Chapter-2

Multimedia System (Pokhara University)

Chapter 2 - Sound and Audio System

Sound is the physical phenomenon produced by the vibration of matter. When a matter vibrates, pressure variations are created in the air surrounding it. This alteration of high and low pressure is propagated through the air in a wave like motion. When the wave reaches the human ear, a sound is heard.

Basic Sound Concepts

The pattern of the oscillation is called a ***waveform.*** The waveform repeats the same shape at regular intervals and this point is called a ***period.*** Since sound wave forms occur naturally, Sound waves are never perfectly smooth or uniformly periodic.
Periodic sound: E.g. Musical instruments, vowel sounds, whistling wind, bird songs etc.
Non-periodic sound: E.g. Unpitched percussion Instruments, coughs, sneezes, rushing water, Consonants, such as "t," "f," and "s"etc.

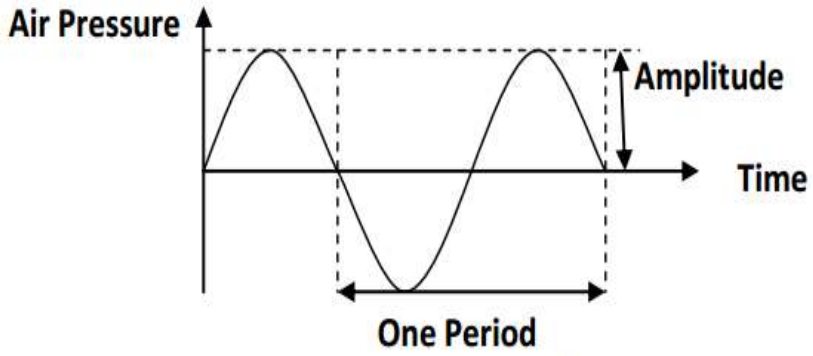


Figure: Oscillation of an air pressure wave

Frequency

The frequency of a sound is the reciprocal value of the period. It represents the number of periods in a second. It is measured in hertz (Hz) or cycles per second (cps).
1 KHz = 1000 Hz

Some of the frequency ranges are:

- ✓ Infra sound – 0 - 20 Hz
- ✓ Human audible sound – 20 Hz - 20KHz
- ✓ Ultra sound – 20KHz - 1GHz
- ✓ Hyper sound – 1GHz - 10THz

Human audible sound is also called audio or acoustic signals (waves). Speech is an acoustic signal produced by the humans.

Amplitude

The amplitude of the sound is the measure of the displacement of the air pressure wave from its mean or quiescent state.

Computer representation of sound

The smooth, continuous curve of a wave form is not directly represented in a computer. A computer measures the amplitude of the waveform at regular time intervals to produce a series of numbers called samples.

Audio signals are converted into digital samples through Analog-to-Digital Converter (ADC). The reverse mechanism is performed by a Digital-to Analog Converter (DAC). E.g. of ADC is AM79C30A digital subscriber controller chip.

Sampling Rate

The rate at which a continuous waveform is sampled is called the sampling rate. It is measured in Hz.

Nyquist Sampling Theorem:

"For lossless digitization, the sampling rate should be at least twice the maximum frequency responses".

E.g. CD standard sampling rate of 44100Hz means that the waveform is sampled 44100 times per second. A sampling rate of 44100 Hz can only represent frequencies up to 22050 Hz.

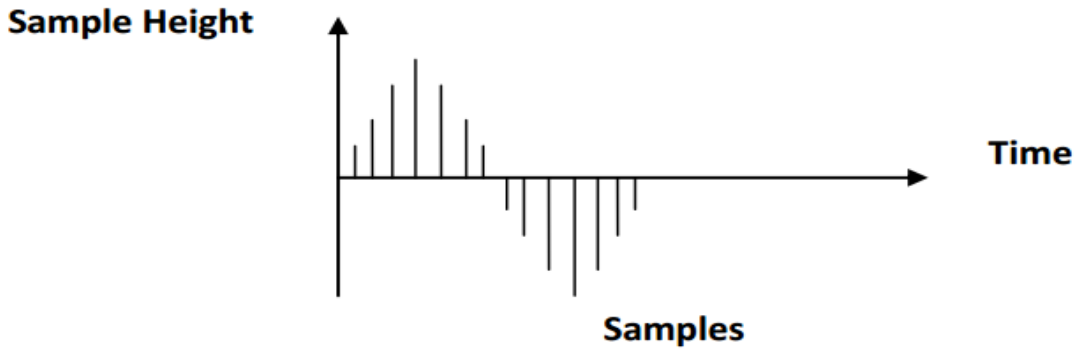


Figure: Sampled Waveform

Quantization:

The value of sample is discrete. Resolution/Quantization of a sample value depends on the no. of bits used in measuring the height of a waveform. For E.g. An 8-bit quantization yields 256 values. 16-bit CD quality quantization results over 65536 values.

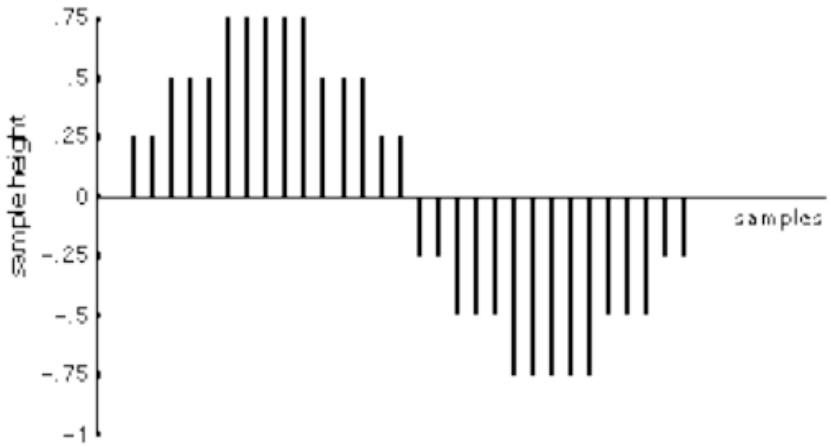


Figure: Three-bit Quantization

Lower the quantization; lower the quality of the sound. For 3-bit quantization, values are 8. i.e. 0.75, 0.5, 0.25, 0, -0.25, -0.5, -0.75 & 1

As you can see, the shape of the waveform becomes less discernible with a coarser quantization. The coarser the quantization, the "buzzier" the sound.

Sound Hardware
Some of the hardware regarding to sound are microphone jacks, built-in speakers, Headsets etc.

Music

The relationship between music and computers has been more and more important especially considering the development of MIDI (Musical Instrument Digital Interface). The MIDI interface between electronic musical instruments and computers is a small piece of equipment that plugs directly into the computers serial port and allows the transmission of musical signal.

MIDI Basic Concepts
Musical Instrument Digital Interface is a standard that manufacturers of electronic musical instrument have agreed upon. It is set of specifications used for building the instrument so that the instrument of one manufacturer can without difficulty communicate musical information between one another.

A MIDI interface has two different components:

The hardware that connects the equipment. It specifies that the physical connection between musical instruments, stipulates that a MIDI port is built into on instrument, specifies a MIDI cable and deals with electronic signals that are sent over the cable.

A data format encodes the information travelling though the hardware. The MIDI data format is digital i.e. data are grouped into MIDI messages. Each MIDI message communicates one musical event between machines.

If the musical instrument satisfies both components of a MIDI standard, the instrument is a ***MIDI device*** (E.g. a synthesizer). MIDI device is capable of communicating with other MIDI devices through channels.

MIDI standard specifies 16 channels and identifies 128 instruments. E.g.
0 - Acoustic grand piano
12 - Marimba
40 – Violin
73 – Flute

Some instruments allow only one note to be played at a time such as *flute.* Other instruments allow more than 1 note to be played at a time such as *organ.*

MIDI Reception Mode:
MIDI reception mode is used for tuning the MIDI devices to one/more channel. There are 4 modes:

- ✓ Mode 1 (Omni On/Poly)
- ✓ Mode 2 (Omni On/Mono)
- ✓ Mode 3 (Omni Off/Poly)
- ✓ Mode 4 (Omni Off/Mono)

Omni On/Off:

If Omni if turned on, the MIDI device monitors all the MIDI channels and responds to all channels messages. If it is turned off, the MIDI device responds only to channel messages sent on the channels the device is set to receive.

Omni Poly/Mono:
Of Poly is set, the device can play several notes at a time. If the mono is set, the device plays notes like monophonic synthesizer-one note at a time.

MIDI Devices
MIDI synthesizer device is the heart of MIDI system. Most synthesizers have following components.

Sound generators:
It synthesizes the sound. It produces an audio signal that becomes sound when fed into a loud speaker. It can change quality of sound by varying the voltage oscillation of the audio. Sound generation is done in 2-ways:

- Storing acoustic signals as MIDI data in advance
- Creating acoustic signals synthetically

Microprocessor:
Microprocessor communicates with the keyboard to know which notes the musician is playing. Microprocessor communicates with the control panel to know what commands the musician wants to sent to the microprocessor. The microprocessor then specifies note and sound commands to the sound generators (i.e. microprocessor sends and receives the MIDI message).

Keyboard:
It affords the musician's direct control of the synthesizer. Pressing keys means signaling microprocessor what notes to play and how long to play them. Keyboard should have at least 5 octaves and 61 keys.

Control panel:
Controls those function that are not directly concerned with notes and duration. Control panel includes a slider, a button and a menu.

Auxiliary controllers:
Gives more control over the notes played on keyboard. Pitch bend and modulation are the 2 common variables on the synthesizer

Memory:
Stores patches for the sound generation and settings on the control panel.

Drum machine:
Specialize in percussion sounds and rhythms.

Master keyboard:
Increases the quality of the synthesizer keyboard, *Guitar Synthesizer, Drum pad controllers, Guitar controllers* and many more

Sequencer:
Sequencer is the important MIDI device. It is used as storage server for generated MIDI data. It is also used as music editor. Musical data are represented in musical notes. Sequencer transforms the notes into MIDI message.

MIDI messages:
MIDI messages transmit information between MIDI devices and determine what kinds of musical events can be passed from device to device.

Formats of MIDI messages

- ✓ *Status byte:* First byte of any MIDI message. It describes the kind of message
- ✓ *Data byte:* The following bytes.

There are two types of MIDI messages

- ✓ Channel message
- ✓ System message

Channel Message

Since, channel message are specified, the channel messages go only to specified devices. There are 2 types of channel messages:

Channel voice messages: Sends actual performance data between MIDI devices, describing keyboard action, controller action and control panel changes. E.g. note on, Note off, channel pressure, control change etc.

Channel mode messages: Determine the way that a receiving MIDI device responds to channel voice messages. E.g. local control, All note off, Omni mode off etc.

System Message:

System messages go to all devices in a MIDI system because no channel numbers are specified. There are three types of system messages:

System real time messages: These messages are short and simple (one byte). It synchronizes the timing of MIDI devices in performance. To avoid delay, they are sent in the middle of other messages. E.g. System reset, Timing clock i.e. MIDI clock etc.

System common messages: Commands that prepare sequencer and synthesizer to play a song. E.g. song select, tune request etc.

System exclusive messages: Allows MIDI manufacturers to create customized MIDI messages to send between their MIDI devices.

MIDI Software

The software applications generally fall into 4 major categories:

1. *Music recording and performance applications:* Provides function as recording of MIDI messages. Editing and playing the messages in performance.
2. *Musical notations and printing applications:* Allows writing music using traditional musical notation. User can play and print music on paper for live performance or publication.
3. *Synthesizer path editor and librarians:* Allows information storage of different synthesizer patches in the computer's memory and disk drives. Editing of patches in computer.
4. *Music education applications:* Teaches different aspects of music using the computer monitor, keyboard and other controllers of attached MIDI instruments.

Processing chain of interactive computer music systems

- ✓ *Sensing stage:* Data is collected from controllers reading the gesture information from human performers on stage.
- ✓ *Processing stage:* Computer reads and interprets information coming from the sensors and prepares data for the response stage.
- ✓ *Response stage:* Computer ad some collection of sound producing devices share in realizing a musical output.

Some of the interactive music systems:

- ✓ Max (OOP language)
- ✓ Cypher (It has a listener and a player)
- ✓ NeXT Computer (It has a music kit)
- ✓ M & Jam Factory (It has graphics control panel)

Speech

Speech can be 'perceived', 'understood', and 'generated' by humans and by machines too.
Human speech signal comprises a subjective lowest spectral component known as the pitch.
Pitch is not proportional to the frequency. Human ear is sensitive in range from 600Hz-6000Hz.

Properties of Speech Signals:

- ✓ Voiced speech signals show periodic behavior at certain time intervals. So, these signals are considered as quasi-stationary signals for around 30 ms.
- ✓ Spectrum of audio signals shows characteristic maxima, which are mostly 3-5 frequency bands. These maxima i.e. formants occurs due to resonances of the vocal tract.

Speech Generation:

Mid 19th century, Helmholtz built a mechanical vocal tract coupling together several mechanical resonators with which sound could be generated. In 1940, Dudley produced the 1st speech synthesizer through imitation of mechanical vibration using electrical oscillation.
Requirement for speech generation is real time signal generation.

- ✓ Speech output system transforms the text into speech automatically.
- ✓ Not lengthy preprocessing is required.

Generated speech must be understandable and must sound natural.

Basic Notions:

Lowest periodic spectral component of the speech signal is called the ***fundamental frequency.*** It is present in the voiced sound.

A ***phone*** is the smallest speech unit, such as the m of mat and the b of bat in English that distinguishes one utterance or word from another in a given language.

Allophones mark the variants of a phone. For e.g. the aspirated p of pit and the un-aspirated p of spit are allophones of the English phoneme p.

The ***morph*** marks the smallest speech unit which carries a meaning itself. Therefore, consider is a morph, but reconsideration is not.

The ***voiced sound*** is generated through the vocal cords, m, v, l are the examples of voiced sounds.

During the generation of an ***unvoiced sound*** the vocal cords are opened. F and s are unvoiced sounds.

Vowels

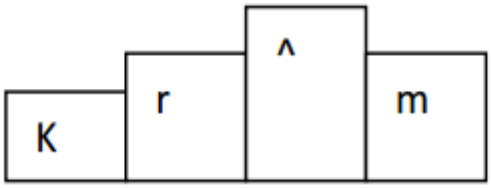
A speech sound created by the relatively free passage of breath through the larynx and oral cavity, usually forming the most prominent and central sound of syllable. E.g. a, e, i, o, u.

Consonants
A speech sound produced by a partial or complete obstruction of the air stream by any of the various constrictions of the speech organs. E.g. b, c, d, f, g, h, j etc.

Voiced consonants (e.g. m)
Fricative voiced consonants (e.g. v)
Fricative voiceless consonants (e.g. s)
Plosive consonants (e.g. d)
Affricate consonants (e.g. dg, ch)

Reproduced Speech Output
The easiest method of speech generation output is to use prerecorded speech and play it back in timely fashion. Speech can be stored as PCM samples.

Time Dependent Sound Concatenation
Individual speech units are composed like building blocks, where the composition can occur at different levels. Individual phones are understood as speech units. Example, crumb phones word phone are shown individually as -



Two phones can construct a *di-phone.* To make transition problem easier, *syllables* can be created. Speech is generated through the set of syllables.

Transition between individual sound units create and essential problem, called *co-articulation,* which is the mutual sound influence throughout several sounds.

Prosody should be considered during speech generation or output. Prosody means the stress and melody course.

Frequency Dependent Sound Concatenation
Speech generation/output can also be based on a frequency dependent sound concatenation. This can be done through a *formant synthesis.* Formants are frequency maxima in the spectrum of the speech signal. Formants synthesis simulates the vocal tract through a filter. A pulse signal with a frequency, corresponding to the fundamental speech frequency, is chosen as a simulation for voiced sounds. Unvoiced sounds are created through a noise generator.

Human speech can be generated using a multi-pole lattice filter. The 1st four or five formants, occurring in human speech are modeled correctly with this filter type. Unvoiced sounds, created by vocal chords are simulated through a noise and tone generator. The method used for the sound synthesis in order to simulate human speech is called the linear predictive coding (LPC) method. Using speech synthesis, an existent text can be transformed into an acoustic signal.

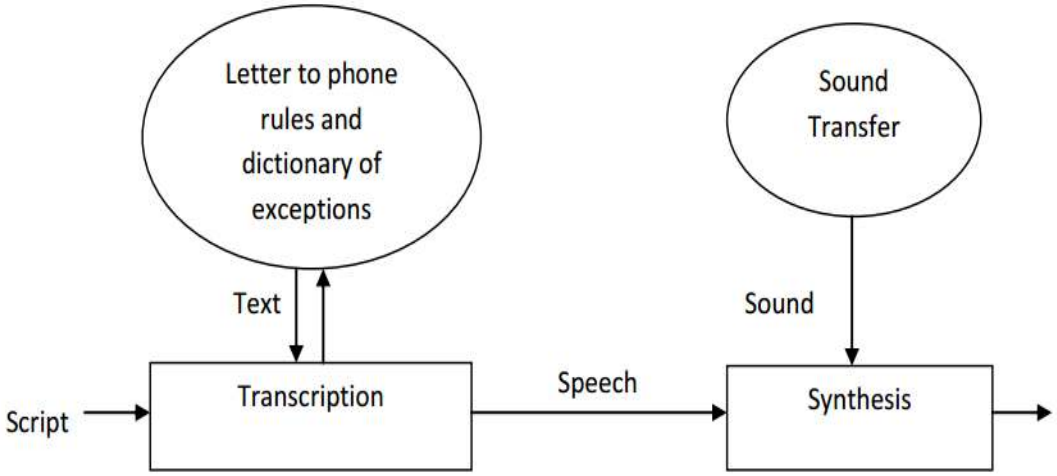


Figure: Components of a speech synthesis system with time dependent sound concatenation

Step 1:

- ✓ Performs transcription
- ✓ Text is translated into sound script
- ✓ This process is done using letter-to-phone rules and dictionary of exceptions
- ✓ User recognizes the formula deficiency in the transcription and improves the pronunciation manual

Step 2:

- ✓ Sound script is translated into a speech signal.
- ✓ Time or frequency dependent concatenation can follow

Speech Analysis

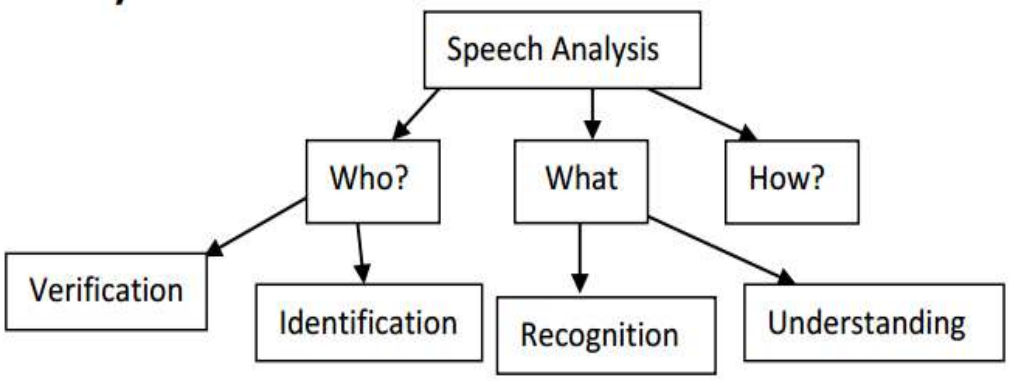


Figure: Research Areas of speech analysis

Speech analysis can serve to analyze who is speaking i.e. to recognize a speaker for his identification and verification. The computer identifies and verifies the speaker using an acoustic fingerprint. An acoustic fingerprint is a digitally stored speech probe of a person.
Speech analysis what has been said i.e. to recognize and understand the speech signal itself. Based on speech sequence, the corresponding text is generated (e.g. speech-controlled typewriter)

Speech analysis also tries to research speech patterns with respect to how a certain statement was said. E.g. a spoken sentence sounds differently if a person is angry or calm which can be used for lie detector.

References:

- ✓ Multimedia: Computing, Communications and Applications", Ralf Steinmetz and Klara Nahrstedt, Pearson Education Asia
- ✓ "Multimedia Communications, Applications, Networks, protocols ad Standards", Fred Halsall, Pearson Education Asia
- ✓ "Multimedia Systems", John F. Koegel Buford, Pearson Education Asia

Assignments:

(1)Explain the different components of a MIDI device.
(2)Illustrate the importance of MIDI. Explain the significance of MIDI messages.
(3)What is MIDI? What features of MIDI make it suitable for multimedia applications?
(4)How can speech be generated from a digital device? Explain in detail.

Gentle Advice:

Please go through your text books and reference books for detail study!!! Thank you all.