University of Technology

الجامعة التكنولوجية

Computer Science Department

قسم علوم الحاسوب

Digital Audio

الصوت الرقمي

Dr. Inam S. Naser

م.د. إنعام صكر ناصر

cs.uotechnology.edu.iq

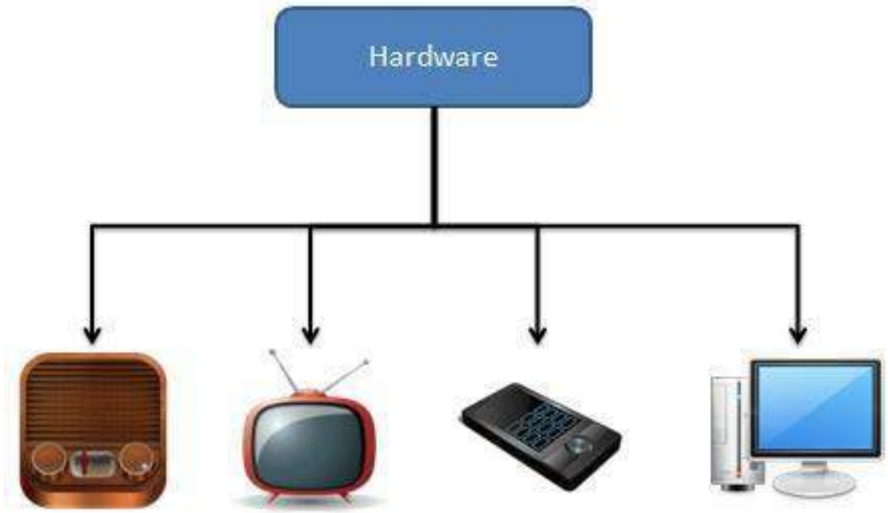
Introduction to Multimedia

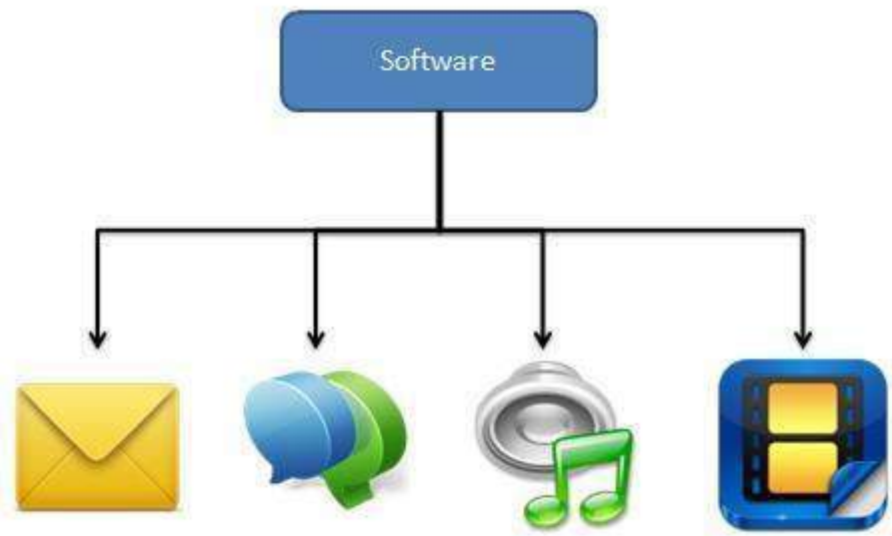
Multimedia is an interactive media and provides multiple ways to represent information to the user in a powerful manner. It provides an interaction between users and digital information. It is a medium of communication. Some of the sectors where multimedia is used extensively are education, training, reference material, business presentations, advertising and documentaries.

Definition of Multimedia

By definition Multimedia is a representation of information in an attractive and interactive manner with the use of a combination of text, audio, video, graphics and animation. In other words we can say that Multimedia is a computerized method of presenting information combining textual data, audio, visuals (video), graphics and animations. For examples: E-Mail, Yahoo Messenger, Video Conferencing, and Multimedia Message Service (MMS).

Multimedia as name suggests is the combination of Multi and Media that is many types of media (hardware/software) used for communication of information.


Hardware


Software

Components of Multimedia

Following are the common components of multimedia:

- **Text-** All multimedia productions contain some amount of text. The text can have various types of fonts and sizes to suit the profession presentation of the multimedia software.
- **Graphics-** Graphics makes the multimedia application attractive. In many cases people do not like reading large amount of textual matter on the screen. Therefore, graphics are used more often than text to explain a concept, present background information etc. There are two types of Graphics:
    - **Bitmap images-** Bitmap images are real images that can be captured from devices such as digital cameras or scanners. Generally bitmap images are not editable. Bitmap images require a large amount of memory.
    - **Vector Graphics-** Vector graphics are drawn on the computer and only require a small amount of memory. These graphics are editable.
- **Audio-** A multimedia application may require the use of speech, music and sound effects. These are called audio or sound element of multimedia. Speech is also a perfect way for teaching. Audio are of analog and digital types. Analog audio or sound refers to the original sound signal. Computer stores the sound in digital form. Therefore, the sound used in multimedia application is digital audio.
- **Video-** The term video refers to the moving picture, accompanied by sound such as a picture in television. Video element of multimedia application gives a lot of information in small duration of time. Digital video is useful in multimedia application for showing real life objects. Video have highest performance demand on the computer memory and

on the bandwidth if placed on the internet. Digital video files can be stored like any other files in the computer and the quality of the video can still be maintained. The digital video files can be transferred within a computer network. The digital video clips can be edited easily.

- **Animation-** Animation is a process of making a static image look like it is moving. An animation is just a continuous series of still images that are displayed in a sequence. The animation can be used effectively for attracting attention. Animation also makes a presentation light and attractive. Animation is very popular in multimedia application

Applications of Multimedia

Following are the common areas of applications of multimedia.

- **Multimedia in Business-** Multimedia can be used in many applications in a business. The multimedia technology along with communication technology has opened the door for information of global wok groups. Today the team members may be working anywhere and can work for various companies. Thus the work place will become global. The multimedia network should support the following facilities:
    - Voice Mail
    - Electronic Mail
    - Multimedia based FAX
    - Office Needs
    - Employee Training
    - Sales and Other types of Group Presentation
    - Records Management

- **Multimedia in Marketing and Advertising-** By using multimedia marketing of new products can be greatly enhanced. Multimedia boost communication on an affordable cost opened the way for the marketing and advertising personnel. Presentation that have flying banners, video transitions, animations, and sound effects are some of the elements used in composing a multimedia based advertisement to appeal to the consumer in a way never used before and promote the sale of the products.
- **Multimedia in Entertainment-** The uses of multimedia is more often used in the entertainment industry which is mainly used in creating movies, short films, 2D animations, and 3D animations. This modern multimedia is used in video games that need heavy graphics, sound, and videos for better performance and better viewing. The presence of multimedia in the games gets a real-life feeling when playing them, and creates more excitement and thriller. In some movies nowadays multimedia are used to its full potential to create a supernatural occurrence that should look like a natural one.
- **Multimedia in Education-** Many computer games with focus on education are now available. Consider an example of an educational game which plays various rhymes for kids. The child can paint the pictures, increase reduce size of various objects etc apart from just playing the rhymes. Several other multimedia packages are available in the market which provide a lot of detailed information and playing capabilities to kids.
- **Multimedia in Bank-** Bank is another public place where multimedia is finding more and more application in recent times. People go to bank to open saving/current accounts, deposit funds, withdraw money, know various financial schemes of the bank, obtain loans etc. Every bank has a lot of information which it wants to impart to in customers. For this purpose, it can use multimedia in many ways. Bank also displays information about its various schemes on a PC monitor placed in the rest area for customers. Today

on-line and internet banking have become very popular. These use multimedia extensively. Multimedia is thus helping banks give service to their customers and also in educating them about banks attractive finance schemes.

- **Multimedia in Hospital-** Multimedia best use in hospitals is for real time monitoring of conditions of patients in critical illness or accident. The conditions are displayed continuously on a computer screen and can alert the doctor/nurse on duty if any changes are observed on the screen. Multimedia makes it possible to consult a surgeon or an expert who can watch an ongoing surgery line on his PC monitor and give online advice at any crucial juncture.

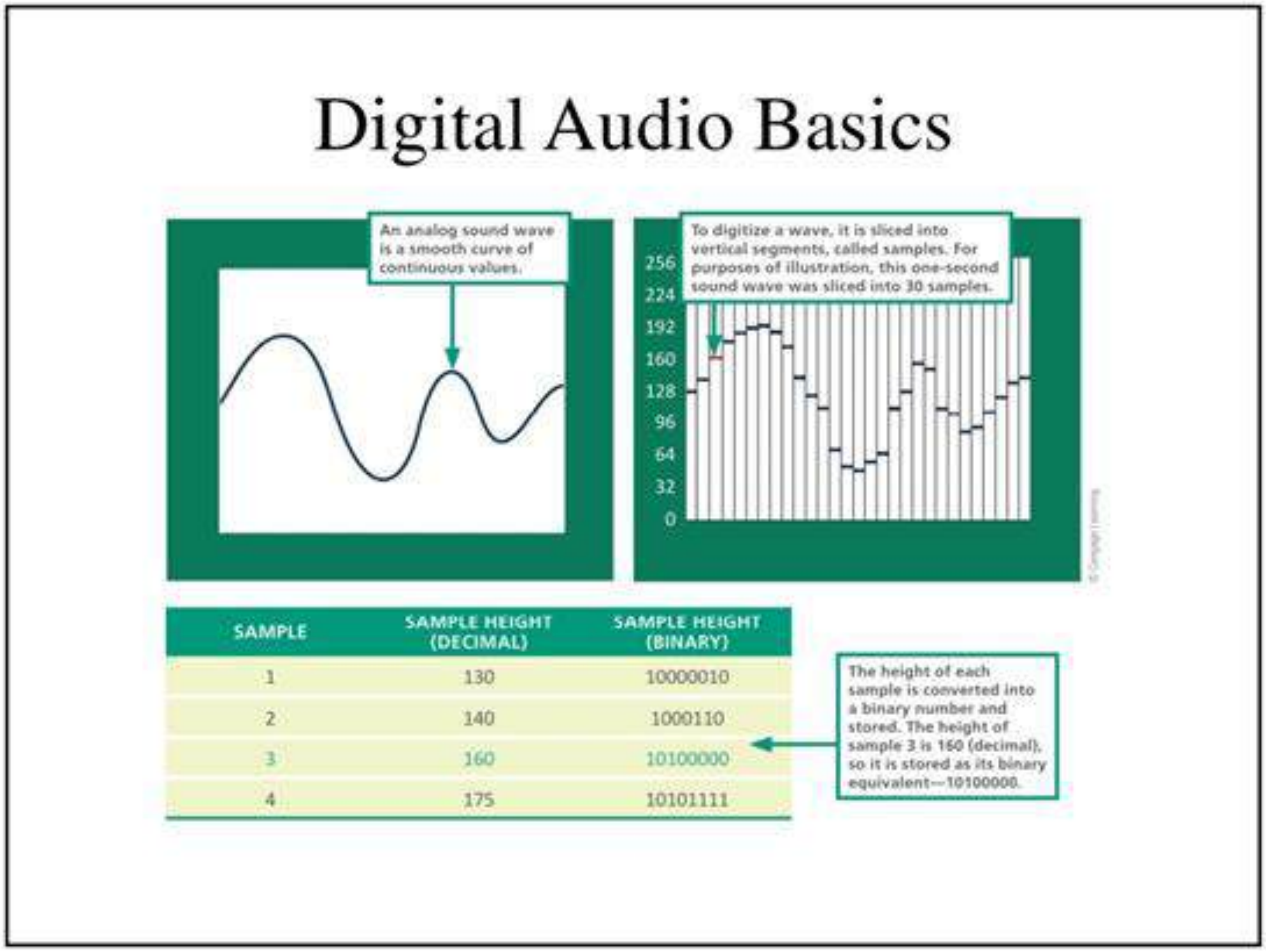
  In hospitals multimedia can also be used to diagnose an illness with CD-ROMs/ Cassettes/ DVDs full of multimedia based information about various diseases and their treatment. Some hospitals extensively use multimedia presentations in training their junior staff of doctors and nurses. Multimedia displays are now extensively used during critical surgeries.

- **Communication Technology and Multimedia Services-** The advancement of high computing abilities, communication ways and relevant standards has started the beginning of an era where you will be provided with multimedia facilities at home. These services may include:
  - Basic Television Services
  - Interactive entertainment
  - Digital Audio
  - Video on demand
  - Home shopping

- Financial Transactions
- Interactive multiplayer or single player games
- Digital multimedia libraries
- E-Newspapers, e-magazines

Basics of digital audio

- ***Digital audio*** is music, speech, and other sounds represented in binary format for use in digital devices.
- Most digital devices have a built-in microphone and audio software, so recording external sounds is easy.
- To digitally record sound, samples of a sound wave are collected at periodic intervals and stored as numeric data in an audio file.
- Sound waves are sampled many times per second by an ***analog-to-digital converter***.
- A ***digital-to-analog converter*** transforms the digital bits into analog sound waves.
- ***Sampling rate*** refers to the number of times per second that a sound is measured during the recording process.
- Higher sampling rates increase the quality of the recording but require more storage space.



Digital Audio File Formats

- A digital file can be identified by its type or its file extension, such as Thriller.mp3 (an audio file).
- The most popular ***digital audio formats*** are: **AAC, MP3, Ogg, Vorbis, WAV, FLAC,** and **WMA.**

AUDIO FORMAT	EXTENSION	ADVANTAGES	DISADVANTAGES
AAC (Advanced Audio Coding)	.aac, .m4p, or .mp4	Very good sound quality based on MPEG-4; lossy compression; used for iTunes music	Files can be copy protected so that use is limited to approved devices
MP3 (also called MPEG-1 Layer 3)	.mp3	Good sound quality; lossy compression; can be streamed over the Web	Might require a standalone player or browser plugin
Ogg Vorbis	.Ogg	Free, open standard; lossy compression; supported by some browsers	Slow to catch on as a popular standard; part of Google's WebM format
WAV	.Wav	Good sound quality; supported in browsers without a plugin	Audio data is stored in raw, noncompressed format, so files are very large
FLAC (Free Lossless Audio Compression)	.flac	Excellent sound quality; lossless compression	Open source format support ported by many devices
WMA (Windows Media Audio)	.wma	Lossy or lossless compression; very good sound quality; used on several music download sites	Files can be copy protected; requires an add-on player for some devices

- To play a digital audio file, you must use some type of audio software, such as:

- **Audio players:** small standalone software application or mobile app.
- **Audio plugins:** software that works in conjunction with your computer's browser to manage and play audio from a Web page.
- **Audio software:** general-purpose software and apps used for recording, playing, and modifying audio files, such as iTunes, Windows Media Player, and Audacity.

Digitized Speech

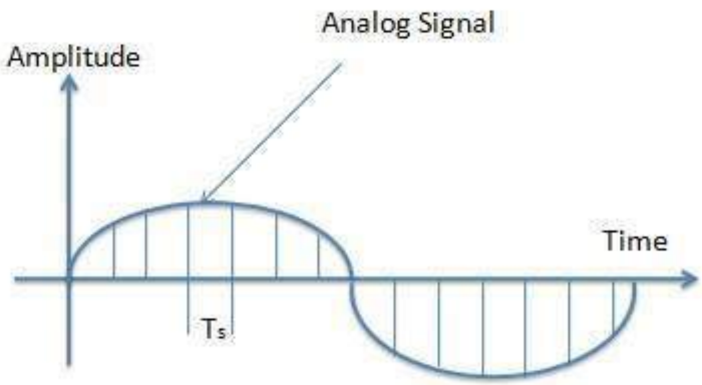
- ***Speech synthesis*** is the process by which machines produce sound that resembles spoken words. It is the technique of generating sound, using electronic hardware or software, from scratch.
- ***Speech recognition*** (or voice recognition) refers to the ability of a machine to understand spoken words.
- Speech recognition software analyzes the sounds of your voice and converts each word into groups of phonemes (basic sound units).
- The software then compares the groups to the words in a digital dictionary to find a match.
- When a match is found, the software can display the word on the screen or use it to carry out a command.

Digitization of Sound

Digitization is a process of converting the analog signals to a digital signal by a method called

PCM (Pulse code modulation) independent of the complexity of the analog waveform. All types of analog data like video, voice; music etc. can be transferred using PCM. It is the standard form for digital audio in computers. It is used by Blu-ray, DVD and Compact Disc formats and other systems such as digital telephone systems.

There are three steps of digitization of sound.



- **Sampling** - Sampling is a process of measuring air pressure amplitude at equally spaced moments in time, where each measurement constitutes a sample.
    - ✓ A sampling rate is the number of times the analog sound is taken per second.
    - ✓ A higher sampling rate implies that more samples are taken during the given time interval and ultimately, the quality of reconstruction is better.
    - ✓ The sampling rate is measured in terms of Hertz, Hz in short, which is the term for Cycle per second.

    Suppose analog signal is sampled every Ts seconds. Ts are referred to as the sampling interval.

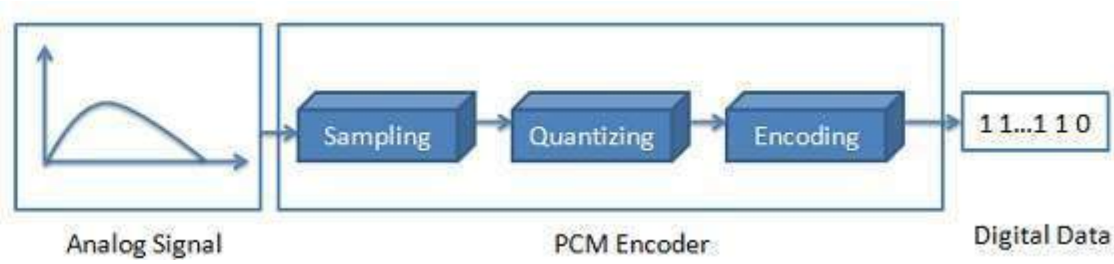
    $$Fs=1/Ts$$

    It is called the sampling rate or sampling frequency.

- **Quantization** - Quantization is a process of representing the amplitude of each sample as integers or numbers. How many numbers are used to represent the value of each sample known as sample size or bit depth or resolution.

- ✓ Commonly used sample sizes are either 8 bits or 16 bits.
- ✓ The larger the sample size, the more accurately the data will describe the recorded sound.
- ✓ An 8-bit sample size provides 256 equal measurement units to describe the level and frequency of the sound in that slice of time.
- ✓ A 16-bit sample size provides 65,536 equal units to describe the sound in that sample slice of time.
- ✓ The value of each sample is rounded off to the nearest integer (quantization) and if the amplitude is greater than the intervals available, clipping of the top and bottom of the wave occurs.

- **Encoding -** Encoding converts the integer base-10 number to a base-2 that is a binary number. The output is a binary expression in which each bit is either a 1(pulse) or a 0(no pulse).



Quantization of Audio

Quantization is a process to assign a discrete value from a range of possible values to each sample. Number of samples or ranges of values are dependent on the number of bits used to represent each sample. Quantization results in stepped waveform resembling the source signal.

- **Quantization Error/Noise** - The difference between sample and the value assigned to it is known as quantization error or noise.

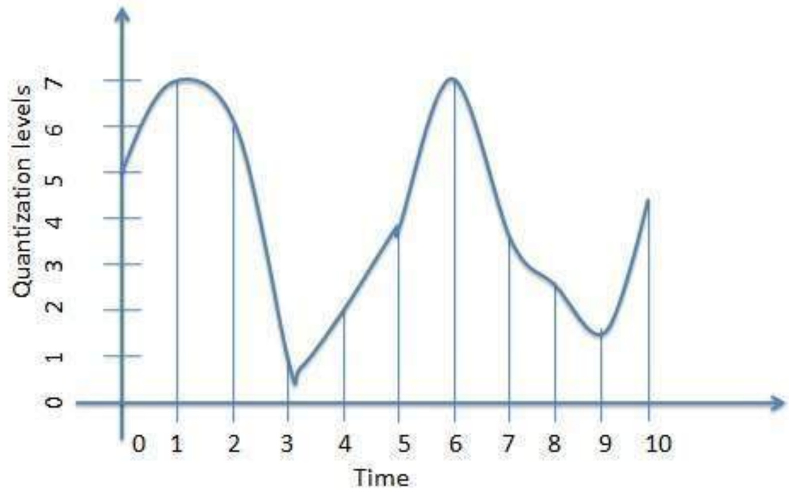
  Quantization noise can be reduced by increasing the number of quantization intervals or levels because the difference between the input signal amplitude and the quantization interval decreases as the number of quantization intervals increases.

- **Signal to Noise Ratio (SNR)** - Signal to Noise Ratio refers to signal quality versus quantization error.

  Higher the Signal to Noise ratio, the better the voice quality.

  Working with very small levels often introduces more error. So instead of uniform quantization, non-uniform quantization is used as compounding.

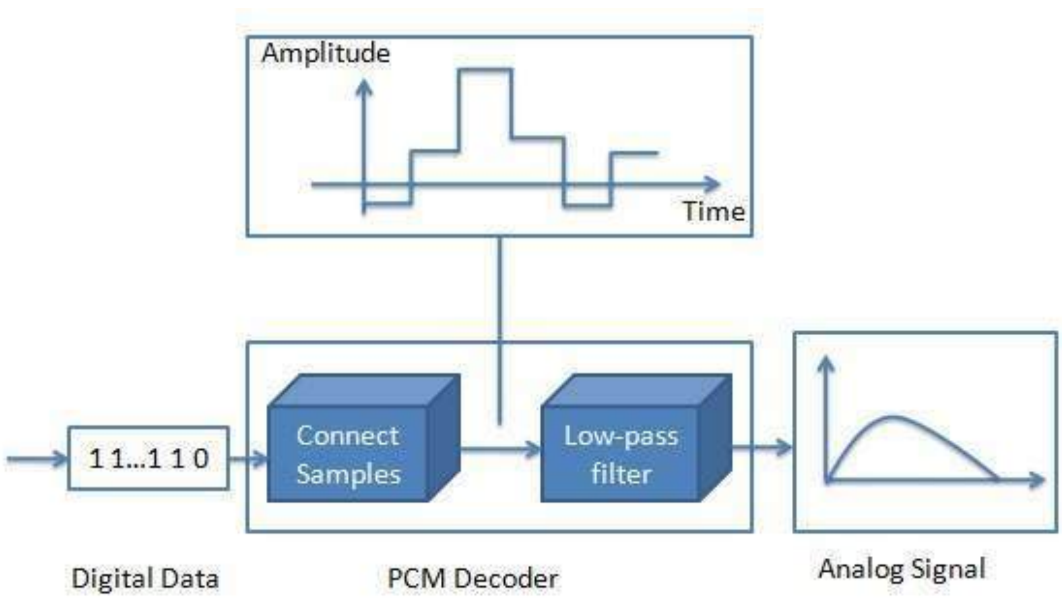
  Compounding is a process of distorting the analog signal in controlled way by compressing large values at the source and then expanding at receiving end before quantization takes place.



Transmission of Audio

In order to send the sampled digital sound/ audio over the wire that it to transmit the digital audio, it is first to be recovered as analog signal. This process is called de-modulation.

- **PCM Demodulation** - PCM Demodulator reads each sampled value then apply the analog filters to suppress energy outside the expected frequency range and outputs the analog signal as output which can be used to transmit the digital signal over the network.

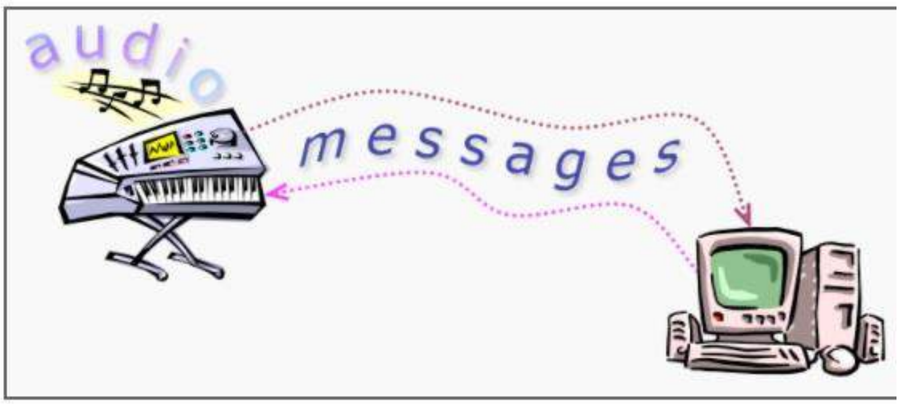


Introduction to the MIDI Standard

- MIDI (Musical Instrument Digital Interface) is a technical standard that describes a communications protocol, digital interface, and electrical connectors that connect a wide variety of electronic musical instruments, computers, and related audio devices for playing, editing, and recording music.
- Virtual instruments — computer programs that simulate hardware synthesizers and samplers — also communicate with computer sequencing software running on the same computer using MIDI messages.
- A single MIDI link through a MIDI cable can carry up to sixteen channels of information, each of which can be routed to a separate device or instrument. This could be sixteen different digital instruments, for example. MIDI carries event messages; data that specify the instructions for music, including a note's notation, pitch, velocity (which is heard typically as loudness or softness of volume); vibrato; panning; and clock signals.
- When a musician plays a MIDI instrument, all of the key presses, button presses, knob turns and slider changes are converted into MIDI data.
- One common MIDI application is to play a MIDI keyboard or other controller and use it to trigger a digital sound module (which contains synthesized musical sounds) to generate sounds, which the audience hears produced by a keyboard amplifier.
- MIDI data can be transferred via MIDI or USB cable, or recorded to a sequencer or digital audio workstation to be edited or played back.[

Communication by Message

- The most important thing to understand about MIDI is that it is based on the idea of message-passing between devices (pieces of equipment or software). Imagine a common situation: you have a keyboard synthesizer and would like to record a sequence using the sounds that are in that synthesizer. You connect the computer and synthesizer so that they can communicate using the MIDI protocol, and start recording. What happens?



Messages, not audio, flows between a synthesizer and a computer running MIDI sequencing software.

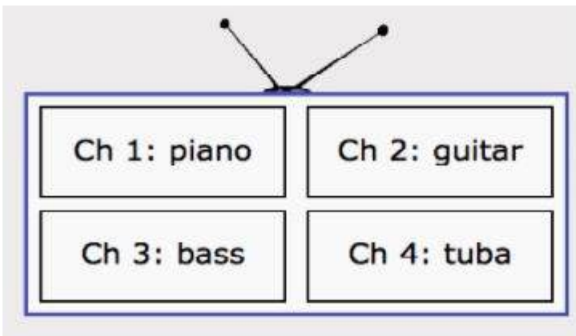
- When you play notes on the synthesizer, all your physical actions (except the dance moves) are transmitted as MIDI messages to the computer sequencing software, which records the messages.

- MIDI messages are brief numeric descriptions of an action. Keys you press, knobs you turn, the joystick you wiggle — all these actions are encoded as MIDI messages. You hear the sound you're making, but that sound comes out of the synthesizer, directly to your speakers. The computer does not record the sound itself.

- When you play your recorded sequence, the computer sends MIDI messages back to the synthesizer, which interprets them and creates audio in response.
- Because the music handled by the computer is in the form of encoded messages, rather than acoustic waveforms, it's possible to change the sound of a track from a piano to a guitar after having recorded the track. That would not be possible if you were recording the sound that the synthesizer makes.

MIDI Channels

- The concept of channels is central to how most MIDI messages work.
- A channel is an independent path over which messages travel to their destination.
- There are 16 channels per MIDI device.
- A track in your sequencer program plays one instrument over a single channel.
- The MIDI messages in the track find their way to the instrument over that channel.



Four separate MIDI channels, one for each instrument

- MIDI channels are a bit like channels on your TV set: each channel is independent of the others, and, on some models of TV, can even be watched simultaneously in separate boxes that appear on the screen.

Applications of MIDI

1. ***Instrument control***

MIDI was invented so that electronic or digital musical instruments could communicate with each other and so that one instrument can control another.

2. ***Composition***

MIDI events can be sequenced with computer software, or in specialized hardware music workstations. Many digital audio workstations (DAWs) are specifically designed to work with MIDI as an integral component. MIDI piano rolls have been developed in many DAWs so that the recorded MIDI messages can be easily modified. These tools allow composers to audition and edit their work much more quickly and efficiently than did older solutions, such as multitrack recording.

3. ***Use with computers***

The personal computer market stabilized at the same time that MIDI appeared, and computers became a viable option for music production.

3.1. ***Computer files:***

- ***Standard files:*** he Standard MIDI File (SMF) is a file format that provides a standardized way for music sequences to be saved, transported, and opened in other systems.

- ***RMID files:*** Microsoft Windows bundles SMFs (Standard MIDI Files) together with Downloadable Sounds (DLS) in a Resource Interchange File Format (RIFF) wrapper, as RMID files with a .rmi extension. RIFF-RMID has been deprecated in favor of Extensible Music Files (XMF).

3.2. Software:

The main advantage of the personal computer in a MIDI system is that it can serve a number of different purposes, depending on the software that is loaded. Multitasking allows simultaneous operation of programs that may be able to share data with each other.

3.3. Sequencers:

Sequencing software allows recorded MIDI data to be manipulated using standard computer editing features such as cut, copy and paste and drag and drop. Keyboard shortcuts can be used to streamline workflow, and, in some systems, editing functions may be invoked by MIDI events. The sequencer allows each channel to be set to play a different sound and gives a graphical overview of the arrangement.

3.4. Notation software:

With MIDI, notes played on a keyboard can automatically be transcribed to sheet music.[13]:213 Scorewriting software typically lacks advanced sequencing tools, and is optimized for the creation of a neat, professional printout designed for live instrumentalists.

These programs provide support for dynamics and expression markings, chord and lyric display, and complex score styles.

3.5. Editor/librarians:

Patch editors allow users to program their equipment through the computer interface. These became essential with the appearance of complex synthesizers such as the Yamaha FS1R, which contained several thousand programmable parameters, but had an interface that consisted of fifteen tiny buttons, four knobs and a small LCD.

3.6. Auto-accompaniment programs:

Programs that can dynamically generate accompaniment tracks are called auto-accompaniment programs. These create a full band arrangement in a style that the user selects, and send the result to a MIDI sound generating device for playback. The generated tracks can be used as educational or practice tools, as accompaniment for live performances, or as a songwriting aid.

3.7. Synthesis and sampling:

Computers can use software to generate sounds, which are then passed through a digital-to-analog converter (DAC) to a power amplifier and loudspeaker system. The number of sounds that can be played simultaneously (the polyphony) is dependent on the power of the computer's CPU, as are the sample rate and bit depth of playback, which directly affect the quality of the sound.

3.8. Game music:

Early PC games were distributed on floppy disks, and the small size of MIDI files made them a viable means of providing soundtracks. Games of the DOS and early

Windows eras typically required compatibility with either Ad Lib or Sound Blaster audio cards. These cards used FM synthesis, which generates sound through modulation of sine waves.

4. *Other applications*

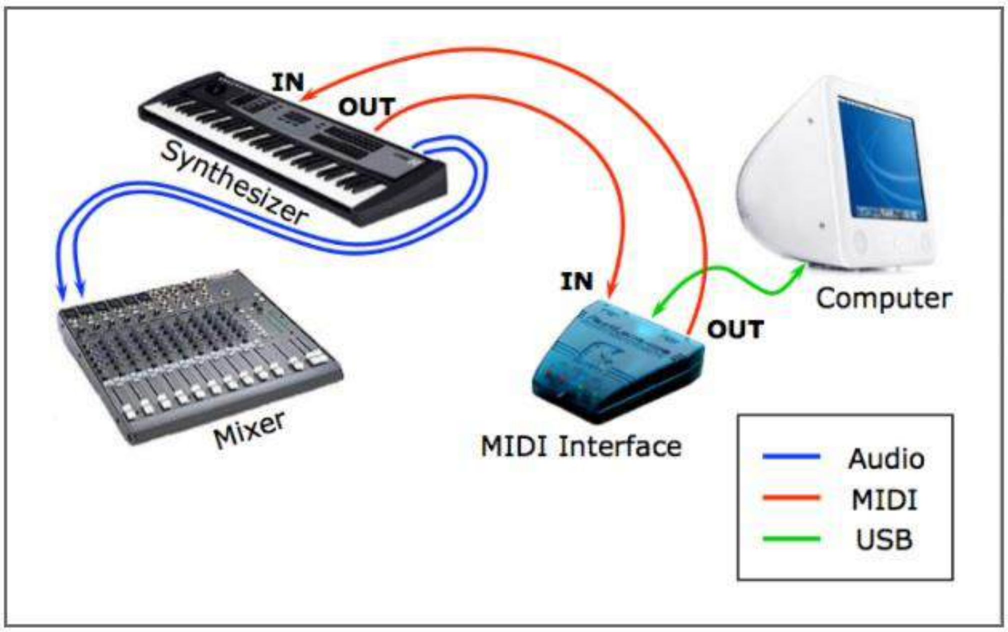
Despite its association with music devices, MIDI can control any electronic or digital device that can read and process a MIDI command. MIDI has been adopted as a control protocol in a number of non-musical applications.

- ✓ MIDI Show Control uses MIDI commands to direct stage lighting systems and to trigger cued events in theatrical productions.
- ✓ VJs and turntablists use it to cue clips, and to synchronize equipment, and recording systems use it for synchronization and automation.
- ✓ Apple Motion allows control of animation parameters through MIDI.

Basic MIDI Hardware Setup

Present day software is capable of performing the sound-making function formerly available only in external hardware-based synthesizers. It’s just as likely now to see, connected to a computer, a keyboard that can’t make any sound at all. Its function is to trigger and control, via MIDI messages, sounds made by the computer. But the sound-making part of the computer software still communicates with the sequencing part using the MIDI protocol.

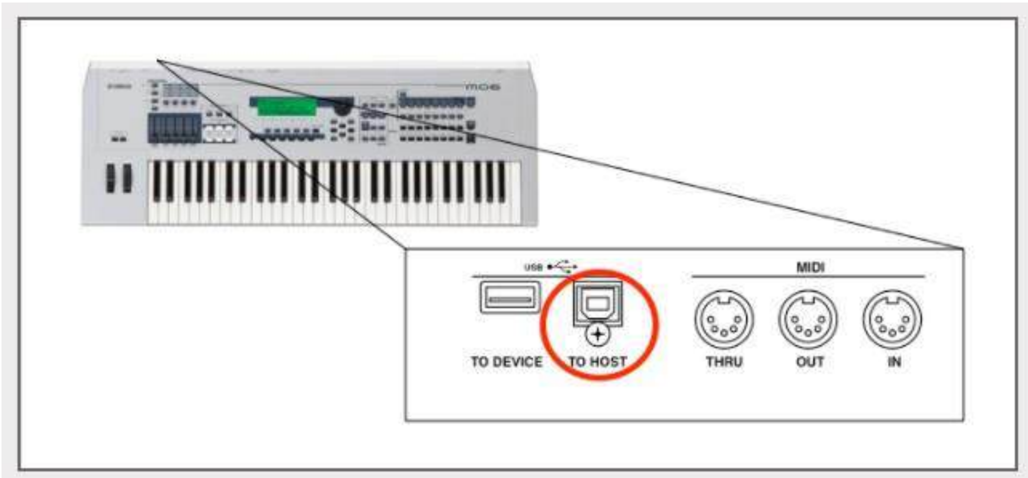
There are still plenty of MIDI setups that work in the traditional way, with the computer just recording and playing MIDI messages, and the sound created by an external synthesizer. These are especially useful in live setups, where the reliability and faster response of hardware synthesizers are distinct advantages. In such a system, you use MIDI cables to connect the synthesizer to a MIDI interface, which then connects to the computer with the same sort of USB cable you use to connect a printer. MIDI cables are unidirectional — they transport messages in only one direction. So you need two MIDI cables. USB is bidirectional. The sound made by the synthesizer goes to a mixer, which then feeds an amplifier and speakers (not shown below).



USB, MIDI, an audio cable connections among computer, synthesizer, and mixer MIDI ports on the interface and synthesizer are labeled IN and OUT. You connect the MIDI OUT jack of the synthesizer to the MIDI IN jack of the interface, and vice versa.

Current Trends

For simpler setups, it's more common today to find keyboards with a USB port that allows for direct connection to a computer, bypassing the MIDI interface. The keyboard in the picture below has both USB (circled) and traditional MIDI ports (to the right).



USB B and MIDI jacks on the back of a synthesizer

As mentioned above, a lot of the action formerly taking place in external boxes is now happening in the computer, avoiding the need for complex hardware setups. For many situations, all you need is an inexpensive MIDI controller keyboard (without internal sounds), with a USB connection to the computer.

General MIDI

Synthesizers and samplers have large numbers of sounds (which we call patches or programs). The patches appear in banks of 128 or fewer, and your computer software selects the patches by number, even if you choose the patches from a list of names and never notice the patch numbers. Types of sounds — pianos, guitars, violins — are assigned to numbers in a way that is not compatible between different synthesizers. That means that a sequence recorded using one type of synthesizer will not sound remotely the same when played using a different type of synthesizer.

To address this problem, the MIDI standard includes the General MIDI (or GM) specification. The most important part of this is a standard assignment of instrument types to patch numbers. For example, in a General MIDI compatible sequence, a violin sound will always be patch number 41. The violins on two different keyboards will not sound exactly the same, but at least they will sound like violins.

A similar problem affects drum kit patches: the assignment of individual drum sounds to keys on the keyboard is not guaranteed to be compatible between different synthesizers. General MIDI specifies a map of typical drum sounds to keys. It also declares that channel 10 is the drum channel, so that a sequence can depend on finding drum sounds there.

Standard MIDI Files

To enhance compatibility between different MIDI sequencing and music notation programs, even those running on different operating systems, the MIDI standard defines a specification for the Standard MIDI File. This type of file (usually having the file extension ".mid") represents multi-track sequences, complete with patch selections, notes, pitch bend, and other controls. A wide variety of programs can read and write SMF files. The format is especially useful in conjunction with the GM patch set, to enhance portability between different systems.

Audio signal and Sampling rate

Audio Signal

- An ***audio signal*** is a representation of sound, typically using either a changing level of electrical voltage for analog signals, or a series of binary numbers for digital signals. Audio signals have frequencies in the audio frequency range of roughly 20 to 20,000 Hz, which corresponds to the lower and upper limits of human hearing.

- Audio signals may be synthesized directly, or may originate at a transducer such as a microphone, musical instrument pickup, phonograph cartridge, or tape head. Loudspeakers or headphones convert an electrical audio signal back into sound.

- Digital audio systems represent audio signals in a variety of digital formats.

- An audio channel or audio track is an audio signal communications channel in a storage device or mixing console, used in operations such as multi-track recording and sound reinforcement.

- Signal flow is the path an audio signal will take from source to the speaker or recording device. Signal flow may be short and simple as in a home audio system or long and convoluted in a recording studio and larger sound reinforcement system as the signal may pass through many sections of a large mixing console, external audio equipment, and even different rooms.

- Audio signals may be characterized by parameters such as their ***bandwidth, nominal level, power level in decibels (dB),*** and ***voltage level*****.** The relationship between power and voltage is determined by the impedance of the signal path. Signal paths may be single-ended or balanced.

Bandwidth is the difference between the upper and lower frequencies in a continuous band of frequencies. It is typically measured in hertz.

Nominal level is the operating level at which an electronic signal processing device is designed to operate.

The decibel (symbol: dB) is a relative unit of measurement equal to one tenth of a bel (B). It expresses the ratio of two values of a power or root-power quantity on a logarithmic scale.

- Audio signals have somewhat standardized levels depending on the application. Outputs of professional mixing consoles are most commonly at line level. Consumer audio equipment will also output at a lower line level. Microphones generally output at an even lower level, commonly referred to as mic level.

- The digital form of an audio signal is used in audio plug-ins and digital audio workstation (DAW) software. The digital information passing through the DAW (i.e. from an audio track through a plug-in and out a hardware output) is an audio signal.

- A digital audio signal can be sent over optical fiber, coaxial and twisted pair cable. A line code and potentially a communication protocol is applied render a digital signal for a transmission medium.

Sampling rates

Sampling rate or sampling frequency defines the number of samples per second (or per other unit) taken from a continuous signal to make a discrete or digital signal. For time-domain signals like the waveforms for sound (and other audio-visual content types), frequencies are measured in hertz (Hz) or cycles per second.

Nyquist–Shannon sampling theorem (Nyquist principle) states that perfect reconstruction of a signal is possible when the sampling frequency is greater than twice the maximum frequency of the signal being sampled. For example, if an audio signal has an upper limit of 20,000 Hz (the approximate upper limit of human hearing), a sampling frequency greater than 40,000 Hz (40 kHz) will avoid ***aliasing*** and allow theoretically perfect reconstruction.

Nyquist principle states that the sampling rate must be at least twice the maximum bandwidth of the analog signal in order to allow the signal to be completely represented.

Aliasing: a sampling effect that leads to spatial frequencies being falsely interpreted as other spatial frequencies.

- Audio sampling is the process of transforming a musical source into a digital file. Digital audio recording does this by taking samples of the audio source along the soundwaves at regular intervals. The more samples you take - known as the 'sample rate' - the more closely the final digital file will resemble the original. A higher sample rate tends to deliver a better-quality audio reproduction.

- Sample rates are usually measured per second, using kilohertz (kHz) or cycles per second. CDs are usually recorded at 44.1 kHz - which means that every second, 44,100 samples were taken.

- If you can master sample rates, you can create more accurate recordings. Once you have this digital copy you can manipulate, mix and edit without losing any sound quality.

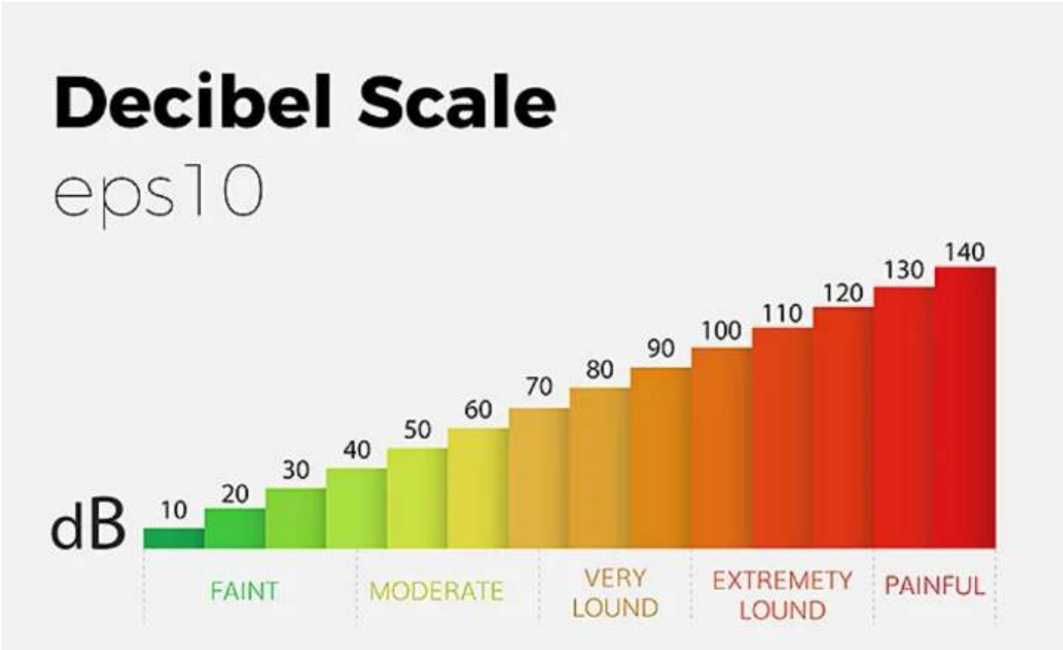
- The final sound quality of your recording relies on more than just sample rate - bit depth also plays its part.

Bit Depth/rate

- Every sample you take when making an audio recording needs to be stored within your computer's 'bits'. The more bits you use to record each sample, the better the sound reproduction.

- In other words, a high sample rate along with a high bit depth (AKA 'bit rate' or 'sample format') will deliver the best audio quality in your recording. The higher the bit depth, the higher the dynamic range.

- Dynamic range is the difference between the low volume and high volume sections of your recording. This is measured in decibels - or dBs. The human ear can hear sounds up to 90dB, but recording over this level allows softer sounds to be amplified for high fidelity audio.



8-bit audio.

This is a fairly low-quality reproduction, producing audio at 46dB - around half the top level of human hearing.

16-bit audio.

This is where the human ear can usually hear to at 96dB.

24-bit audio.

While at 145dB it's well above the human hearing range, it can be useful to work at this level to reduce the 'noise floor' - essentially the digital white noise.

32-bit float audio.

This can offer nearly infinite decibel levels and is really only used for super-high-quality audio - for example, sudden loud noises that need capturing without the use of limiters.

Why is the standard audio sample rating 44.1 kHz?

- Many modern-day recordings use the 44.1 kHz sample rate and it's the standard rate for CDs. This is partly down to how sample rates work and how we hear as humans.
- Sample rates were first discussed in the 1940s, as part of the Nyquist-Shannon theorem. This states that any sampling rate must have twice the frequency of the original recording; otherwise the sound is not faithfully reproduced.
- The human ear can hear between 20 hertz (20Hz) and 20 kilohertz (20 kHz). 44.1 kHz is more than twice the top range of human hearing, so will provide a very accurate reproduction according to the theory.

- Some people still record in higher sample rates to capture all sounds. While we can't hear these in the original recordings, if an audio sample recorded at 192 kHz was pitched down, some hitherto inaudible frequencies would become audible. If it was recorded at a lower sample rate and then pitched down, some of the highs in that audio would be lost.
- However, even if you're recording at a higher rate, it will be likely converted back to 44.1 kHz - the rate to which many modern audio systems are set.

What other sample rates are used - and what for?

- With 44.1 kHz the standard for CD audio, you might wonder why other sample rates exist. As we've mentioned, higher sampling rates can provide clearer audio with no white noise. They can also be useful for mastering and mixing audio. Even if we can't hear some sounds, they do exist in the higher sampling rate recording and so can still be manipulated.

48 kHz.

This sample rate is also used as a standard rate alongside 44.1 kHz. Do check though, as audio recorded in one rate and played at another will be either speeded up or slowed down.

88.2 kHz.

This is now the gold standard for hi-res recordings. Using this sample rate produces less distortion (called 'aliasing') when converting from analogue to digital and allows greater freedom when mixing and mastering.

96 kHz.

Similar to 88.2 kHz, this sample rate provides more options when mixing and mastering the audio. But working at these higher rates could be an issue if your computer can't handle the added information and storage needed.

192 kHz.

Some reports have suggested that recording at such a high sample rate can produce issues in your audio, such as jittering. It's also hard to find computers that can handle it. Really, it's only useful for slowing down high-frequency audio.

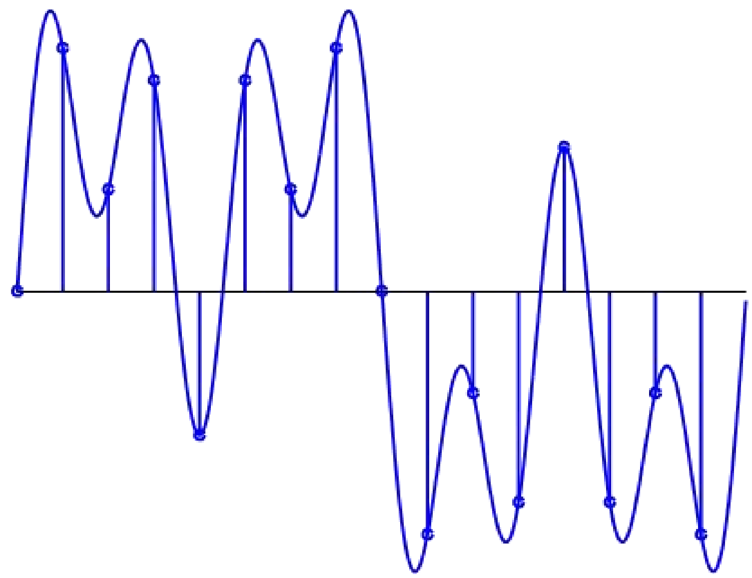
Audio signal and Sampling rate_Part2

Continuous-Time Signals

- Analog signals are continuous in time.
- A continuous-time signal is an infinite and uncountable sequence of numbers, as are the possible values each number can have.
  - Between a start and end time, there are infinite possible values for time t and the waveform's instantaneous amplitude x (t).
- A continuous signal cannot be stored, or processed, in a computer since it would require infinite data.
- Analog signals must be discretized (digitized) to produce a finite set of numbers for computer use.

Discrete-Time, Digital Signals

- When analog signals are brought into a computer, they must be made discrete (finite and countable).
- A **discrete-time signal** is a finite sequence of numbers, with finite possible values for each number.
  - number values are limited by how many bits are used to represent them (bit depth);
  - can be stored on a digital storage medium.



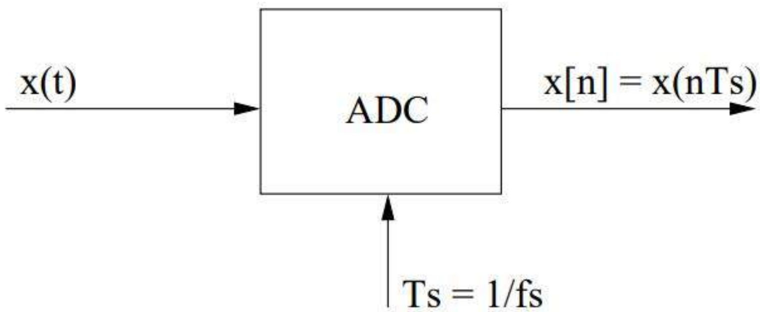
- **Sampling:** the process of taking individual values of a continuous-time signal (at regular time intervals).

Analog to Digital Conversion

- The process by which an analog signal is digitized is called analog-to-digital or "a-to-d" conversion
- Analog -to- digital converter (ADC): the device that discretizes (digitizes) an analog signal.
- The ADC must accomplish two (2) tasks:
  - ✓ **Sampling:** taking values (samples) at regular time intervals;
  - ✓ **Quantization:** Assign a number to the value (using limited computer bits).

Sampling:

- ***Sampling:*** process of taking values (samples) of the analog waveform at regularly spaced time intervals.



The ideal analog-to-digital converter.

- ***Sampling Rate***: number of samples taken per second (Hz);

  – fs = 48 kHz (professional studio)

  – fs = 44.1 kHz (CD)

  – fs = 32 kHz (broadcasting)

- ***Sampling Period***: time interval (in seconds) between samples:

$$Ts = 1/fs \text{ seconds.}$$

Sampled Sinusoids:

- Sampling corresponds to transforming the continuous time variable *t* into a set of discrete times that are integer *n* multiples of the sampling period T_S:

$$t \rightarrow nT_S$$

- Integer *n* corresponds to the index in the sequence.

- Continuous sinusoid:

$$X(t) = A\sin(\omega t + \varphi).$$

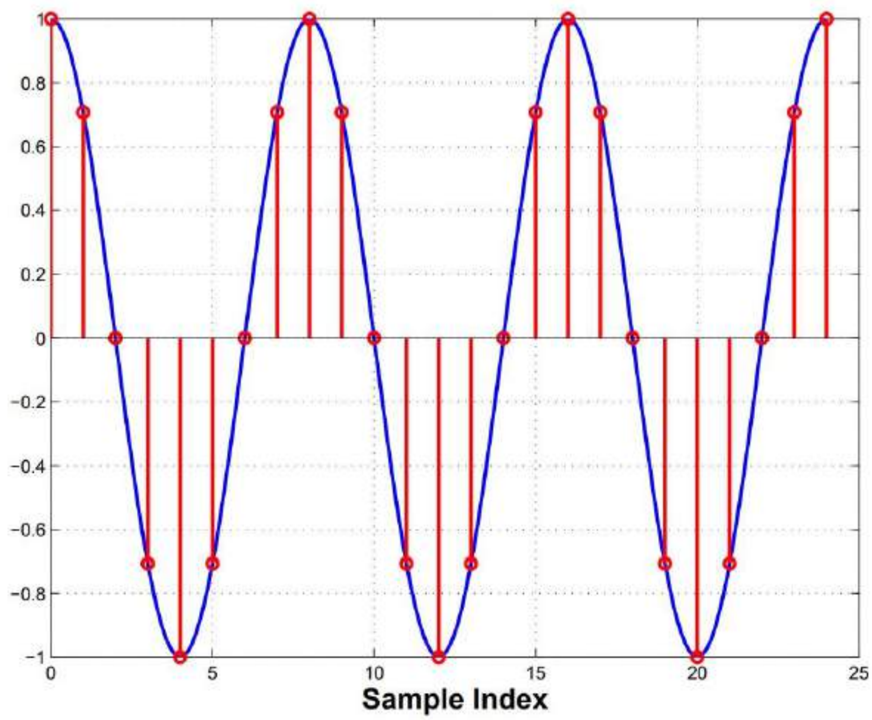
- Discrete sinusoid

$$X(n) = A\sin(\omega n T_S + \varphi),$$

a sequence of numbers that may be indexed by n.

Example1:

If the following sinusoid was sampled at fs = 16 Hz, what is the duration of the signal shown?



Answer:

The sampling period (time between samples) is

$$T_S = 1/16\ s.$$

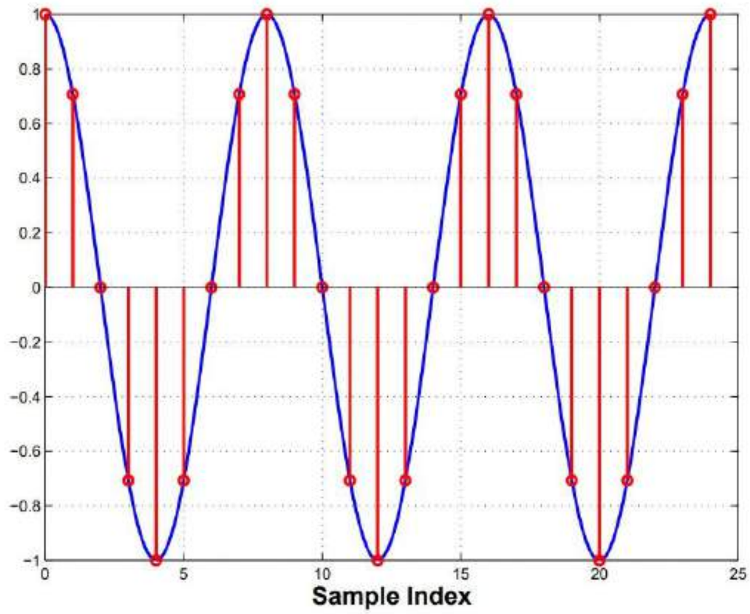
Since 24 samples are shown, the duration is

$$24 \times T_s = \frac{24}{16} = \frac{3}{2} = 1.5\ \text{s}.$$

What would the duration be if fs = 32?

Example2:

If the following sinusoid was sampled at $f_S = 16$ Hz, what is the *frequency* of the sinusoid?



Answer: (Method1)

16 samples corresponds to 1 second,

there are 2 cycles in 1 second (after 16 samples),

the frequency is 2 Hz.

Method2:

The period of the sinusoid is:

$$T = 8 \times T_s = \frac{8}{16} = \frac{1}{2},$$

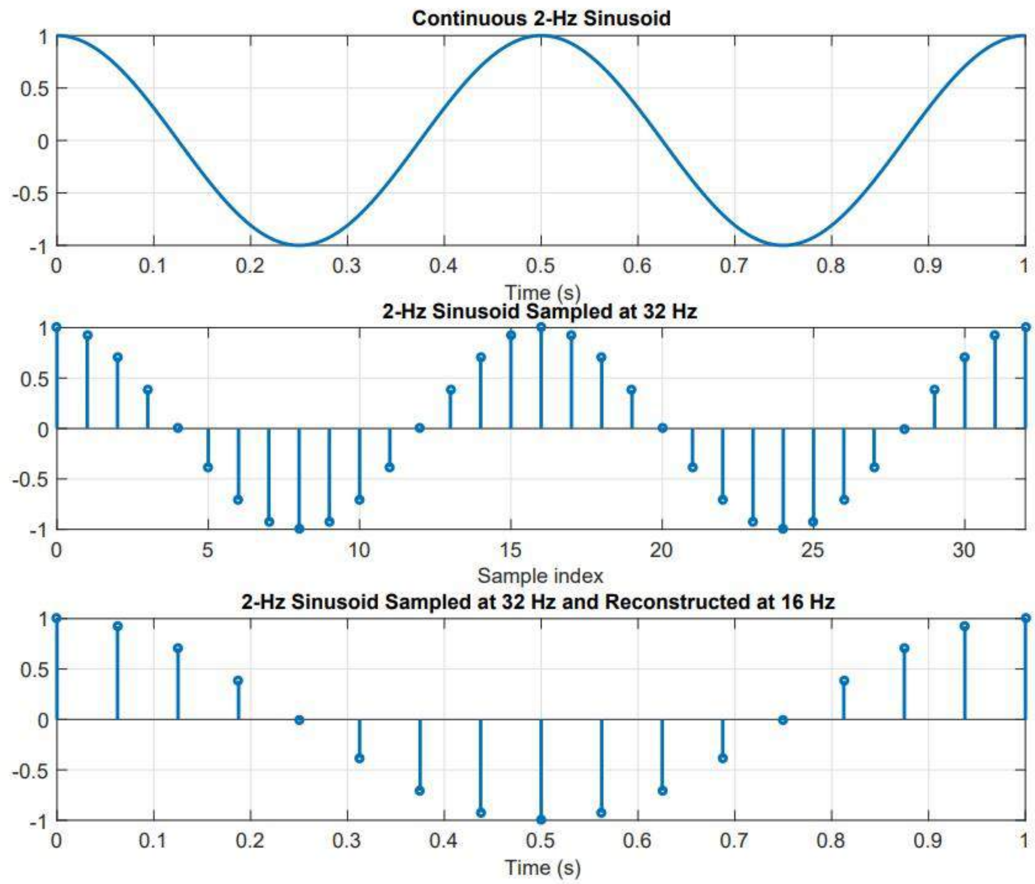
and the frequency is:

$$f = \frac{1}{T} = \frac{1}{1/2} = 2 \text{ Hz.}$$

If fs = 32, what is the frequency of the sinusoid?

Sampling and Reconstruction

- Once x (t) is sampled to produce x (n), time scale information is lost.
- X (n) may represent a number of possible waveforms.



- Reconstructing at half the sampling rate (Fs/2) will double the time between samples (2/Fs), making the sinusoid twice as long and halving the frequency.

Importance of Knowing Sampling Rate

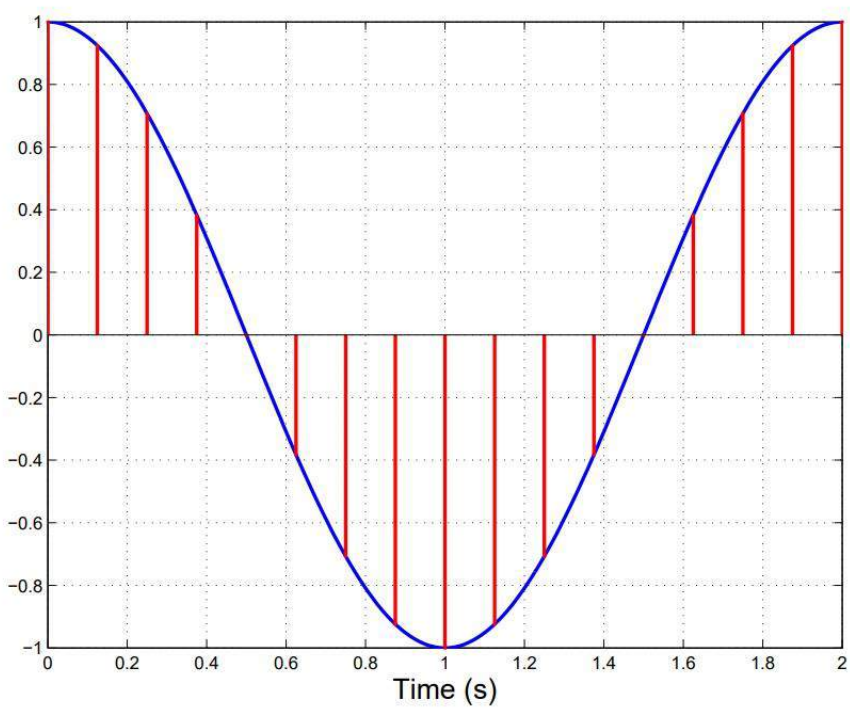
- If the signal is digitized and reconstructed using the same sampling rate, the frequency and duration will be preserved.

- If reconstruction is done using a different sampling rate, the time interval between samples changes resulting in a change in

  - ✓ overall signal duration,
  - ✓ time to complete one cycle (period) and thus sounding frequency.

Homework:

What is the sampling rate and frequency of the following sinusoid?



Nyquist Sampling Theorem

The ***Nyquist Sampling Theorem*** states that:

A bandlimited continuous-time signal can be sampled and perfectly reconstructed from its samples if the waveform is sampled over twice as fast as its highest frequency component.

Nyquist limit: the highest frequency component that can be accurately represented:

$$fmax < fs/2.$$

Nyquist frequency: sampling rate required to accurately represent up to *fmax*:

$$fs > 2fmax.$$

No information is lost if sampling above 2fmax.

No information is gained by sampling much faster than 2fmax.

Audio modulation

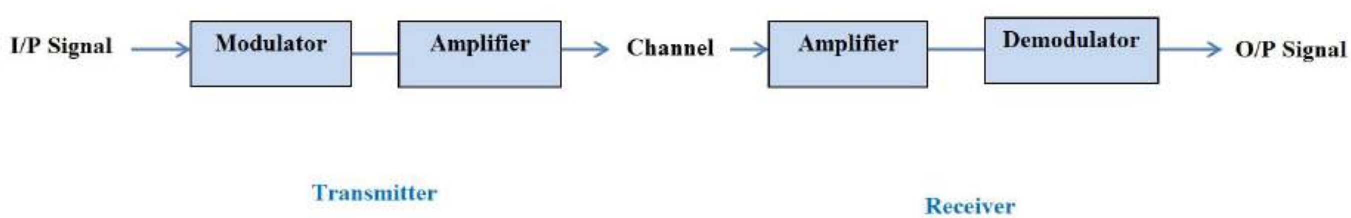
Basic System

The basic communications system has:

Transmitter: The sub-system that takes the information signal and processes it prior to transmission. The transmitter modulates the information onto a carrier signal, amplifies the signal and broadcasts it over the channel.

Channel: The medium which transports the modulated signal to the receiver. Air acts as the channel for broadcasts like radio. It also may be a wiring system like cable TV or the internet.

Receiver: The sub-system that takes in the transmitted signal from the channel and processes it to retrieve the information signal. The receiver must be able to discriminate the signal from other signals which may use the same channel (called tuning), amplify the signal for processing and demodulate (remove the carrier) to retrieve the information. It also then processes the information for reception (for example, broadcast on a loudspeaker).



Basic Communication System

Modulation

- Whenever a parameter of a sound or audio signal called the *carrier* is varied systematically, the signal is said to be modulated.
  - *Carrier*: an electromagnetic or audio signal, one or more of whose parameters is modulated by another signal.
- The signal whose waveform is being used to control the carrier is called the *modulator* or *program signal*.
- Modulation is a standard technique for both radio transmission and sound synthesis, although the frequencies used are in very different ranges.
- There are three common types of modulation: *frequency modulation, amplitude modulation, and ring modulation*. All of which are non-linear approaches to sound synthesis.
  - The response of a system is *linear* when the output is directly proportional to the input, that is, any change in the input produces a proportional change in the output. When plotted on a graph, a straight line results.
- Modulation also refers to the signal level on a recording, transmission or reproduction system. Full modulation or 100% modulation refers to the maximum permissible (i.e. distortion-free) level of such a system.
- The information signal can rarely be transmitted as is, it must be processed.
- In order to use electromagnetic transmission, it must first be converted from audio into an electric signal. The conversion is accomplished by a transducer. After conversion it is used to modulate a carrier signal.

- A carrier signal is used for two reasons:

  - ✓ *To reduce the wavelength for efficient transmission and reception* (the optimum antenna size is ½ or ¼ of a wavelength). A typical audio frequency of 3000 Hz will have a wavelength of 100 km and would need an effective antenna length of 25 km! By comparison, a typical carrier for FM is 100 MHz, with a wavelength of 3 m, and could use an antenna only 80 cm long.

  - ✓ *To allow simultaneous use of the same channel, called multiplexing.* Each unique signal can be assigned a different carrier frequency (like radio stations) and still share the same channel. The phone company actually invented modulation to allow phone conversations to be transmitted over common lines.

- The process of modulation means to systematically use the information signal (what you want to transmit) to vary some parameter of the carrier signal.

- The carrier signal is usually just a simple; single-frequency sinusoid (varies in time like a sine wave).

- The basic sine wave goes like $X(t) = A \sin(\omega t + \varphi)$, where the parameters are defined below:

X (t): the voltage of the signal as a function of time.

A: the amplitude of the signal (represents the maximum value achieved each cycle)

$\omega = 2\pi f$, when ***f:*** the frequency of oscillation, the number of cycles per second (also known as Hertz = 1 cycle per second)

φ: the phase of the signal, representing the starting point of the cycle.

- To modulate the signal just means to systematically vary one of the three parameters of the signal: *amplitude, frequency or phase*. Therefore, the type of modulation may be categorized as either

AM: amplitude modulation

FM: frequency modulation or

PM: phase modulation

- PM may be an unfamiliar term but is commonly used. The characteristics of PM are very similar to FM and so the terms are often used interchangeably.

FM Modulation

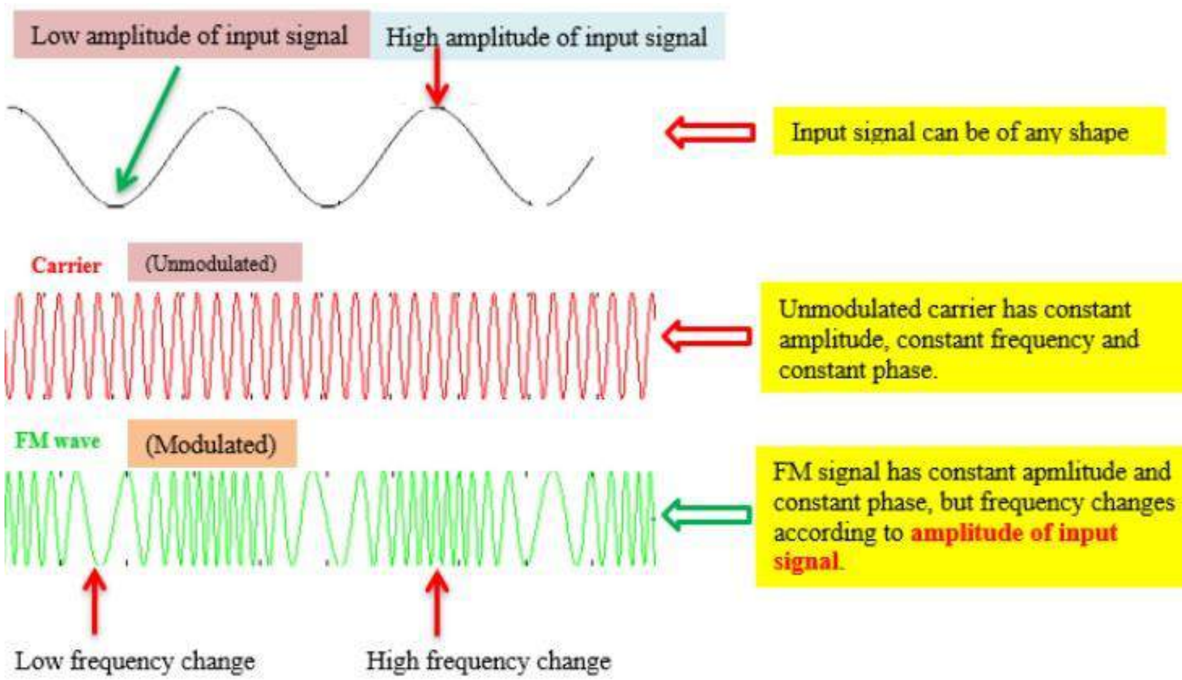
- Frequency Modulation (FM) is the encoding of information in a carrier wave by changing the instantaneous frequency of the wave.

- FM technology is widely used in the fields of computing, telecommunications, and signal processing.

- Frequency modulation uses the information signal, $X_m(t)$ to vary the carrier frequency within some small range about its original value.

- Here are the three signals in mathematical form:

  - ***Information Signal:*** $X_m(t)$
  - ***Carrier Signal:*** $X_c(t)= A_{co} \sin (2 \pi f_c t + \phi)$
  - ***FM Signal:*** $X_{FM} (t) = A_{co} \sin (2 \pi [f_c + (\Delta f/A_{mo}) X_m (t)]t + \phi)$



- The carrier frequency term has been replaced with a *time-varying frequency*. Also a new term: Δf, the *peak frequency deviation* have been introduced.

  In this form, you should be able to see that the *carrier frequency* term: $f_c + (\Delta f/A_{mo}) X_m(t)$ now varies between the extremes of $fc - \Delta f$ and $fc + \Delta f$.

  The interpretation of Δf becomes clear: it is the farthest away from the original frequency that the FM signal can be. Sometimes it is referred to as the "swing" in the frequency.

- A *modulation index* for FM can be defined, analogous to AM:

$$\beta = \Delta f / f_m,$$

where f_m is the maximum modulating frequency used.

- The simplest interpretation of the *modulation index*, β, is as a measure of the peak frequency deviation, Δf. In other words, β represents a way to express the peak deviation frequency as a multiple of the maximum modulating frequency, f_m, i.e.

$$\boxed{\Delta f = \beta f_m}$$

- **Example:** suppose in FM radio that the audio signal to be transmitted ranges from 20 to 15,000 Hz (it does). If the FM system used a maximum modulating index, β, of 5.0, then the frequency would "swing" by a maximum of 5 x 15 kHz = 75 kHz above and below the carrier frequency.

- FM radio is example of using frequency modulation. The frequency band for FM radio is about 88 to 108 MHz. The information signal is music and voice which falls in the audio spectrum. The full audio spectrum ranges from 20 to 20,000 Hz, but FM radio limits the upper modulating frequency to 15 kHz.

  A spectrum represents the relative amounts of different frequency components in any signal.

FM Performance

- **Bandwidth:** the bandwidth of a FM signal may be predicted using:

$$BW = 2\,(\beta + 1)\, f_m$$

β is the modulation index and

f_m is the maximum modulating frequency used.

- ✓ FM radio has a significantly larger bandwidth than AM radio, but the FM radio band is also larger. The combination keeps the number of available channels about the same.
- ✓ The bandwidth of an FM signal has a more complicated dependency than in the AM case (recall, the bandwidth of AM signals depend only on the maximum modulation frequency). In FM, both the modulation index and the modulating frequency affect the bandwidth. As the information is made stronger, the bandwidth also grows.

Efficiency

- ✓ The efficiency of a signal is the power in the side-bands as a fraction of the total.
- ✓ In FM signals, because of the considerable side-bands produced, the efficiency is generally high.
- ✓ Recall that conventional AM is limited to about 33 % efficiency to prevent distortion in the receiver when the modulation index was greater than 1. FM has no analogous problem.
- ✓ The side-band structure is fairly complicated, but it is safe to say that the efficiency is generally improved by making the modulation index larger (as it should be).
- ✓ But if you make the modulation index larger, so make the bandwidth larger (unlike AM) which has its disadvantages. As is typical in engineering, a compromise between efficiency and performance is struck.
- ✓ The modulation index is normally limited to a value between 1 and 5, depending on the application.

Noise

- ✓ FM systems are far better at rejecting noise than AM systems. Noise generally is spread uniformly across the spectrum (the so-called white noise, meaning wide spectrum).

- ✓ The amplitude of the noise varies randomly at these frequencies.
- ✓ The change in amplitude can actually modulate the signal and be picked up in the AM system. As a result, AM systems are very sensitive to random noise.
- ✓ An example might be ignition system noise in your car.
- ✓ Special filters need to be installed to keep the interference out of your car radio.
- ✓ FM systems are inherently immune to random noise.
- ✓ In order for the noise to interfere, it would have to modulate the frequency somehow.
- ✓ But the noise is distributed uniformly in frequency and varies mostly in amplitude.
- ✓ As a result, there is virtually no interference picked up in the FM receiver.
- ✓ FM is sometimes called "static free" referring to its superior immunity to random noise.

Frequency Modulation in Communication Systems

- There are two different types of frequency modulation used in telecommunications: *analog frequency modulation* and *digital frequency modulation.*

- In *analog modulation*, a continuously varying sine carrier wave modulates the data signal. The three defining properties of a carrier wave -- frequency, amplitude, and phase -- are used to create AM, FM, and Phase Modulation.

- *Digital modulation*, categorized as either *Frequency Shift Key*, *Amplitude Shift Key*, or *Phase Shift Key* functions similarly to analog, however where analog modulation is typically used for AM, FM, and short-wave broadcasting, digital modulation involves transmission of binary signals (0 and 1).

Summary

- In FM signals, the efficiency and bandwidth both depend on both the maximum modulating frequency and the modulation index.
- Compared to AM, the FM signal has a higher efficiency, a larger bandwidth and better immunity to noise.

Audio modulation_Part2

Amplitude modulation

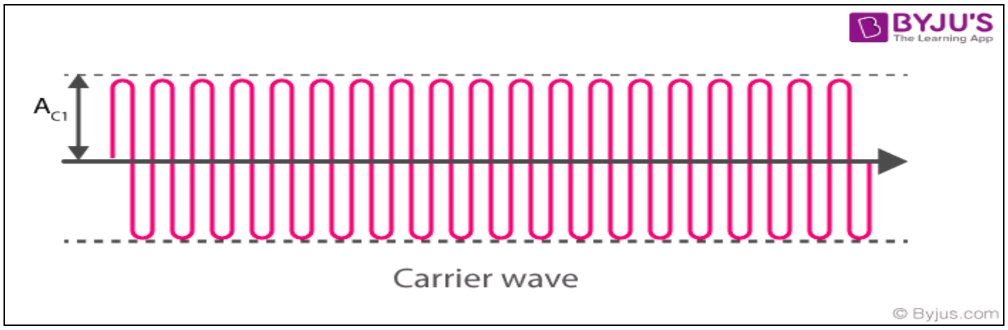
- ***Amplitude modulation (AM)*** is a process by which the wave signal is transmitted by modulating the amplitude of the signal. It is a modulation technique used in electronic communication, most commonly for transmitting messages with a radio wave.

- In amplitude modulation, the amplitude (signal strength) of the wave is varied in proportion to that of the message signal, such as an audio signal.

- There are three main types of amplitude modulation. They are;
  - Double sideband-suppressed carrier modulation (DSB-SC).
  - Single Sideband Modulation (SSB).
  - Vestigial Sideband Modulation (VSB).

Common Terms

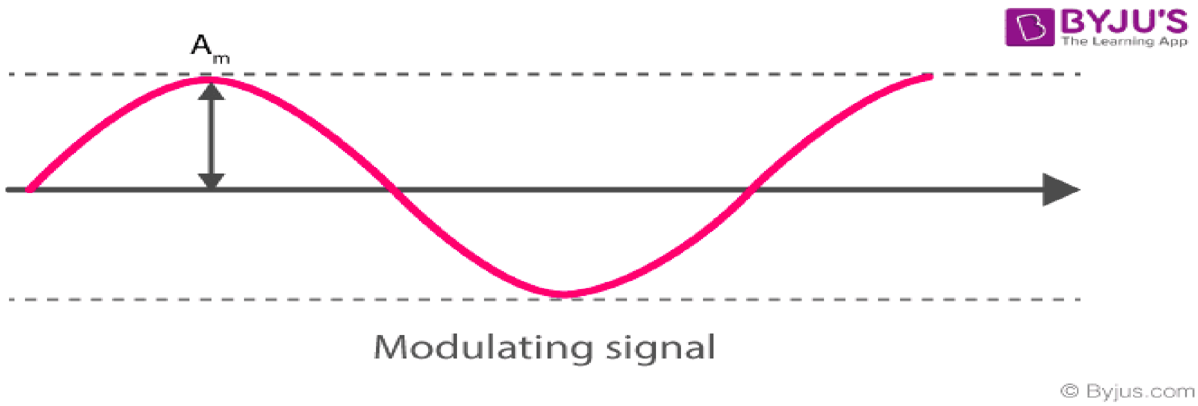
- ***Carrier Wave (High frequency)***



Amplitude and frequency of a carrier wave remain constant. Generally, it will be high frequency and it will be sine (or) cosine wave of electronic signal it can be represented as

$X_c(t) = A_c \sin \omega_c t$

Modulating Signal



Modulating signal is nothing but the input signal (electronic signal), which has to be transmitted. It is also a sine (or) cosine wave it can be represented as

$X_m(t) = A_m \sin \omega_m t$

Where:

- ✓ *Ac and Am* : Amplitude of the carrier wave and the modulating signal.
- ✓ $\sin \omega_c t$: phase of the carrier wave
- ✓ $\sin \omega_m t$: phase of the modulating signal

Amplitude modulated wave

Modulating signal is superimposed into carrier wave and also the amplitude of the carrier wave is varied in accordance with the amplitude of the modulating signal to get the amplitude modulated wave.

$X_{AM}(t) = (A_c + A_m \sin \omega_m t) \sin \omega_c t$

Frequencies of Amplitude Modulated Wave

- There are three frequencies in amplitude modulated wave f_1, f_2 and f_3 corresponding to ω_c, $\omega_c + \omega_m$ and $\omega_c - \omega_m$ respectively.

$\omega_1 = \omega_c$ → it is corresponding $f_1 = f_c$

$\omega_2 = \omega_c + \omega_m$ → it is corresponding $f_2 = f_c + f_m$

$\omega_3 = \omega_c - \omega_m$ → it is corresponding $f_3 = f_c - f_m$

Where:

f_c → carrier wave frequency

$f_c + f_m$ → upper side band frequency

$f_c - f_m$ → lower side band frequency

f_m → modulating signal frequency

In general $f_c >> f_m$

Bandwidth: (*BW*)

It is the difference between the highest and lowest frequencies of the signal.

BW = upper sideband frequency – lower sideband frequency $(f_c - f_m)$

Or

$BW = f_{max} - f_{min}$

$BW = f_c + f_m - f_c + f_m = 2 f_m$

$BW = 2f_m$ = ***twice the frequency of modulating signal***

Modulation Index

Is the ratio of Amplitude of modulating signal to the amplitude of the carrier wave.

$$\boxed{\mu = \frac{A_m}{A_c} = \frac{Amplitude\ of\ modulating\ signal}{Amplitude\ of\ carrier\ wave}}$$

Amplitude Modulated Waveform

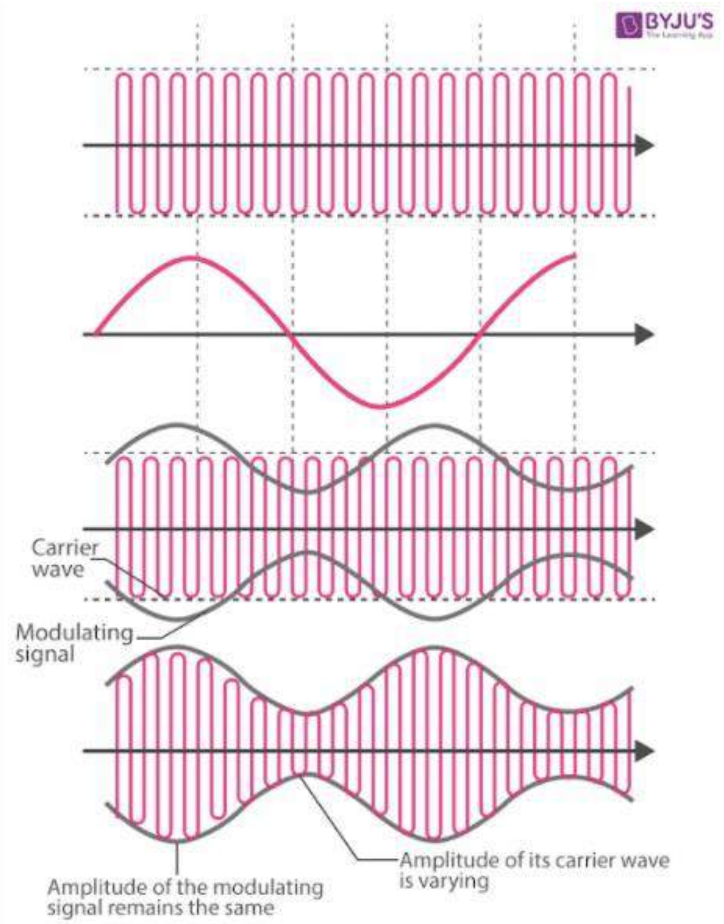
Waveform representation of Amplitude modulated wave:

Carrier wave →

Modulating signal →

Superposition of the carrier wave and modulating signal→

Amplitude modulated wave →



Advantages and Disadvantages of Amplitude Modulation

Advantages	Disadvantages
Amplitude Modulation is easier to implement.	When it comes to power usage it is not efficient.
Demodulation can be done using few components and a circuit.	It requires a very high bandwidth that is equivalent to that of the highest audio frequency.
The receiver used for AM is very cheap.	Noise interference is highly noticeable.

Applications of Amplitude Modulation

While amplitude modulation use has decreased over the years it is still present and has several applications in certain transmission areas.

- ***Broadcast Transmissions:*** AM is used in broadcasting transmission over the short, medium and long wavebands. Since AM is easy to demodulate radio receivers for amplitude modulation are therefore easier and cheaper to manufacture.
- ***Air-band radio:*** AM is used in the VHF transmissions for many airborne applications such as ground-to-air radio communications or two-way radio links for ground staff personnel.

- ***Single sideband:*** Amplitude modulation in this form is used for HF radio links or point-to-point HF links. AM uses a lower bandwidth and provides more effective use of the transmitted power.
- ***Quadrature amplitude modulation:*** AM is used extensively in transmitting data in several ways including short-range wireless links such as Wi-Fi to cellular telecommunications and others.

These are some of the important applications of amplitude modulation.

Ex 1

Carrier wave of frequency f = 1mHz with pack voltage of 20V used to modulate a signal of frequency 1kHz with pack voltage of 10v. Find out the following

1. μ?
2. Frequencies of modulated wave?
3. Bandwidth

Solution:

1. $$\mu = \frac{A_m}{A_c} = \frac{10v}{20v} = \frac{1}{2} = 0.5$$

2. Frequencies of modulated wave

$f \rightarrow f_c$, $f_c + f_m$ and $f_c - f_m$

$f_c = 1mHz$, $f_m = 1kHz$

$f_c + f_m = 1\times10^6 + 1\times10^3 = 1001 \times10^3 = 1001$ kHz

$f_c - f_m = 1\times10^6 - 1\times10^3 = 999 \times 10^3 = 999$ kHz

3. Band width: (***W***)

(W) = upper side band frequency – lower side band frequency

$= f_c + f_m - (fc - fm)$

$= 2f_m = 1001$ kHz – 999 kHz = 2 kHz

Summary

- **AM** (or ***Amplitude Modulation***) and **FM** (or ***Frequency Modulation***) are ways of broadcasting radio signals.
- Both transmit the information in the form of electromagnetic waves.
- AM works by modulating (varying) the amplitude of the signal or carrier transmitted according to the information being sent, while the frequency remains constant.
- This differs from FM technology in which information (sound) is encoded by varying the frequency of the wave and the amplitude is kept constant.

Comparison chart

	AM	**FM**
Stands for	AM stands for Amplitude Modulation	FM stands for Frequency Modulation
Origin	AM method of audio transmission was first successfully carried out in the mid 1870s.	FM radio was developed in the United states in the 1930s, mainly by Edwin Armstrong.

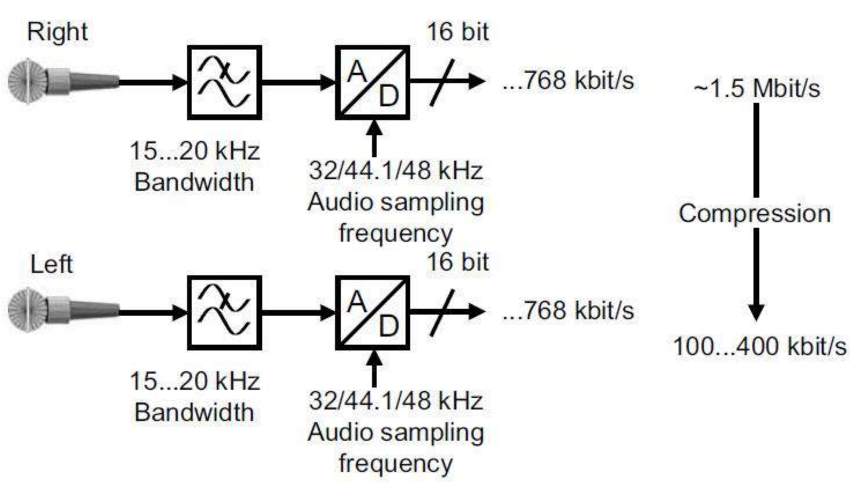
Modulating differences	In AM, a radio wave known as the "carrier" or "carrier wave" is modulated in amplitude by the signal that is to be transmitted. The frequency and phase remain the same.	In FM, a radio wave known as the "carrier" or "carrier wave" is modulated in frequency by the signal that is to be transmitted. The amplitude and phase remain the same.
Pros and cons	AM has poorer sound quality compared with FM, but is cheaper and can be transmitted over long distances. It has a lower bandwidth so it can have more stations available in any frequency range.	FM is less prone to interference than AM. However, FM signals are impacted by physical barriers. FM has better sound quality due to higher bandwidth.
Frequency Range	AM radio ranges from 535 to 1705 KHz (OR) Up to 1200 bits per second.	FM radio ranges in a higher spectrum from 88 to 108 MHz. (OR) 1200 to 2400 bits per second.
Bandwidth Requirements	Twice the highest modulating frequency. In AM radio broadcasting, the modulating signal has bandwidth of 15kHz, and hence the bandwidth of an amplitude-modulated signal is 30kHz.	Twice the sum of the modulating signal frequency and the frequency deviation. If the frequency deviation is 75kHz and the modulating signal frequency is 15kHz, the bandwidth required is 180kHz.
Zero crossing in modulated signal	Equidistant	Not equidistant
Complexity	Transmitter and receiver are simple but synchronization is needed in case of SSBSC AM	Transmitter and receiver are more complex as variation of modulating signal has to be

	carrier.	converted and detected from corresponding variation in frequencies.(i.e. voltage to frequency and frequency to voltage conversion has to be done).
Noise	AM is more susceptible to noise because noise affects amplitude, which is where information is "stored" in an AM signal.	FM is less susceptible to noise because information in an FM signal is transmitted through varying the frequency, and not the amplitude.

Audio Compression_ Part I

Digital Audio Source Signal

- The human ear has a dynamic range of about 140 dB and a hearing bandwidth of up to 20 kHz. High-quality audio signals must, therefore, match these characteristics.
- Before the analog audio signals are sampled and digitized, they have to be band-limited by means of a low-pass filter. Then analog-to-digital conversion is performed at a sampling rate of 32 kHz, 44.1 kHz or 48 kHz (and now also at 96 kHz), and with a resolution of at least 16 bits.
- The 44.1 kHz sampling rate corresponds to that of audio CDs, 48/96 kHz are studio quality. While the 32 kHz sampling frequency is still provided for in the MPEG standard, it is in fact obsolete.
- A sampling rate of 48 kHz at 16 bit resolution yields a data rate of 786 kbit/s per channel, which means approximately 1.5 Mbit/s for a stereo signal (*Figure below: Digital audio source signal*).



Digital audio source signal

- The objective of audio compression is to reduce the 1.5 Mbit/s data rate to between about 100 kbit/s and 400 kbit/s.
- MP3 audio files, which are very widely used today, often have a data rate as low as 32 kbit/s.
- Similarly as with video compression, this is achieved by way of redundancy reduction and irrelevance reduction.
- In redundancy reduction, superfluous information is simply omitted; there is no loss of information.
- By contrast, in irrelevance reduction information is eliminated that cannot be perceived at the receiving end, in this case the human ear.
- All audio compression methods are based on a psychoacoustic model, i.e. they make use of the "imperfection" of the human ear to remove irrelevant information from the audio signal.
- The human ear is not capable of perceiving sound events close to strong sound pulses in frequency or in time. This means that, to the ear, certain sound events will mask other sound events of lower amplitude.

Digital Audio Compression

- Digital audio compression is the process of representing a sampled digital audio signal using fewer bits while preserving signal quality.
- Digital audio compression enables more efficient storage and transmission of audio data.
- Efficient compression of high quality digital audio information is a necessity for many of the emerging digital information services.

- Although digital compression of speech audio signals has been used in practice for at least a decade1, compression of generic audio signals has become possible only more recently.
- High quality audio compression can be achieved by using knowledge of the audio signal (the signal source), the human perception mechanism (the signal sink), or both, to avoid transmitting information.
- The strategy is to identify components of the audio signal that are either redundant (i.e. over-described), or are irrelevant to the sink.
- Signal components which are redundant can be removed before transmission and then recovered in the decoder from the remaining information.
- Irrelevant signal components can be dropped completely with no loss to the sink, in a perceptual sense.
- Redundancy and irrelevancy reduction are complementary methods of compression.
- An important difference between these two methods is that using redundancy reduction; the original signal can be reconstructed exactly.
- When irrelevancy is exploited, some information is permanently discarded. Thus schemes using irrelevancy reduction are known as lossy, while redundancy reduction provides lossless compression.
- All perceptual compression schemes are lossy.
- High quality digital audio is broadly defined as having a minimum of 20 kHz audio bandwidth and 96 dB dynamic range.

- Typical digital systems use 44.1 kHz or 48 kHz sampling frequency, and 16 to 20 bits per uncompressed audio sample.

- It is most often presented in one of the following audio channel configurations:

  • Single channel (mono)
  • Two channel (stereo)
  • Four or five channel (multichannel stereo)

Digital Audio Compression Algorithms

- High quality digital audio compression algorithms are a compelling application for digital signal processing (DSP).

- At audio data rates, inexpensive, high performance device technology allows complex algorithms using models of human hearing to optimize compression.

- The basic algorithmic building blocks of compression algorithms are digital signal processing (DSP) operations.

- Fundamental components include linear filtering, and families of fast algorithms such as the Fast Fourier Transform (FFT) and those for the Discrete Cosine Transform (DCT).

- Audio compression algorithms are implemented in software as audio codecs.

- In both lossy and lossless compression, information redundancy is reduced, using methods such as coding, quantization, discrete cosine transform and linear prediction to reduce the amount of information used to represent the uncompressed data.

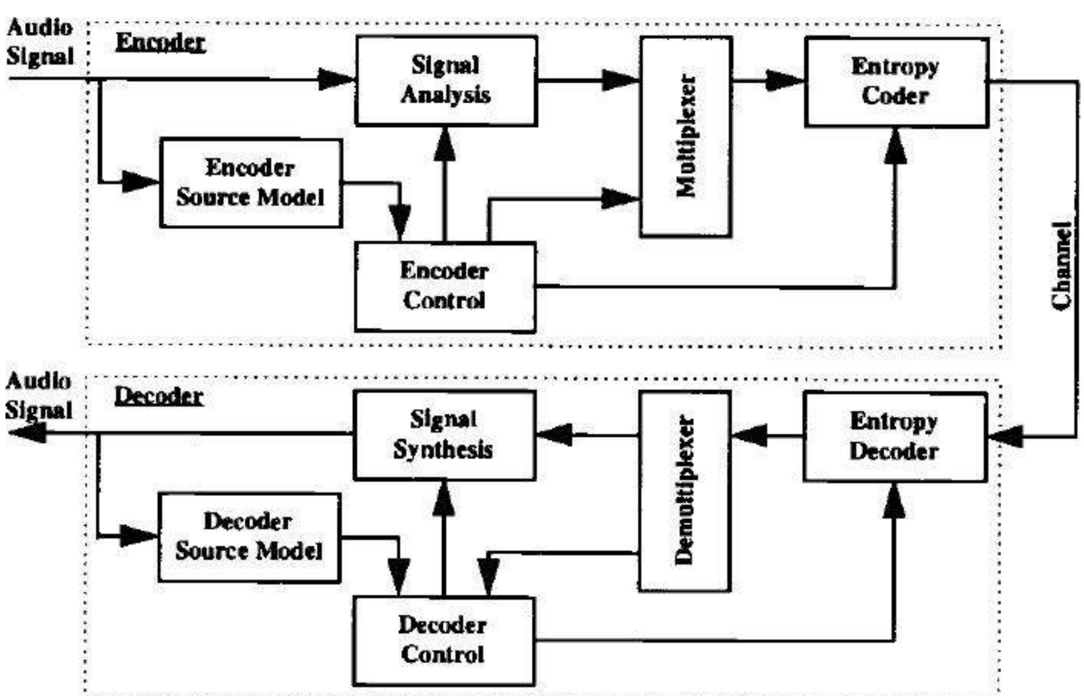
Lossless Audio Compression

- Almost all audio signals contain redundant information.
- Redundancy exists in a digital audio signal when the bits of each sample can be predicted by some algorithm with better accuracy than by a fair coin toss.
- If this is the case then the signal, in a specific sense, contains too many bits for the amount of information it actually conveys, and compression is possible.
- At least two kinds of redundancy are immediately apparent in audio signals. *One* is that signals normally do not have the maximum possible loudness, so samples with small amplitude are more common.
- Most signals also contain identifiable tonal components, and so knowledge of the signal frequency spectrum can be used to predict sample values from previous samples.
- Stereo or multichannel signals have additional redundancy between the audio channels, since these are likely to contain related signals.
- Redundancy can be quantified by an information-theoretic property of the signal, its entropy, presented here in units of bits.
- The entropy H of a random signal x is related to its probability distribution by:

$$H(\boldsymbol{x}) \equiv E[-\log_2 P(\boldsymbol{x})] \qquad (1)$$

- Entropy is strictly a consequence of the source signal statistics.

- It can be interpreted as the smallest number of bits that any lossless compression scheme for *x* can achieve, averaged over all possible signals.
- The maximum compression is obtained by applying a variable-length code, or entropy code, to the signal, such as Huffman or arithmetic coding.
- This is often assisted by transforming the signal to remove correlation between the audio samples, so that the probability distribution has a simpler structure.
- Both these tasks require a model of the audio signal source to estimate $P(x)$.
- The figure below gives a general scheme for lossless audio coding.
- In this and subsequent diagrams, the audio signal is taken to be vector-valued in the case of stereo and multichannel audio.



Structure of a lossless audio codec

- The purpose of the encoder source model is to estimate the audio signal probability distribution.

- This information is used to determine an appropriate signal analysis transformation to de-correlate the audio samples, and an entropy code to apply to the result.

- Control parameters deduced by the encoder can be transmitted as side-information, or derived again using an identical signal model in the decoder if it depends strictly on previous audio data.

- In general the functional blocks in the control path of the codec correspond to processes in the codec implementation.

- However simple codecs may assume some fixed policy for these functions, in which case the blocks represent only algorithm design tasks.

- Purely lossless compression presents several problems when applied to generic audio coding:

• The achievable compression depends on the signal content

• Lack of a widely applicable audio signal source model

• Low signal dimensionality

- Since many physical digital transmission channels have a predetermined constant bit rate limitation, the dependence of compression ratio on the audio signal content is frequently unacceptable.

- Generic audio codecs cannot assume a particular source mechanism, such as human speech, because of the diversity of their applications.
- Effective, widely applicable source models are currently unknown, so models for generic audio are therefore limited at present to those based on measurement of the signal power spectrum, and interchannel correlation in the case of stereo and multichannel coding.
- In general, audio signals contain less redundant information than higher-dimensional signals such as video, so compression ratios are intrinsically lower.
- Because of these factors, it is normally not viable to employ only lossless audio compression. Some lossy compression is highly desirable.

Audio Compression_ Part II

Digital rights management (DRM)

Lossy Audio Compression

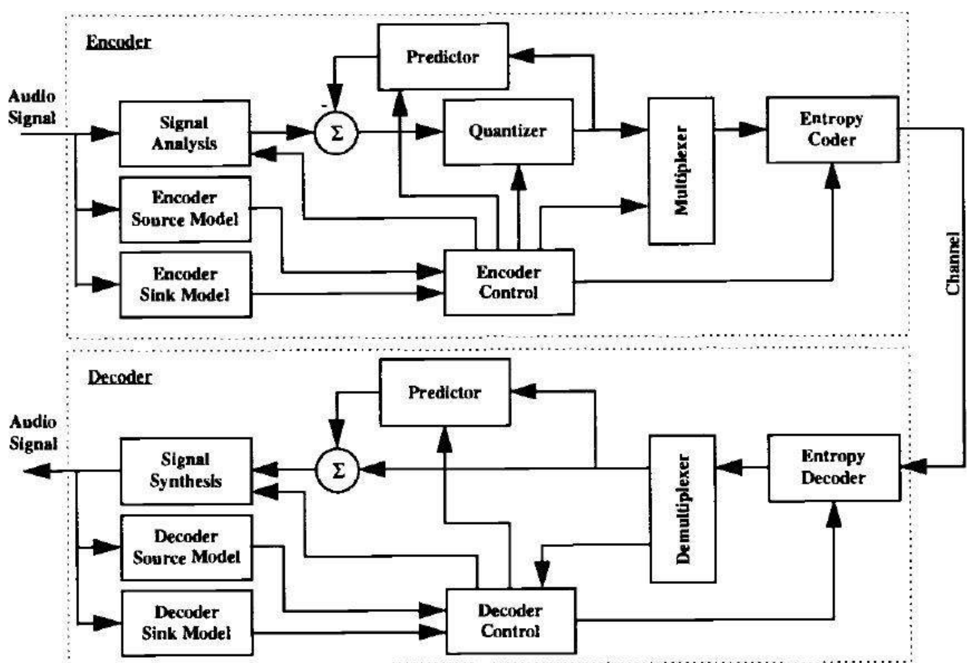
- When lossless compression cannot meet the bit rate requirement the encoder has no choice but to discard information, however the properties of the signal sink can normally be exploited to minimize the effect of this loss.
- Almost all signal sinks detect some information more readily than others.
- In the best case, the source signal contains irrelevant information which is of negligible importance to the sink.
- If all the information the encoder discards is irrelevant, then the compression scheme appears to the sink to be lossless in a perceptual sense.
- This is the goal of high quality generic audio compression.
- For audio the signal sink is normally taken to be the human ear.
- At least two types of irrelevancy are then apparent. One is that a signal may be too quiet to be audible. Another is that the signal may be outside the audible frequency range.
- In stereo and multichannel signals, there is also irrelevant information in the spatial field. These effects typically depend on the input signal.
- Perceptually-based, lossy, generic audio compression has become successful recently for several reasons:

- High compression and predictable bit rate are not adequately addressed by lossless audio compression

- Research into human auditory perception has resulted in computational psychoacoustic models which predict useful irrelevancy

- The computational models are tractable for existing semiconductor device technology and can be implemented cost-effectively using DSP techniques

- A general structure for perceptually-based lossy audio codecs is shown in the figure below.



Structure of a lossy audio codec

- The fundamental difference between a lossy and lossless codec is the inclusion of a quantization stage in the lossy encoder.

- Its function is to replace the signal by an approximation chosen from an agreed subset of the possible source signals.
- There are fewer approximations than signals and therefore less information, so the subsequent lossless compression stage can achieve a lower bit rate.
- The selection of the signal approximation is guided by a psychoacoustic model which predicts the effect of the distortion introduced.
- A numerical optimization algorithm in the encoder chooses an approximation that minimizes the distortion while satisfying bit rate constraints.
- The strategy used is to concentrate the distortion introduced by the approximation into signal features which the ear is least sensitive to.
- The signal analysis and synthesis blocks have a dual role in lossy compression.
- In addition to decollating the input signal in the encoder, the signal synthesis block of the decoder determines the transfer function from the quantizer to the audio signal.
- These functions typically have conflicting requirements.
- The signal prediction block overlaps in function with the signal analysis stage. Its purpose is also to remove correlation in the audio signal.
- However the prediction block differs in that for a well-behaved quantizer, it does not affect the shape of the quantization distortion in the reconstructed audio signal.
- This is primarily because it uses only information that is also available to the decoder.

- This allows the prediction block to implement functions which are prohibited to the signal analysis and synthesis blocks of a lossy codec because of undesirable interactions with the quantization distortion, such as noise amplification.

- The prediction block does, however, strongly affect the audio signal during channel errors, since its internal state can become inconsistent between the encoder and decoder during errors.

- As in a lossless codec, the control information produced by the source and sink models in the encoder can be transmitted to the decoder as side-information, or re-derived in the decoder using identical models. An important difference is that in the lossy codec, the decoder does not have access to information which was discarded by the quantizer.

- Thus any algorithm block in the encoder which must be exactly duplicated in the decoder can depend only on the quantized audio signal.

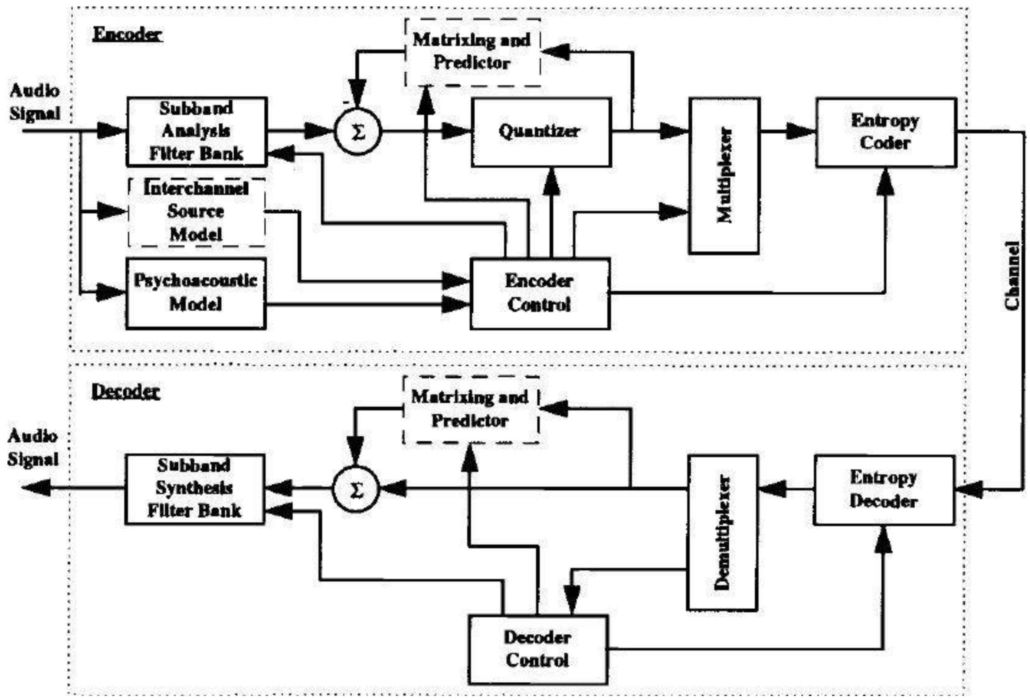
- Finally once again, not all functions are implemented in all compression algorithms, and some functions may implicitly be present only as algorithm design tasks.

Overview of MPEG Audio Algorithm

- The MPEG-1 and MPEG-2 standards define perceptually-based lossy audio compression algorithms.

- In this section we briefly introduce the MPEG Audio algorithms in terms of the previously described structures to facilitate later examples.

- The figure bellow illustrates the structure of an MPEG audio codec.
- The dashed lines indicate functional blocks which are only present in an MPEG-2 codec.



Structure of an MPEG audio codec

- Recall that MPEG-1 Audio is for two-channel stereo, and MPEG-2 Audio for multichannel stereo audio.
- Both algorithms define three levels, termed layers 1, 2 and 3.
- The layers have very similar functionality but differ in coding efficiency, delay, implementation complexity, structure and syntax.
- Layer 3 is the most efficient in terms of achieving the highest compression for the same sound quality, but also the most complex.

- The algorithms support sampling frequencies of 32, 44.1 and 48 kHz, with sample resolutions up to about 18 bits.
- The available transmission bit rates are from 32 to 192 kbit/s per audio channel, with a corresponding drop in audio quality at the low rates.
- MPEG-2 additionally supports up to two-channel stereo operation at half these sample and bit rates for high-quality speech compression.
- The MPEG algorithms are subband-based, meaning that the encoder separates the components of the audio signal according to frequency using a bank of bandpass filters.
- There are 32 subbands per audio channel for layers 1 and 2, and up to 384 subbands per channel for layer 3.
- The component subbands are independently quantized according to a psychoacoustic model present only in the encoder.
- The model predicts the sensitivity of the ear to distortion in each subband.
- The quantization parameters are transmitted to the decoder as side-information.
- Both MPEG-1 and MPEG-2 rely heavily on the irrelevancy predicted by the model, and MPEG-2 additionally defines a source prediction model to exploit redundancy between the audio channels.
- Layers 1 and 2 have very simple entropy coding schemes and provide a constant transmission bit rate output.

- Layer 3 includes extensive Huffman coding and also provides a constant bit rate, but allows buffering to absorb short-term peaks in the bit rate demand.

Digital rights management (DRM)

- Digital rights management is the management of legal access to digital content.
- Various tools or technological protection measures (TPM) such as access control technologies can restrict the use of proprietary hardware and copyrighted works.
- DRM technologies govern the use, modification, and distribution of copyrighted works (such as software and multimedia content), as well as systems that enforce these policies within devices.
- Laws in many countries criminalize the circumvention of DRM, communication about such circumvention, and the creation and distribution of tools used for such circumvention.
- Such laws are part of the United States' Digital Millennium Copyright Act (DMCA), and the European Union's Information Society Directive (the French DADVSI is an example of a member state of the European Union implementing the directive).
- DRM techniques include licensing agreements and encryption.
- The industry has expanded the usage of DRM to various hardware products, such as Keurig's coffeemakers, Philips' light bulbs, mobile device power chargers, and John Deere's tractors.
- For instance, tractor companies try to prevent farmers from making repairs under via DRM.

- DRM users argue that the technology is necessary to protect intellectual property, just as physical locks prevent personal property from theft, that it can help the copyright holder maintain artistic control, and to support licensing modalities such as rentals.
- DRM is not without controversy.
- Critics of DRM contend that no evidence proves that DRM helps prevent copyright infringement, arguing that it serves only to inconvenience legitimate customers, and that DRM can stifle innovation and competition.
- Furthermore, works can become permanently inaccessible if the DRM scheme changes or if a required service is discontinued.
- DRM technologies have been criticized for restricting individuals from copying or using the content legally, such as by fair use or by making backup copies.
- DRM is in common use by the entertainment industry (*e.g.*, audio and video publishers).
- Many online stores such as OverDrive, use DRM technologies, as do cable and satellite service operators.
- Apple removed DRM technology from iTunes around 2009.
- Typical DRM also prevents lending materials out through a library, or accessing works in the public domain.