CHAPTER 6
AUDIO COMPRESSION

hungtanle@0913030731

WHAT IS AUDIO ?

Audio is the sound system that comes with or can be added to a computer.

An audio file is a record of captured sound that can be played back.

An audio card contains a special built-in processor and memory for processing audio files and sending them to *speakers* in the computer.

2

WHAT DO COMPRESSION MEAN ?

What is ?

Compression is performed by a program that uses an algorithm to determine how to compress or decompress data.

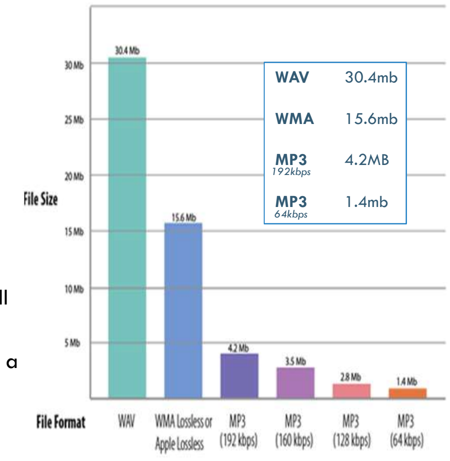
Why ?

✓Reduce File Size
✓Save disk space
✓Reduce transmission time

AUDIO COMPRESSION

Audio compression is a form of data compression designed to reduce the size of audio files.

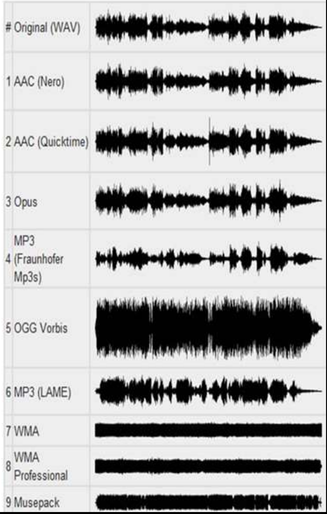
- the audio file must still be playable after compression
- If the file is compressed 'too much' there will be loss of quality.
- The compression is done with a thing called a codec (compressor and decompressor)
- This codec is a special algorithm to reduce the size.



4

SOME AUDIO FORMATS

- Wav: WAVE form audio file (Microsoft & IBM)
- MP3: is an audio-specific format that was designed by MPEG as part of its MPEG-1, MPEG-2.
- Wma: Windows Media Audio
- Aac: Advanced Audio Coding/is the default or standard audio format for YouTube,iPhone, iPod ,iPad
- Flac: Free Lossless Audio Codec
- Ogg: Ogg is a free



INTERNATIONAL ORGANIZATIONS DEALING WITH AUDIO COMPRESSION STANDARDIZATION

ISO/IEC: International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC)

MPEG-1 Layer III (MP3)

MPEG-1 Layer II MPEG-1 Layer I

AAC

MPEG-4 ALS MPEG-4 SLS

MPEG-D USAC

ITU International Telecommunication Union

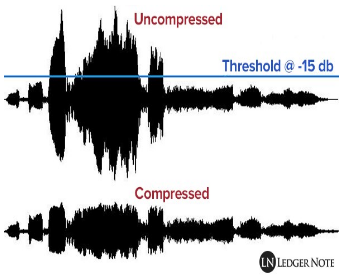
ITU became a specialized agency of the United Nations in 1947. The International Telegraph and Telephone Consultative Committee was created in 1956, and was renamed ITU-T in 1993

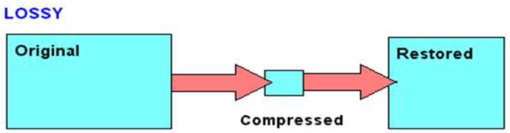
G.711	G.718	G.719
G.722	G.723	G.726
G.728	G.729	

6

TYPES OF COMPRESSION
LOSSY COMPRESSION

- A compression technique that does not decompress data back to 100% of the original.
- Lossy methods provide high degrees of compression and result in smaller compressed files, but there is a certain amount of visual loss when restored.





7

AUDIO LOSSY COMPRESSION

- Audio can tolerate loss, it may not be noticeable to the human ear.
- In other cases, it may be noticeable, but not that critical to the application.
- The more tolerance for loss, the smaller the file can be compressed, and the faster the file can be transmitted over a network.

8

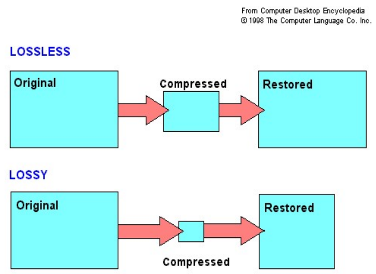
TYPES OF COMPRESSION
LOSSY COMPRESSION

The most popular type of audio compression is lossy due to encoded files being smaller while still acceptable sound quality.

- difficult to compress using lossless methods, except for special cases.
- Some compression can be obtained by run-length encoding samples that fall below a threshold that can be considered to represent silence.
- Companding uses non-linear quantization to compress speech.
  - μ-law and A-law companding are used for telephony

LOSSLESS COMPRESSION

- A compression technique that decompresses data back to its original form without any loss.
- The decompressed file and the original are identical.
- For example, the ZIP archiving technology (WinZip...) is the most widely used lossless method.



LOSSLESS COMPRESSION

- Lossless audio files typically require more storage space than lossy encoded ones.
- However this type of format is often favored by users wanting to backup original audio CDs.
- a perfect copy can be restored in the event of loss or damage to the CD.
- FLAC, Apple Lossless (ALAC) and WMA Lossless are examples of lossless compression formats.

kind	Apple Lossless audio file
duration	23:03
size	366.8 MB
bit rate	2,224 kbps
sample size	24 bit
sample rate	88.200 kHz

STANDARD CODECS
FOR LOSSLESS COMPRESSION

- LPAC (Lossless predictive audio compression):is an improved lossless audio compression algorithm developed by Tilman Liebchen, Marcus Purat and Peter Noll.
- ALAC (Apple Lossless Audio Codec):is an audio coding format, and its reference audio codec implementation, developed by Apple.
- FLAC(Free Lossless Audio Codec): can typically reduce the original size of audio file to 50–60%, developed by Josh Coalson.
- WMA Lossless (Windows Media Audio Lossless): developed by Microsoft.

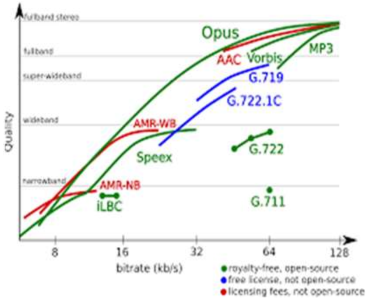
AUDIO COMPRESSION

Lossless Audio Compression

- Apple Lossless
- apt-X Lossless
- Audio Lossless Coding
- Direct Stream Transfer (DST)
- Dolby TrueHD
- DTS-HD Master Audio
- Free Lossless Audio Codec

Lossy Audio Compression

- Nero AAC Codec(Nero "advanced audio coding" codec): Nero AG.
- FAAC(Freeware Advanced Audio Coder):
- ADPCM ATRAC
- Dolby AC-3
- MP3



MPEG

MOTION PICTURE EXPERTS GROUP

WHAT IS MPEG/AUDIO COMPRESSION?

MPEG/audio is the international standard for digital compression of high-fidelity audio.

MPEG/audio is the result of over 3 years of research by the Motion Picture Experts Group.

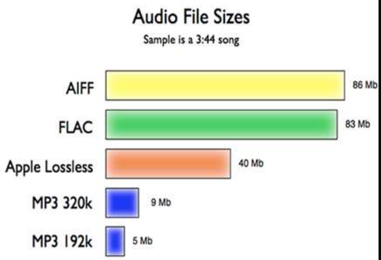
MPEG/audio standard adopted by International Organization for Standards and the International Electrotechnical Commission (ISO/IEC) in 1992.

MPEG AUDIO CHARACTERISTICS

- An ISO standard for high-fidelity audio compression.
- A lossy compression, but losses are perceptually irrelevant
- Compresses without regard to the source of the audio data.
- Removes distortion and other features that are imperceptible to the human ear.
- 6-to-1 compression ratio.

WHAT IS MPEG/AUDIO COMPRESSION?

- The standard defines strict rules to ensure inter-operability.
- Defines the syntax for encoding and decoding.
- Tests accuracy of decoding.
- Ensures any fully compliant decoder can decode any MPEG/audio file with a predictable result.



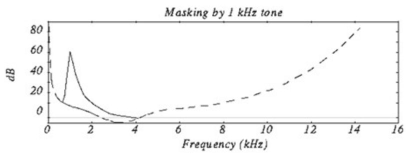
Three different layers of compression:

- Layer I
- Layer II
- Layer III

HOW DOES MP3 SAVE SO MUCH SPACE?

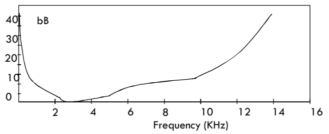
- Compression algorithms used our knowledge of *psychoacoustics* to manage the data bandwidth.
- Psychoacoustics refers to how our brain interprets sounds.
- The brain uses certain tricks like auditory masking to allocate resources and attention to what is the most important sound happening at any given time.
- The input audio stream simultaneously passes through a psychoacoustic model.

THRESHOLD OF HEARING



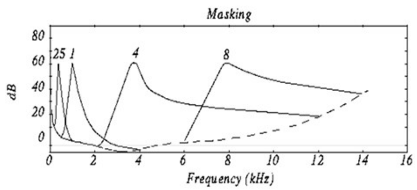
Experiment1: *Put a person in a quiet room. Raise level of 1 kHz tone until just barely audible. Vary the frequency and plot*

- The ear is most sensitive to frequencies between 1 and 5 kHz, where we can actually hear signals below 0 dB.
- Two tones of equal power and different frequencies will not be equally loud.
- Sensitivity decreases at low and high frequencies.

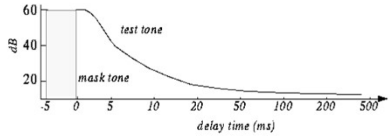


Experiment2: *Play 1 kHz tone (masking tone) at fixed level (60 dB). Play test tone at a different level (e.g., 1.1 kHz), and raise level until just distinguishable.*

Vary the frequency of the test tone and plot the threshold when it becomes audible:



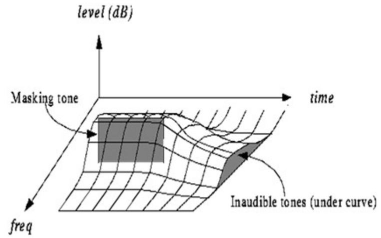
TEMPORAL MASKING



If we hear a loud sound, and then it stops, it takes a little while until we can hear a soft tone nearby (in frequency).

Experiment 3:

- *Play 1 kHz masking tone at 60 dB, plus a test tone at 1.1 kHz at 40 dB. Test tone can't be heard (it's masked).*
- *Stop masking tone, then stop test tone after a short delay.*
- *Adjust delay time to the shortest time when test tone can be heard (e.g., 5 ms).*
- *Repeat with different level of the test tone and plot:*



PSYCHOACOUSTIC MODEL

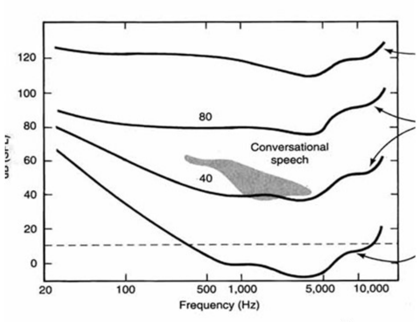
Psychoacoustics is the study of human perception of sound.

Psychoacoustics deals with relations between perception of sound and physical properties of sound waves.

How sensitive is human hearing?

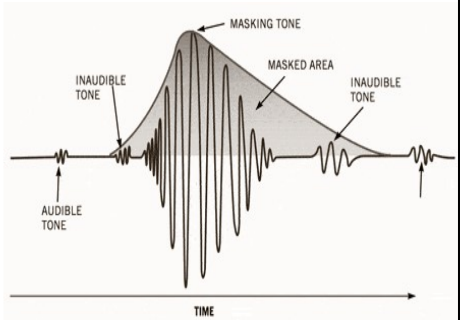
To answer this question we look at the following concepts:

- Threshold of hearing
  - Describes the notion of "quietness"
- Frequency Masking
  - A component (at a particular frequency) masks components at neighboring frequencies. Such masking may be partial.
- Temporal Masking
  - When two tones (samples) are played closed together in time, one can mask the other.



TEMPORAL MASKING

- if two sound events occur within milliseconds of each other, we're only going to be able to focus on the loudest one.
- It's how we've been evolutionarily primed to react. Our ears and minds can't separate events that close in time.
- So what the encoder algorithm does is ignore or at least allocate much less data to the quieter sound since we won't perceive it anyways

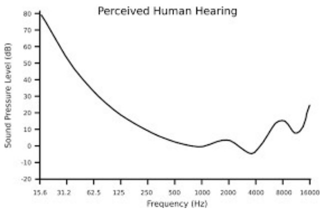


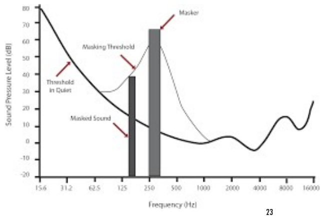
DE-EMPHASIZE THE QUIET

This refers to something our ears and brains do called *simultaneous masking*.

if a loud sound is blaring out over the top of a lot of low-volume sounds, you're naturally going to focus on the loud sound.

we can spend lot less data on the quiet sounds. They don't need as much detail encoded in them during those times.



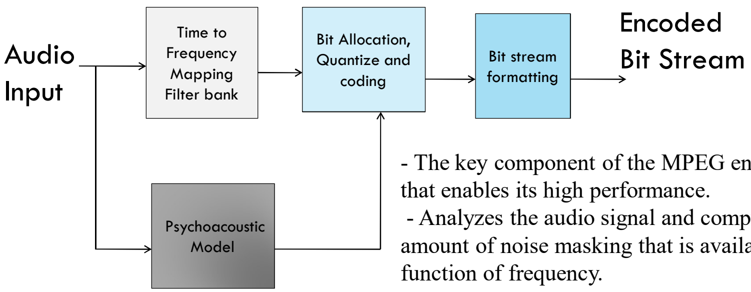


APPLING MINIMUM AUDITION THRESHOLD

The minimum audition threshold refers to volume.

- As a voice or sound becomes quieter and quieter, we're able to make out less and less detail.
- The encoder knows this and chooses to not save every single detail of quiet sounds since we can't use it anyways.
- And if a sound dips below a certain volume threshold where the human ear can't hear it, then it gets tossed out completely.

MPEG ENCODER ARCHITECTURE



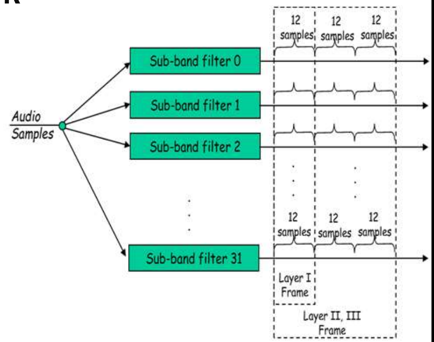
- The key component of the MPEG encoder that enables its high performance.
- Analyzes the audio signal and computes the amount of noise masking that is available as a function of frequency.

25

THE POLYPHASE FILTER BANK

The filter bank divides the audio signal into 32 equal-width frequency subbands.

- The equal widths of the subbands do not accurately reflect the human auditory system's frequency dependent behavior.
- The filter bank and its inverse are not lossless transformations.
- adjacent filter bands have a major frequency overlap*(A signal at a single frequency can affect two adjacent filter bank outputs)*



EQUATION FOR THE FILTER BANK OUTPUTS

$$s_t[i] = \sum_{k=0}^{63} \sum_{j=0}^{7} M[i][k] * (C[k+64j] * x[k+64j])$$

i is the subband index and ranges from 0 to 31,

st[i] is the filter output sample for subband i at time t, where t is an integer multiple of 32 audio sample intervals,

C[n] is one of 512 coefficients of the analysis window defined in the standard,

x[n] is an audio input sample read from a 512 sample buffer, and

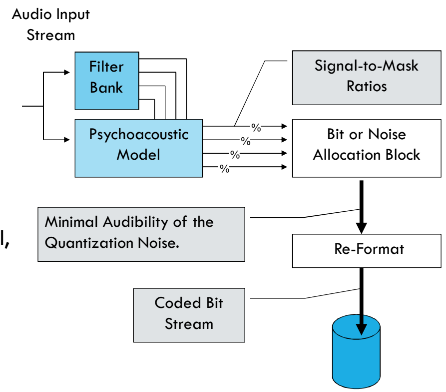
M[i][k] = cos[((2*i+1)*(k-16)*p) / 64] are the analysis matrix coefficients.

MP3

MPEG audio coding is comprised of three independent layers.

Each layer is a self-contained SBC (Sub-Band Coding) coder with its own time-frequency mapping, psychoacoustic model, and quantizer.

- Layer I: Uses sub-band coding
- Layer II: Uses sub-band coding (longer frames, more compression)
- Layer III: Uses both sub-band coding and transform coding.

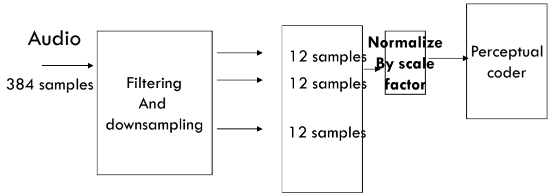


MPEG-1 Audio is intended to take a PCM audio signal sampled at a rate of 32, 44.1 or 48 kHz, and encode it at a bit rate of 32 to 192 kbps per audio channel (depending on layer).

MPEG AUDIO LAYER 1

MPEG (1 and 2) audio allows sampling rate at 44.1 48, 32, 22.05, 24 and 16KHz.

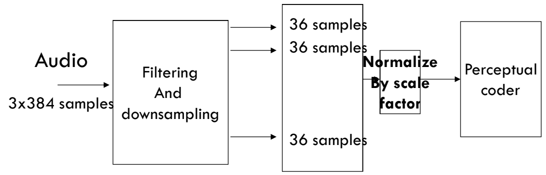
MPEG filters the input audio into 32 bands.



MPEG AUDIO LAYER 2

- Layer 2 is very similar to Layer 1, but groups 3 x 12-samples together in coding.
- Improvement of the scaling factor quantization and also groups 3 audio samples together in bit assignment.

The Layer II encoder forms frames of 1152 samples per audio channel



LAYER3 CODING

Layer	Application
I	Digital audio cassette
II	Digital audio broadcast
III	CD quality

Alias reduction. Layer III specifies a method of processing the MDCT (Modified Discrete Cosine Transform) values to remove some artifacts

Non uniform quantization.

Scalefactor bands. Layer III uses scalefactor bands.

Entropy coding of data values. Layer III uses variable-length Huffman codes to encode the quantized samples to get better data compression.

Use of a "bit reservoir". The design of the Layer III bitstream better fits the encoder's time-varying demand on code bits

MPEG AUDIO MODES

Supports one or two audio channels in one of the four modes:

1. Monophonic -- single audio channel
2. Dual-monophonic -- two independent channels, e.g., English and French
3. Stereo -- for stereo channels that share bits, but not using Joint-stereo coding
4. Joint-stereo -- takes advantage of the correlations between stereo channels

32

EFFECTIVENESS OF MPEG AUDIO

Layer	Target bit-rate	Ratio	Quality* at 64 kbps	Quality at 128 kbps
Layer I	192 kbps	4:1	--	--
Layer II	128 kbps	6:1	2.1 to 2.6	4+
Layer III	64 kbps	12:1	3.6 to 3.8	4+

*Quality factor:

- 5 – perfect
- 4 - just noticeable
- 3 - slightly annoying
- 2 – annoying
- 1 - very annoying

Standard	Audio Sampling rate(kHz)	Compressed bit rate (kbits/sec)	Channels
MPEG - 1 Layer I	32, 44.1, 48	32- 448	1- 2 channels
MPEG - 1 Layer I	32, 44.1, 48	32 – 384	1-2 channels
MPEG - 1 Layer III	32, 44.1, 48	32 – 320	1-2 channels
MPEG - 2 Layer I	32, 44.1, 48	32 - 448	1 – 5.1 channels
	16, 22.05, 24	32 - 256	
MPEG - 2 Layer II	32, 44.1, 48	32 - 384	1 – 5.1 channels
	16, 22.05, 24	8 - 160	
MPEG - 2 Layer III	32, 44.1, 48	32 - 384	1 – 5.1 channels
	16, 22.05, 24	8 - 160	

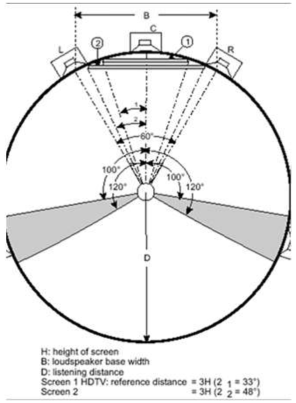
33

FUTURE MPEG/AUDIO STANDARDS PHASE 2/ MPEG-2

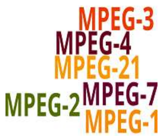
MPEG-2 audio became an international standard in November 1994.

This further extends the original MPEG/audio standard, widely used in DVD and Digital TV

- Multichannel audio support: 5 high frequency channels and 1 low frequency enhanced channel (5.1 channels) usable for High Definition Television and digital movies.
- Multilingual audio support: 7 additional channels for commentary.



EVOLUTION OF MPEG

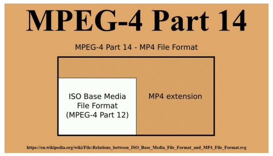


MPEG-3

- Originally developed for HDTV, but abandoned when MPEG-2 was determined to be sufficient

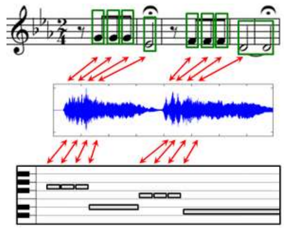
MPEG-4

- Includes support for AV "objects", 3D content, low bitrate encoding, and DRM
- In practice, provides equal quality to MPEG-2 at a lower bitrate, but often fails to deliver outright better quality
- MPEG-4 Part 10 is H.264, which is used in HD-DVD and Blu-Ray



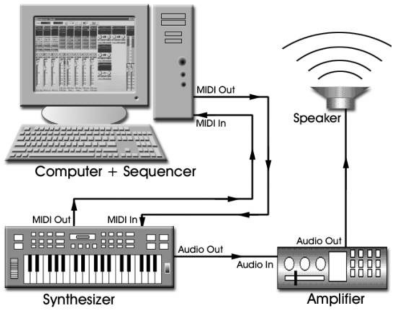
MIDI AUDIO

MIDI is a protocol designed for recording and playing back music on digital synthesizers



MUSICAL INSTRUMENT DIGITAL INTERFACE

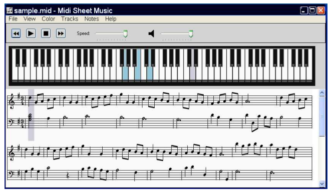
- Provides a standardized and efficient means of conveying musical performance information as electronic data.
- we have knowledge of musical instrument and composing
- It is in the form of music score and not samples or recording.

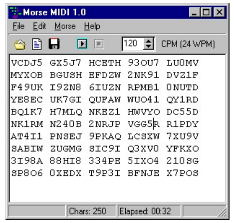


MIDI AUDIO: REQUIREMENTS

To make MIDI score, we need:

1. Midi keyboard / Midi keyboard software
2. Sequencer software
3. Sound synthesizer (built-in in to sound card)





MIDI COMPONENT

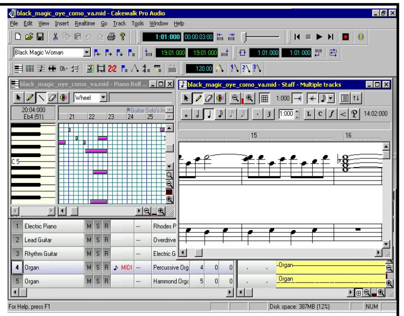
A **MIDI sequencer** software lets us to record and edit MIDI data like a word processor

- Cut and paste
- Insert / delete

Recording MIDI Files

MIDI files can be generated:

- by recording the MIDI data from a MIDI instrument (electronic keyboard) as it is played.
- by using a MIDI sequencer software application.



MIDI
***.MID, *.KAR, *.MIDI, *.SMF**

ADVANTAGES/DISADVANTAGES OF MIDI

- MIDI files smaller that digital audio files.
- MIDI files embedded in web pages load and play more quickly.
- If MIDI sound source are high quality – sound better.
- Can change the length of MIDI files without changing the pitch of the music or degrading the audio quality.

MIDI data does not represent the sound but musical instruments, playback will be accurate only if the MIDI playback (instrument) is identical to the device used in the production.

- Higher cost and requires skill to edit.
- Cannot emulate voice, other effects.

SUMMARY: ADDING SOUND TO MM PROJECT

- **File formats** compatible with multimedia authoring software being used along with delivery mediums, must be determined.
- Sound **playback capabilities** offered by end user's system must be studied.
- The **type of sound**, whether background music, special sound effects, or spoken dialog, must be decided.
- **Digital audio or MIDI** data should be selected on the basis of the location and time of use.

SUMMARY

- audio compression is a key technology
- many algorithms ➔ many applications
- Better algorithms ➔ better quality, more compression

The MPEG audio standard has three layers of successive complexity for improve compression performance.

MPEG audio has been extended in following ways

- Multichannel audio support
- compressed audio rates
- Lower audio sampling rates

42