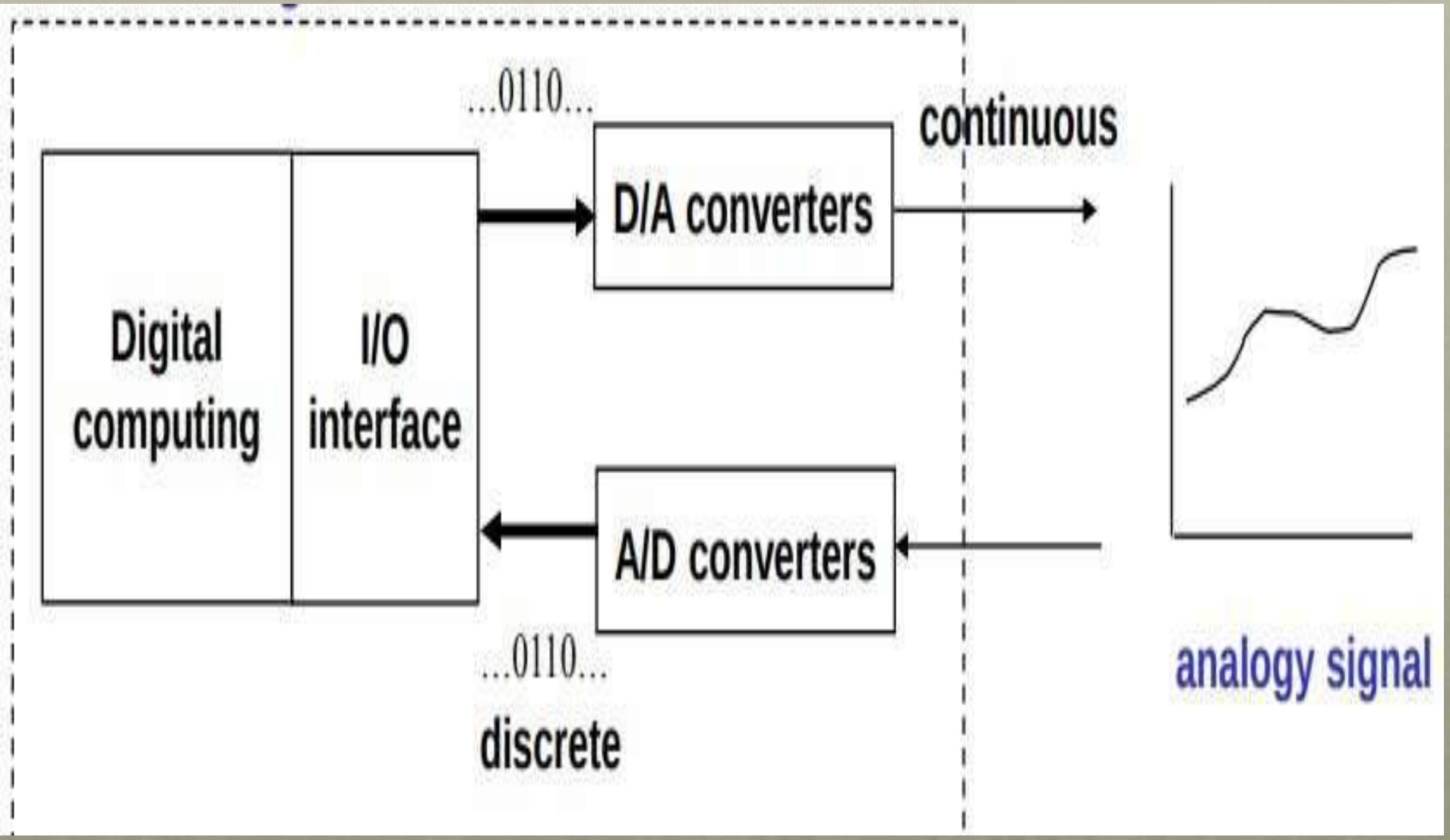
Amplitude

- The maximum displacement of the molecule from the mean position
- It measures the loudness
- It is measured in decibel(db).
- Human can perceive amplitude in range (0 - 120db)

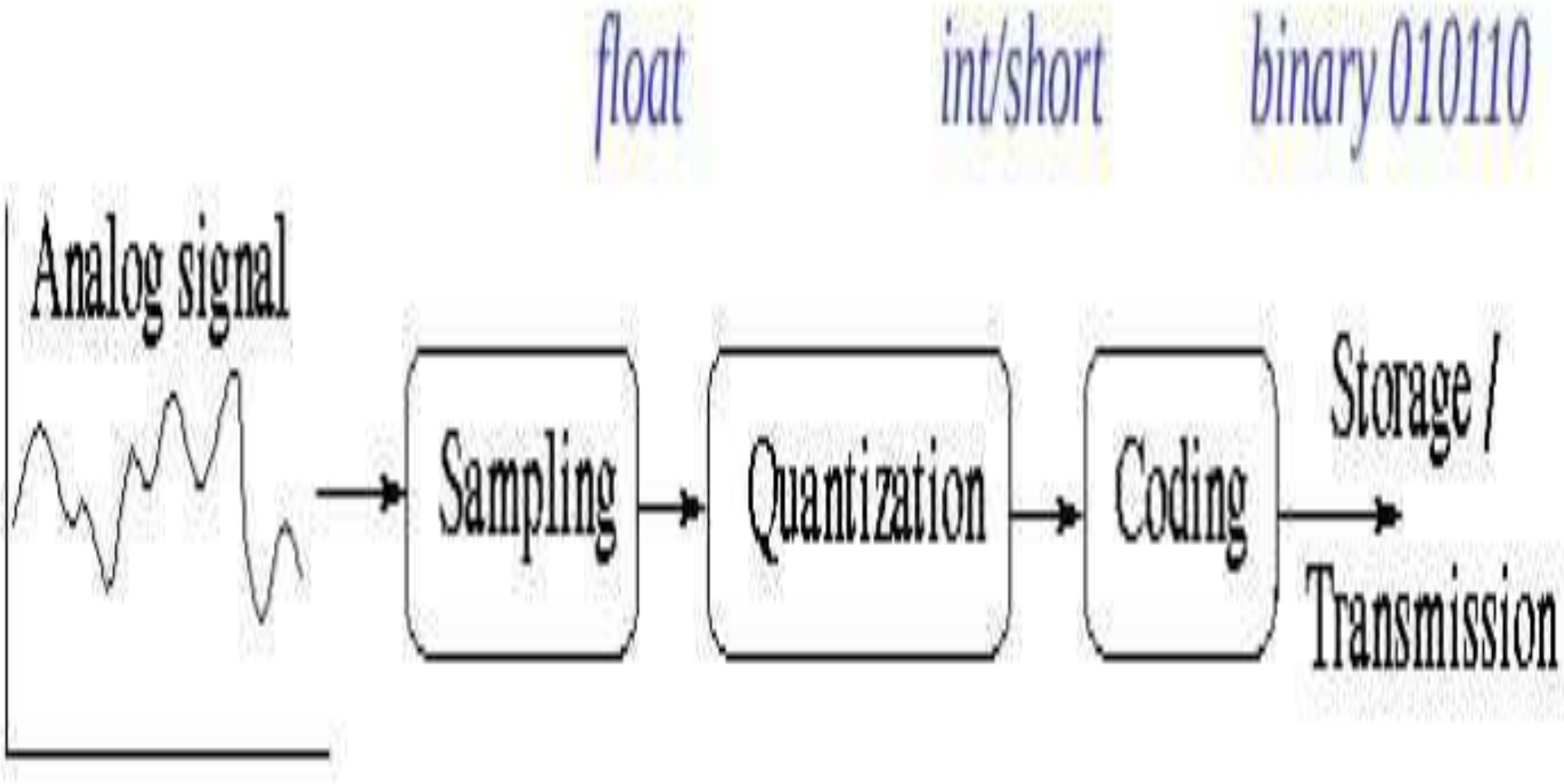
Computer representation of sound

- Analog sound signal from microphone is converted to digital signal to store in computer
- This is done by Analog-to-Digital converter.
- Reverse is done by Digital-to-Analog converter.

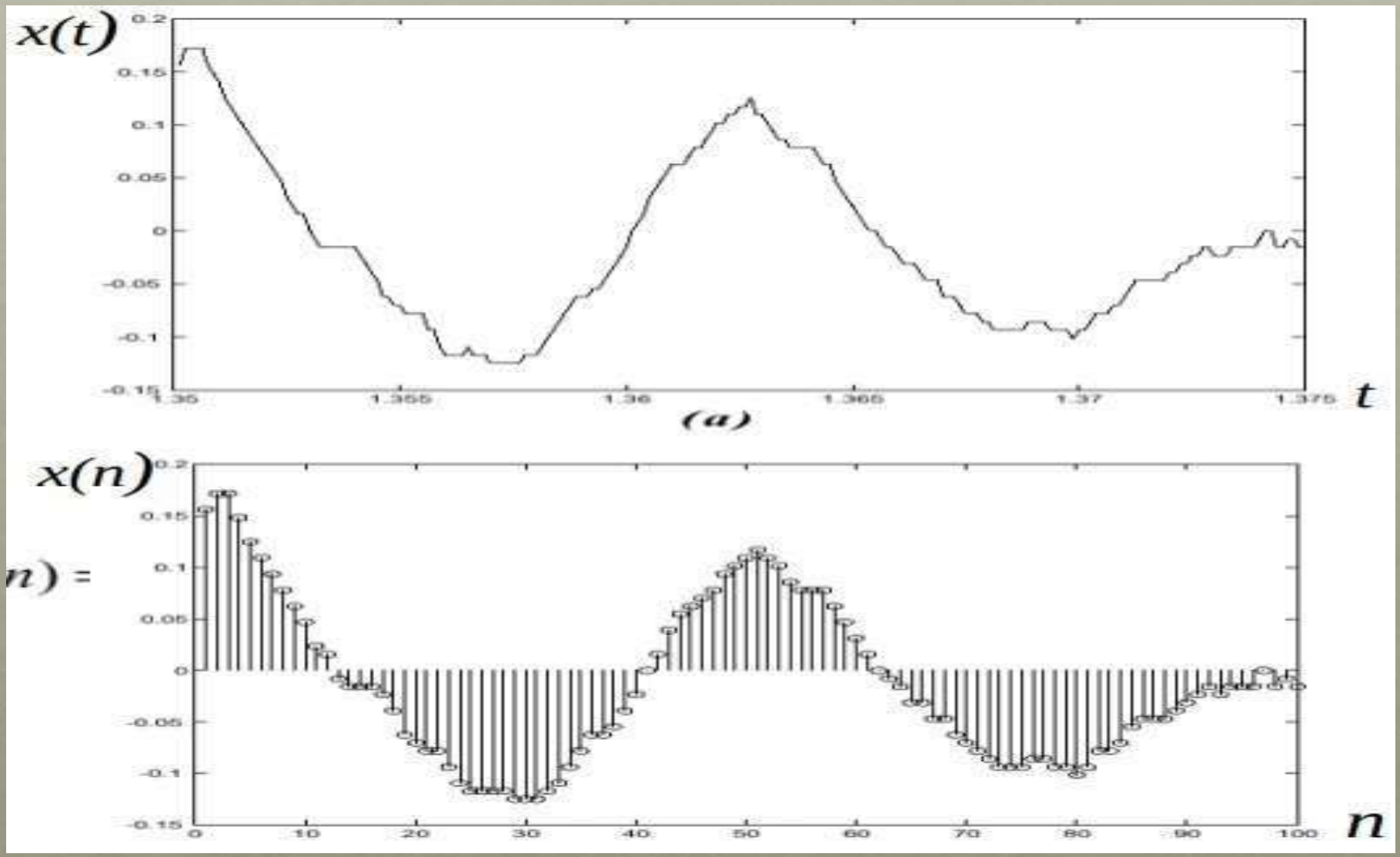
Analog to Digital conversion



Analog to Digital conversion

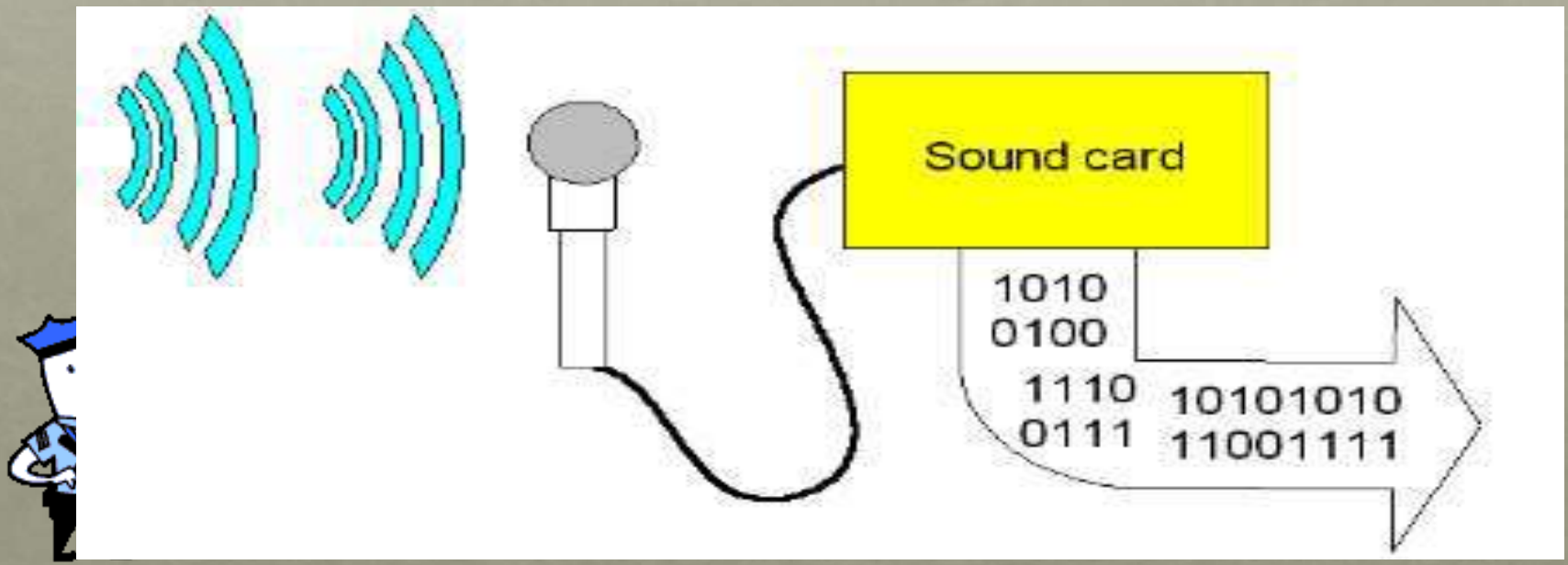


Analog to Digital conversion



Digitization

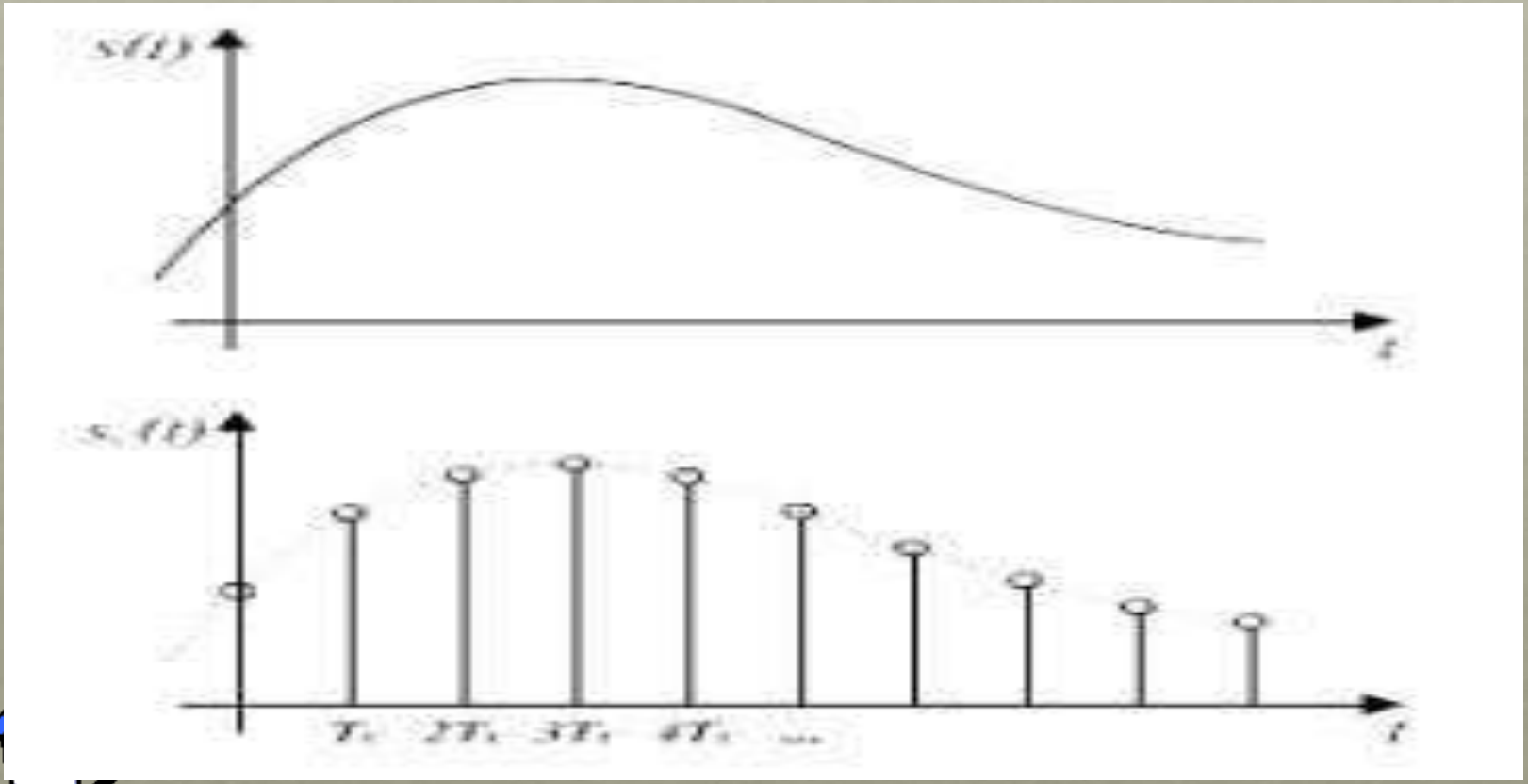
- Process of converting analog signal to digital signal
- It is done by sampling of the analog signal followed by quantization.



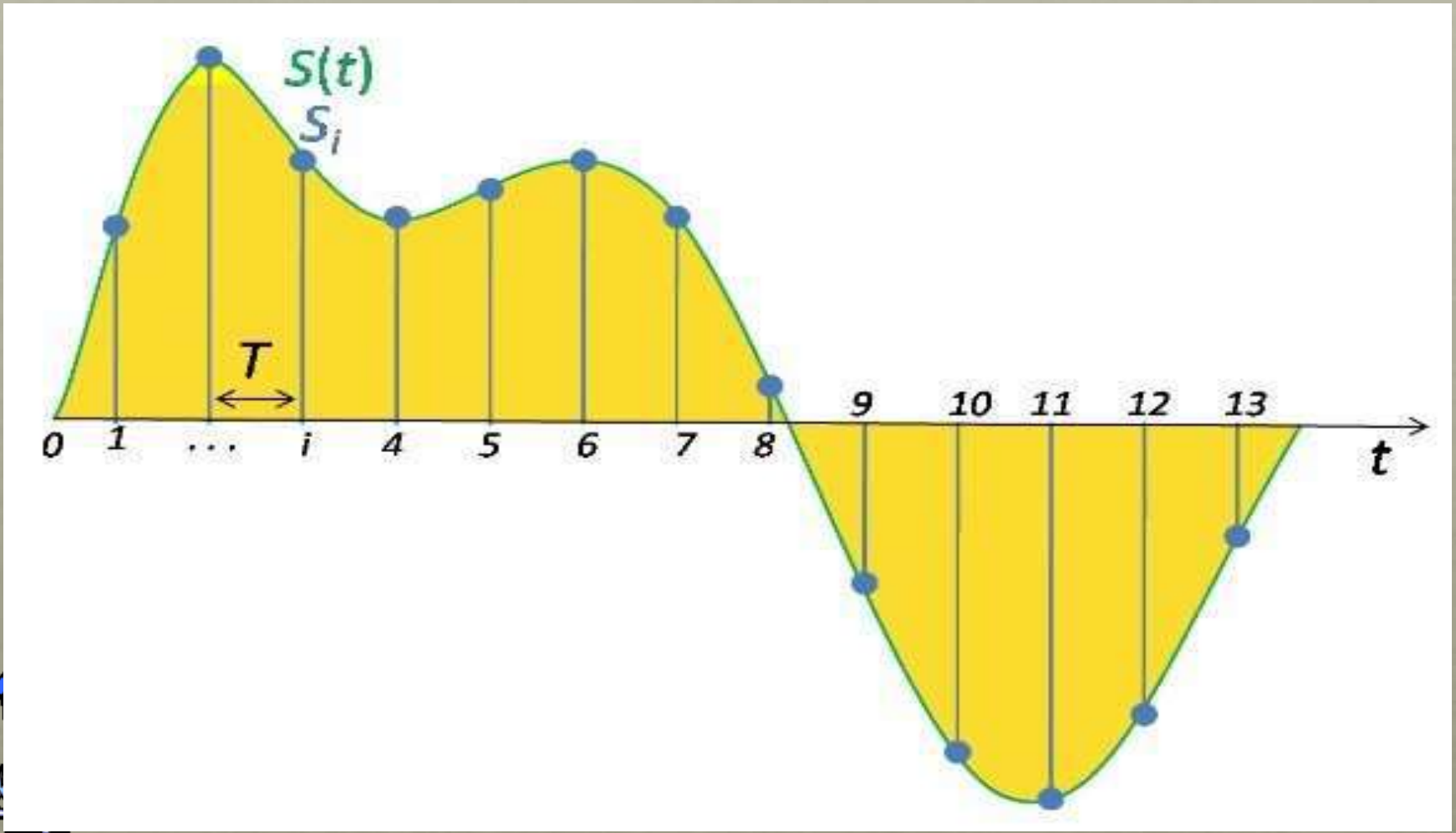
Sampling Rate

- Fixed samples are taken at fixed rate from analog signal
- Measured in hz
- Sampling process records amplitude at fixed interval
- How many samples are taken within a cycle
- High sampling rate, high quality of sampled signal.

Sampling

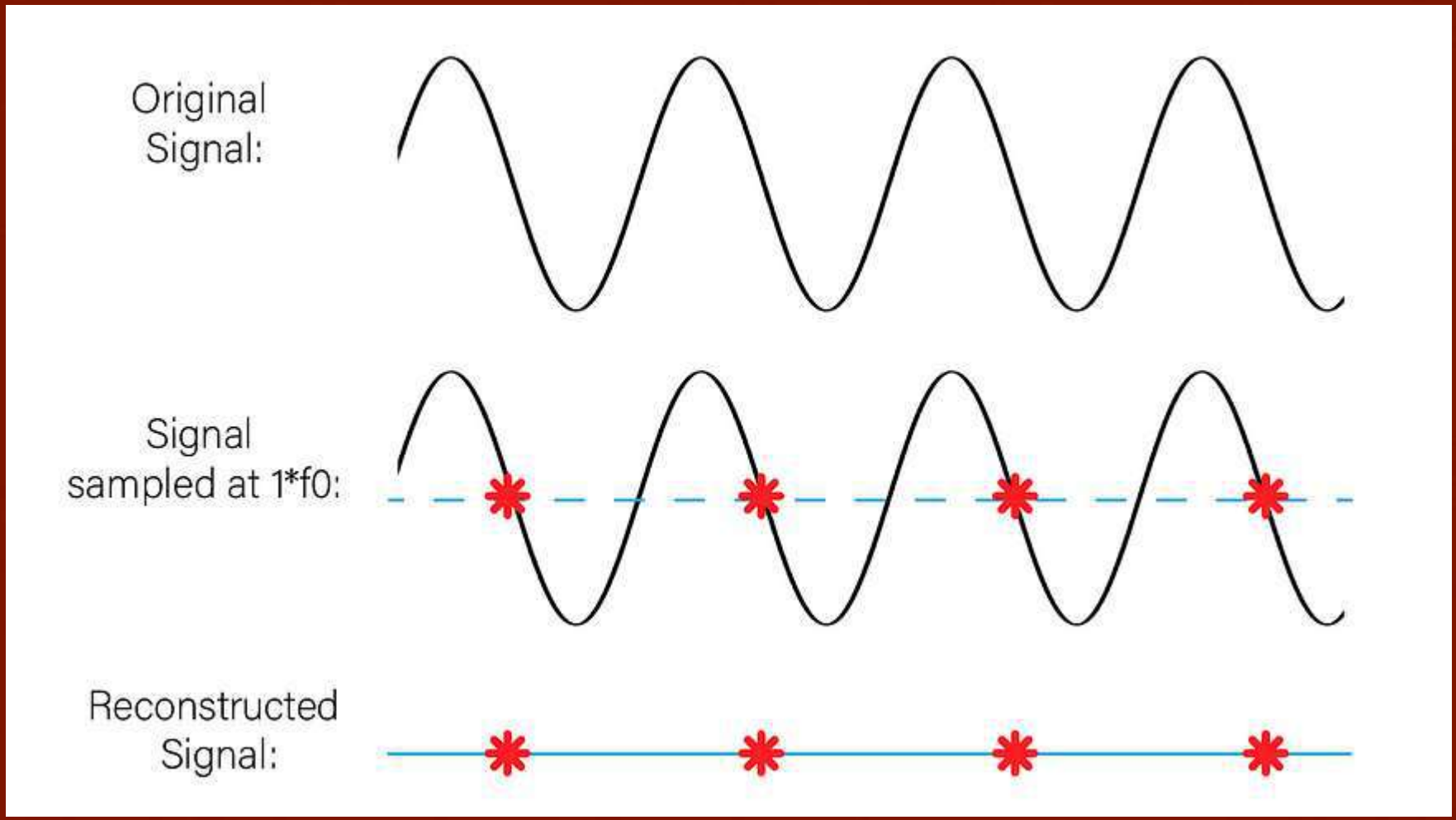


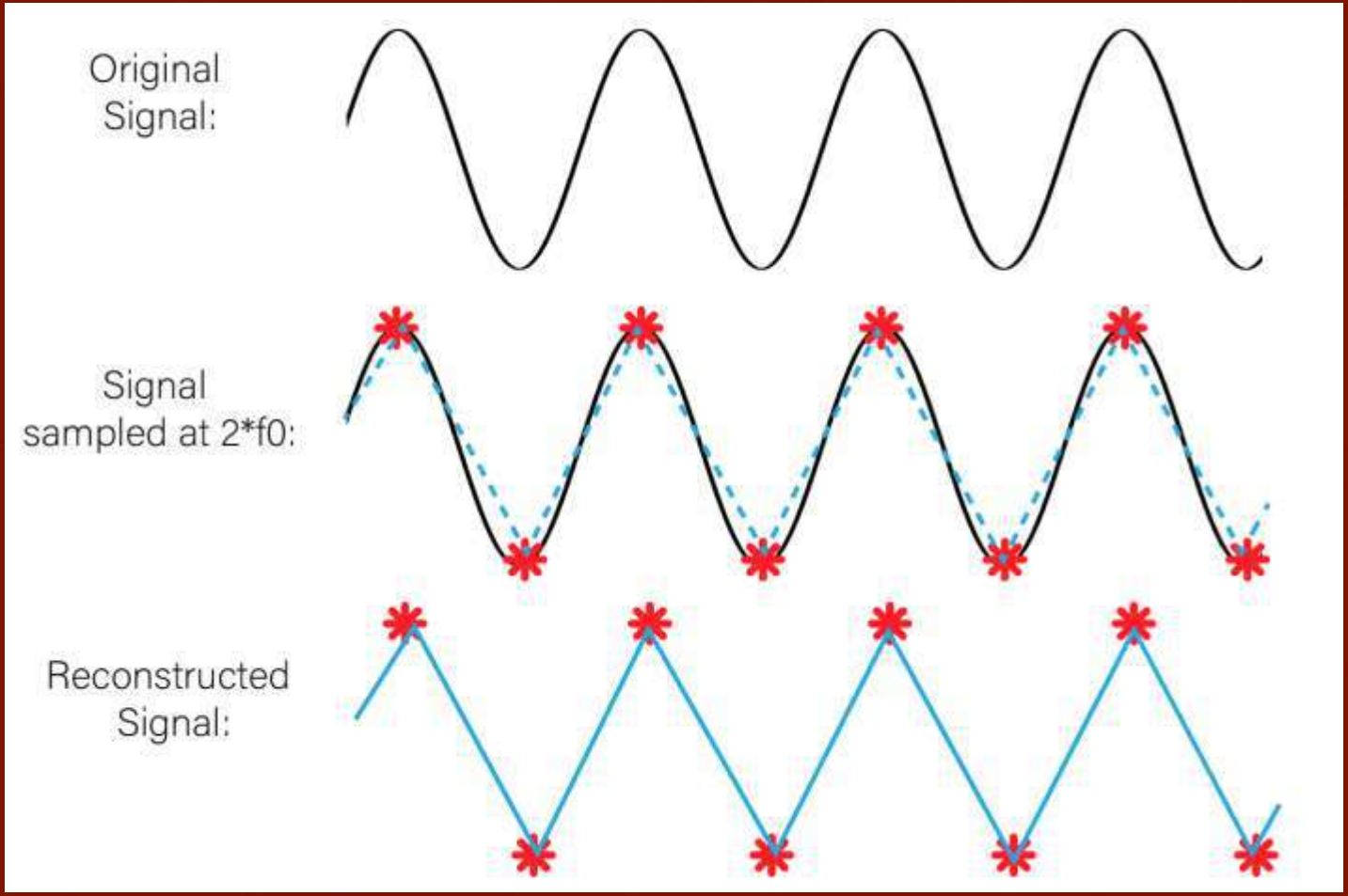
Sampling

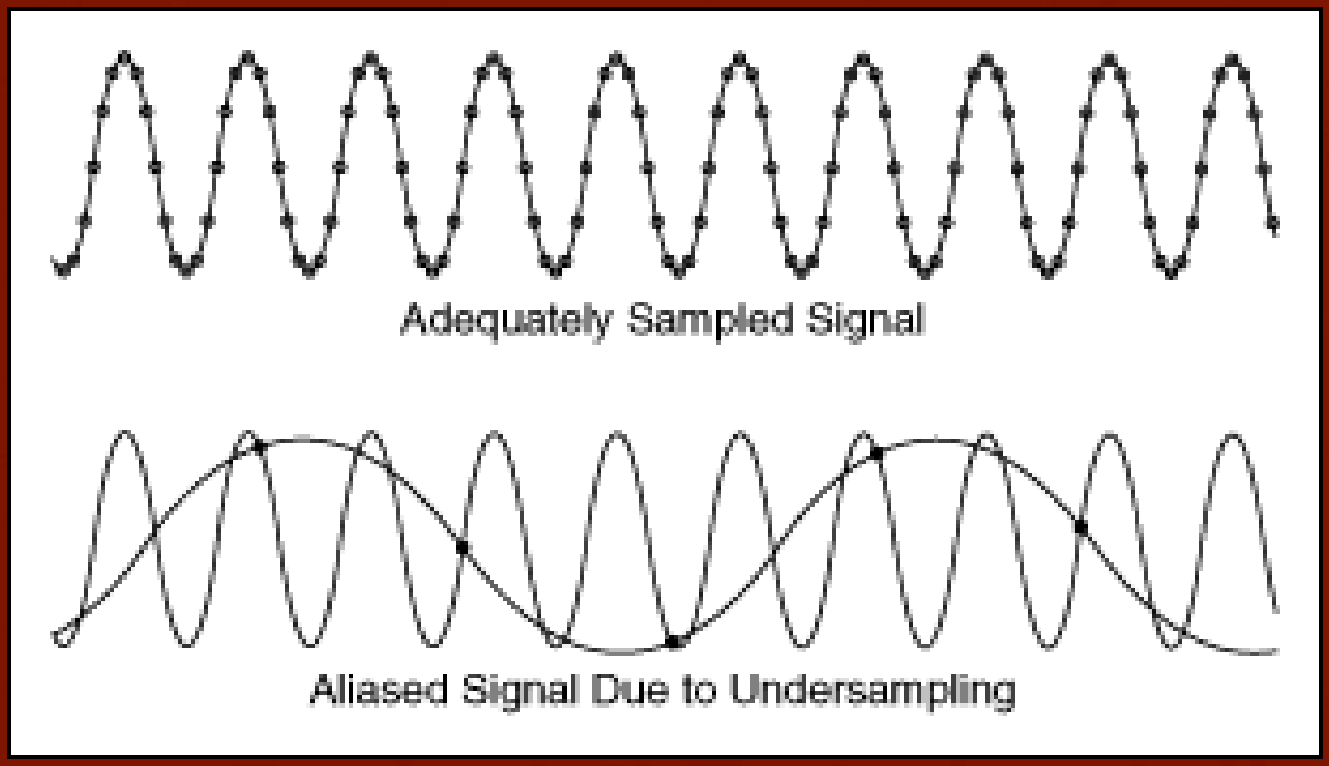


Nyquist sampling theorem

- The Nyquist theorem is also known as the sampling theorem.
- It is the principle to accurately reproduce a pure sine wave measurement, or sample, rate, which must be at least twice its frequency.
- The Nyquist theorem underpins all analog-to-digital conversion and is used in digital audio and video to reduce aliasing.
- The Nyquist theorem is also known as the Nyquist-Shannon theorem or the Whittaker-Nyquist-Shannon sampling theorem


Original Signal:
Signal sampled at 1*f0:
Reconstructed Signal:


Original
Signal:
Signal
sampled at 2*f0:
Reconstructed
Signal:

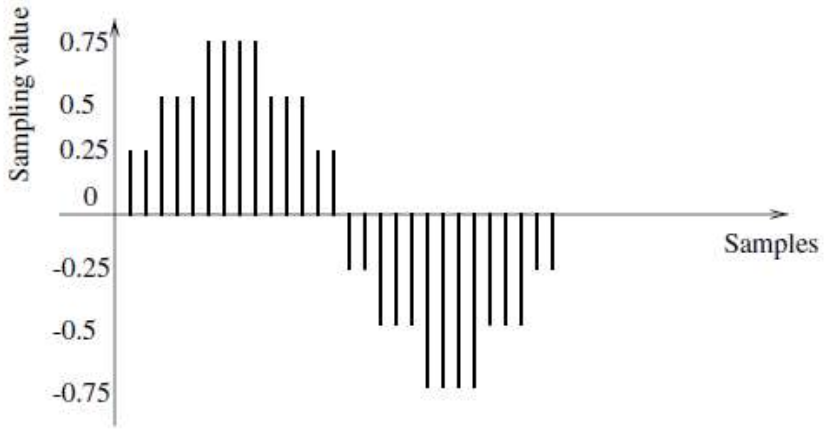

Adequately Sampled Signal
Aliased Signal Due to Undersampling

Quantization

- Sampled amplitude values are mapped into different levels or binary codes
- If n is bit size of quantized values, then number of levles, $L = 2^n$
- If xmx and xmn are maximum and minimum sampled values, then (xmx-xmn)/L is step size
- Each levels are represented by distinct binary numbers

Quantized value of samlpled value is the binary value of nearest level.

Quantization



3-bit quantization.

The values transformed by a 3-bit quantization process can accept eight different characteristics: 0.75, 0.5, 0.25, 0, -0.25, -0.5, -0.75, and -1, so that we obtain an “angular-shape” wave.

This means that the lower the quantization (in bits), the more the resulting sound quality deteriorates.

Bit Rate

We want to digitize the human voice. What is the bit rate, assuming 8 bits per sample?

Solution

- *The human voice normally contains frequencies from 0 to 4000 Hz. So the sampling rate and bit rate are calculated as follows:*

$$\text{Sampling rate} = 4000 \times 2 = 8000 \text{ samples/s}$$
$$\text{Bit rate} = 8000 \times 8 = 64{,}000 \text{ bps} = 64 \text{ kbps}$$

Bit rate

- Number of bits received or transmitted per sec
- Number of bits used to represent sound signal
- Higher the bit rate, higher the quality and size of recording

Bit Rate

- An analog signal carries 4 bits/signal elements. If 1000 signal elements(bauds) are sent per second, find the bit rate
- n=4 bits/element
- r(baud rate)=1000 baud (elements/sec)
- R=nr=4*1000=4000bits/sec
-

MIDI

- Musical instruments digital interface
- Protocol adopted by electronic music industry that enables computers, synthesizers, keyboards and other musical devices to communicate with each other
- MIDI interface is incorporated into most sound cards
- MIDI generates small data also known as events for the production of sounds.

MIDI event may include pitch, volume of a single note and the instrument that play sound

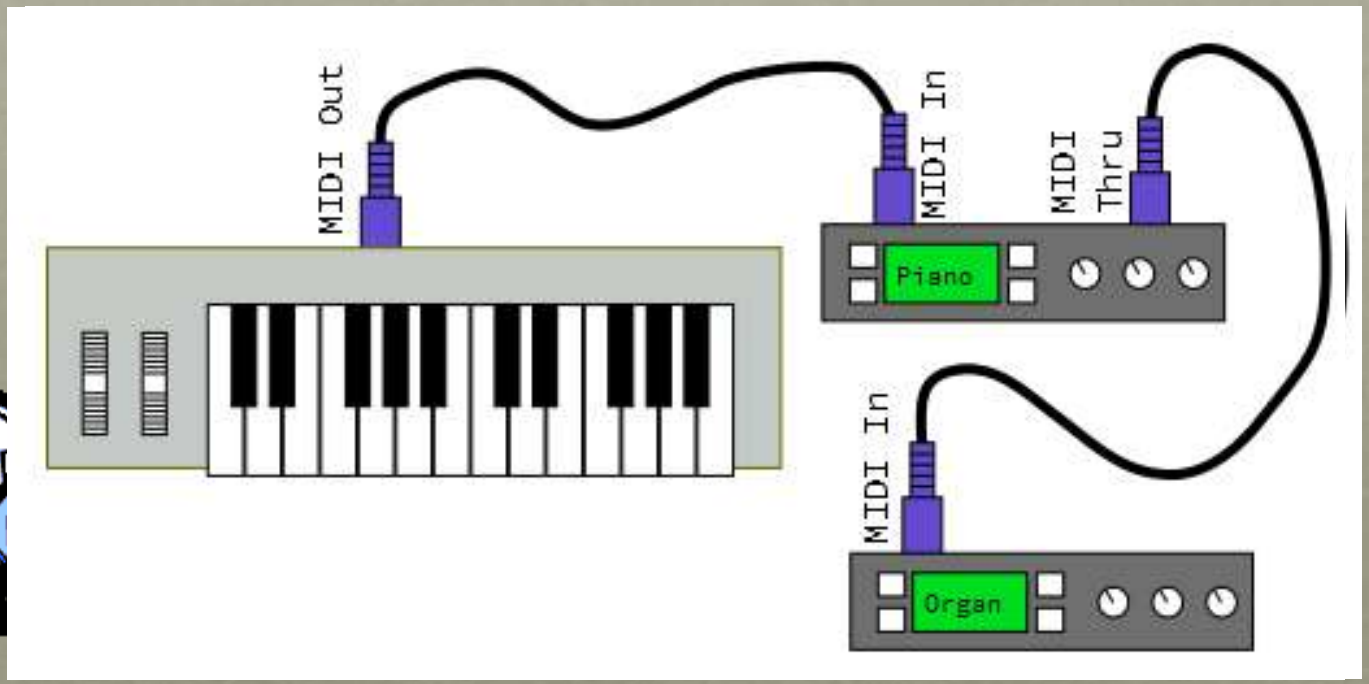
MIDI

- MIDI events are very small thus requires less bandwidth
- Sound effects can be produced(like bass, echo)

MIDI

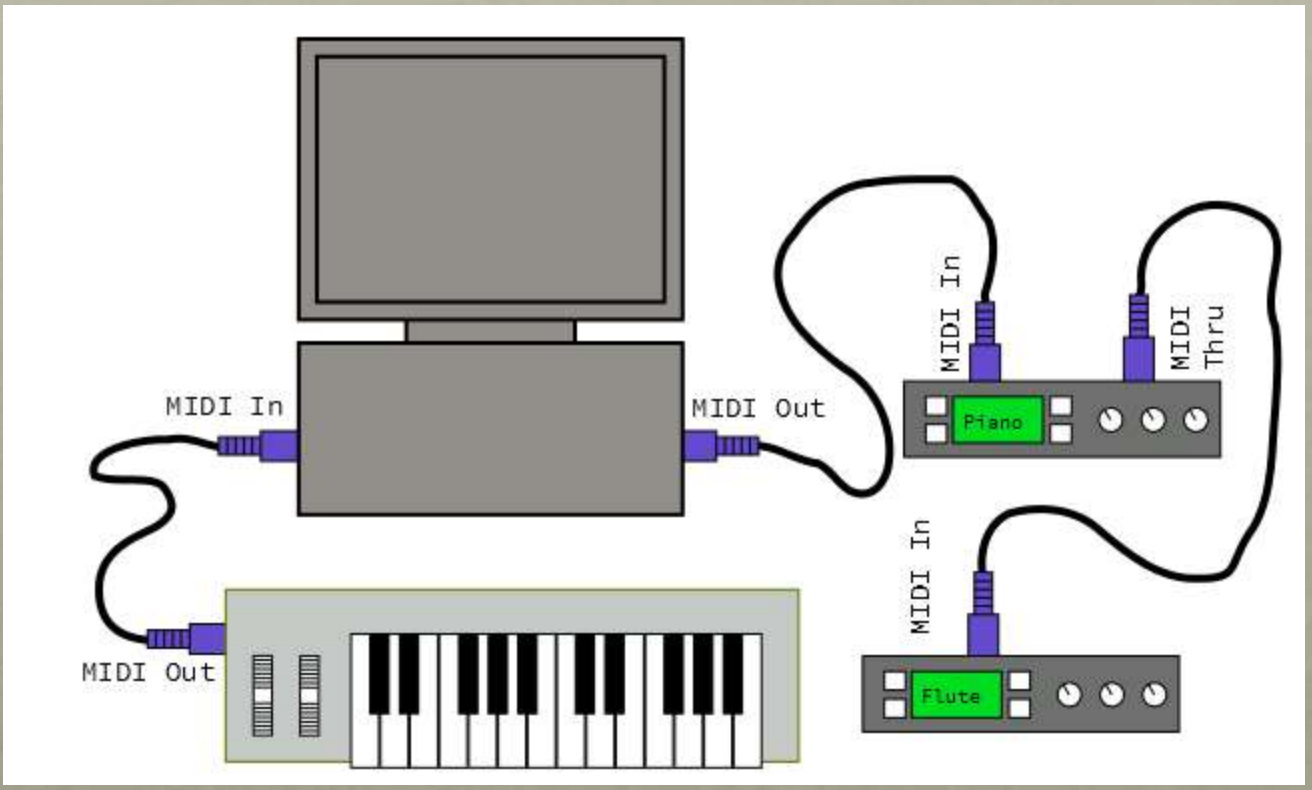
MIDI

- One controller and two tone generator (sound generator)
- This sound module generates according to MIDI messages from controller



Organ

MIDI



- Sequencer software in computer records MIDI messages from controller and is played on sound modules
- It also allows to edit the messages

Components of MIDI system

- Synthesizer
- Sequencer
- Computer system
- MIDI control Input devices
- MIDI interfaces
- MIDI control output devices

Synthesizer

- Is a stand alone sound generator
- Uses sample based synthesis to generate sound (like lightening sound in middle of music).
- It has keyboard to generate MIDI events
- These MIDI events are fed converted to audio signal
- Which is fed to speaker to generate audio.
- It has microprocessor, keyboard, control panels, memory, many ins and outs

Synthesizer

Kurzweil K2000 Synthesizer (Image courtesy wikimedia commons)

Yamaha synthesizer

Synthesizer

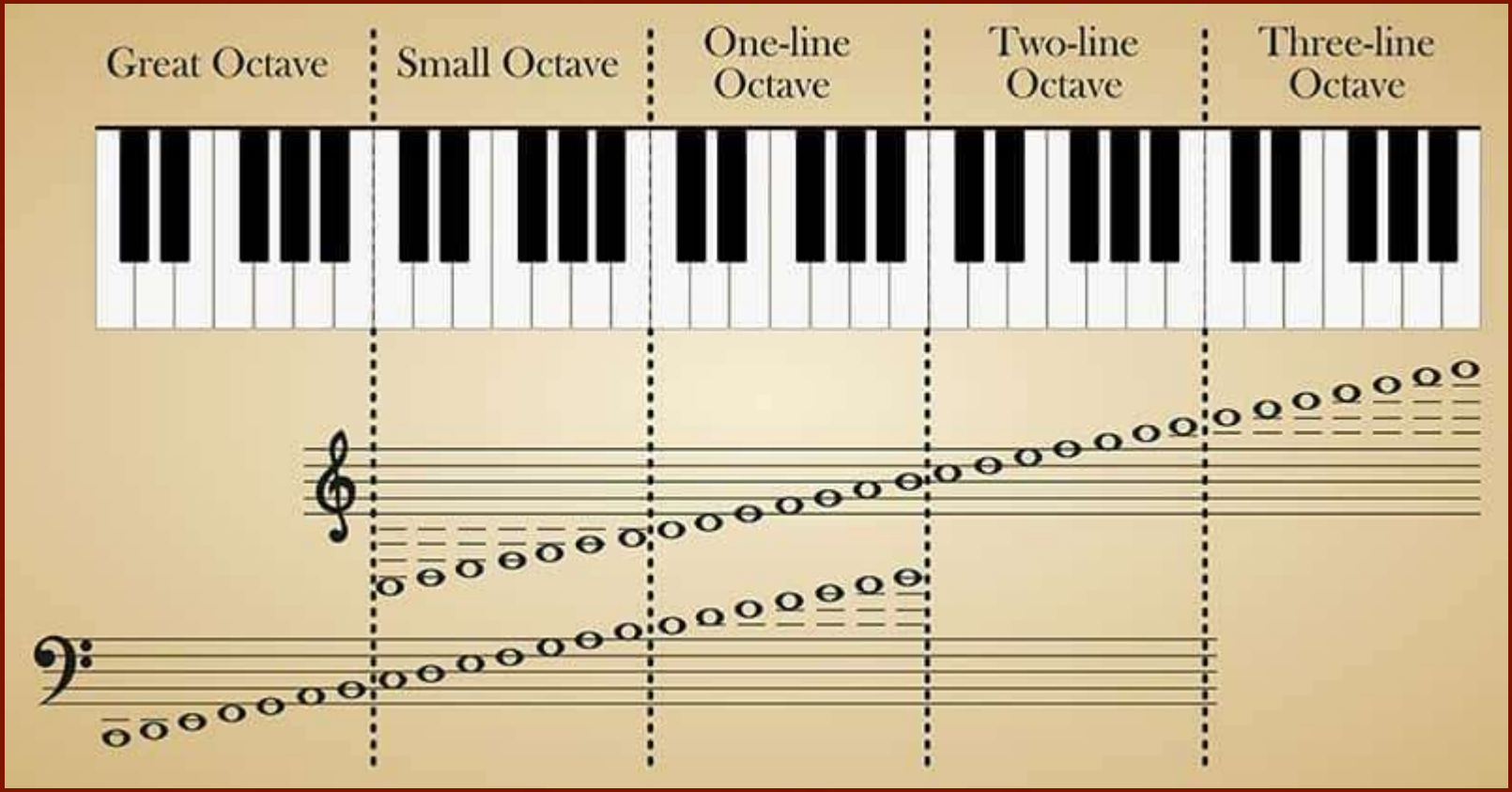
Microprocessor

- Communicates with keyboard to know the inputs like notes(music), commands(effect,brightness of screen)
- The notes and commands are send to sound generator
- It means it sends and receives MIDI messages

Synthesizer

Keyboard

- Commands like playing notes, loudness of note, add vibrato
- Loudness of tone depends upon speed and acceleration of keys pressed
- Keyboard should have at least five octaves with 61 keys


Great Octave
Small Octave
One-line Octave
Two-line Octave
Three-line Octave

Synthesizer

Control Panel

- Controls notes and its duration
- It includes slider, a button, and a menu
- Slider sets volume
- Button turns synthesizer on or off
- Menu calls up different patches

Synthesizer

Auxiliary Controllers

- Gives more control over notes
- Pitch bend controllers can bend pitch up and down
- Modulation controllers can increase or decrease effects such as vibrato

Synthesizer

Memory

Stores patches for sound generator

- Multiple memory cartridges are used for different sound synthesizer.

Controller

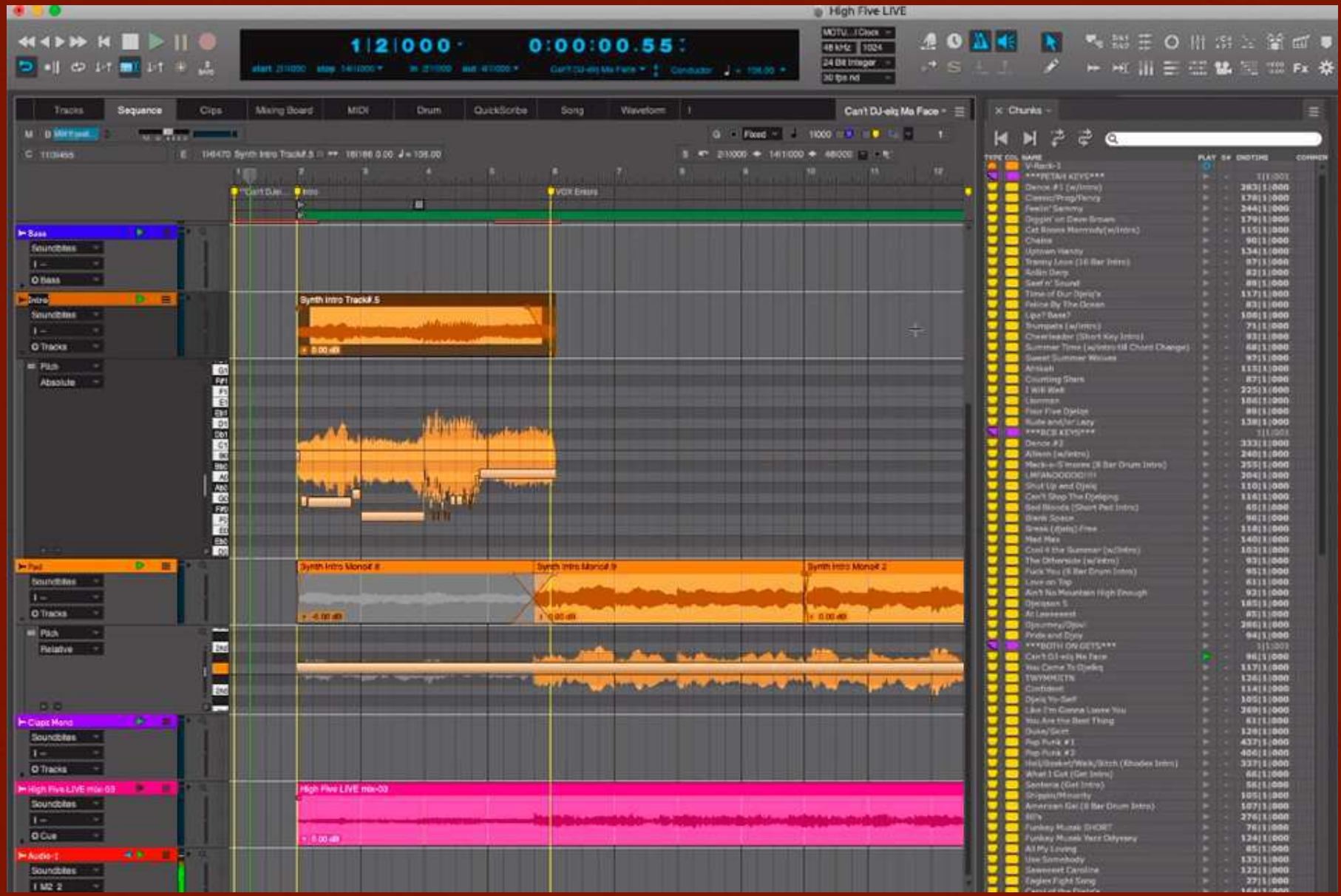
- Generate midi messages
- But can’t generate sound
- Midi messages are send to sound modules that generate the sound

Sequencer

- Stand alone hardware or software running on a computer
- Records the MIDI message and play back.
- Has MIDI ins, outs, wifi, blue tooth connection
- Allows user to edit and rearrange MIDI data

Computer System

- Heart of a MIDI system
- Controls the scheduling, synchronization and recording of all data.


High Five LIVE
1|2|000
0:00:00.55
Tracks
Sequence
Clips
Mixing Board
MIDI
Drum
QuickScribe
Song
Waveform
Can't DJ-elq Ma Face
Chunks
Synth Intro Track#.5
Synth Intro Mono# 8
Synth Intro Mono# 9
Synth Intro Mono# 2
High Five LIVE mix-03

MIDI Control Input Devices:

- Usually a Keyboard with additional control: sustain, pitch bend,modulation, aftertouch and other controllers
- Can be another musical device e.g. Customised Guitar, Wind Controller
- Can be just a bunch of controllers.
- Can be even more strange: Motion Capture, or Virtual Input or Mind Control!!

MIDI Interfaces:

MIDI devices (still) need to connect to computer with some interface

- MIDI Interface — USB or Firewire
- Often functionality bundled with Keyboard or controller
- Audio Interface via USB or Firewire common
- Even Wireless Keyboards

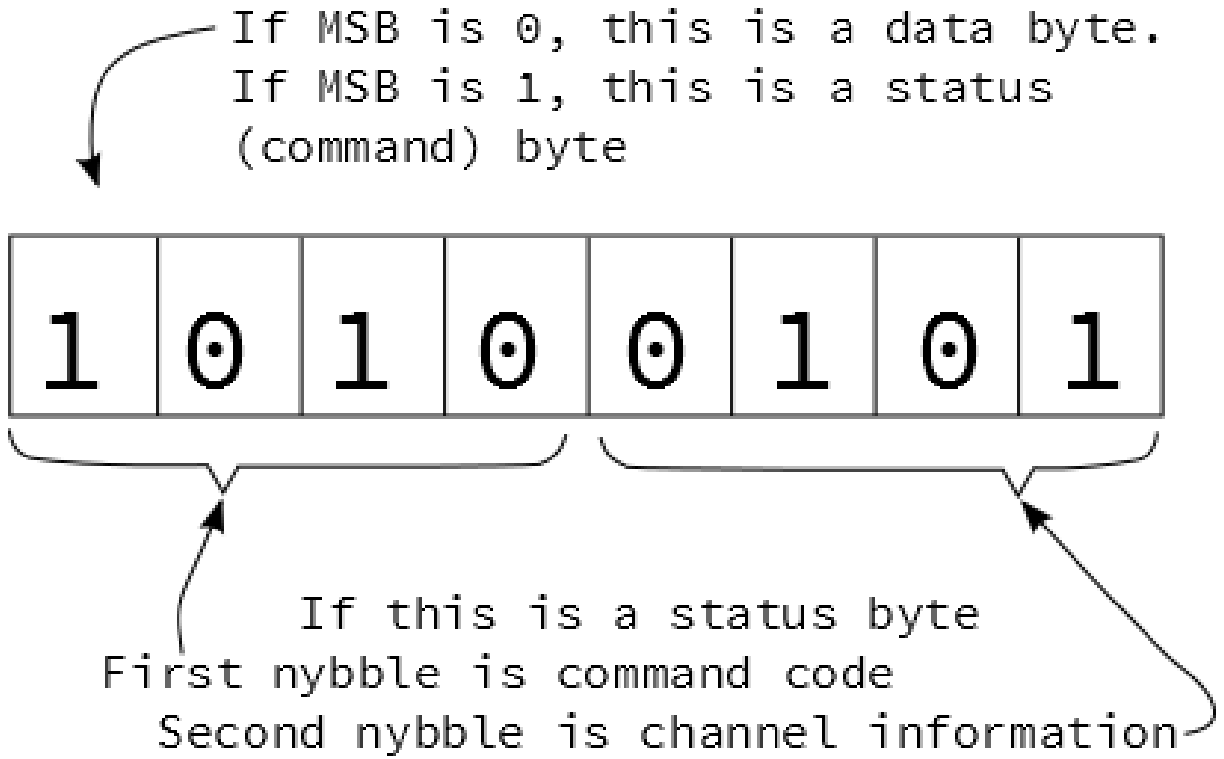
Structure of MIDI message

- MIDI message includes
- 1 status byte and upto 2 data bytes.
- The most significant bit of status byte is set to 1
- The most significant bit of data byte is set to 0

Status Byte

- The most significant bit is set to 1
- Range is from 128 to 255
- Last 4 bit identifies the channel message belongs to
- It means single MIDI cable can be assigned with 16 different channels.
- Remaining 3 bits identifies the message.

Status byte



Status Byte Dissection

Status byte

Voice Message	Status Byte	Data Byte1	Data Byte2
Note off	8x	Key number	Note Off velocity
Note on	9x	Key number	Note on velocity
Polyphonic Key Pressure	Ax	Key number	Amount of pressure
Control Change	Bx	Controller number	Controller value
Program Change	Cx	Program number	None
Channel Pressure	Dx	Pressure value	None
Pitch Bend	Ex	MSB	LSB

Notes: 'x' in status byte hex value stands for a channel number.

MIDI Channels

- Last four bits of status byte represent MIDI channels
- So, there are 16 channels from 0 to 15.
- Channels separates MIDI messages
- Each channel is associated with particular instrument
- Channel 1 is the piano, channel 10 is the drums.

Ch 1: piano	Ch 2: guitar
Ch 3: bass	Ch 4: tuba

MIDI software

- Variety of MIDI applications run on MIDI system
- Music recording and performance application
- Musical notations and printing applications
- Music education applications

Speech

- Speech is generated, perceived and understood by human
- Dialect and pronunciation differ human by human
- Human brain can recognize speech and noise
- Human ear is sensitive in the range 600hz to 6000hz(human audible 20hz to 20000hz)
- Machine support speech generation and recognition

At the Atlanta Airport, artificial voice can be heard

Speech

- Natural form of human communication
- Relate to language; branch of social science
- Related to human physiological capability; branch of medical science
- Related to sound and acoustics; branch of physical science
-

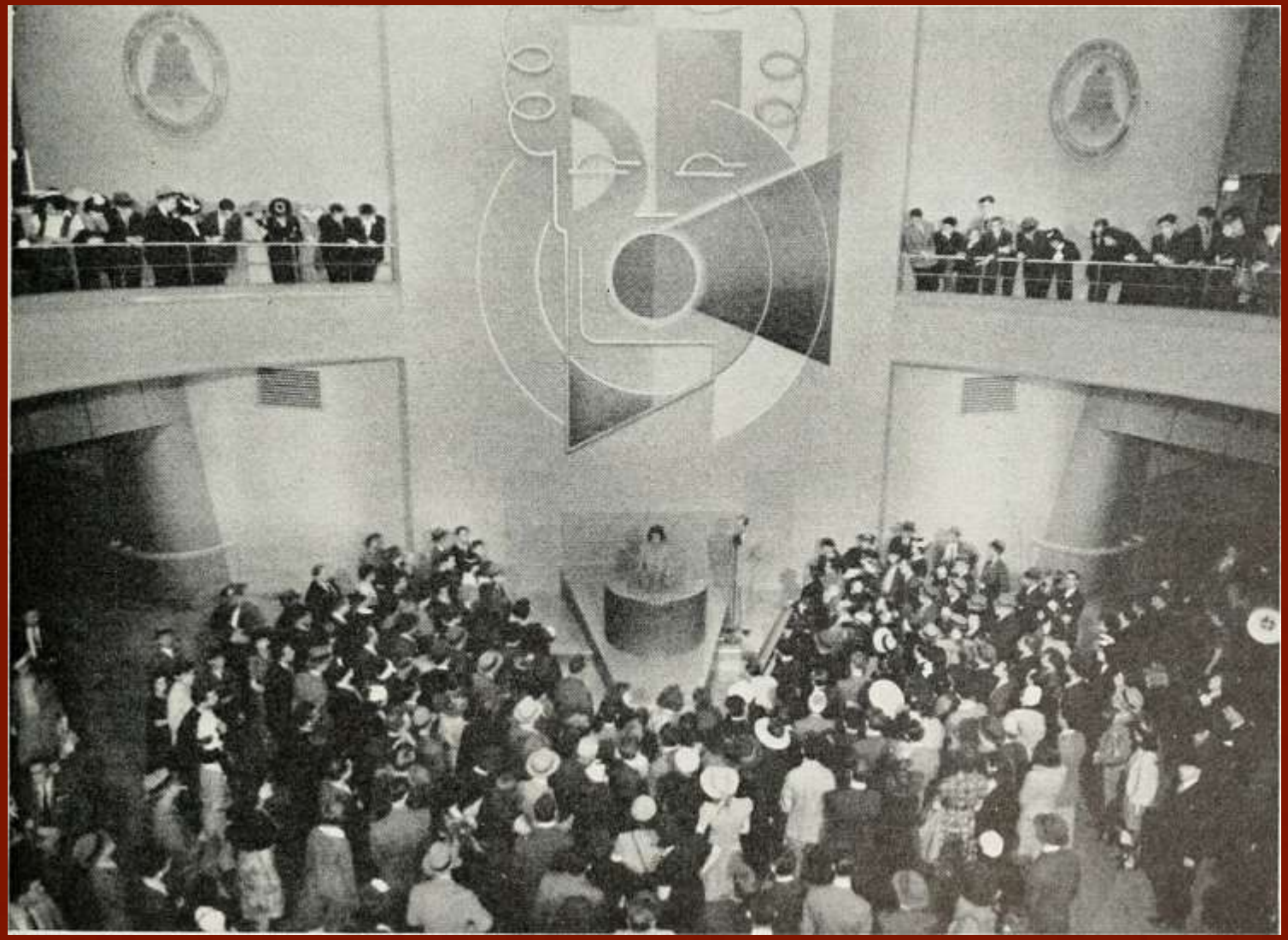
Speech Processing

- Purpose of speech processing includes speech coding (digitization, compression)

1. Speech synthesis
2. Speech recognition
3. Speaker verification
4. Speech enhancement

Speech Generation

- In 19 century, Helmholt introduced mechanical vocal tract to generate speech.
- In 1940, Dudley introduced first speech synthesizer
- Real time signal generation is very important for speech generation
- Generating speech need large vocabulary
- Must be natural



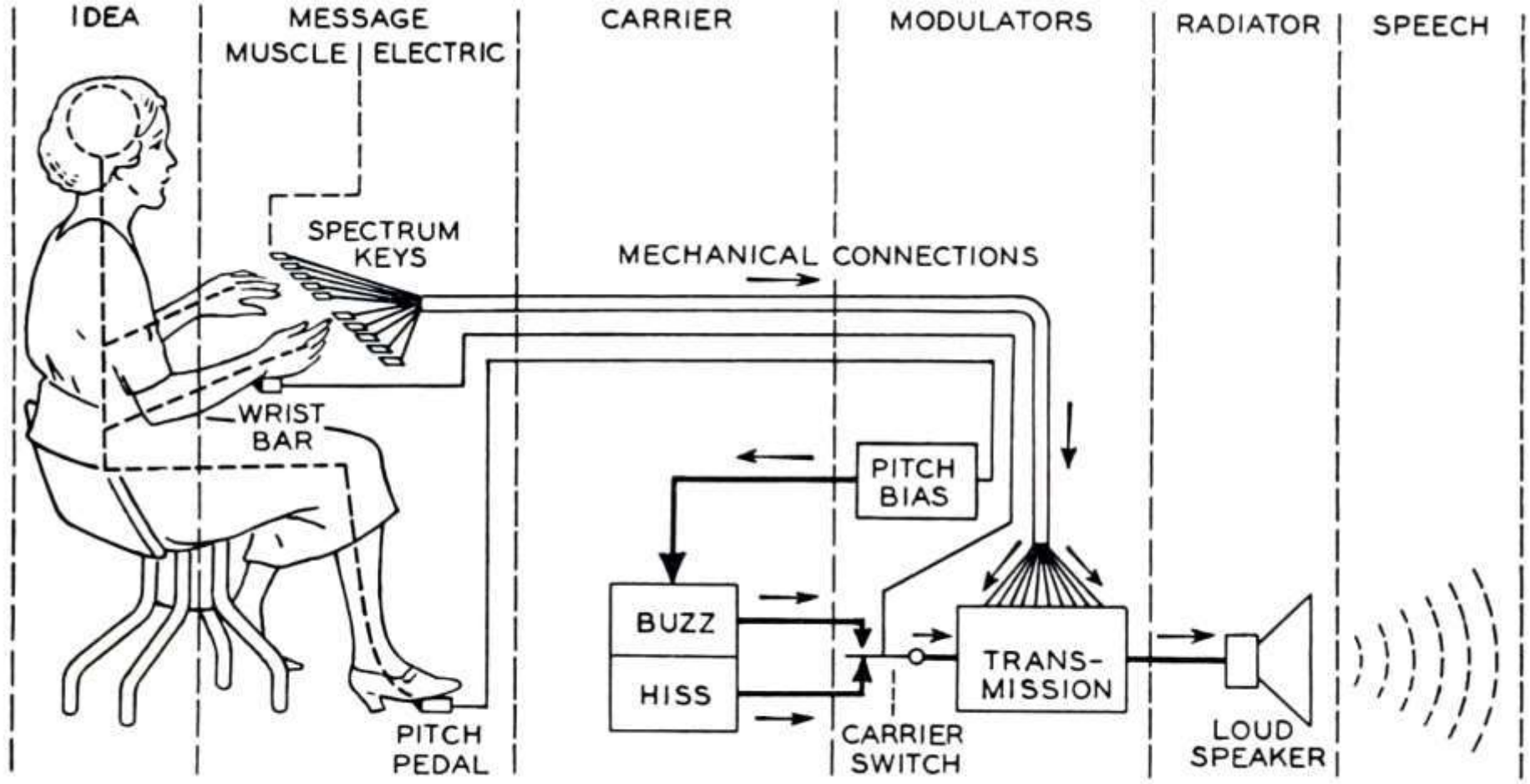


Fig. 8—Schematic circuit of the voder.

Speech Synthesis

- ***Synthesis of Speech*** is the process of generating a speech signal using computational means for effective human-machine interactions
  - machine reading of text or email messages
  - telematics feedback in automobiles
  - talking agents for automatic transactions
  - automatic agent in customer care call center
  - handheld devices such as foreign language phrasebooks, dictionaries, crossword puzzle helpers
  - announcement machines that provide information such as stock quotes, airlines schedules, weather reports, etc.

Speech Synthesis

- Computer system used to produce speech is called speech synthesizer
- Text to speech converts text into human speech.
- Two methods of speech synthesis
- Time-dependent concatenation
- Frequency-dependent concatenation

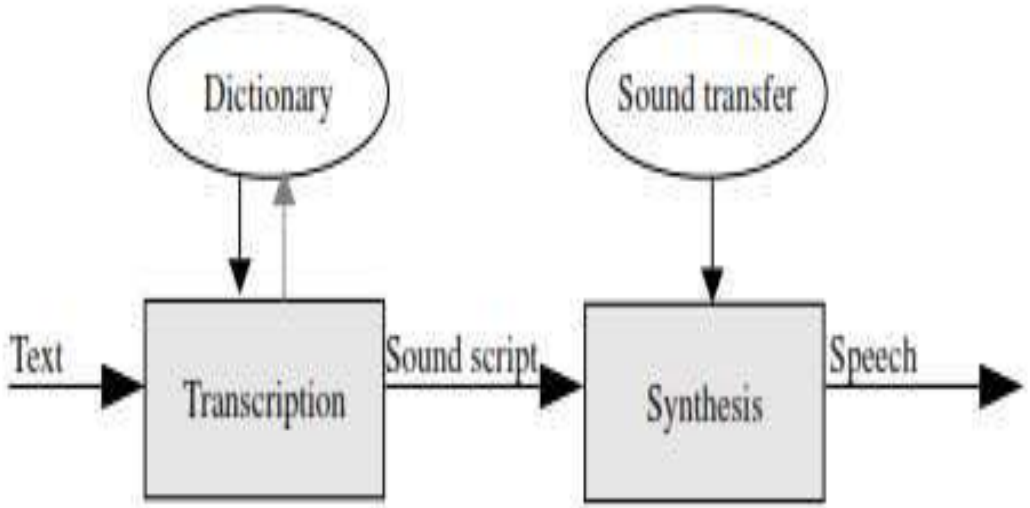
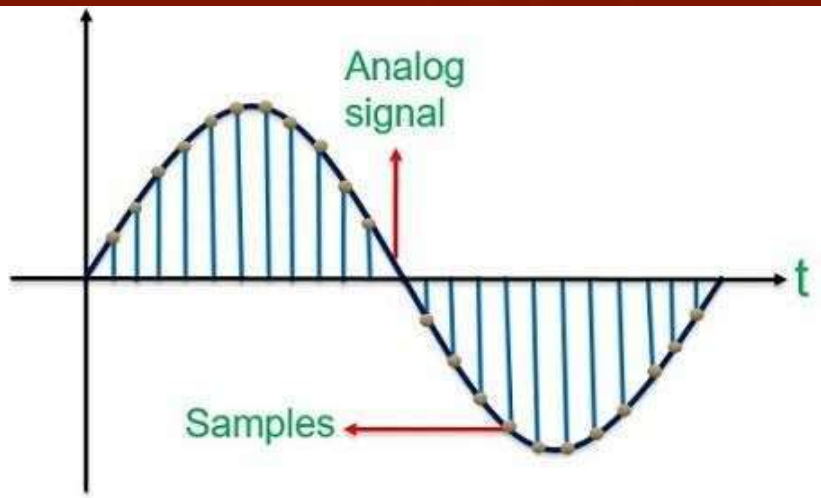


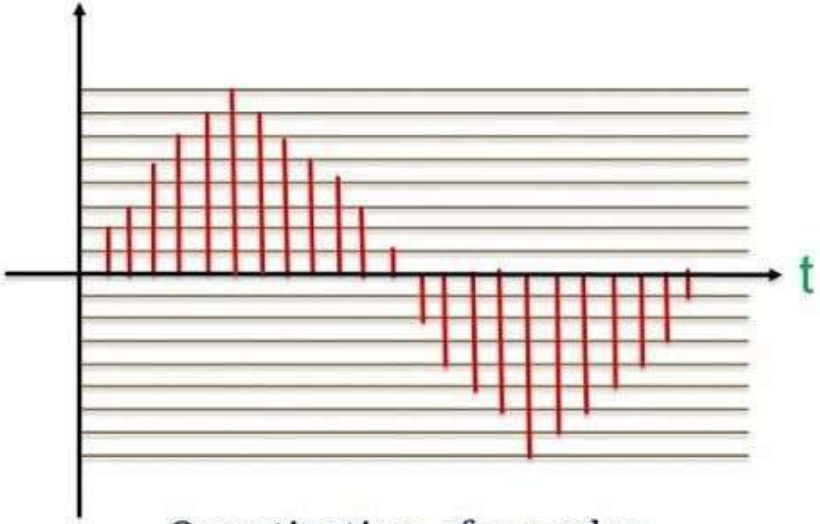
Figure 3-11 Components of a speech synthesis system, using sound concatenation in the time range.

Time-dependent Concatenation

- The easiest method is to use prerecorded speech and play it back in timely fashion.
- Speech can be stored as PCM samples
- Compression techniques can be applied to such recorded speech.
- But If a word is not recorded, then it can’t be used.



Sampling of analog signal



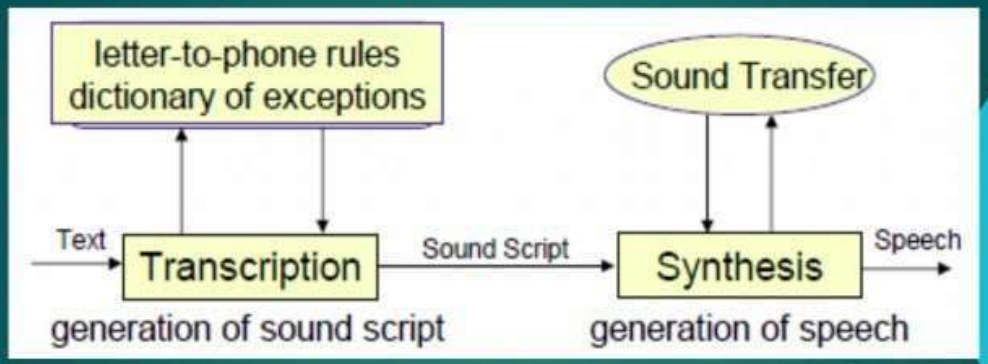
Quantization of samples

Frequency Dependent Concatenation

- Based upon vocal tract simulation i.e. formant synthesis.
- Formant is the high amplitude frequencies.
- Formant synthesis uses filters to simulate the vocal tract

Components of a speech synthesis system

Components of a speech synthesis system:



Step 1: Generation of a Sound Script

Transcription from text to a sound script using a library containing (language specific) letter-to-phone rules. A dictionary of exceptions is used for word with a non-standard pronunciation.

Step 2: Generation of Speech

The sound script is used to drive the time- or frequency-dependent sound concatenation process

Components of a speech synthesis system

- In the first step sound script is produced by the method of transcription.
- For this, letter to phones rules and dictionary of Exceptions are used
- In second step, sound script is translated into a speech signal.
- For this time dependent concatenation is used.
- First step is done by software while second is done by signal processor

Speech Analysis

- Human speech is always distinct.
- Speech analysis helps to recognize the speaker
- Thus computer can verify the speaker based upon the speech
- It also help to recognize and understand speech signal.
- Corresponding text can be generated
- Speech controlled typewriter

Speech Analysis

- Study of speech signals and processing methods of these signals
- Processed usually in digital form
- It includes acquisition(input), manipulation, storage, transfer, speech recognition, coding etc
- For example computer may accept a verbal command, computer may synthesis speech from text etc
-

Speech Analysis

- Based upon speech, mood of speaker can be classified
- A person sounds different when he/she is angry or calm.

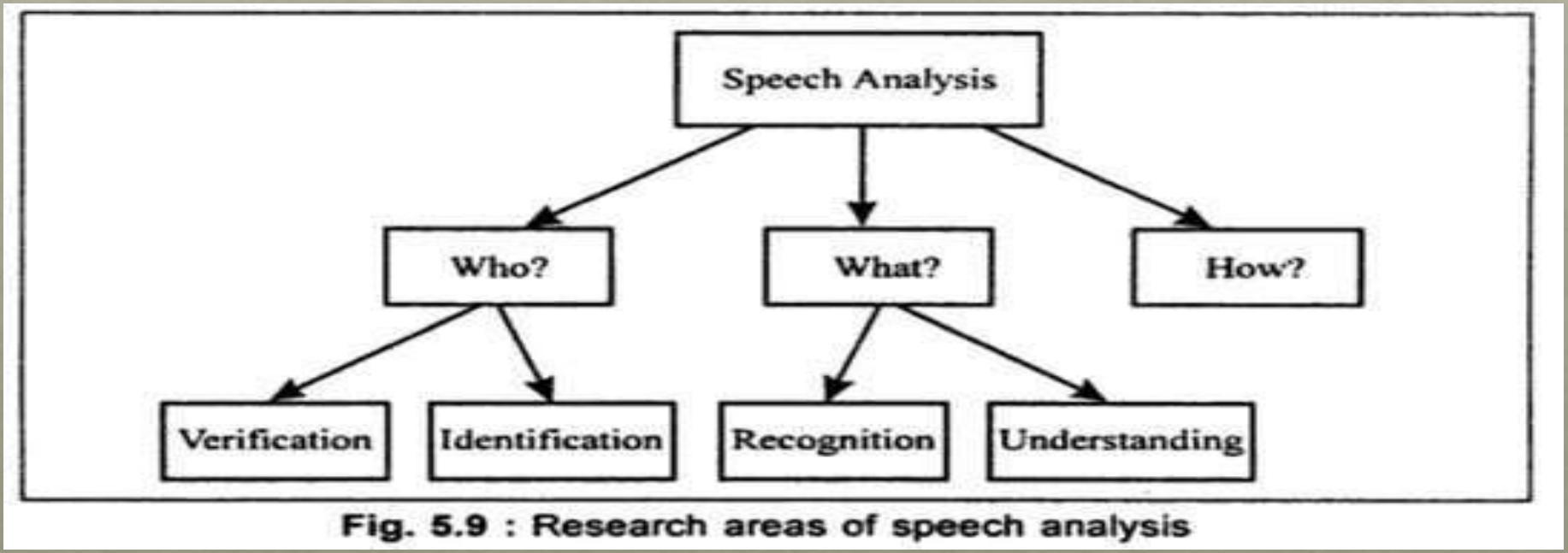


Fig. 5.9 : Research areas of speech analysis

Speech Analysis

- Primary goal is to determine individual words with probability <=1 .
- Due to ambient noise, sense ambiguity (there, their) dialect, stress, systems are not much accurate

We multiply probability of recognizing individual words for n times to calculate probability of recognizing a sentence.

Where n is number of words in the sentence.

Speech Analysis

- If probability of recognizing individual words is .95 and sentence has 3 words, then
- Probability of recognizing sentence is .95*.95*.95 = .875
- Problems in recognizing sentence include word boundaries identification (accent), semantics, normalization (quickly, slowly)

Components of Speech

- Sound patterns and word models are applied to perform acoustical and phonetically analysis.
- Next, syntactical analysis is performed.
- This helps to recognized errors in previous step.
- Syntactical analysis provides additional aid to recognized the speech
- Deals with semantic of the language.
- We can implement this step with AI methods

Components of Speech

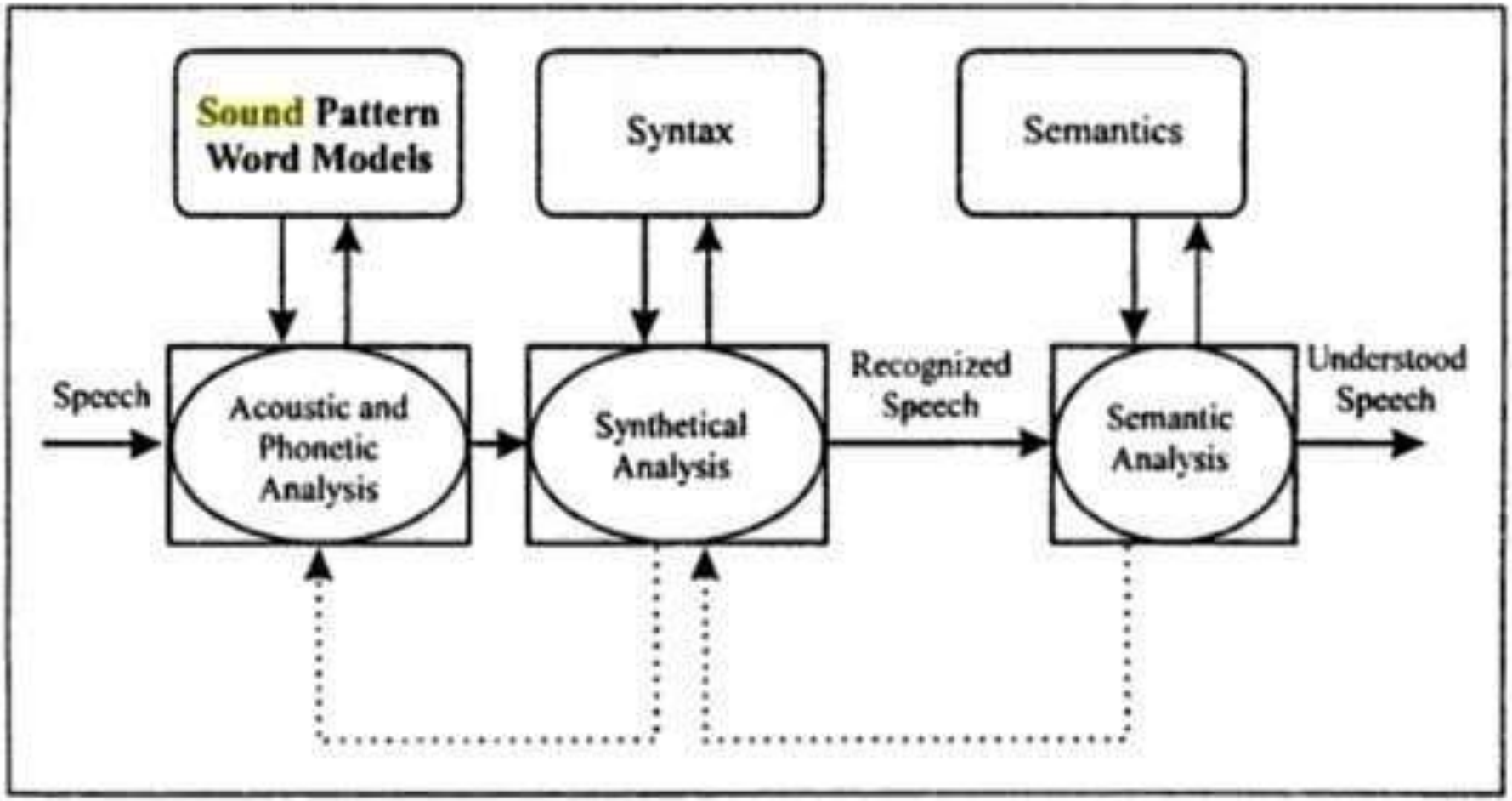


Fig. 5.11 : Components of speech recognition and understanding

Speech recognition system types

- Speaker independent recognition system
- Speaker dependent recognition system

Speaker Independent system

- Can be used without training.
- But limited number of words can be recognized
- But speaker dependent system gets training to recognize extensive vocabularly

Speech Transmission

- Speech is usually large in size.
- So, it must be sampled, quantized and coded before we transfer it.
- Coded in such a way that Quality is almost same at both sender and receiver sides
- Differences in signals between present and previous time can effectively reduce size.

Signal form coding

- One of coding used to allow efficient transmission
- It does not use specific properties and parameters
- Tries to achieve most efficient coding of audio signal
- PCM coded Cd quality speech needs only 1411200 bits/s data rate for transmission
- DPCM coded telephony quality requires only 56Kbits/s
- ADPCM coded speech signal requires only 32Kbits/s data rate

Source coding

- Specific speech parameters are used for speech data rate reduction
- The probability of symbols that source produce can be exploited for generating the source code.
- The frequent symbols can be coded with more bits than less frequent.
- This is called huffman coding or variable length coding.
- Fixed length codes also can be used.

Source coding

- Deals with efficient coding of speech signal for transmission
- Quality is almost same at both sender and receiver sides
- Data rate is about 3 Kbits/s(nowadays 256)
- Quality is of speech satisfactory