

Digital Audio and MIDI

6.1 Speech Signals

Speech can be processed by humans or machines, although it is the dominant form of communication of human beings. The field of study and handling the digitized speech is called **digital speech processing**.

Speech is based on spoken languages, which means that it has a **semantic content**.

Speech understanding means the efficient adaptation to speakers and their speaking habits. Speech signals have two important characteristics that can be used by speech processing applications:

1. Voiced speech signals (in contrast to unvoiced sounds) have an almost **periodic structure** over a certain time interval, so that these signals remain quasi-stationary for about 30ms.
2. The spectrum of some sounds have characteristic maxima that normally involve **up to five frequencies**. These frequency maxima, generated when speaking, are called formants.

6.2 Speech Transmission

Speech transmission is a field relating to highly efficient encoding of speech signals to enable low-rate data transmission, while minimizing noticeable quality losses.

Signal form encoding does not consider speech-dependent properties or parameters.

The technologies applied are merely expected to offer efficient encoding of audio

signals. A straightforward technique for digitizing an analog signal (waveform) is **Pulse Code Modulation (PCM)**. This method is simple, but it still meets the high quality demands stereo-audio signals in the data rate used for CDs:

$$rate = 2 \times \frac{44,100}{s} \times \frac{16 \text{ bits}}{8 \text{ bits/byte}} = 176,400 \text{ bytes/s}$$

Differential Pulse Code Modulation (DPCM) achieves 56Kbit/s in at least equal quality, while Adaptive Pulse Code Modulation (ADPCM) enables a further reduction to 32Kbit/s.

6.3 Speech Synthesis

Computers can translate an encoded description of a message into speech. This scheme is called speech synthesis. A particular type of synthesis is **text-to-speech** conversion. Fair-quality text-to-speech software has been commercially available for various computers and workstations, although the speech produced in some lacks naturalness.

Speech recognition is normally achieved by drawing various comparisons. With the current technology, a speaker-dependent recognition of approximately 25,000 words is possible. The problems in speech recognition affecting the recognition quality include dialects, emotional pronunciations, and environmental noise. It will probably take some time before the considerable performance discrepancy between the human brain and a powerful computer will be bridged in order to improve speech recognition and speech generation.

6.4 Common Audio Formats

There are two basic types of audio files: the traditional **discrete audio** file that you can save to a hard drive or other digital storage medium, and the **streaming audio** file that you listen to as it downloads in real time from a network/internet server to your computer.

6.4.1 Discrete Audio File Formats

Common discrete audio file formats include **WAV**, **AIF**, **AU** and **MP3**. A fifth format, called **MIDI** is actually not a file format for storing digital audio, but a system of instructions for creating electronic music.

1. WAV

The **WAV** format is the **standard audio file format for Microsoft Windows** applications, and is the default file type produced when conducting digital recording within Windows. It supports a variety of bit resolutions, sample rates, and channels of audio. This format is very popular upon IBM PC (clone) platforms, and is widely used as a basic format for saving and modifying digital audio data.

2. AIF/AIFF

The **Audio Interchange File Format (AIFF)** is the **standard audio format** employed by computers using the **Apple Macintosh operating system**. Like the WAV format, it supports a variety of bit resolutions, sample rates, and channels of audio and is widely used in software programs used to create and modify digital audio.

3. AU

The **AU file format** is a compressed audio file format developed by Sun Microsystems and popular in the **UNIX** world. It is also the standard audio file format for the Java programming language. Only supports 8-bit depth thus cannot provide CD-quality sound.

4. MP3

MP3 stands for **Motion Picture Experts Group**, Audio Layer 3 Compression. MP3 files **provide near-CD-quality sound** but are only about 1/10th as large as a standard audio CD file. Because MP3 files are small, they can easily be transferred across the Internet and played on any multimedia computer with MP3 player software.

6.4.2 Streaming Audio File Formats

Streaming is a network technique for transferring data from a server to client in a format that can be continuously read and processed by the client computer. Using this method, the client computer can start playing the initial elements of large time-based audio or video files before the entire file is downloaded. As the Internet grows, streaming technologies are becoming an increasingly important way to deliver time based audio and video data.

For streaming to work, the client side has to receive the data and continuously 'feed' it to the 'player' application. If the client receives the data, more quickly than required, it has to temporarily store or 'buffer' the excess for later play.

On the other hand, if the data doesn't arrive quickly enough, the audio or video presentation will be interrupted.

There are three primary streaming formats that support audio files: Real Network's RealAudio (**RA, RM**), Microsoft's Advanced Streaming Format (**ASF**) and its audio subset called Windows Media Audio 7 (**WMA**) and Apple's QuickTime 4.0+ (**MOV**).

1. RA/RM

For audio data **on the Internet**, Real Network's RealAudio (RA) compressed streaming audio format is standard. These files require a RealPlayer program or browser plug-in. The latest versions of Real Networks' server and player software can handle multiple encodings of a single file, allowing the quality of transmission to vary with the available bandwidth. Webcast radio broadcast of both talk and music frequently uses RealAudio. Streaming audio can also be provided in conjunction with video as a combined Real Media (RM) file.

2. ASF

Microsoft's Advanced Streaming Format (ASF) is similar to designed to Real Network's Real Media format, in that it provides a common definition for internet streaming media and can accommodate not only synchronized audio, but also video and other multimedia elements, all while supporting multiple bandwidths within a single media file. Also like Real Network's Real Media format, Microsoft's ASF requires a program or browser plug-in.

The pure audio file format used in Windows Media Technologies is Windows Media Audio 7 (WMA files). Like MP3 files, WMA audio files use sophisticated audio compression to reduce file size. Unlike MP3 files, however, WMA files can function as either discrete or streaming data and can provide a security mechanism to prevent unauthorized use.

3. MOV

Apple QuickTime movies (MOV files) can be created without a video channel and used as a sound-only format. Since version 4.0, Quicklime provides true streaming capability.

QuickTime also accepts different audio sample rates, bit depths, and offers full functionality in both Windows as well as the Mac OS. Popular audio file formats are:

- au (*Unix*)
- aiff (*MAC*)
- wav (*PC*)
- mp3

4. MIDI

MIDI stands for Musical Instrument Digital Interface. MIDI is a protocol that enables computer, synthesizers, keyboards, and other musical device to

communicate with each other. This protocol is a language that allows interworking between instruments from different manufacturers by providing a link that is capable of transmitting and receiving digital data. MIDI transmits only commands; it does not transmit an audio signal.

6.5 Components of a MIDI System

1. Synthesizer

It is a sound generator (various pitch, loudness, tone color). A good (musician's) synthesizer often has a microprocessor, keyboard, control panels, memory, etc.

2. Sequencer

It can be a stand-alone unit or a software program for a personal computer. (It used to be a storage server for MIDI data. Nowadays it is more a software music editor on the computer.). It has one or more MIDI INs and MIDI OUTs

1.6 Basic MIDI Concepts

1. Track:

Track in sequencer is used to organize the recordings. Tracks can be turned on or off on recording or playing back.

2. Channel: MIDI channels are used to separate information in a MIDI system.

There are 16 MIDI channels in one cable. Channel numbers are coded into each MIDI message.

3. Timbre: The quality of the sound, e.g., flute sound, cello sound, etc.

Multitimbral – capable of playing many different sounds at the same time (e.g., piano, brass, drums, etc.)

4. Pitch: The Musical note that the instrument plays

5. Voice: Voice is the portion of the synthesizer that produces sound.

Synthesizers can have many (12, 20, 24, 36, etc.) voices. Each voice works

independently and simultaneously to produce sounds of Different timbre and pitch.

6. Patch: The control settings that define a particular timbre

6.7 Hardware Aspects of MIDI

6.7.1 MIDI connectors:

Three 5-pin ports found on the back of every MIDI unit:

- **MIDI IN:** the connector via which the device receives all MIDI data.
- **MIDI OUT:** the connector through which the device transmits all the MIDI data it generates itself.
- **MIDI THROUGH:** the connector by which the device echoes the data receives from MIDI IN.

6.7.2 MIDI Messages

MIDI messages are used by MIDI devices to communicate with each other. MIDI messages are very low bandwidth:

6.7.3 Advantages

Because MIDI is a digital signal, it's very easy to interface electronic instruments to computers, and then do manipulations on the MIDI data on the computer with software. For example, software can store MIDI messages to the computer's disk drive. Also, the software can playback MIDI messages upon all 16 channels with the same rhythms as the human who originally caused the instrument(s) to generate those messages.

6.8 How is MIDI file Different from a WAV or MP3 Files?

A MIDI file stores MIDI messages. These messages are commands that tell a musical device what to do in order to make music. For example, there is a MIDI message that tells a device to play a particular note. There is another MIDI

message that tells a device to change its current "sound" to a particular patch or instrument.

The MIDI file also stores timestamps, and other information that a sequencer needs to play some "musical performance" by transmitting all of the MIDI messages in the file to all MIDI devices. In other words, a MIDI file contains.