Unit-2: Sound/audio system

- Ankit Pangeni DATE

Concepts of sound system

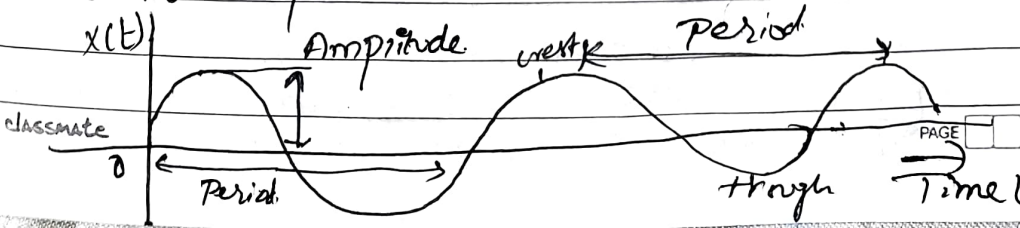
Sound is defined as vibrations that travel through air or any other medium as an audible mechanical wave. It involves three systems

- The source which emits sound
- The medium through which sound propagates
- The detector which receives & interprets the sound

* Frequency: Frequency, sometimes referred to as pitch, is the number of times per second that a sound pressure wave repeats itself.
In other words, the no. of vibrations made by a particle of the medium in one second is called the frequency of sound wave. It is measured in Hz.
$F = 1/p$, where p = period is the interval at which a periodic signal repeats

Infra-sound => 0-20 Hz (Earthquake)
Human hearing range => 20Hz - 20 kHz (used by mankind)
Ultrasound => 20kHz - 1 GHz (Sonograms)
Hypersound => 1 GHz - 10 THz

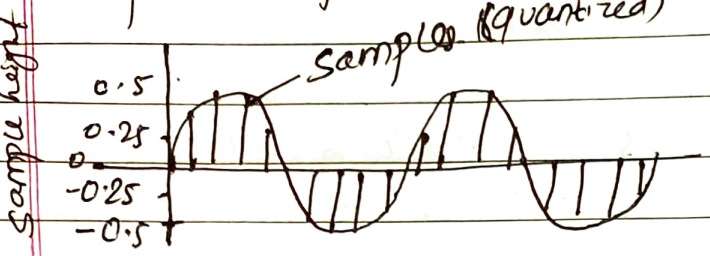
* Amplitude: It is the measure of height of the sound wave. It can be defined as the loudness or the amount of maximum displacement of vibrating particles of the medium from their mean position when the sound is produced



PAGE

DATE

* Computer representation of sound



- Sound waves are continuous while the computers are good at handling discrete numbers
- In order to store a sound wave in a computer, samples of the wave are taken. Each sample is represented by a number, the 'code'.
- The analog signal is converted into a digital stream by discrete sampling.
- The analogous signal is sampled in regular time intervals ie. amplitude is measured.
- Discretization both in time & amplitude (quantization) to get representative values in a limited range (Ex: quantization with 8 bit: 256 possible values)
- Result: series of values

0	0.25	0.5	0	-0.25	-0.5

....

* Sampling:

Sound waves are continuous while the computers are good at handling discrete number. In order to store a sound wave in a computer, sample of the wave is taken. These samples are the measurement of amplitude of wave in the regular time interval. This process is called sampling.

classmate

PAGE

Note: Sampling process → digitization
converting " ⇒ quantization

DATE

* Sampling rate:

The rate at which a continuous wave form is sampled is called sampling rate. The rate is measured in Hz. The sampling process is known as digitization/ discretization & for loseless digitization, the sampling rate should be twice the max frequency response.

* Quantization: (dividing amplitude)

It is the process of converting a continuous analog audio signal to a digital signal with discrete values. The quantization of the sample value depends on the number of bits used in measuring the height of wave form. The lower quantization, lower the quality of sound, higher quantization, higher the quality of sound.

* Sound hardware

Devices that are connected to ADC & DAC (Analog to digital converter & Digital to Analog converter) for input & output of audio to the computers are known as sound hardware.

classmate

DATE

Music and Speech

* Basic MIDI Concept

Musical Instrument Digital Interface (MIDI) is a communication standard developed in the early 1980s for electronic instruments & computers. It is a protocol for building the musical instrument so that the instrument of different manufacturers can easily communicate.

* MIDI devices:-

- **Synthesizers** : It converts the MIDI note messages to an audio signal. It is basically a device or software that synthesizes sounds in response to incoming MIDI data. Ex: Sound generator, Microprocessor, keyboard, Control pannel, Memory. (combines audio from diff. devices)

- **Sequencers** : It is an electronic device in cooperating with both hardware & software, which is used as Storage Server for generated MIDI data. It allows the user to record & edit a musical performance without using an audio-based input source. ex: launchpad

- **Controllers** : MIDI Controllers are the devices for manipulating the generated MIDI Software messages. They are commonly integrated with synthesizers to maximize the application. Ex: Alesis &

classmate

PAGE

DATE

* Network: MIDI network is the combination of hardware & software to interconnect group of MIDI devices such as synthesizer, controller & sequencer.

MIDI messages

MIDI messages are used by the MIDI devices to communicate with each other. It is made up of an 8 bit status byte followed by one or two data bytes.

Structure: They include a status byte & upto 2 data bytes.

- Status byte: The MSB of status byte is set to 1. The 4 lower order bits identify which channel it belongs to. (2^4=16 possible channels) & 3 remaining bits identify the message.
- The MSB of data byte is set to 0.

Classification of MIDI messages:

- Channel messages: They are channel specific and consist of voice & mode messages. Channel messages are received by only specific devices. Channel voice messages are used to send musical performance information like pitch, note on, note off etc. & Channel mode message includes information like Omni on, omni off which determines the way that a receiving MIDI device respond to channel voice messages.
- System messages: They carry information that is not Channel specific, such as timing signal for synchronization, detailed setup information for destination device. They are received by all devices in a system, including Common, real time & Exclusive.

classmate

→ System real time messages, used to synchronize all MIDI clock based equipment like sequencer, drum machine

PAGE

DATE

* MIDI and SMPTE Timing Standard:-

SMPTE Stands for Society of Motion Picture Television Engineers. This group has defined Cooperative Standards called SMPTE timecode which is used to synchronize film, video or audio material.

MIDI Sync synchronizes via relative time (24 clock messages per quarter note), whereas MIDI Timecode (MTC) synchronizes via absolute time (sending "Quarter Frame" messages based on SMPTE frame rate of, for example, 30 fps). The SMPTE timecode appears as hour: minute: second: frame. (for ex: One houre would be written as 01:00:00:00). The frame rate is derived directly from the data of the recorded medium & it can differ for film vs. digital, video vs. audio, & color vs black and white.

* MIDI Software:-

Once a Computer is connected to MIDI System, a variety of MIDI software applications can run on it. The software applicatin generaly falls under four major categories.

- Music recording & performance application
- Musical notation & printing application
- Synthesizer patch editor & library patch
- Music education application

classmate

PAGE

DATE

Speech Generation.

Speech-generation is the computer-generated simulation of human speech. Speech can be perceived, understood and generated by humans and by machines. Generated speech must be understandable and must sound natural

* Basic Notions

- The lowest periodic spectral component of the speech signal is called the fundamental frequency
- A phone is the smallest speech unit, such as m of the mat and the b of bat in English, that distinguish one utterance or word from another in a given language
- Allophones mark the variants of a phone. Ex, the aspirated p of pit & the unaspirated p of spit are allophones of the English phoneme p.
- The morph marks the smallest speech unit which carries the meaning itself. i.e, consider is a morph, but reconsideration is not
- A voiced sound is generated through the vocal chords m, v and I are examples of voiced sounds. Its pronunciation depends strongly on each speaker
- During generation of unvoiced sound, the vocal chords are opened. F and s are unvoiced sounds. They are relatively independent from the speaker.

* Reproduced Speech Output

The easiest method of speech generation/output is to use pre-recorded speech & play it back in a timely fashion.

classmate

PAGE

DATE

The speech can be stored as PCM (pulse code Modulation) samples.

There are two way of speech generation/output:

1. (Time-dependent sound concatenation)

→ Individual speech units are composed like building blocks, e.g. phones.
→ Transition between speech units via allophones ie variants of phones depending on previous & following phone
→ Creations of syllables as building blocks for words and sentences
→ Prosody: ie. Stress melody course of a spoken phrase.

2. (Frequency-dependent Sound Concatenation)

→ Speech generation can also be based on a frequency-dependent sound concatenation Ex. through a formant synthesis. Formants are frequency maxima in the spectrum of the speech signal. Formant syntheses simulates the vocal tract through a filter.

→ Individual speech elements (e.g. phones) are defined through the characteristics values of the formants

→ A pulse signal frequency is chosen as a simulation for voiced sounds. On the other hand, unvoiced sounds are created through a noise generation.

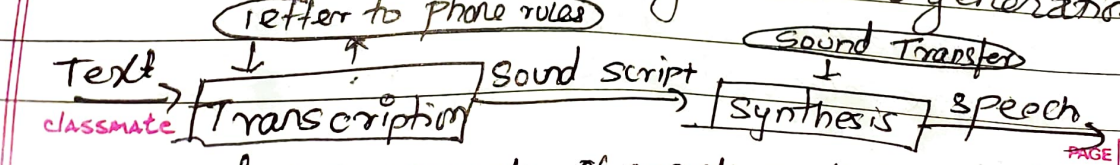


fig: Components of speech synthesis system

classmate

PAGE

with time independent sound co

DATE

Speech Analysis

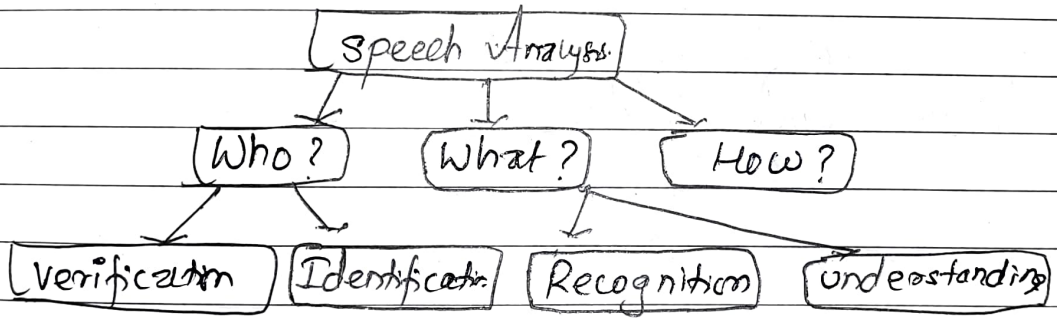
The primary goal of speech analysis is to correctly determine individual word.

Purpose of speech analysis

→ Who is speaking : Speaker identification.
→ What is being said: automatic transcription of speech into text
→ How was a statement said: understanding psychological factors of a speech pattern (angry or calm or lying)

* Research area of speech analysis:

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



You can write above purposes:

* Speech Recognition

It is the capability of electronic devices to understand the spoken words. The system which provides recognition and understanding of a speech signal ~~applies~~ is a speech recognition system. Speech recognition is the ability of a machine to identify words spoken aloud & convert them into readable text.

classmate

PAGE

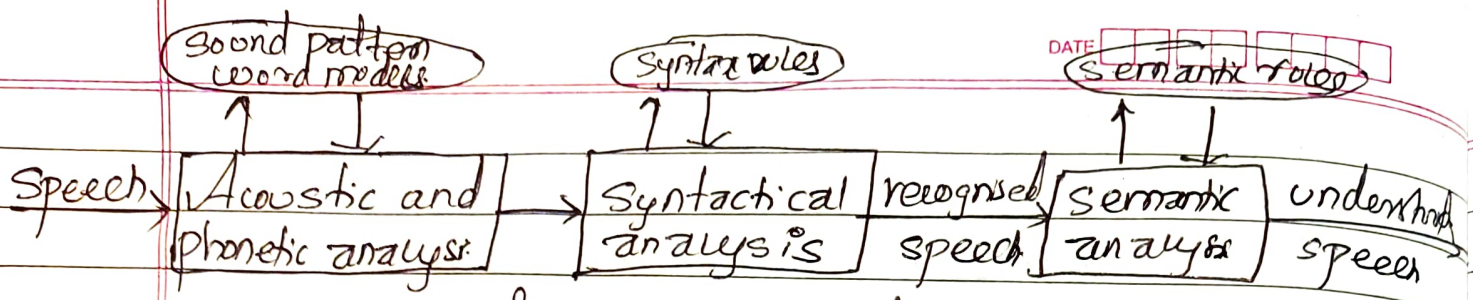


fig. components of speech recognition system

- Initially the speech generated goes for acoustic and phonetical analysis where the characteristics of speech including description of speech in terms of its physical properties such as frequency, intensity & duration is analysed
- Then goes for syntactical analysis so that the errors of previous step can be recognised & syntactical analysis is made providing additional decision to produce recognized speech.
- 3rd step deals with semantic analysis to the recognized speech that further filters the errors of previous & finally understandable speech is generated

DATE

Speech Transmission

Speech transmission is the process of sending speech/audio from sender to receiver with fundamental goal to provide the same speech/sound quality as was generated at the sender's side. It deals with efficient coding of the speech signal to allow speech/sound transmission at low transmission rates over networks.

Some principles that are connected to speech generation & recognition are:

- Signal form coding: It is the technique to achieve most efficient coding of audio signal without considering speech property & parameters. Here, the goal is to achieve the most efficient coding of the audio signal.
- Source coding: Parameterized systems work with source coding algorithms. Here, the specific speech characteristics are used for data reduction. Channel vo-coder is an example of such a parameterized system.

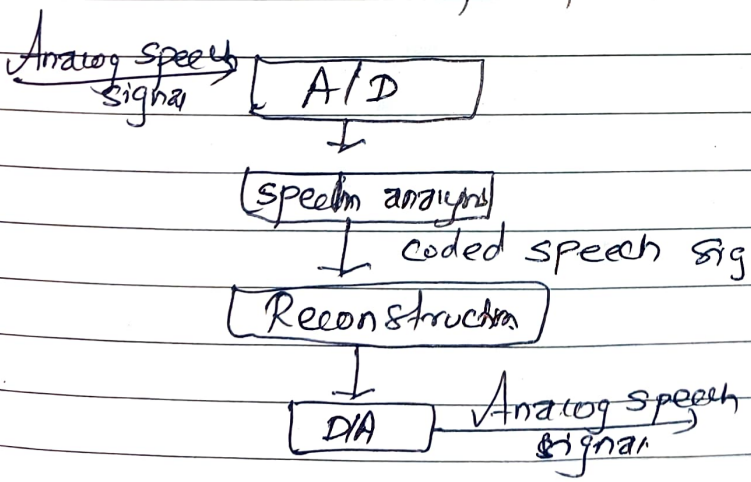


fig: Source coding

- Recognition/synthesis methods: There have been attempts to reduce the transmission rate using pure

DATE

Recognition/synthesis methods. Speech analysis (recognition) follows on the sender side of a speech transmission system & speech synthesis (generation) follows on the receiver's side

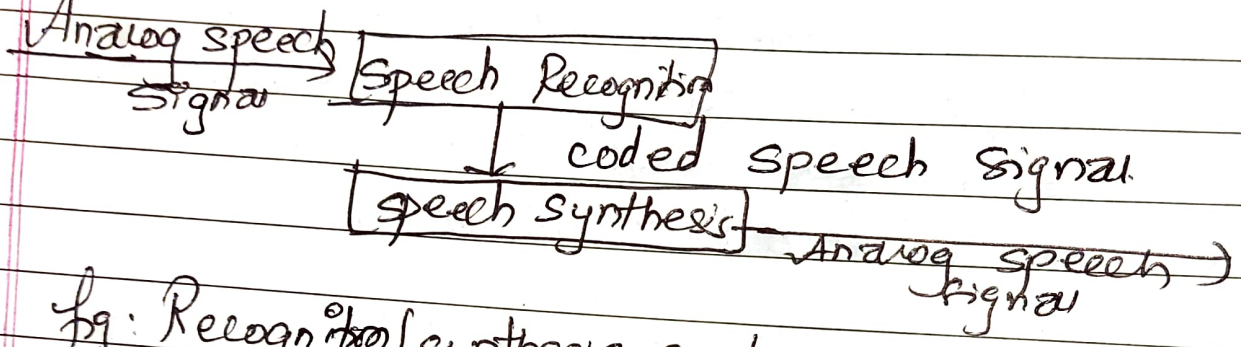


fig: Recognition/synthesis system.

Questions asked from this chapter

Q. Explain the speech generation method (2076-5 marks)

Q. Compare between MIDI & Digital Audio. Calculate the file size in bytes for a 60 second recording at 44.1 KHz, 8 bits resolution stereo sound (2078 - 10 marks) (2076 - 5 marks)

Assignment-01

: Ankit Pangeni

DATE

Q-1. Describe the data stream characteristics for continuous media.

=> Continuous media is data where there is a timin relationship between source and destination. The most common example are audio and motion video

The data stream characteristics for continuous media are related to any audio and video data transfer. can be diswssed on the basis of 3 properties or fact

a. According to time intervals between consecutive packets
- On the basis of this factor we have 3 properties

- Strongly periodic data stream
  → If the time intervals are of same length between two consecutive packets that is a constant, then the stream is called strongly periodic & Example is PCM coded speech in traditional telephone switchin

- Weakly periodic data stream
  → If the time intervals between two consecutive packets is not constant but are of periodic nature with finite period then the data stream is called weakly period

- A-periodic data stream
  → If the sequence of time intervals is neither strongly weakly periodic, then it is called aperiodic data strea. Here, time period varies between packets to packets.

DATE

b. According to variation of consecutive packet amount.

- Strongly regular data stream:
→ If the amount of data stays constant during the lifetime of a data stream, for ex:- video stream of camera in uncompressed form, then it is strongly regular.

- Weakly regular data stream
→ If the amount of datastream varies periodically with time then it is weakly regular. ex: Compressed video stream

- Irregular data stream
→ If the amount of data is neither constant nor changes according to a periodic function, then the data stream are irreg. Transmission & processing of this category is more complicate

C. According to Continuity or connection between consec. packet

- Continuous data stream
→ If consecutive packets are directly transmitted one after another without any time gap, then it is called continuous d ex: audio data use for B channel of ISDN with transm ission rate of 64 kbps.

- Unconnected data stream
→ A data stream with gaps between information units is called unconnected data stream.