Lecture notes of Image Compression and Video Compression series

6. Perceptual Audio Coding

Prof. Amar Mukherjee
Weifeng Sun

Topics

- Introduction to Image Compression
- Transform Coding
- Subband Coding, Filter Banks
- Introduction to Wavelet Transform
- Wavelet Image Compression
- Perceptual Audio Coding
- Video Compression

#2

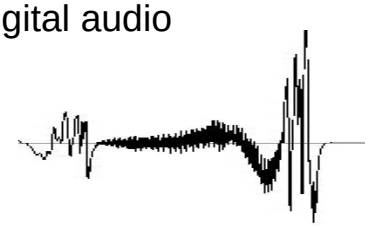
Contents

- Sound
- Physiology of the Ear
- Human Sound Perception
  - Psychoacoustics
  - Auditory Masking
    - Frequency masking
    - Temporal masking
- MPEG Audio Compression
- Example

#3

Motivation

- Many applications need digital audio
  - Storing
    - ipod
    - games
  - Streaming
    - Internet radio
  - Interactive multimedia session like VOIP, conferencing.

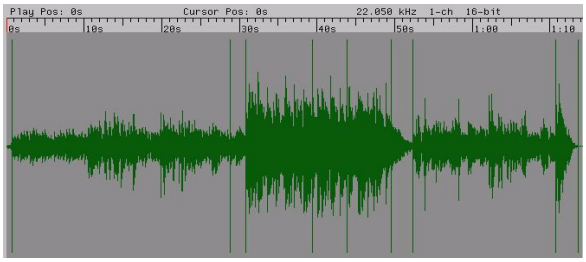


- However, cannot apply entropy coder directly.
  - Very random, missing repeated sequence of byte pattern.

#4

Sound

- Sound is a continuous signal that is created by compressing and decompressing the air.
- This changing air pressure causes the eardrum to vibrate.



#5

Describing Sound

- Intensity
  - Defined as power per unit area.
  - The level of perceived sound by a human is referred to as *loudness*.
    - Measured in terms of *phons*.
    - Defined as the perceive intensity at 1000 Hz by normal human population.
- Pitch
  - frequency of sound (tone)
- Quality (timbre)

#6

Decibels

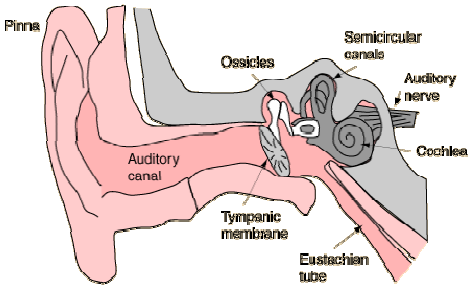
- For most audio specifications, sound levels are presented in terms of decibels (dB),
  - A **log** scale for intensity changes.
  - The softest 1000 Hz tone heard by the normal ear is I_0 = 1 picowatt/meter^2.
  - With I_0 as a reference, all other intensities, I, can be given in terms of dB by:

$$I(\text{dB}) = \text{Log10}\left(\frac{I}{I_0}\right)$$

#7

Physiology of the Ear

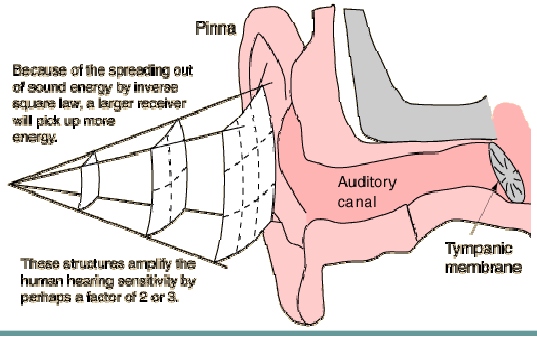
- Thousands of "microphones"
  - hair cells in cochlea
- Automatic gain control
  - muscles around transmission bones
- Directivity
  - pinna
- Boost of middle frequencies
  - auditory canal
- Nonlinear processing
  - auditory nerve



#8

The Outer Ear

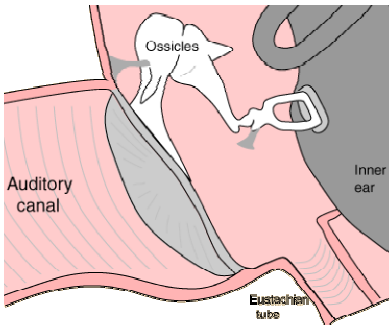
- For a point source of sound, it spreads out according to the inverse square law.
- For a given sound intensity, a larger ear captures more of the wave and hence more sound energy.



#9

The Tympanic Membrane

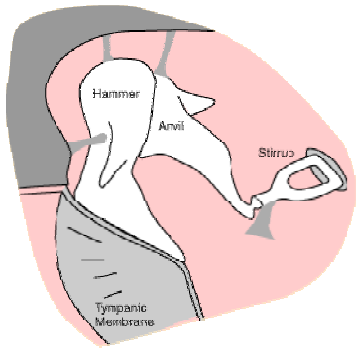
- The tympanic membrane or "eardrum" receives vibrations traveling up the auditory canal and transfers them through the tiny ossicles to the oval window, the port into the inner ear.
- The eardrum is some fifteen times larger than the oval window, giving an amplification of about fifteen compared to the oval window alone.



#10

The Ossicles

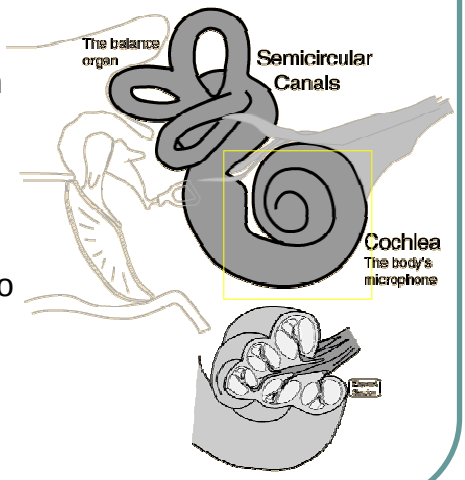
- The three tiniest bones in the body form the coupling between the vibration of the eardrum and the forces exerted on the oval window of the inner ear.
- Serves as an amplifier with a factor of about three under optimum conditions.
- Can be adjusted by muscle action to actually attenuate the sound signal for protection against loud sounds.



#11

The Inner Ear (Two Organs)

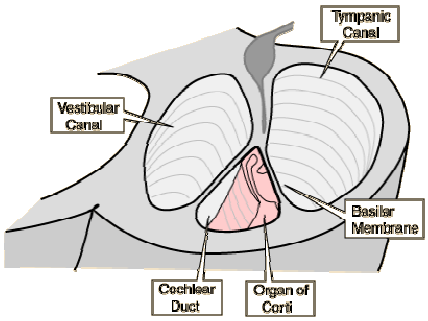
- Semicircular canals
  - body's balance organ
- Cochlea
  - body's microphone
  - converting sound pressure impulses from the outer ear into electrical impulses which are passed on to the brain via the auditory nerve.



#12

The Cochlea

- Snail-shell like structure.
- Three fluid filled sections.

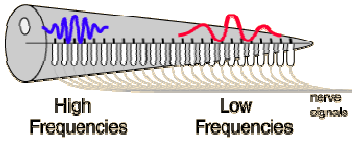


- The organ of Corti is the sensor of pressure variations.

#13

Place Theory

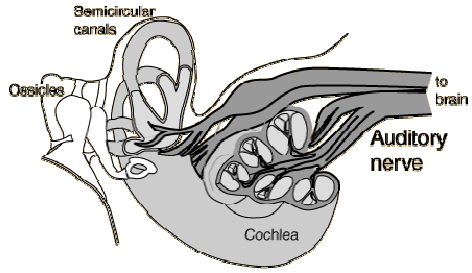
- High frequency sounds selectively vibrate the basilar membrane of the inner ear near the entrance port (the oval window).
- Lower frequencies travel further along the membrane before causing appreciable excitation of the membrane.
- The basic pitch determining mechanism is based on the location along the membrane where the hair cells are stimulated.



#14

The Auditory Nerve

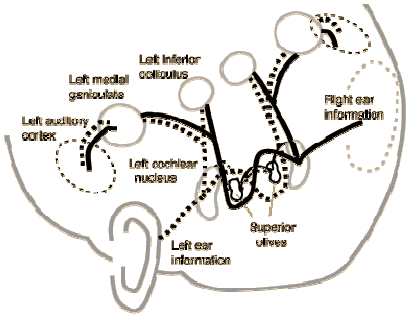
- Taking electrical impulses from the cochlea and the semicircular canals, the auditory nerve makes connections with both auditory areas of the brain.



#15

Auditory Area of Brain

- Binaural relay
  - When the auditory nerve from one ear takes information to the brain, that information is directly sent to both the processing areas on both sides of the brain.



#16

Perceptual Audio Compression

- The basis of the Perceptual Codecs is Psychoacoustic Masking.
- An audio file will contains sounds that are not heard by us, even though these sounds lie within the human audible range.
- Masking Techniques:
  - Frequency (Concurrent) Masking
  - Temporal Masking

#17

Dynamic Range of Hearing

- The practical dynamic range could be said to be from the threshold of hearing to the threshold of pain.

Threshold of Hearing	Threshold of Pain
I_0	$10^{13} I_0$ = 10,000,000,000,000 I_0
0 decibels	130 decibels

- This remarkable dynamic range is enhanced by an effective amplification structure which extends its low end and by a protective mechanism which extends the high end.

#18

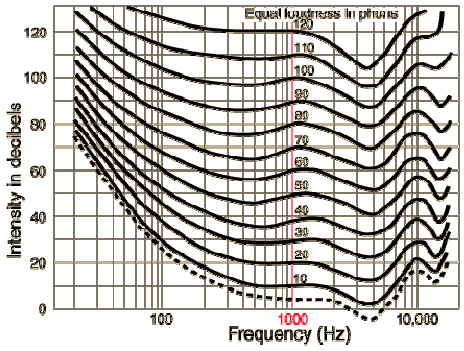
Frequency Response of the Ear

- The ear has a non-flat frequency response.
  - Tones played at the same volume with different frequencies can sound like they are being played at different volume levels.
  - In order to maintain the same loudness over the hearing spectrum, the intensity level must vary.

#19

Equal-loudness Curve

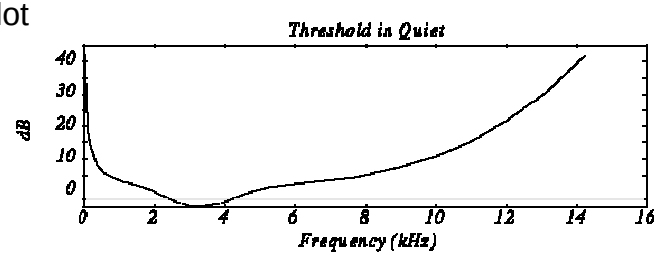
- Audible frequency range
  - 20 to 20,000 Hz
- Normal voice range
  - 500 Hz to 2 kHz
  - Low frequencies are vowels and bass
  - High frequencies are consonants



#20

Human Hearing Sensitivity

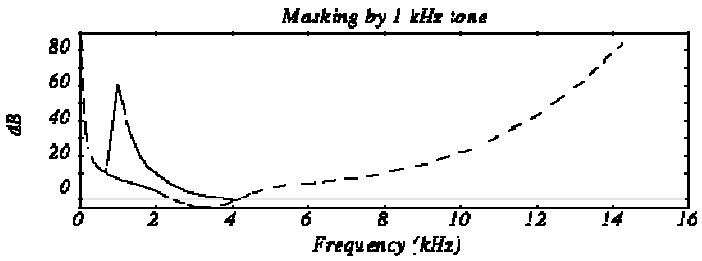
- Experiment
  - Put a person in a quiet room.
  - Raise level of 1 kHz tone until just barely audible.
  - Vary the frequency
  - Plot



#21

Frequency (Concurrent) Masking

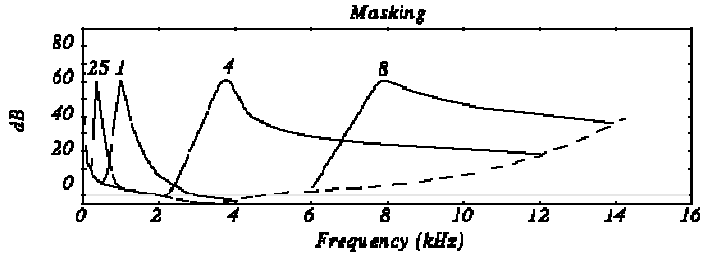
- Experiment: Play 1 kHz tone (*masking tone*) at fixed level (60 dB). Play *test tone* at a different level (e.g., 1.1 kHz), and raise level until just distinguishable.
- Vary the frequency of the test tone and plot the threshold when it becomes audible



#22

Frequency (Concurrent) Masking

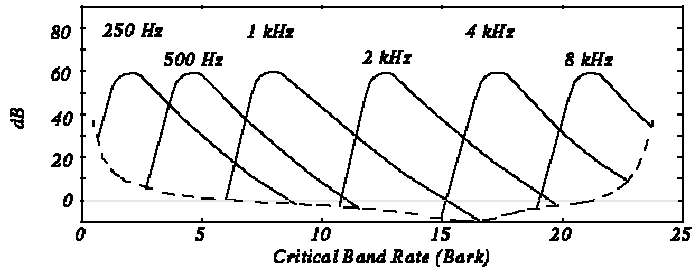
- Frequency masking effect dependent on frequency.
- As shown people can detect lower frequency test tones closer than higher frequency test tones.



#23

Frequency (Concurrent) Masking

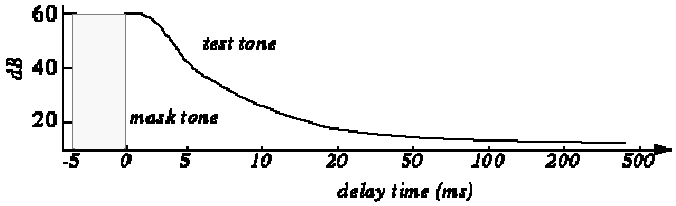
- Sub bands are uniform, but human hearing sensitivity isn't, so psychoacoustic model and MPEG encoder compensates by preserving more data in lower filtered sub bands.
- There are about 25 critical bands.



#24

Temporal Masking

- Experiment: 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.
- Repeat with different level of the test tone and plot.



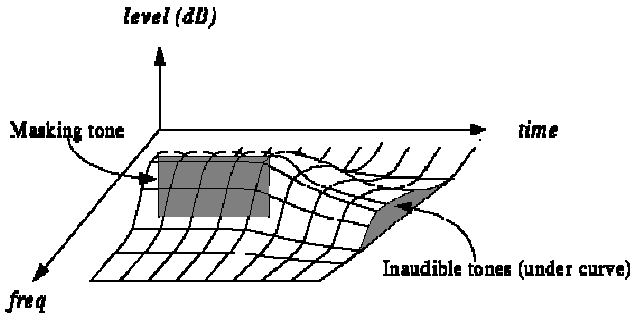
#25

Temporal Masking

- Postmasking
  - If the volume drops sharply then the human ear takes a few milliseconds to adjust, during this period the lower volume sounds are not heard and these can be discarded.
- Premasking
  - If the volume rises sharply then the human ear discards the last few milliseconds of the quieter sound and immediately starts processing the louder sound. The last few milliseconds of data before a sharp rise in volume can also be discarded.
- The discarded sounds are replaced by silence, which can be run length encoded.

#26

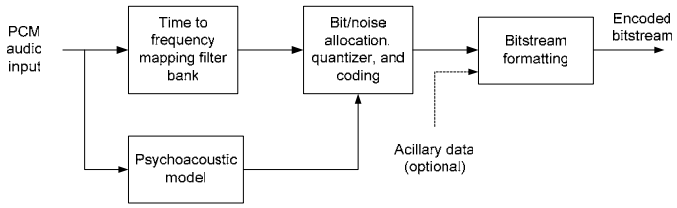
Combination



#27

MPEG Audio Coding Algorithm

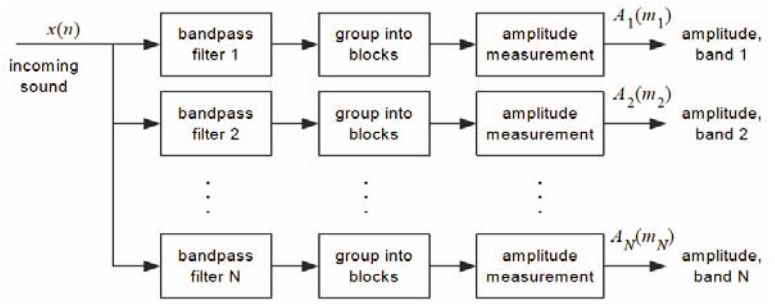
- Subband filtering
- Masking for each band caused by nearby band using the psychoacoustic model
- Discard those bands If their power are below the masking threshold
- Quantization/bit-allocation/coding
- Format bitstream



#28

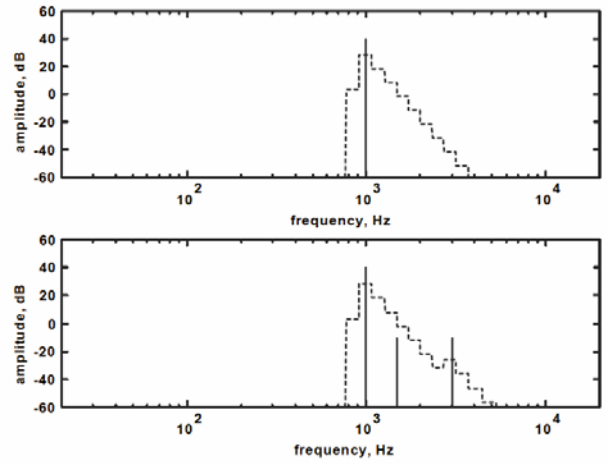
Filter Banks

- Explains frequency-domain masking



#29

Frequency-domain Masking



#30

Example

After analysis, the first levels of 16 of the 32 bands are:

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

If the level of the 8th band is 60dB, it gives a masking of 12 dB in the 7th band, 15dB in the 9th.

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

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

Only the amount above the masking level needs to be sent, so instead of using 6 bits to encode it, we can use 4 bits saving 2 bits (= 12 dB).

#31

MPEG Audio Compression

- Three MPEG codecs
  - layer 1
    - suitable for consumer recording
    - 384kbps for a stereo signal, compression ratio 4:1
  - layer 2
    - suitable for professional recording and Broadcasting
    - 256..192 kbps for a stereo signal, compression ratio 6:1..8:1
  - layer 3 (mp3)
    - suitable for Internet transmission
    - 128..112 kbps for a stereo signal, compression ratio 10:1..12:1
  - More on MP3
    - Better critical band filter is used (non-equal frequencies)
    - psychoacoustic model includes temporal masking effects
    - takes into account stereo redundancy
    - uses Huffman coder

#32