Multimedia Software Systems CS4551

Psychoacoustics and

MPEG Audio Compression

Sound as Perceived

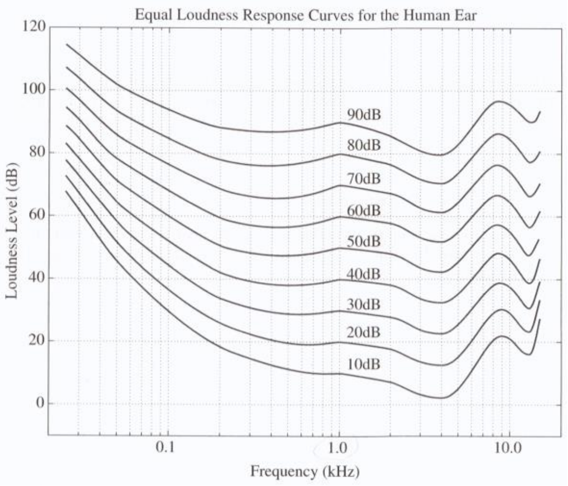
- Compression attained by variations of PCM coding techniques alone are not sufficient to attain data rates for modern applications to transmit CD and Surround Sound quality music.
- Perception of Sound additionally can help in compression by studying
  - What frequencies we hear
  - When do we hear them
  - When do we not hear them
- This branch of study –Psychoacoustics –deals with sound perception science. “Auditory Masking” is a perceptual weakness of the ear, which can be used to exploit compression without compromising quality

Psychoacoustics - Human Hearing

- Human hearing
  - Frequency range of human hearing: 20Hz ~ 20 KHz
    * Ultrasonic are sounds at higher frequencies.
  - Frequency range of the voice: 500Hz to 4kHz
  - Human can hear up to 120dB (sound amplitude).
  - The reference point for 0dB is the quietest sound that human can hear measured at 1KHz. Technically, it is the sound that creates a barely audible sound intensity of 10^{-12} Watt per square meter.
  - The range of hearing depends on frequency.

Psychoacoustics - Human Hearing

- **Equal-Loudness Relations**
- Fletcher-Munson equal-loudness curve
  - Frequencies (x-axis)
  - Sound pressure level in dB (y-axis)
  - Perceived loudness plotted

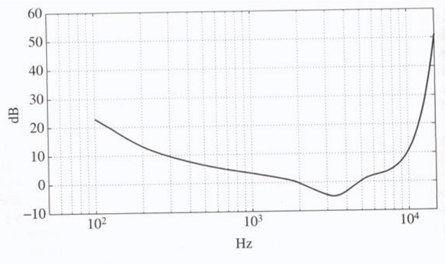


Psychoacoustics - Human Hearing

- **Equal-Loudness Relations**
  – Two tones with the same amplitude but different frequencies sound louder than the other.
  – At the 4kHz dip, we perceive the sound as 10dB when the actual sound pressure is 2dB. At 10kHz, the same 10dB sound is produced by the actual 20dB sound pressure.
  – Ear is most sensitive to frequencies b/w 1kHz and 5kHz. (because the ear canal amplifies frequencies from 2.5kHz to 4kHz). Ear is not usually sensitive to low frequencies.
  – As the overall loudness increases, the curves flattens. => If the sound level is loud enough, we are equally sensitivities to low frequencies.

Psychoacoustics - Human Hearing

- Threshold of hearing



Threshold of human hearing, for pure tones.

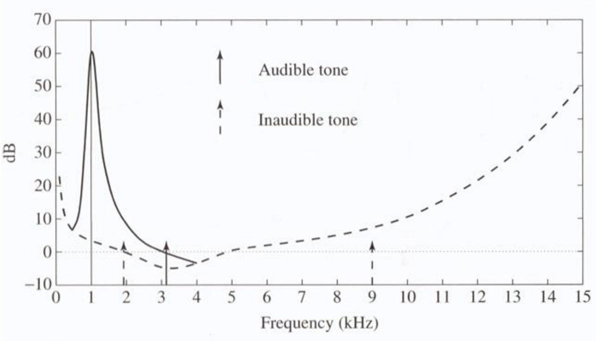
- Masking Threshold Curve
  – A tone is *audible* only if its power is above the absolute threshold level
  – 2dB for 6kHz tone

Psychoacoustics - Human Hearing

- **Frequency Masking**
  - A lower tone can effectively mask (make us unable to hear) a higher tone. Reverse is not true.
  - The greater the power in the masking tone, the wider its influence – the broader the range of frequencies it can mask.
  - If two tones are widely separated in frequency, little masking occurs.
  - Experiments
    - Play a tone (frequency) at a loud volume and check how the tone affects out ability to hear tones at nearby frequencies.

Psychoacoustics - Human Hearing

- **Frequency Masking**

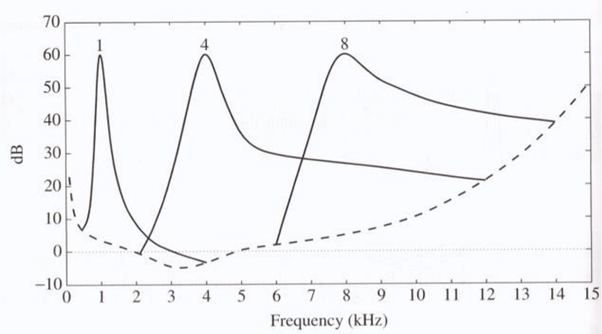


Effect on threshold of human hearing for a 1 kHz masking tone.

Generate a masking tone (1kHz tone at sound level 60dB), raise the level of nearby tones until it is audible and plot the level.

Psychoacoustics - Human Hearing

- **Frequency Masking**



Effect of masking tones at three different frequencies.

•Masking diagram changes according to the frequency of a masking tone.

•The higher frequency of the masking tone, the broader a range of influence.

=> If a signal can be decomposed into frequencies, then we can detect masked frequencies and only audible part can be quantized.

Psychoacoustics - Human Hearing

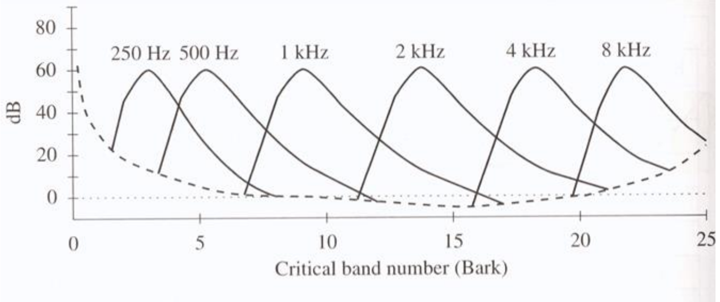
- **Frequency masking and critical band**
- Human auditory system can be roughly modeled as a filterbank, consisting of 25 bands.
- Each band is called a critical band
- The ear cannot distinguish sounds within the same band that occur simultaneously.
- Critical bandwidth is non-uniform.

Critical bands and their bandwidths.

Band #	Lower bound (Hz)	Center (Hz)	Upper bound (Hz)	Bandwidth (Hz)
1	-	50	100	-
2	100	150	200	100
3	200	250	300	100
4	300	350	400	100
5	400	450	510	110
6	510	570	630	120
7	630	700	770	140
8	770	840	920	150
9	920	1000	1080	160
10	1080	1170	1270	190
11	1270	1370	1480	210
12	1480	1600	1720	240
13	1720	1850	2000	280
14	2000	2150	2320	320
15	2320	2500	2700	380
16	2700	2900	3150	450
17	3150	3400	3700	550
18	3700	4000	4400	700
19	4400	4800	5300	900
20	5300	5800	6400	1100
21	6400	7000	7700	1300
22	7700	8500	9500	1800
23	9500	10500	12000	2500
24	12000	13500	15500	3500
25	15500	18775	22050	6550

Psychoacoustics - Human Hearing

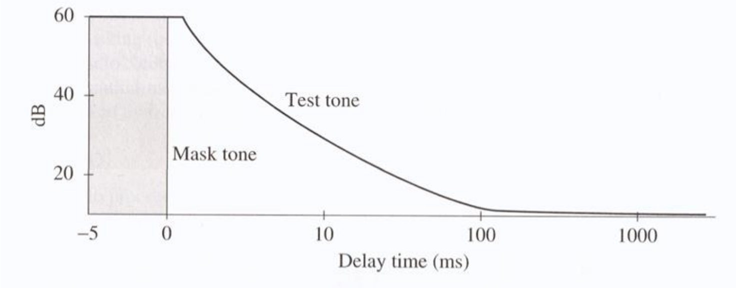
- **Frequency masking and critical band**
- Critical bandwidth represented in frequency unit is non-uniform. So a new unit *Bark* is introduced in order to define a more perceptually uniform unit for critical band.
- Critical bands: The widths of the masking bands for different masking tones are different, increasing with the frequency of the masking tone.



Effect of masking tones, expressed in Bark units.

Psychoacoustics - Human Hearing

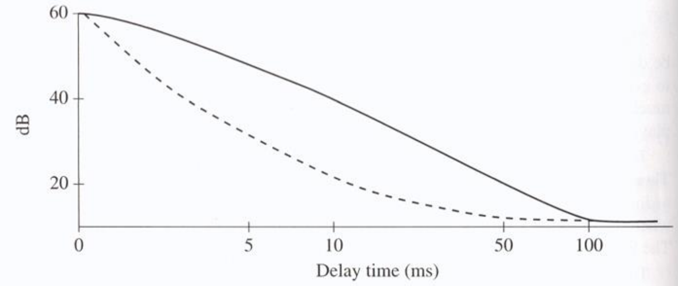
- **Temporal masking**
  - After the dance, it takes time for our hearing system return to normal. How long does it take to come back to normal?



The louder the test tone, the shorter the amount of time required before the test tone is audible once the masking tone is removed.

Psychoacoustics - Human Hearing

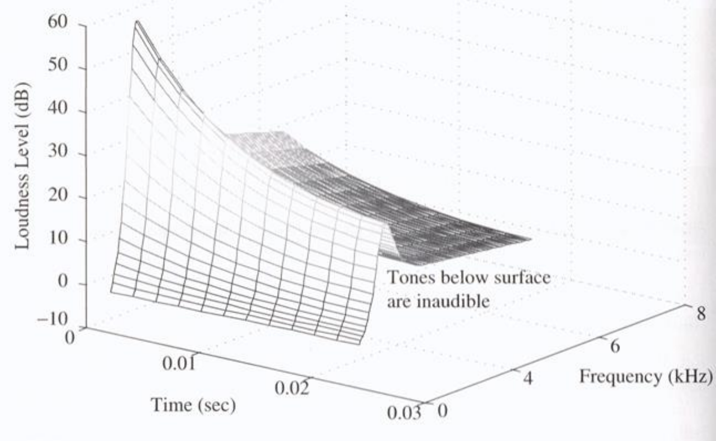
- **Temporal masking**



Effect of temporal masking also depends on the length of time the masking tone is applied. Solid curve: masking tone played for 200 msec; dashed curve: masking tone played for 100 msec.

CSULA CS4551 Multimedia Software Systems by Eun-Young Kang

Psychoacoustics - Human Hearing



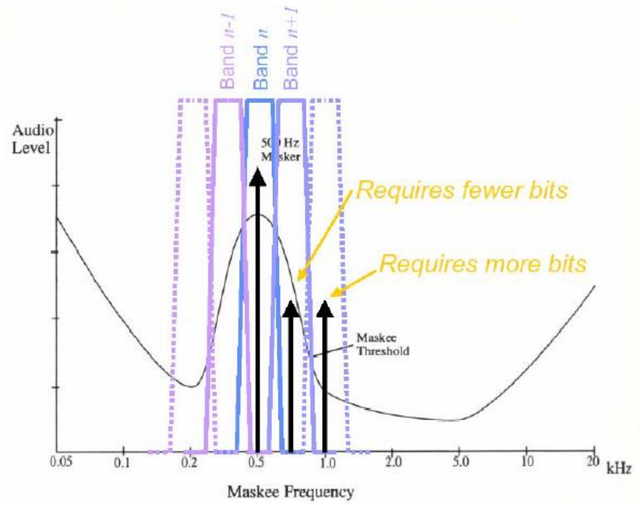
Effect of temporal masking depends on both time and closeness in frequency

CSULA CS4551 Multimedia Software Systems by Eun-Young Kang

Perceptual Coding

- *Perceptual coding* is a coding based on *psychoacoustic* model of human hearing.
- *Perceptual coding* tries to minimize the perceptual distortion in a transform coding scheme by carrying out the psychoacoustic model of hearing.
- Basic concept: allocate more bits (more quantization levels, less error) to those frequency channels that are most audible, fewer bits (more error) to those channels that are the least audible.
- Needs to continuously analyze the signal to determine the current audibility threshold curve using a perceptual model.

Perceptual Coding



Perceptual Coding Example

- Assume that the levels of 16 are:

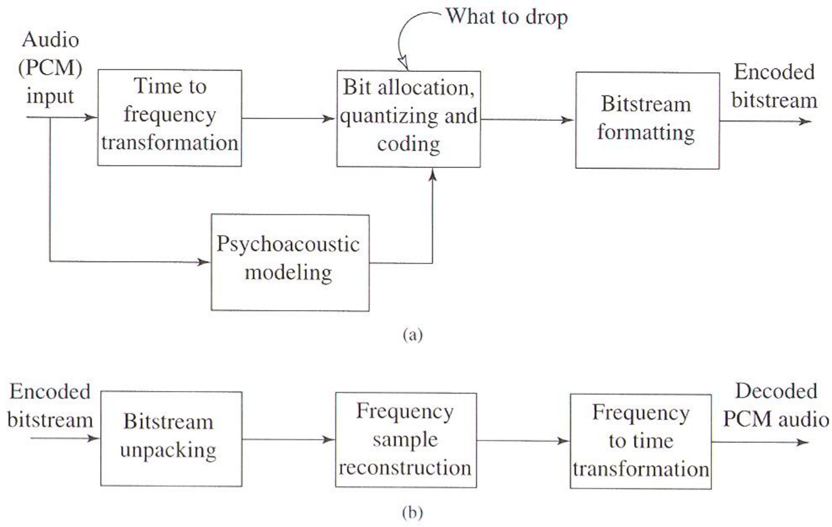
Band	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Level (db)	0	8	12	10	6	2	10	60	35	20	15	2	3	5	3	1

- Assume that if the level of the 8th band is 60dB, it gives a masking of 12dB 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 encode it.

Audio Coding : Main Standards

- ISO MPEG (Moving Picture Expert Group) family
  - MPEG1-Layer 1, Layer 2, **Layer 3 (MP-3)**
  - MPEG2-Backward compatible with MPEG1, AAC(non-backward compatible)
  - MPEG4–CELP (Code Excited Linear Prediction) and AAC (Advanced Audio Coding)

Basic MPEG-1 Audio Encoder/Decoder



MPEG-1 Coding Steps

1. Divide the audio signal into frequency subbands.
2. Determine amount of masking for each band based on its frequency and the energy of its neighboring band in frequency and time (the psychoacoustic model).
3. If the energy in a band is below the masking threshold, don't encode it. Otherwise, determine number of bits needed to represent the coefficient in this band.
4. Format bitstream: insert proper headers, code the side information (e.g. quantization scale factors) and code the quantized coefficient indices, using variable length encoding (e.g. Huffman coding).

MPEG-1 Audio Coding

- Layered Audio Compression Scheme, each being backward compatible
- **Layer1**
  - *Transparent (undetectable difference from original signal)* at 384 Kbps/channel
  - Digital Audio Tape typically uses Layer1.
  - Use only *frequency masking* of the psychoacoustinc model
  - Subband coding with 32 channels (12 samples/band)
  - No entropy coding after transform coding
  - Decoder is much simpler than the encoder.

MPEG-1 Audio Coding

- **Layer2**
  - Transparent at 296 Kbps/channel
  - Proposed for use in **digital audio broadcasting**
  - Subband coding with 32 channels (36 samples/band)
  - Use a bit of the *temporal masking* as well
  - Bitrate reduction and quality improvement at the price of an increase in complexity

MPEG-1 Audio Coding

- **Layer 3 (MP3)**
  - Transparent at 96 Kb/s per channel
  - Aimed at audio transmission over ISDN line. Gained popularity via MP3 players
  - Use *temporal masking* as well
  - Use 32 channels more close to true perceptual critical bands
  - Has entropy coder (Huffman)!
  - Much more complex than Layer 1 and 2
  - At the time of MPEG1 audio development (finalized 1992), Layer 3 was considered too complex to be practically useful. But today, layer 3 is the most widely deployed audio coding method (known as MP3).

Sound quality	Bandwidth	Mode	Compression ratio
Telephony	3.0 kHz	Mono	96:1
Better than shortwave	4.5 kHz	Mono	48:1
Better than AM radio	7.5 kHz	Mono	24:1
Similar to FM radio	11 kHz	Stereo	26:1 to 24:1
Near-CD	15 kHz	Stereo	16:1
CD	> 15 kHz	Stereo	14:1 to 12:1

MP3 Performance Comparison

CSULA CS4551 Multimedia Software Systems by Eun-Young Kang