MPEG

MPEG –selected standards

https://mpeg.chiariglione.org/

MPEG-1: Video CD and MP3.

MPEG-2: Digital Television and DVD

MPEG-4: Web (streaming media) and CD distribution, voice (telephone, videophone) and broadcast television applications

MPEG-7: Description and search of audio and visual content

MPEG-21: Multimedia Framework

MPEG-1
MPEG Moving Picture Experts Group

MPEG-1 is a standard for loss compression of video and audio.

The standard consists of the following five parts:

1. Systems (storage and synchronization of video, audio, and other data together)
2. Video (compressed video content)
3. Audio (compressed audio content)
4. Conformance testing (testing the correctness of implementations of the standard)
5. Reference software (example software showing how to encode and decode according to the standard)

MPEG Audio

MPEG audio compression takes advantage of psychoacoustic models, constructing a large multi-dimensional lookup table to transmit masked frequency components using fewer bits.

- Applies a filter bank to the input to break it into its frequency components
- In parallel, a psychoacoustic model is applied to the data for bit allocation block
- The number of bits allocated are used to quantize the info from the filter bank – providing the compression

MPEG 1 Layers

1. Layer 1 quality can be quite good provided a comparatively high bit-rate is available. Digital Audio Tape typically uses Layer 1 at around 192 kbps
2. Layer 2 has more complexity; was proposed for use in Digital Audio Broadcasting
3. Layer 3 (MP3) is most complex, and was originally aimed at audio transmission over ISDN lines

Most of the complexity increase is at the coder, not the decoder – accounting for the popularity of MP3 players

MPEG Audio

Basic MPEG Audio Strategy

MPEG approach to compression relies on:

Human auditory system

Quantization

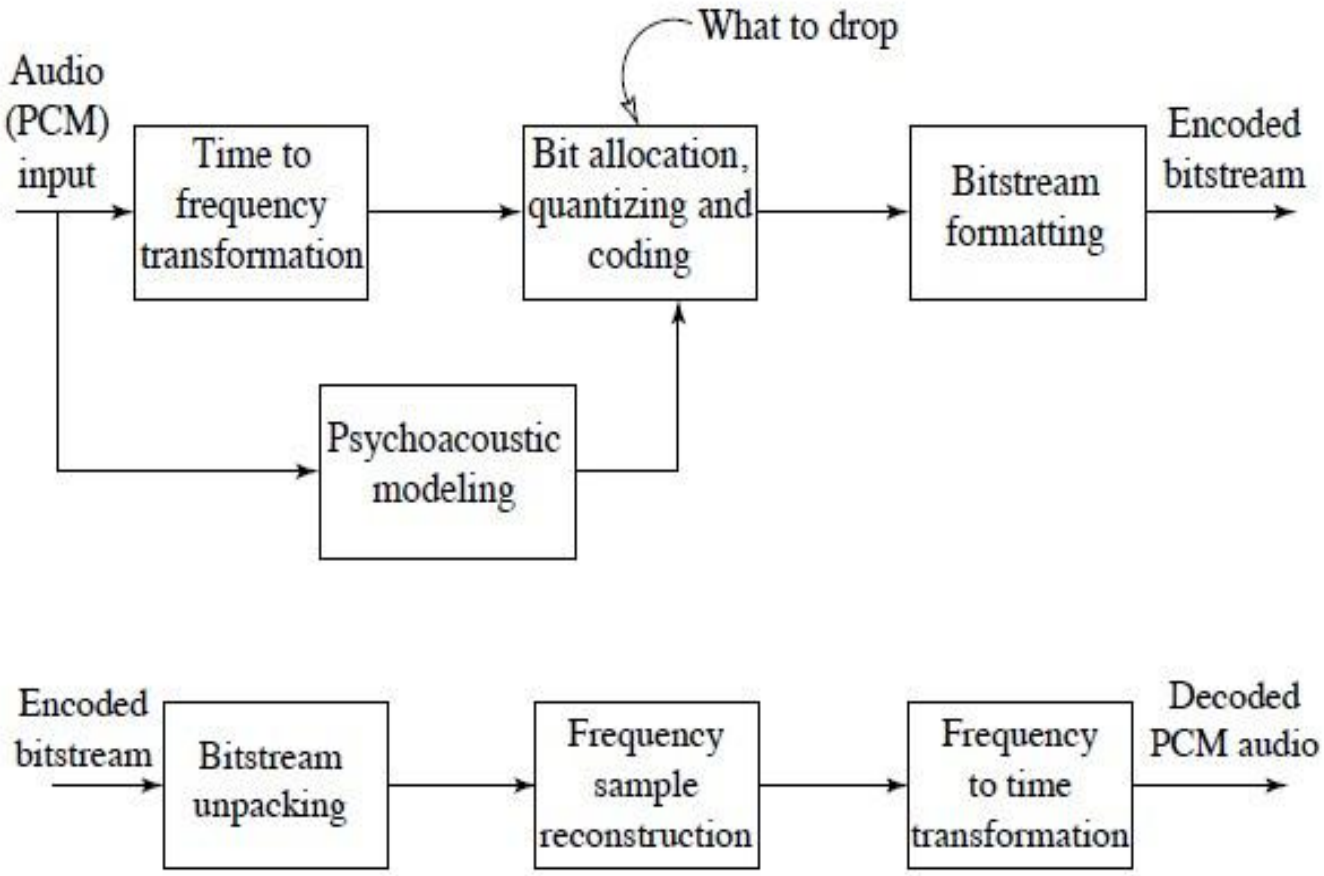
MPEG encoder employs a bank of filters to:

Analyse the frequency (“spectral”) components of the audio signal by calculating a frequency transform of a window of signal values.

Decompose the signal into sub-bands by using a bank of filters
Layer 1 & 2: “quadrature-mirror” filters;

Layer 3: adds a DCT; psychoacoustic model: Fourier transform

Basic MPEG Audio encoder and decoder



Basic MPEG Audio encoder and decoder

1. Use convolution filters to divide the audio signal into 32 frequency sub-bands. (sub-band filtering)

2. Determine amount of masking for each band caused by nearby band using the psychoacoustic model .

3. If the power in a band is below the masking threshold, don't encode it.

4. Otherwise, determine number of bits needed to represent the coefficient such that, the noise introduced by quantization is below the masking effect (Recall that one fewer bit of quantization introduces about 6 dB of noise).

5. Format bitstream

Audio compression

Audio Processing - Motivation

- compression - storage and retrieval
- digital audio effects (delay, equalization, noise reduction, time compression and expansion, pitch shifting ...)
- conversion
- reshape impulse response (to simulate a different room)
- move perceived location from which sound comes
- locate speaker in 3D space using microphone arrays
- cover missing samples
- mix multiple signals (i.e. conference)
- echo cancellation

Why Compression ?

Data rate = sampling rate * quantization bits * channels (+ control information)

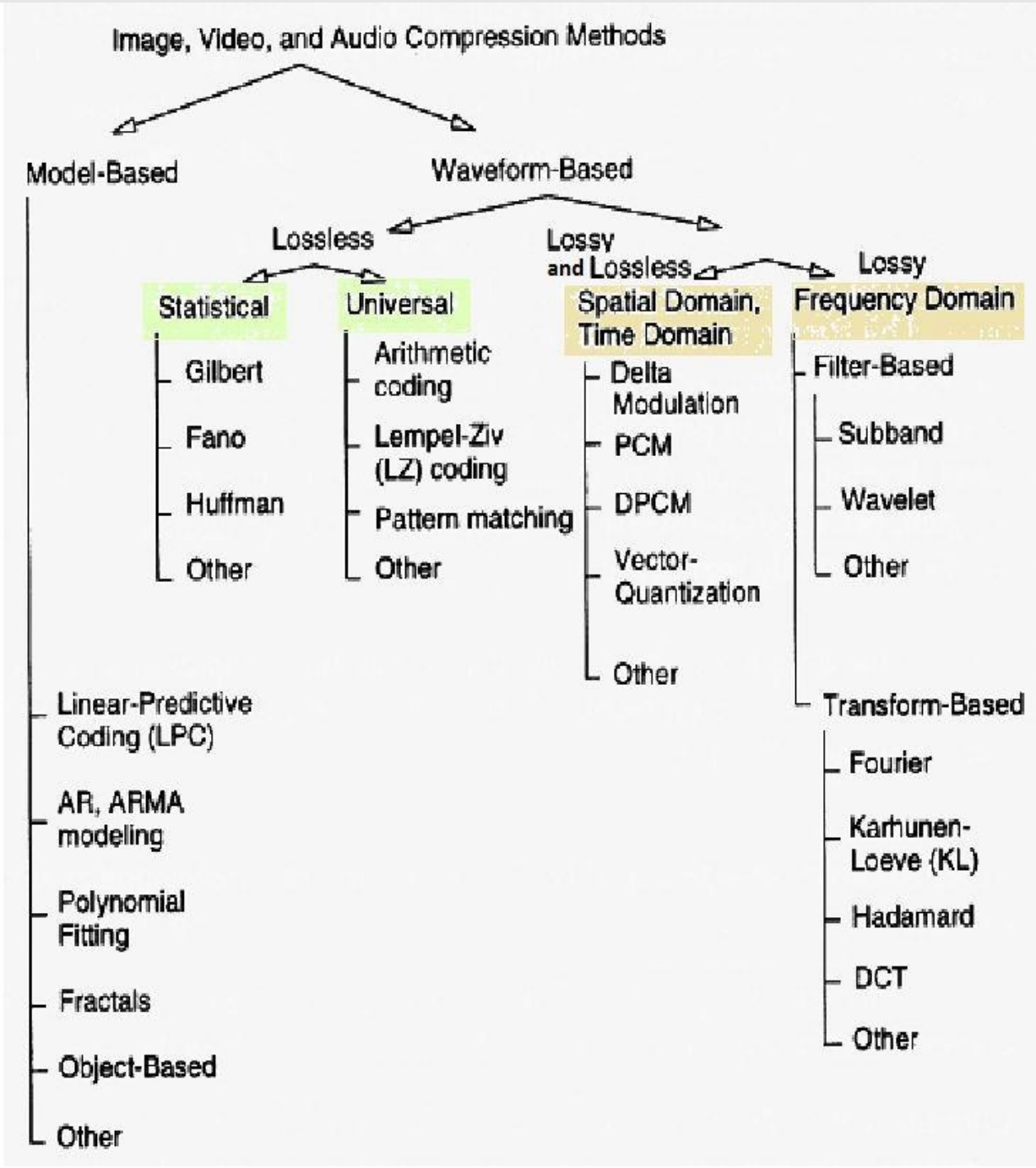
For example (digital audio):

44100 Hz; 16 bits; 2 channels

generates about 1.4Mb of data per second;

84Mb per minute; 5Gb per hour

Compression Taxonomy



Audio Data Compression

Redundant information

- Implicit in the remaining information
- Example. oversampled audio signal

Irrelevant information

- Perceptually insignificant

Audio Data Compression

Lossless Audio **Compression**

- Removes redundant data

Lossy Audio **Encoding**

- Removes irrelevant data

Perceptual audio coding

Human Auditory System
Perceptual audio coding

1. Discard weaker signal if a stronger one exists in the same frequency band (frequency-domain masking)

2. Discard soft sound after a loud sound (time-domain masking)

3. Different quantization for different critical frequency bands Sub-band coding (If you can't hear the sound, don't encode it...)

4. Stereo redundancy: At low frequencies, we can't detect where the sound is coming from. Encode it mono.

Generic Audio Encoder Psychoacoustic Model

Psychoacoustics – study of how sounds are perceived by humans

Uses **perceptual coding**

- eliminate information from audio signal that is inaudible to the ear

Detects conditions under which different audio signal components **mask** each other

Psychoacoustic Model

Signal Masking:

- Threshold cut-off
- Spectral (Frequency / Simultaneous) Masking
- Temporal Masking

Threshold cut-off and spectral masking occur in frequency domain,

temporal masking occurs in time domain

Signal Masking

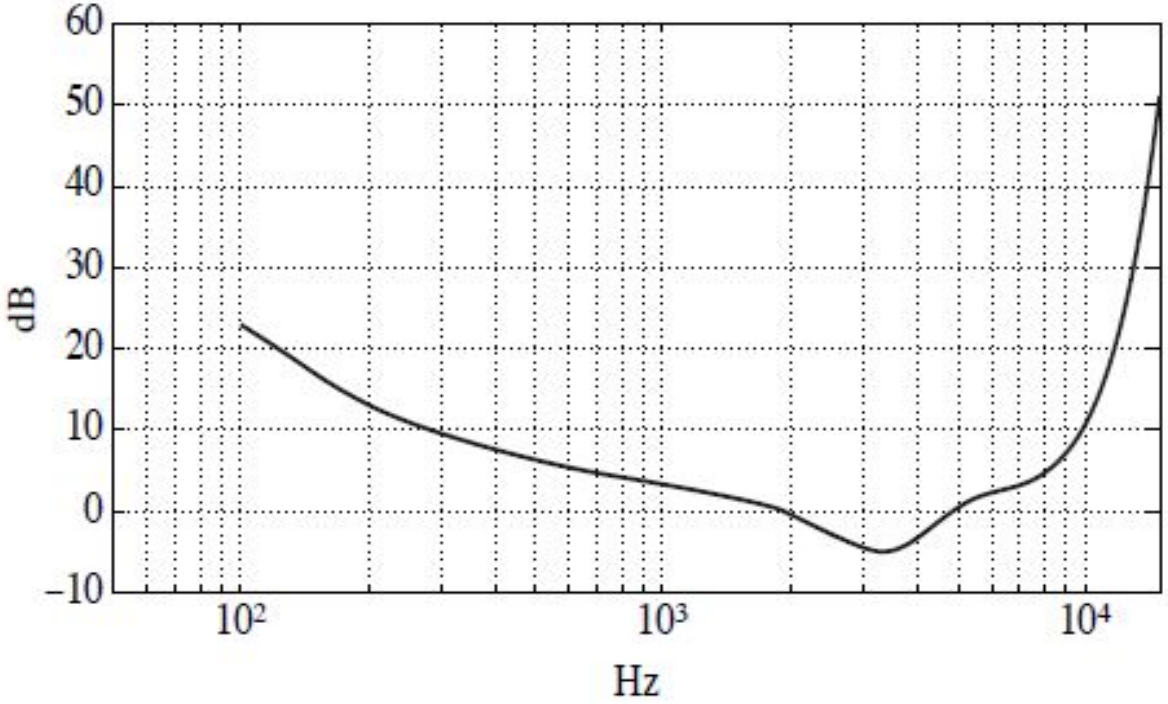
Threshold cut-off

- Hearing threshold level – a function of frequency
- Any frequency components below the threshold will not be perceived by human ear

Human Auditory System
Range of human' hearing: 20Hz – 20kHz
Dynamic range ~ 120 dB

Threshold of Hearing

- A plot of the threshold of human hearing for a pure tone



Threshold of human hearing, for pure tones

Threshold of Hearing

The threshold of hearing curve: if a sound is above the dB level shown then the sound is audible

Turning up a tone so that it equals or surpasses the curvem eans that we can then distinguish the sound

An approximate formula exists for this curve:

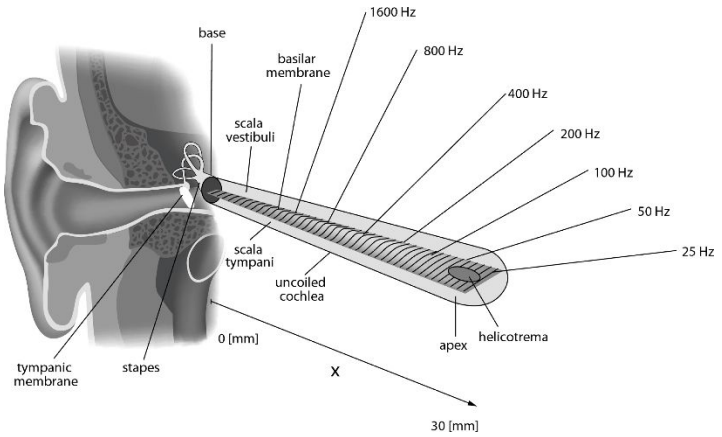
$$\text{Threshold}(f) = 3.64(f/1000)^{-0.8} - 6.5e^{-0.6(f/1000-3.3)^2} + 10^{-3}(f/1000)^4$$

The threshold units are dB; the frequency for the origin (0,0) in formula above is 2000 Hz: Threshold(f) = 0 at f =2 kHz

Loudness and Pitch

More sensitive to loudness at mid frequencies than at other frequencies intermediate frequencies at [500hz, 5000hz]

Perceived loudness of a sound changes based on frequency of that sound basilar membrane reacts more to intermediate frequencies than other frequencies.



Human Auditory System
Equal-Loudness Relations

Fletcher-Munson Curves

– Equal loudness curves that display the relationship between

perceived loudness ("Phons", in dB) for a given stimulus sound volume ("Sound Pressure Level", also in dB), as a function of frequency

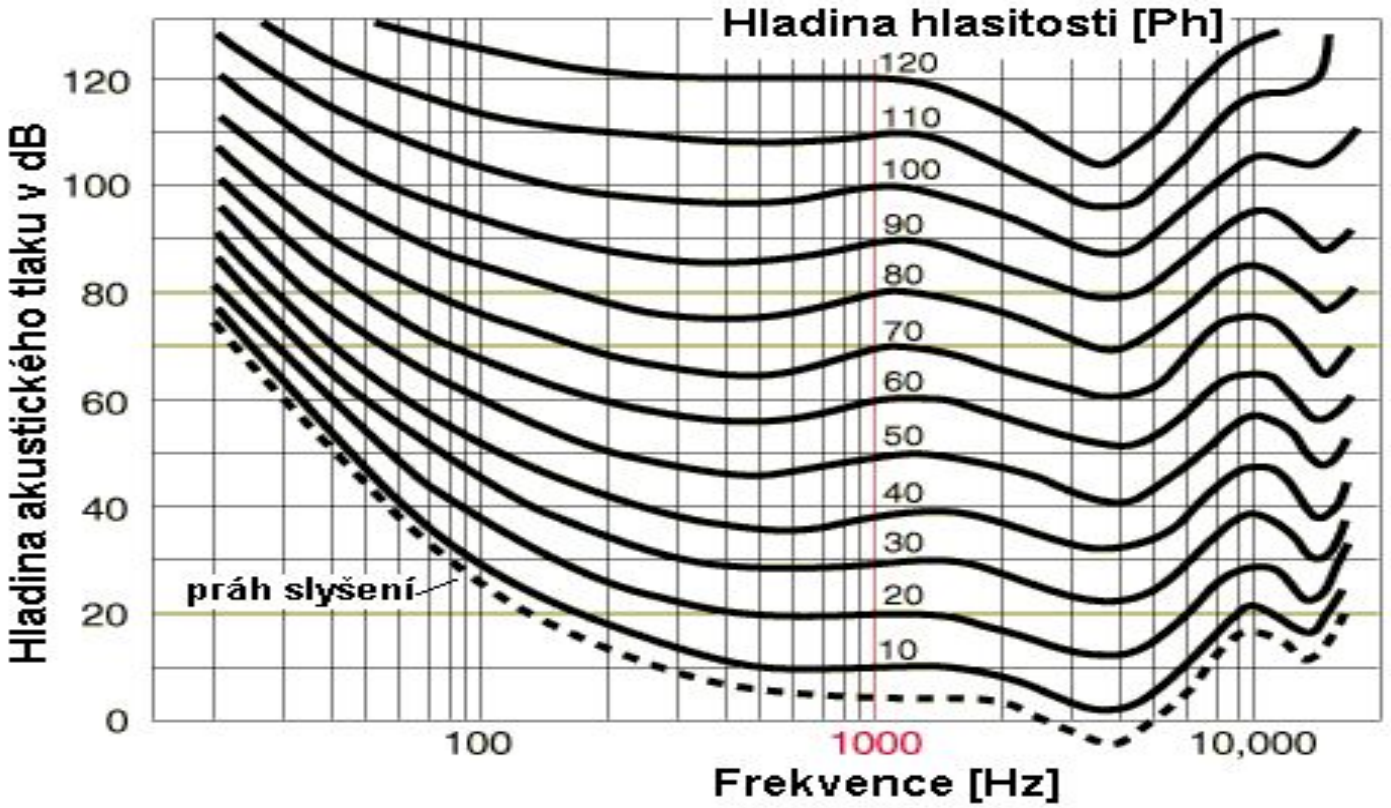
Figure on next slide shows the ear's perception of equal loudness:

The bottom curve shows what level of pure tone stimulus is required to produce the perception of a 10 dB sound

All the curves are arranged so that the perceived loudness level gives the same loudness as for that loudness level of a pure tone at 1 kHz

Fletcher-Munson Contours

Perception sensitivity (loudness) is not linear across all frequencies and intensities



Each contour represents an equal perceived sound

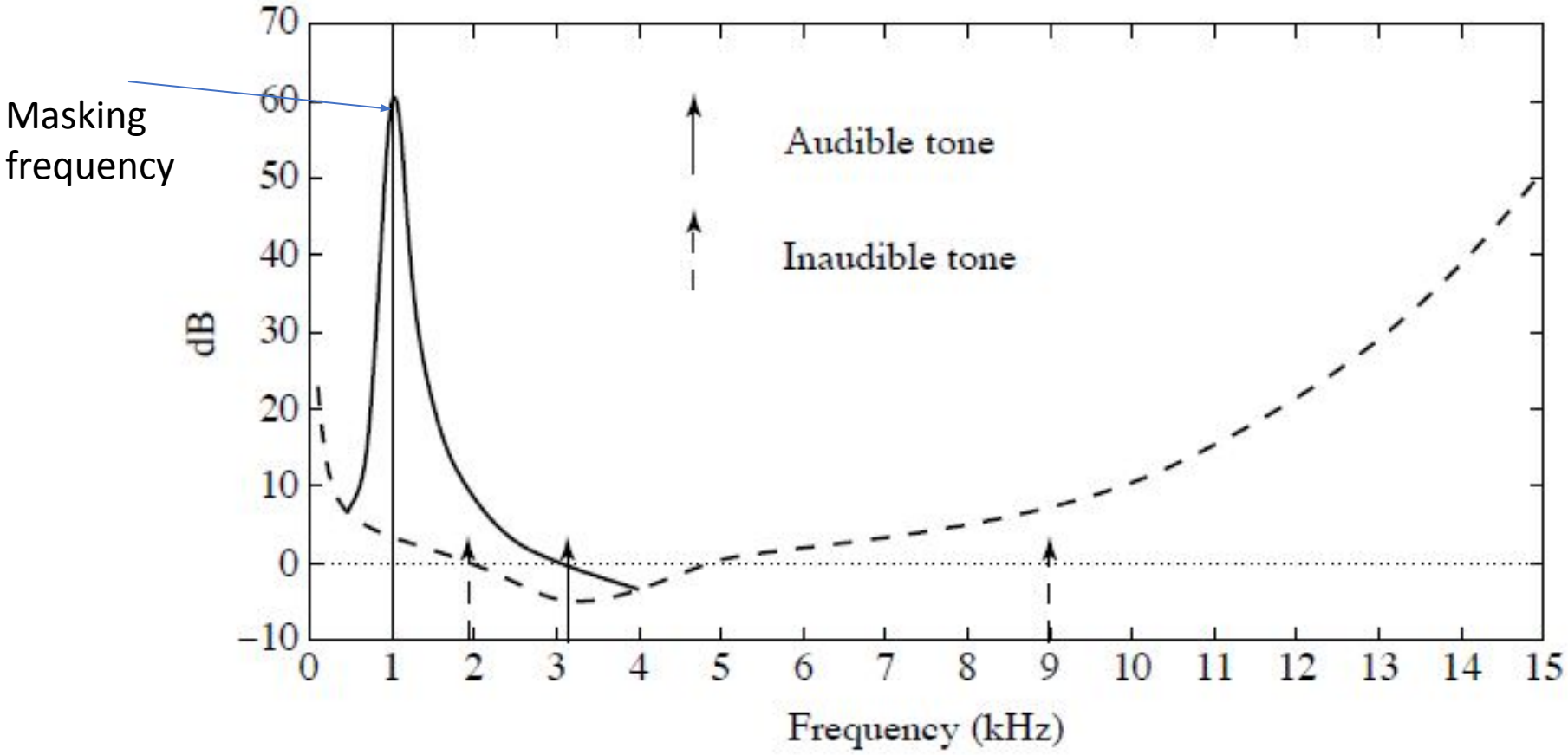
Frequency and temporal masking

Signal Masking
Spectral- frequency masking

- A frequency component can be partly or fully masked by another component that is close to it in frequency.

- This shifts the hearing threshold.

Frequency Masking



Effect on threshold for 1 kHz masking tone

Frequency Masking

Loss audio data compression methods, such as MPEG/Audio encoding, remove some sounds which are masked anyway

The general situation regarding to masking is as follows:

A lower tone can effectively mask (make us unable to hear) a higher tone

The greater the power in the masking tone, the wider is its influence – the broader the range of frequencies it can mask.

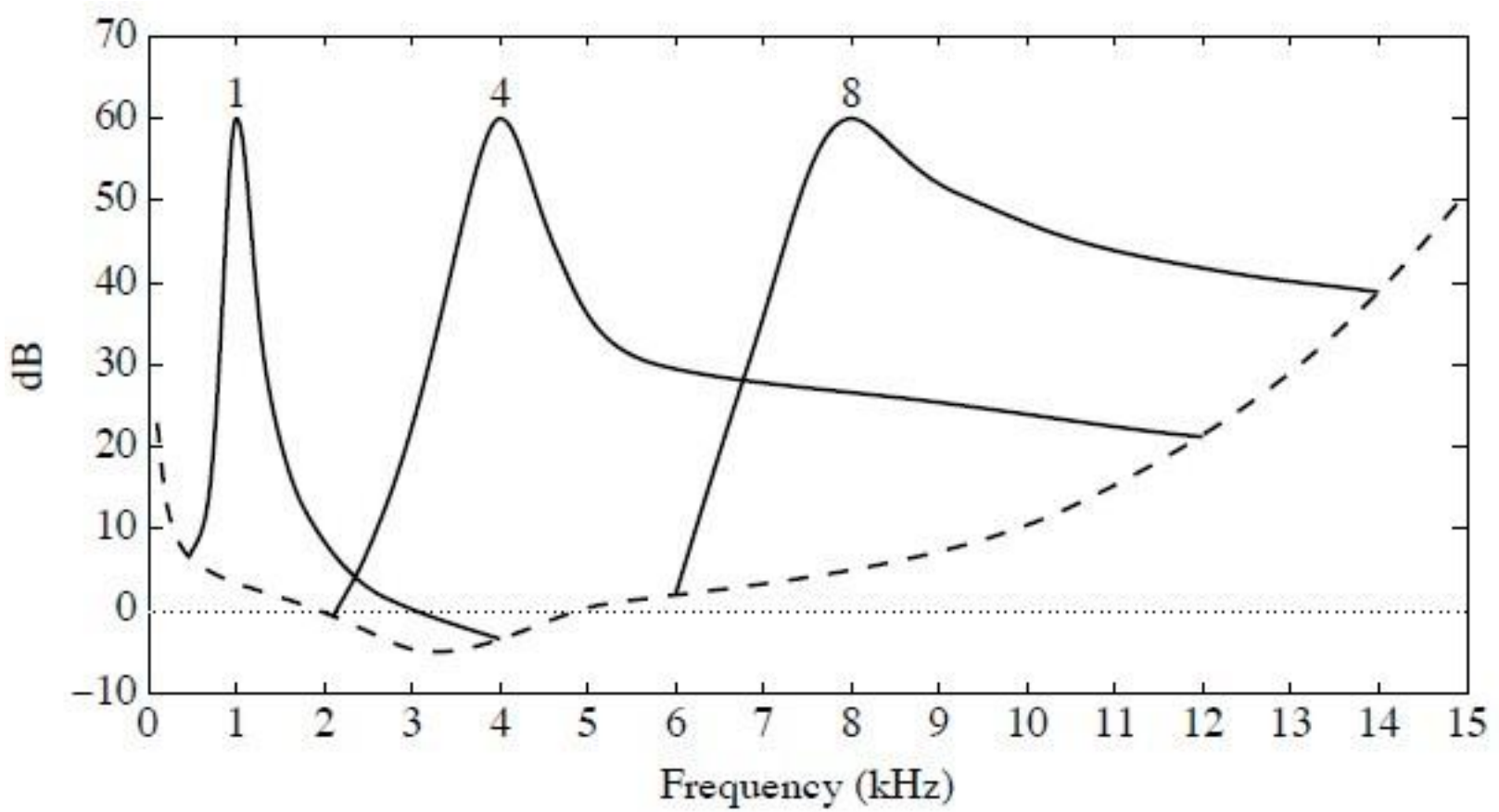
Frequency Masking Curves

Frequency masking is studied by playing a particular pure tone, say 1 kHz again, at a loud volume, and determining how this tone affects our ability to hear tones nearby in frequency

– one would generate a 1 kHz masking tone, at a fixed sound level of 60 dB, and then raise the level of a nearby tone, e.g., 1.1 kHz, until it is just audible

The threshold in the Figure on the next slide plots the audible level for a single masking tone (1,4,8 kHz)

Effect of masking tone at three different frequencies

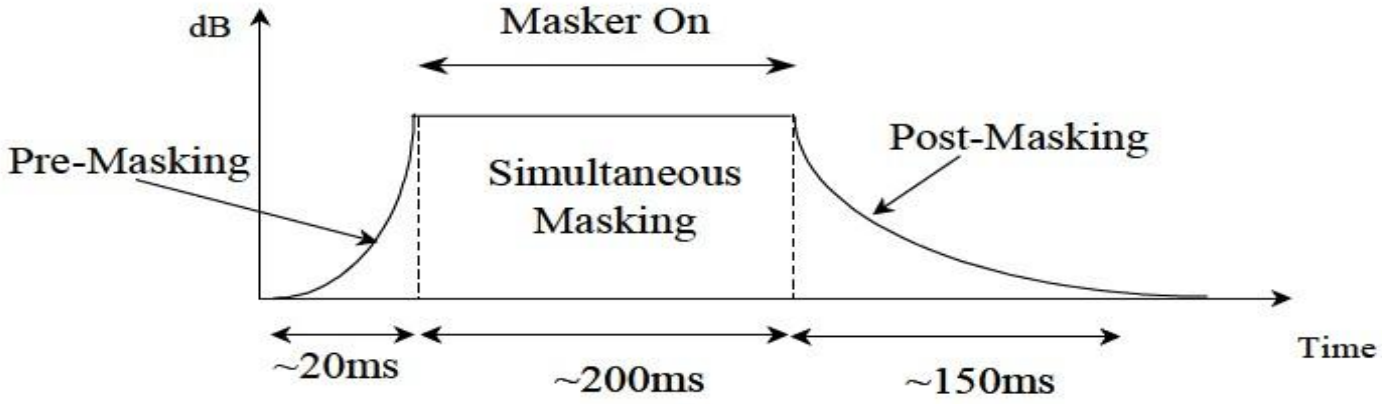


Temporal Masking

A quieter sound can be masked by a louder sound if they are temporally close sounds that occur both (shortly) **before** and **after** volume increase can be masked

Masking effect in the time domain

Simultaneous masking: Two sounds occur simultaneously, and one is masked by the other.



Forward masking (Post):

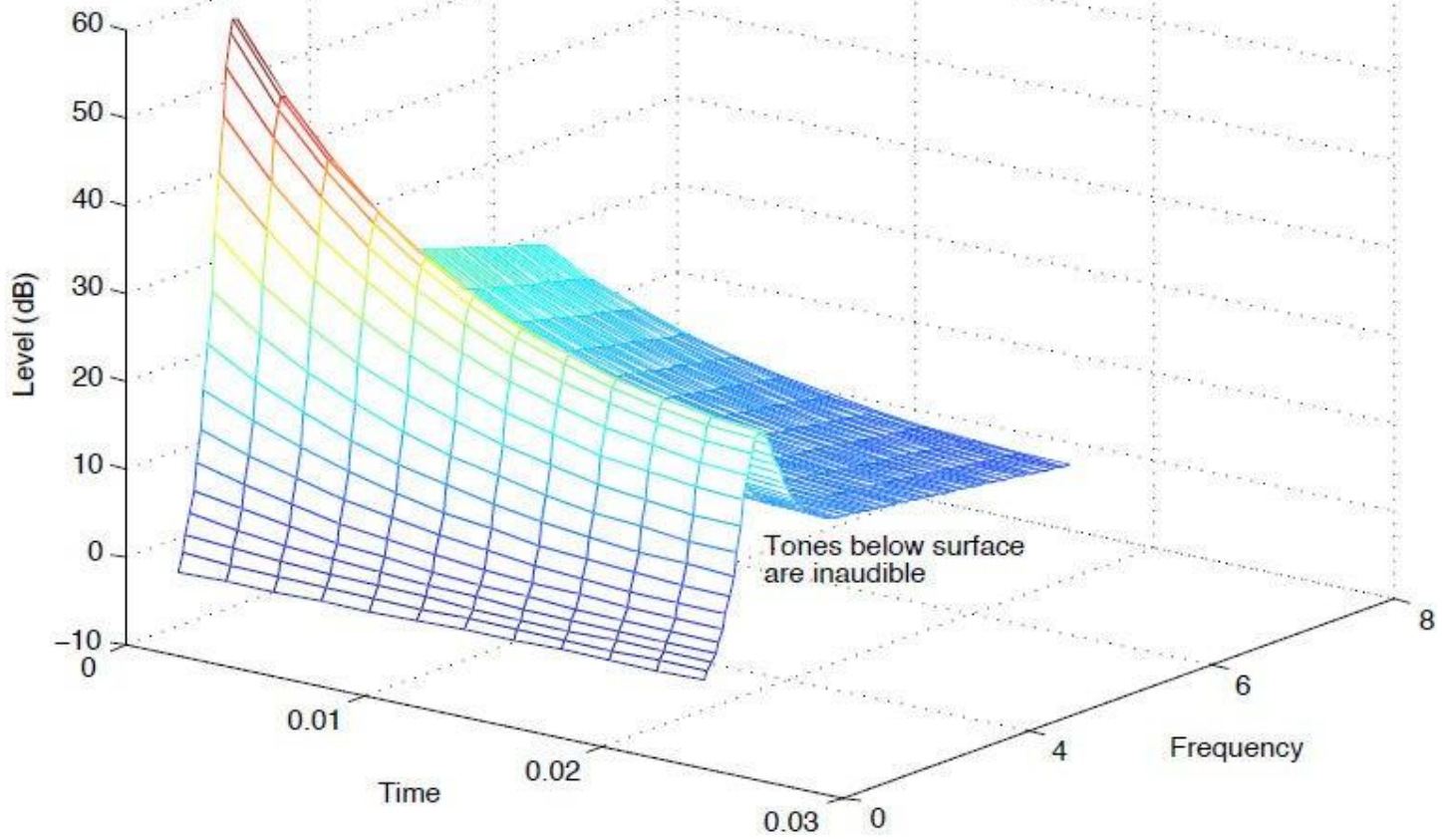
softer sounds that occur as much as 200 milliseconds after the loud sound will also be masked.

Backward masking (Pre):

A softer sound that occurs prior to a loud one will be masked by the louder sound.

Masking effect in the time domain

Effect of temporal and frequency masking's depending on both time and closeness in frequency.



Sub-band coding

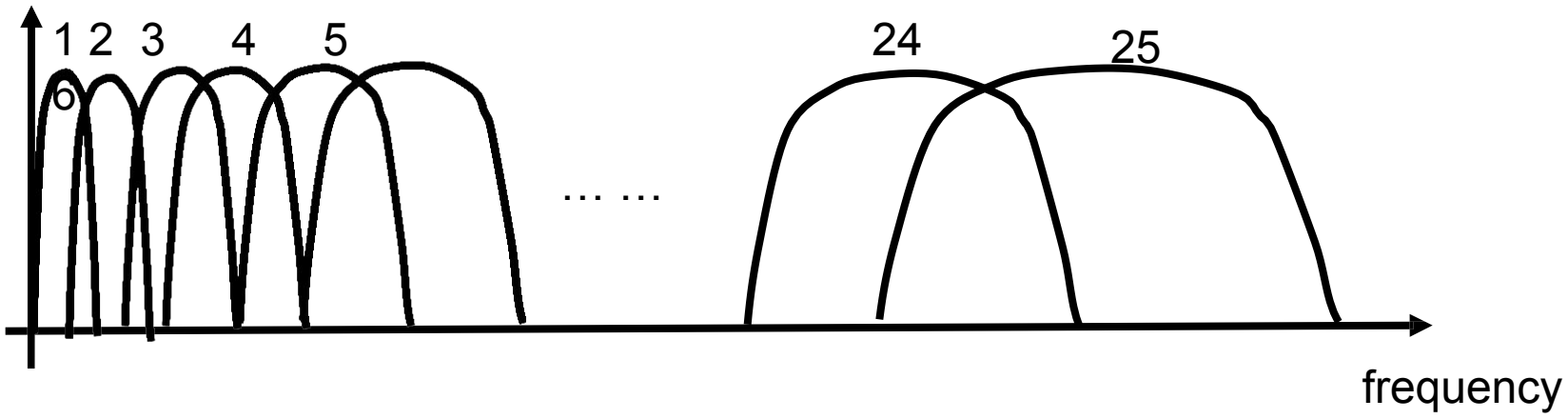
Properties of Human Auditory System
Sub-band coding

Critical Bands (Sub-band filtering):

Our brains perceive the sounds through distinct critical bands, the bandwidth grows logarithmically with frequency.

At 100Hz, the bandwidth is about 160Hz;

At 10kHz it is about 2.5kHz in width.



Spectral Analysis – recap.

Linear Transforms

- Fast Fourier Transform (FFT)
- Discrete Cosine Transform (DCT)
- Modified Discrete Cosine Transform (MDCT)
  - [used by MPEG-1 Layer-III, MPEG-2 AAC, Dolby AC-3]
  - overlapped and windowed version of DCT

Spectral Analysis
Filter Banks

- Time sample blocks are passed through a set of bandpass filters

- Masking thresholds are applied to resulting frequency subband signals

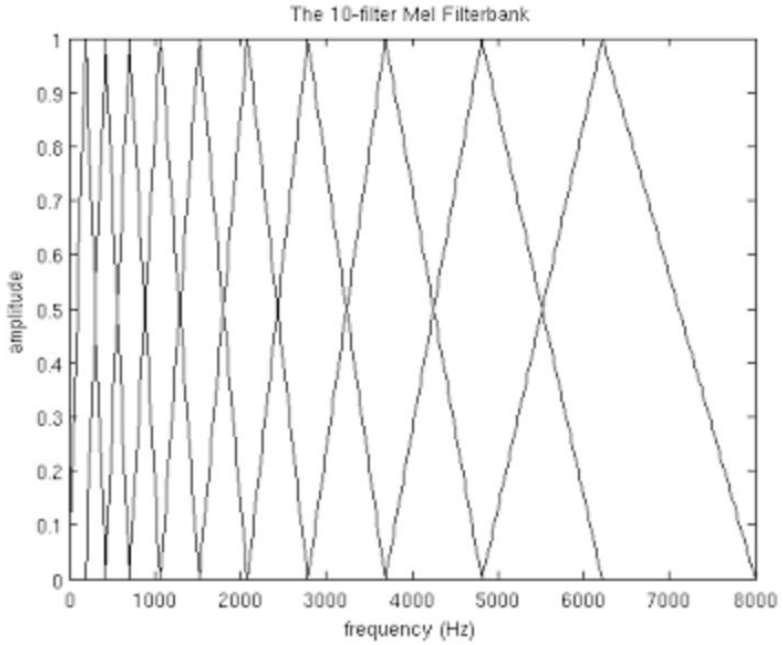
  - *Poly-phase and wavelet banks are most popular filter structures*

Filter Bank Structures

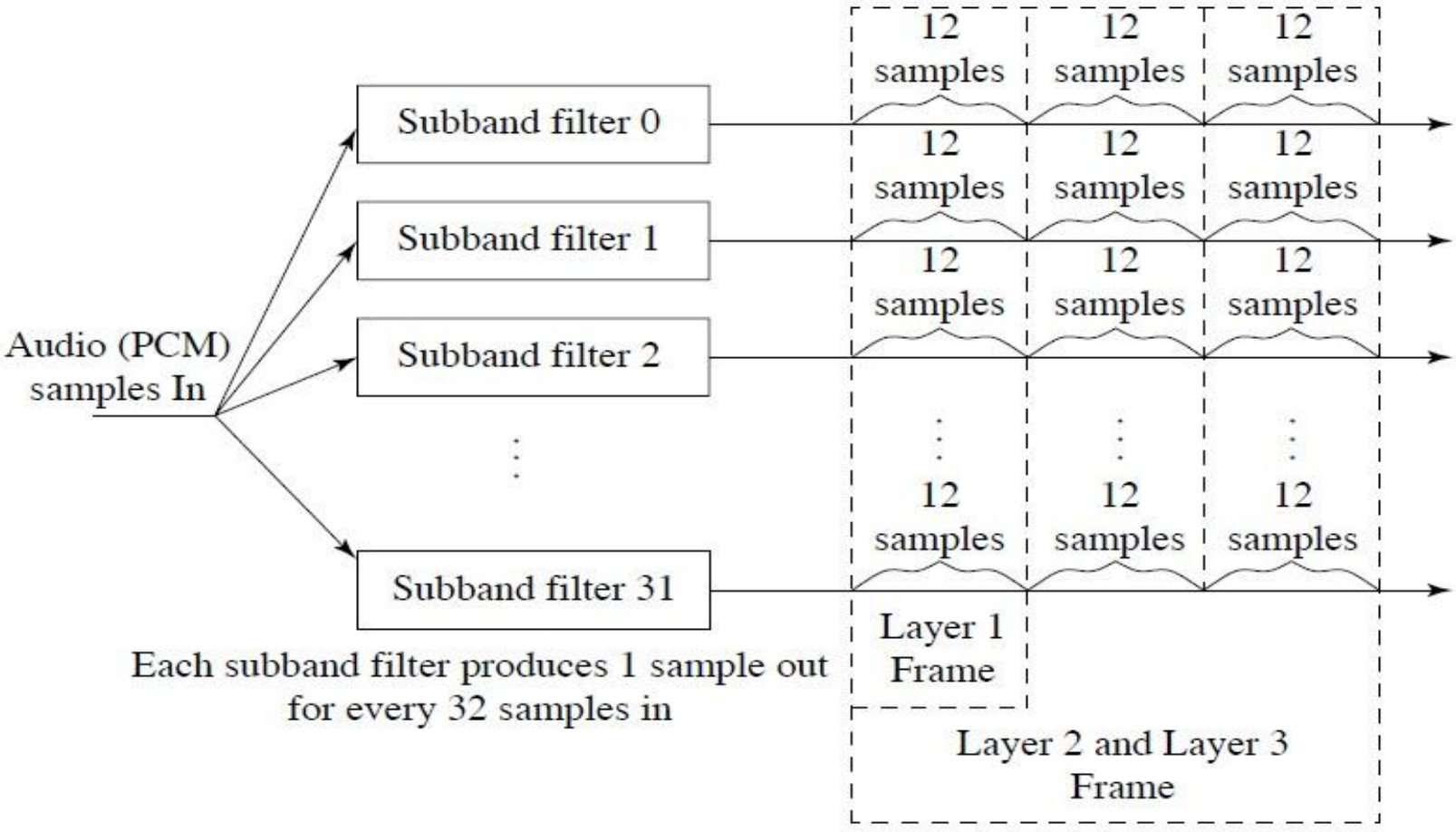
Polyphase Filter Bank

(used in all of the MPEG-1 encoders)

- Signal is separated into subbands



MPEG Audio Frame Sizes



Noise Allocation

Noise Allocation

System Task: derive and apply shifted hearing threshold to the input signal

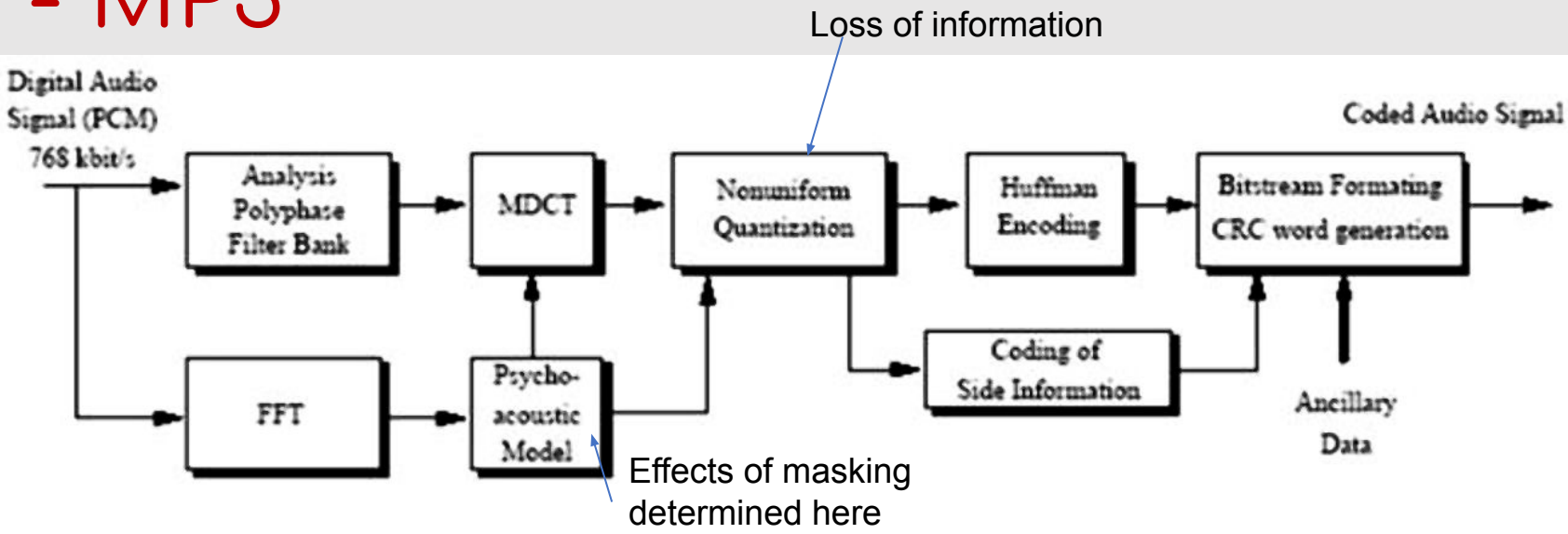
- Anything below the threshold doesn't need to be transmitted
- Any noise below the threshold is irrelevant

Frequency component **quantization**

- Tradeoff
- Encoder saves on space by using just enough bits for each frequency component to keep noise under the threshold - this is known as noise allocation

MPEG-Audio Layer 3 Coding - MP3

MPEG-Audio Layer 3 Coding - MP3



1. Use convolution filters to divide the audio signal (e.g., 48 kHz sound) into 32 frequency sub-bands. (*sub-band filtering*)
2. Determine amount of masking for each band caused by nearby band using the *psychoacoustic model* .
3. If the power in a band is below the masking threshold, don't encode it.
4. Otherwise, determine number of bits needed to represent the coefficient such that, the noise introduced by quantization is below the masking effect (Recall that one fewer bit of quantization introduces about 6 dB of noise).
5. Format bitstream

Masking and Quantization (Example)

Example: performing the subband filtering step on the input results in the following values (for demonstration, we are only looking at the first 16 of the 32 bands):

Band	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Level	0	8	12	10	6	2	10	60	35	20	15	2	3	5	3	1

The 60dB level of the 8th band gives a masking of 12 dB in the 7th band, 15dB in the 9th. (according to the Psychoacoustic model)

The level in 7th band is 10 dB (< 12 dB), so ignore it.

The level in 9th band is 35 dB (> 15 dB), so send it.

We only send the amount above the masking level

Therefore, instead of using 6 bits to encode it, we can use 4 bits -- a saving of 2 bits (= 12 dB).

“determine number of bits needed to represent the coefficient such that, the noise introduced by quantization is below the masking effect” [noise introduced = 12bB; masking = 15 dB]

MP3 Audio Format

An MP3 File

Internal Structure of An MP3 File

MP3 Header
MP3 Data
MP3 Header
MP3 Data
+++ Repeated +++
MP3 Header
MP3 Data
MP3 Header
MP3 Data

Note That The MP3 File Structure Maybe 'encapsulated' within an ID3 Tag

ID3v2x Metadata
MP3 Header
MP3 Data
MP3 Header
MP3 Data
Repeated
MP3 Header
MP3 Data
MP3 Header
MP3 Data

An MP3 Frame

MP3 Header
MP3 Data

Example MP3 Header

FF FB A0 40

Color Coding shows binary bit mapping to hex values below

Detail Of An MP3 Header

Bits	1 2 3 4 5 6 7 8 9 10 11 12	13	14 15	16	17 18 19 20	21 22	23	24	25 26	27 28	29	30	31 32
Binary	1 1 1 1 1 1 1 1 1 1 1 1	1	0 1	1	1 0 1 0	0 0	0	0	0 1	0 0	0	0	0 0
Hex	F F F		B		A		0			4			0
Meaning	MP3 Sync Word	Version	Layer	Error Protection	Bit Rate	Frequency	Pad. Bit	Priv. Bit	Mode	Mode Extension (Used With Joint Stereo)	Copy	Original	Emphasis
Value	Sync Word	1 = MPEG	01 = Layer 3	1=No	1010 = 160	00 = 44100 Hz	0 = Frame is not padded	Unknown	01=Joint Stereo	0 = Intensity Stereo Off / 0 = MS Stereo Off	0=Not Copyrighted	0=Copy Of Original Media	00=None

MP3 compression performance

The audio sampling rate can be 32, 44.1, or 48 kHz

The tests showed that even with a 6-to-1 compression ratio (stereo, 16 bits/sample, audio sampled at 48kHz compressed to 256 kbits/sec) and under optimal listening conditions, expert listeners were unable to distinguish between coded and original audio clips with statistical significance.

Additional Encoding Techniques

Additional Encoding Techniques

Other encoding techniques are available (alternative or in combination)

- Predictive Coding
- Coupling / Delta Encoding
  - Used in cases where audio signal consists of two or more channels (stereo or surround sound) Similarities between channels are used for compression A sum and difference between two channels are derived; difference is usually some value close to zero and therefore requires less space to encode
- Huffman Encoding

Additional Encoding Techniques

Predictive Coding

- Often used in speech and image compression
- Estimates the expected value for each sample based on previous sample values
- Transmits/stores the difference between the expected and received value

- Generates an estimate for the next sample and then adjusts it by the difference stored for the current sample
- *Used for additional compression in MPEG2 AAC*

Additional Encoding Techniques

Huffman Coding

- Information-theory-based technique
- An element of a signal that often re-occurs in the signal is represented by a simpler symbol, and its value is stored in a look-up table
- Implemented using a look-up tables in encoder and in decoder
- Provides substantial lossless compression, but requires high computational power and therefore is not very popular

- *Used by MPEG1 and MPEG2 AAC*

Advanced Audio Coding (AAC)

Advanced Audio Coding (AAC)

Introduced 1997 as MPEG-2 Part 7

In 1999 – updated and included in MPEG-4

AAC has been standardized by ISO and IEC as part of the MPEG-2 and MPEG-4 specifications

Advanced Audio Coding (AAC) – now part of MPEG-4 Audio

Inclusion of 48 full-bandwidth audio channels

Default audio format for iPhone, iPad, Nintendo, PlayStation, Nokia, Android, BlackBerry, Used in Windows Media Audio...

AAC's Improvements over MP3

- More sampling frequencies (8-96 kHz)
- Arbitrary bit rates and variable frame length
- Higher efficiency and simpler filter-bank
- Uses pure MDCT (modified discrete cosine transform)

MPEG-4 Audio (.mp4)

MPEG-4 Audio (.mp4)

- Called MP4 with Extension .mp4
- Multimedia container format
- Stores digital video and audio streams and allows streaming over Internet
- Container or wrapper format
- meta-file format whose spec. describes how different data elements and metadata co-exist in computer file

MPEG-4 Audio Part 3

Variety of applications:

- General audio signals
- Speech signals
- Synthetic audio
- Synthesized speech (structured audio)

Includes variety of audio coding technologie:

- Lossy speech coding (e.g., CELP) CELP – code-excited linear prediction – speech coding
- General audio coding (AAC)
- Lossless audio coding
- Text-to-Speech interface
- Structured Audio (e.g., MIDI)

MPEG-4 Audio - features

- Bit-rate 2-64 kbps
- Scalable for variable rates
- MPEG-4 defines set of coders
- Parametric Coding Techniques Code Excited Linear Prediction Time Frequency Techniques

    - low bit-rate 2-6 kbps, 8kHz sampling frequency
    - medium bit-rates 6-24 kbps, 8 and 16 kHz sampling rate
    - high quality audio 16 kbps and higher bit-rates, sampling rate > 7 kHz

Comparison AAC and MP4

AAC has been standardized by ISO and IEC as part of the MPEG-2 and MPEG-4 specifications and MP4 is MPEG-4 Part 14.

However, MP4 is the container that stores data; that data is encoded in AAC. The container does not affect the quality of the data and is not an encoding format, so AAC cannot be compared with MP4 in this way.

Quality depends on the encoder, not the file type. In other words, MP4 can use the AAC codec or others like AC3, ALS, SLS, or MP3.