UNIT-2

Sound / Audio System

Sound is a physical phenomenon produced by vibration of matter and is transmitted as waves. As matter vibrates, pressure vibration are created in the air surrounding it. The alternation of high and low pressure is propagated through air in a wave like motion. When a wave reaches a human air a sound is heard.

⊛ Basic concepts of sound:

i) Frequency: The wave form repeats the same shape at regular interval and this portion is called period. Frequency of sound is the reciprocal value of period. It represents the number of period in a second and is measured in Hz cycles/sec, simply Hz.

i.e, $\boxed{\text{frequency of sound } (F) = \frac{1}{P}}$

Note: $1\ KH_2 = 1000\ Hz$.

0 to $20 Hz$ = Infra Sound

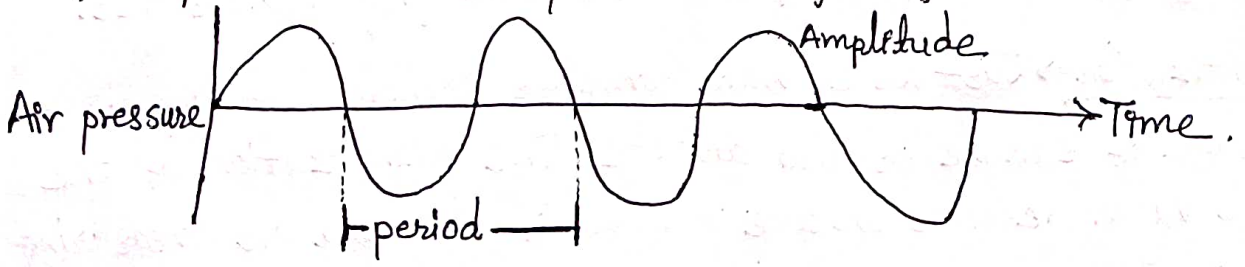
$20 Hz$ to $20 KH_2$ = Human hearing frequency range/Audible sound.

$20 KH_2$ to $1 GH_2$ = Ultra Sound

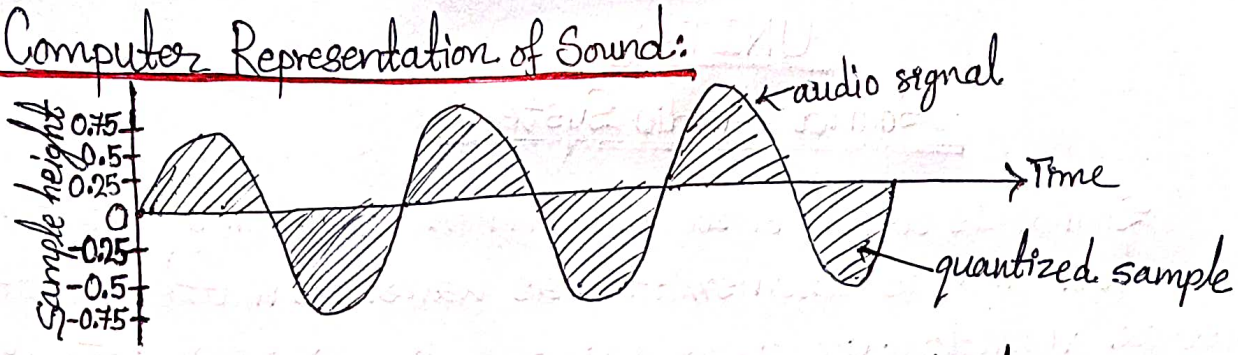
$1 GH_2$ to $10 TH_2$ = Hyper Sound.

⇒ Multimedia system typically makes use of sound only within the frequency range of human hearing.

ii) Amplitude: Amplitude is the loudness of sound. It is the measure of displacement of air pressure waveform from its mean.



⊛. Computer Representation of Sound:



→ A transducer converts pressure to voltage levels.
→ The analog signal is converted into a digital stream by discrete sampling:
→ The analogous signal is sampled in regular time intervals, i.e, the amplitude of the wave is measured.
→ Discretization both in time and amplitude (quantization) to get representative values in a limited range. (e.g. quantization with 8 bit: 256 possible values).
→ Result: series of values:

0	0.25	0.5	0.75	0.75	0.5	0.25	0	-0.25	-0.5	...

⊛. Sampling → Sound wave from the smooth, continuous is not directly represented in the computer. The computer measures the amplitude of wave form in the regular time interval to produce the series the numbers. Each of this measurement is called sample. This process is called sampling.

⊛. Sampling rate → The rate at which a continious wave form is sampled is called sampling rate. Like frequency, sampling rate are measured in Hz. For loss less digitization the sampling rate should be at least twice of the maximum frequency response.

⊛. Quantization → Just as a wave form is sampled at discrete times the value of sample is also discrete. The quantization of the sample value depends on the number of bits used in measuring the height of the wave form. The lower quantization, lower the quality of sound, higher quantization higher quality of sound.

⊛. Sound Hardware → Devices that are connected to ADC and DAC (Analog to digital converter and Digital to analog converter) for input and output of audio to the computers are known as sound hardware.

⊛. Music and Speech:

⊛. Basic MIDI concepts: MIDI stands for Musical Instrument Digital Interface. MIDI is the standard that manufacturer of electronic musical instrument have agreed upon. It is a set of specification, they use in building their instrument. So that the instrument of different manufacturers can communicate without difficulty. MIDI is defined in 1985.

⊛. MIDI Devices:

1) **Synthesizers:** It is a component that generates sound based on MIDI message.
   - → Sound generator
   - → Microprocessor
   - → Keyboard
   - → Control pannel
   - → Memory.

2) **Sequencers:** It is an electronic device in cooperating with both hardware and software, which is used as storage server for generated MIDI data. A sequencer may be computer. Sequencer transforms the note into MIDI message.

3) **Controllers:** MIDI controllers are the devices for manipulating the generated MIDI software messages. MIDI controller are commonly integrated with synthesizer to maximize the application. Example: Ztar, Alesis Q.

4) **Networks:** MIDI network is combination of hardware and software to interconnect group of MIDI devices such as synthesizer, controller and sequencer.

⊛. MIDI Message: MIDI message transmit information between different MIDI devices, which determine the kind of musical event to be passed from device to device. The format of the MIDI message consists of status byte (first byte of any MIDI message) which describe the kind of message and data type. MIDI message are divided into following types:

1) Channel Message: It go only to specific devices. There are two types of channel message:

i) Channel voice message → It sends actual performance data between MIDI devices describing keyboard action, controller action and control pannel changes. They describe music by defining pitch, note on, note off etc.

ii) Channel mode message → It determines the way that a receiving MIDI device respond to channel voice message. Channel mode message includes Omni On, Omni Off etc.

2) System Message: System message go to all devices in a MIDI system because no channel number are specified. There are three types of system message:

i) System real-time message → These are very short and simple, consisting of only one byte. They carry extra data with them. These messages synchronize the timing MIDI devices in performance.

ii) System common message → System common messages are commands that prepares sequencers and synthesizers to play song. E.g. song selected, find the common starting place in the song.

iii) System exclusive message → These allow MIDI manufacturers to create customized MIDI message to send between the MIDI devices.

MIDI Software:

Once a computer is connected to MIDI system, a variety of MIDI software applications can run on it. The software application generally falls into four major categories:

→ Music recording and performance application.
→ Musical notation and printing application.
→ Synthesizer Patch editor and library patch.
→ Music education application.

⊛. MIDI and SMPTE Timing Standard:

MIDI reproduces traditional note length using MIDI clock, which are represented through timing clock message. Using MIDI clock a receiver can synchronize with the clock cycle of sender. To keep the standard timing refrence the MIDI specification state 24 MIDI clock = 1 quarter note.

As an alternative the SMPTE timing standard (Society of Motion Picture and Television Engineer) can be used. The SMPTE timing standard was orginally developed by NASA as a way to make incoming data from different tracking stations so that the receiving computer could fill what time each piece of data was created. SMPTE format consists of hour:second:frames:bits.

⊛. Speech Generation:

Speech generation is the computer-generated simulation of human speech.

⊛ Basic notations:

→ Speech consists of fundamental frequency as it's component as it is used to present a voice of sound.
→ It consists of morph, phone as smallest speech unit.
→ It consists of vowels, consonants as it's component.
→ A voiced sound is generated via vocal cords.
→ Allophones mark the variants of phone.
→ During the generation of an unvoiced sound, the vocal cords are opened.

⊛. Reproduced Speech Output:

Reproduced, time-dependent and frequency dependent are types/methods of speech generation

The easiest method of speech generation/output is to use prerecorded speech and play it back in a timely fashion. The speech can be stored as PCM (Pulse Code Modulation) samples. Further data compression methods, without using language typical properties, can be applied to recorded speech.

⊛. Time dependent sound concatenation:

Speech generation/output can also be performed by sound concatenation in a timely fashion. Individual speech units are composed like building blocks, where the composition can occur at different levels. In simplest case, the individual phones are understood as speech units. Figure1 shows the individual phones of the word crumb.

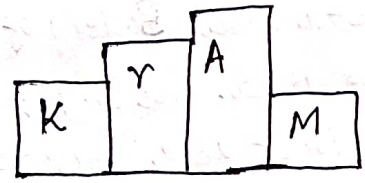


Fig1: Phone sound concatenation

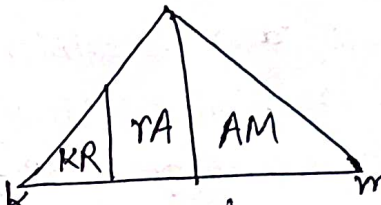


Fig2: Di-phone concatenation.

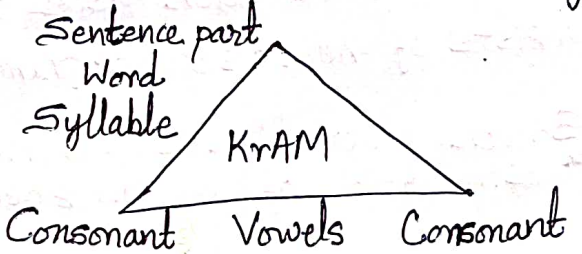


Fig3: Word sound concatenation.

Fig4: Syllable sound.

Fig 2 shows the word crumb, which consists of an ordered set of diphones, Fig 4 shows the syllable sound of the word crumb, through the set of syllables. The best pronounciation of word is achieved through storage of whole word. This leads toward synthesis of the speech sequence as in fig 3.

⊛. Frequency-dependent sound concatenation:

Speech generation/output can also be based on a frequency-dependent sound concatenation. Formats are frequency maxima in the spectrum of the speech signal. Format synthesis simulates the vocal tract through a filter. The characteristic values are the filter's middle frequencies and their bandwidths. A pulse signal with a frequency, corresponding to the fundamental speech frequency is choosen as a simulation for voiced sounds. On the other hand unvoiced sounds are created through a noise generator.

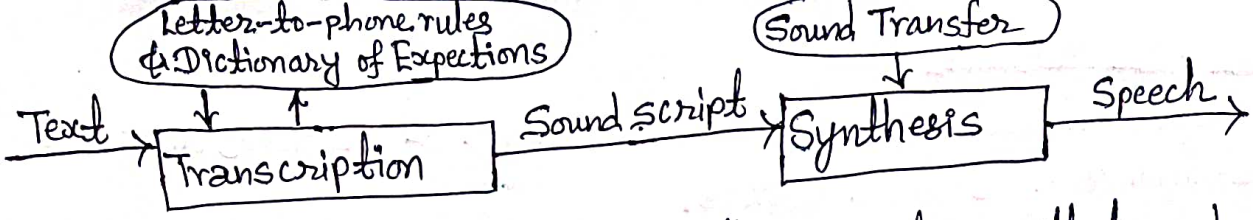


Fig: Components of speech synthesis system with time-dependent sound concatenation.

⊛. Speech Analysis:

The primary goal of speech analysis is to correctly determine individual word. Speech analysis may fail to achieve its goal due to following factors:

→ Ambient noise
→ Word ambiguity (for e.g. there/their)

⊛. Research area of speech analysis:-

Speech analysis/input deals with the research areas as shown in the figure below:-

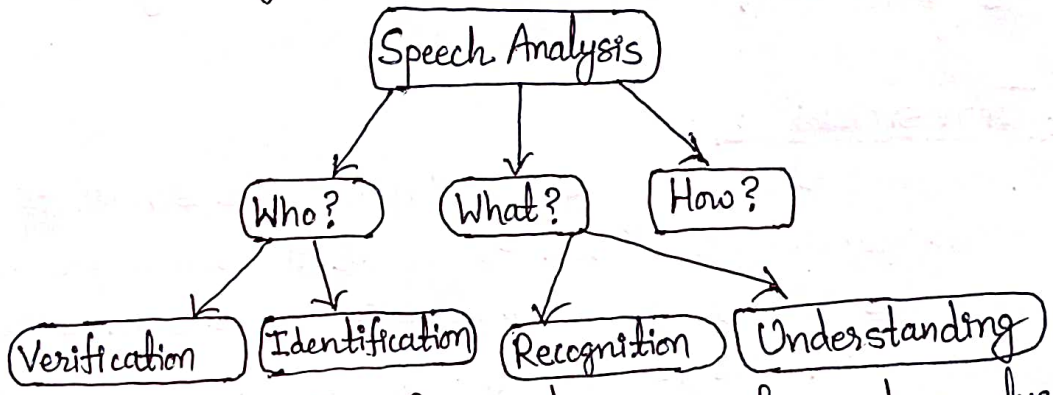


Fig: Research areas of speech analysis.

→ Human speech has certain characteristics determined by speaker. Hence, speech analysis can serve to analyze who is speaking. i.e, to recognize a speaker for his/her identification and verification. The computer identifies and verifies the speaker using an acoustic fingerprint. Acoustic fingerprint is a digitally stored speech probe (i.e, certain statement) of a person.

→ Another main task of speech analysis is to analyze what has been said, i.e, to recognize and understand the speech signal itself.

→ Another area of speech analysis tries to research speech patterns with respect to ~~thigh~~ how a certain statement is said. For example, a spoken sentence sounds differently if a person is angry or calm.

⊛ Speech Recognition:

Speech Recognition is the capability of electronic device to understand the spoken words.

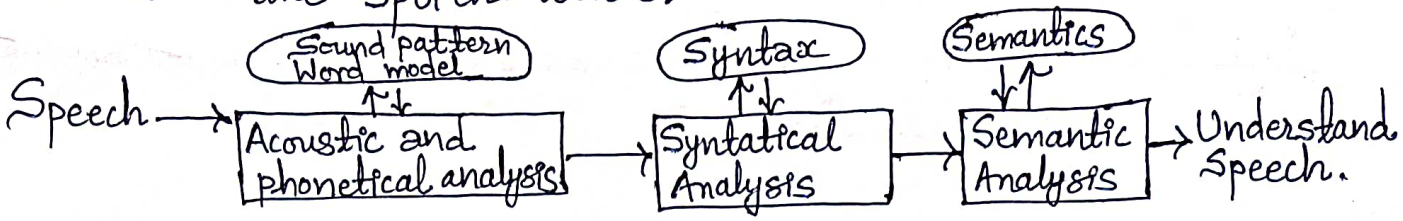


Fig: Components of speech recognition.

→ Initially the speech generated goes for acoustic and phonetical analysis.

→ Then goes for syntatical analysis so that the errors of previous step can be recognized and syntatical analysis is made providing additional decision to produce recognized speech.

→ Third step deals with semantic analysis to the recognized speech that further filters the errors of previous and finally understandable speech is generated.

⊛. Speech Transmission:

The process of sending speech/audio from sender to receiver with fundamental goal to provide the same speech/audio (sound quality) as it was generated at the sender side is called speech transmission. This section includes some principles that are connected to speech generation and recognition.

i) Signal Form Coding → It is a technique to achieve most efficient coding of audio signal without considering speech property & parameter.

ii) Source coding → Parameterized system works with source coding algorithm. It uses the speech/audio characteristics for data rate reduction.

iii) Recognition/Synthesis Methods → In this technique only the characteristics of the speech elements are transmitted. It is used in data rate reduction.