**Assignment 2**

**1. Briefly explain MIDI devices. Explain the term phone, diphones, syllabus, voiced sound and unvoiced sound.**

MIDI stands for Musical Instrument and Digital Interface. MIDI is defined as a technical standard that describes **protocols,** digital **interfaces, and** connectors **that allow various** electronic musical instruments, **computers,** and other related devices to connect and communicate with **each other.** There are many types of MIDI devices. They play different **roles** in making music. MIDI devices are **described** in **detail below.**

1. **Synthesizer/Sampler**: - Well, sampler is a type of sound generator (various pitch, loudness, tone colour). It can use a variety of synthesis or Sample-based synthesis to make sound. A good (musician’s) synthesizer often has a microprocessor, keyboard, control panels, memory, etc. Synthesizers may employ any of a variety of sound generation techniques. They may include an integrated keyboard, or may exist as "sound modules" or "expanders" that generate sounds when triggered by an external controller, such as a MIDI keyboard. A sampler can record and digitize audio, store it in random-access memory (RAM), and play it back. Samplers typically allow a user to edit a sample and save it to a hard disk, apply effects to it, and shape it with the same tools that synthesizers use. For our purposes we define a synthesizer as the tone generation unit. It has one or more MIDI INs and MIDI OUTs and USB/Firewire connectivity.
2. **Sequencer: -** Sequencers is a tool you can compose, record and edit the notes like DAW software (e.g., Logic Pro, Reaper). It is a stand-alone unit or a software program for a personal computer. It used to be a storage server for MIDI data. Furthermore, it is a common in software music editor on the computer. Sequencing software allows recorded MIDI data to be manipulated using standard computer editing features such as cut, copy and paste and drag and drop.
3. **Computer: -** It is the heart of a MIDI system. It helps to control the scheduling, synchronization and recording of all data. Sequencer usually software based and now part of larger applications that control all aspects of Audio and Midi — Digital Audio Workstation packages such as Cubase, Logic, Sonar, Live, and Reason. Nowadays, it also includes many software synthesizers/samplers to make sounds in real time. Computer is one of the important MIDI devices.
4. **Interfaces: -** Midi devices still need to connect to computer with some interface. Each MIDI hardware has an interface like MIDI In and MIDI Out. It’s also very common to have USB port. Some of devices even have more than 8 MIDI in and MIDI out ports to handle a complicate setup. Audio interface via USB or firewire and even wireless keyboard are the interfaces.
5. **Midi Control Input devices: -** Keyboards are by far the most common type of MIDI controller. MIDI was designed with keyboards in mind, and any controller that is not a keyboard is considered an "alternative" controller. This was seen as a limitation by composers who were not interested in keyboard-based music, but the standard proved flexible, and MIDI compatibility was introduced to other types of controllers, including guitars, stringed and wind instruments, drums and specialized and experimental controllers. Other controllers include drum controllers and wind controllers. It can just be a bunch of controllers.
6. **Midi Control Output devices: -** Not just making sounds ,Midi even controls other things like lighting, robotics, video systems, even hamster control as well as lots of other applications.

A **phone** is any distinct speech sound or gesture, regardless of whether the exact sound is critical to the meanings of words.

A **diphone** is an adjacent pair of [phones](https://en.wikipedia.org/wiki/Phone_(phonetics)) in an utterance. For example, in [daɪfəʊn], the diphones are [da], [aɪ], [ɪf], [fə], [əʊ], and [ʊn].

A **syllable** is a unit of organization for a sequence of [speech sounds](https://en.wikipedia.org/wiki/Phone_(phonetics)).

**Voiced sounds**, are sounds which are articulated with participation of the vocal cords. For Example: b.d.g.

**Unvoiced consonants** are consonant sounds that are made without vibrating the vocal cords. Example  
/p/ as in 'pet' /t/ as in 'top' /k/ as in 'cat'

**2. Write short notes on**

**a) Speech generation or speech synthesis:**

Speech synthesis is the computer-generated simulation of human speech. It is used to translate written information into aural information where it is more convenient, especially for mobile applications such as voice-enabled e-mail and Unified messaging. It is also used to assist the vision-impaired so that, for example, the contents of a display screen can be automatically read aloud to a blind user. Speech synthesis is the counterpart of speech or [voice recognition](https://searchcustomerexperience.techtarget.com/definition/voice-recognition-speaker-recognition) . The earliest speech synthesis effort was in 1779 when Russian Professor Christian Katzenstein created an apparatus based on the human vocal tract to demonstrate the physiological differences involved in the production of five long vowel sounds. The first fully functional voice synthesizer, Homer Dudley's VODER (Voice Operating Demonstrator), was shown at the 1939 World's Fair. The VODER was based on Bell Laboratories' vocoder (voice coder) research of the mid-thirties.

Speech generation and recognition are used to communicate between humans and machines. Rather than using your hands and eyes, you use your mouth and ears. This is very convenient when your hands and eyes should be doing something else, such as: driving a car, performing surgery, or (unfortunately) firing your weapons at the enemy. Two approaches are used for computer generated speech: digital recording and vocal tract simulation. In digital recording, the voice of a human speaker is digitized and stored, usually in a compressed form. During playback, the stored data are uncompressed and converted back into an analog signal. An entire hour of recorded speech requires only about three megabytes of storage, well within the capabilities of even small computer systems. This is the most common method of digital speech generation used today. Vocal tract simulators are more complicated, trying to mimic the physical mechanisms by which humans create speech. The human vocal tract is an acoustic cavity with resonate frequencies determined by the size and shape of the chambers. Sound originates in the vocal tract in one of two basic ways, called voiced and fricative sounds. With voiced sounds, vocal cord vibration produces near periodic pulses of air into the vocal cavities. In comparison, fricative sounds originate from the noisy air turbulence at narrow constrictions, such as the teeth and lips. Vocal tract simulators operate by generating digital signals that resemble these two types of excitation. The characteristics of the resonate chamber are simulated by passing the excitation signal through a digital filter with similar resonances. This approach was used in one of the very early DSP success stories, the Speak & Spell, a widely sold electronic learning aid for children.

**b) LDU:**

LDU stands for logical data unit. It is defined as digital signature is used by the data storage system to determine whether or not to actually store each logical data unit. It is a unit of a media stream. The logical data unit (LDU) on the other hand refers to an information unit, which may not be the same as that of the presentation unit. How we define the information unit depends upon the level of granularity. Each frame consists of macro blocks of size 16 x 16 pixels. Each macro block may compose an LDU. If we increase the level of granularity further, each pixel may also form an LDU. Each LDU may not contain the same amount of information. For example, if we consider the compressed frames of a video sequence as LDUs, each compressed frames may require different number of bits to encode. LDUs may also classify into closed and open LDUs. Closed LDUs are the one having predictable duration, like the stored continuous media, e.g. audio and video. Open LDUs typically represent input from a live source. Some typically examples of LDU based synchronization are shown below: Lip synchronization, which requires tight coupling between audio and video.

**3. Explain how to represent sound in computer.**

Sound needs to be converted into binary for computers to be able to process it. To do this, sound is captured - usually by a microphone - and then converted into a digital signal.

An analogue to digital converter will sample a sound wave at regular time intervals. The samples can then be converted to binary. They will be recorded to the nearest whole number. If the time samples are then plotted back onto the same graph, it can be seen that the sound wave now looks different. This is because sampling does not take into account what the sound wave is doing in between each time sample. This means that the sound loses quality as data has been lost between the time samples. The way to increase the quality and store the sound at a quality closer to the original, is to have more time samples that are closer together. This way, more detail about the sound can be collected, so when it’s converted to digital and back to analogue again it does not lose as much quality.

The computer representation of sound are described below in detail: -

I. Sampling: -Sound wave form the smooth, continuous is not directly represented in the computer. The computer measures the amplitude of the 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. Measuring the analog signal at regular discrete interval. Recording the value at this point. The rate at which a continuous wave form is sampled is called sampling rate. Like frequency, sampling rate are measured in Hz.

ii. Quantization: -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 quality of sound, higher quantization higher quality of sound.

**4. What do you mean by MIDI message? Explain its types.**

MIDI messages are used by MIDI devices to communicate with each other. MIDI message includes a status byte and up to two data bytes. In status byte:

* The most significant bit of status byte is set to 1.
* The 4 low-order bits identify which channel it belongs to (four bits produce 16 possible channels).
* The 3 remaining bits identify the message.
* The most significant bit of data byte is set to 0.

There are two types of MIDI message:

**Channel Message:**

It goes only to specified devices. There are two types of channel message.

⚫Channel Voice Message: Channel Voice message send actual performance data between MIDI devices describing keyboard action, controller action and control panel changes. They describe music by defining pitch, note on, note off channel pressure, etc.

⚫Channel Mode Message: Channel Mode message determine the way that a receiving MIDI device respond to channel voice message. It deals with how to play notes coming in over MIDI cables. ◦Channel mode message includes Omni On, Omni Off, note off, note on, etc.

**System message:** System message go to all devices in a MIDI system because no channel number are specified. Not Specific to channel.

5. Discuss briefly on speech generation technique.

Speech is defined as the expression of or the ability to express thoughts and feelings by articulate sounds. It also refers to a form of communication in spoken language, made by a speaker before an audience for a given purpose. The speech generation technique is explained below in detail: -

1. Signal Form Coding: - Here, we do not considers speech-specific properties and parameters Data rate for PCM –coded stereo audio signal with CD-quality: 2\* 44100/s\*16bit/8bit/byte = 176,400 bytes/s=1411200 bits/s Telephone: 64Kbits/s. Using DPCM data rate lowered to 56Kbits/s. Using ADPCM it can be further reduced to 32Kbit/s.
2. Source Coding: -Parameterized system use source coding. Speech characteristics are used for data rate reduction. E.g., Channel Vocoder.
3. Quality: -Dependence of the achieved quality of compressed speech on the data rate. Telephone: 8 Kbit/s sufficient.

**6. Describe on different reception mode and message of MIDI.**

One of several ways in which a device can respond to incoming MIDI information. There are two parts to each mode, one defining whether it is monophonic or polyphonic, and the other determining if it is multitimbral or not. Four modes are included in the MIDI spec, and two others, Multi-Mode and Mono Mode (for MIDI guitar) were developed later.

1. Omni On/Poly – Device responds to MIDI data regardless of channel, and is polyphonic.
2. Omni On/Mono – Device responds to MIDI data regardless of channel, and is monophonic. This mode is rarely, if ever, used.
3. Omni Off/Poly – Device responds to MIDI data only on one particular channel, and is polyphonic. This is the normal mode for most keyboards that are not functioning multitimbrally.
4. Omni Off/Mono – Device responds to MIDI data only on one particular channel, and is monophonic.

The different types of recent mode of MIDI emphasizes the following things: -

The PSR-410 allows any of five reception modes to be individually assigned to MIDI channels 1 through 16.

The five modes are:

1. **Mode “0”**

**RX OFF:** Reception disabled.

1. **Mode “1”**

**MULTI:** Received MIDI note data directly controls the PSR-410 tone generator. Different voices can be played on different channels.

1. **Mode “2”**

**REMOTE:** Received MIDI note data is handled in the same way as data from the  
PSR-410’s own keyboard.

**d. Mode “3”**

**e. CHORD:** Received MIDI note data is interpreted as Auto Accompaniment chord  
commands.

**f. Mode “4”**

**ROOT:** Received MIDI note data is interpreted as Auto Accompaniment bass note  
command