Audio Signal Processing II

Shyh-Kang Jeng
Department of Electrical Engineering/
Graduate Institute of Communication Engineering

Overview

- Psychoacoustics
  - Study the correlation between the physics of acoustical stimuli and hearing sensations
  - Experiments data and models are useful for audio codec
- Modeling human hearing mechanisms
  - Allows to reduce the data rate while keeping distortion from being audible

Sound Pressure Levels

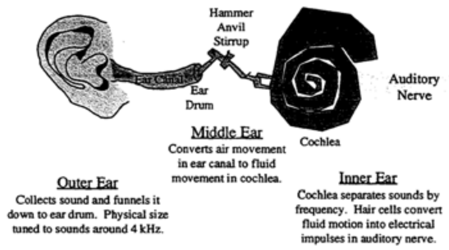
- Definitions

$$\text{SPL} = 20\log_{10}(p/p_o) \text{ in } dB \quad p_o = 20\ \mu Pa$$

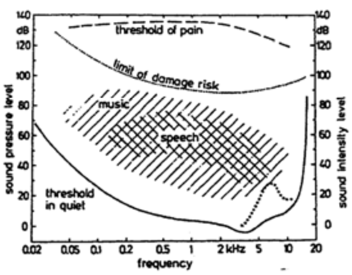
$$10^{-5}\,Pa < p < 10^{2}\,Pa$$

$$\text{SPL} = 10\log_{10}(I/I_o) \text{ in } dB \quad I_o = 10^{-12}\,W/m^2$$

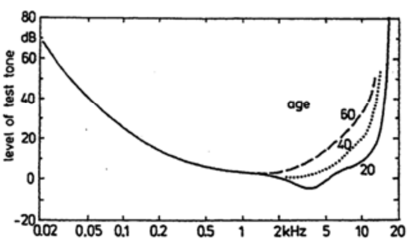
Outer, Middle, and Inner Ear



Hearing Area



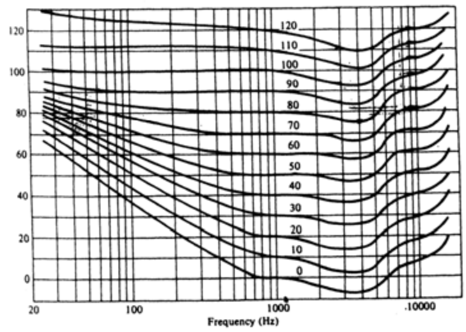
Threshold in Quiet



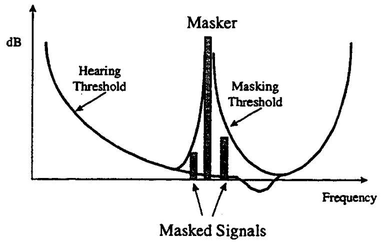
Loudness

- A 1 KHz tone at 100 dB is perceived as loud as a 100 Hz at 100 dB
- A 1 KHz tone at 40 dB is 20 dB louder than a 200 Hz at 40 dB
- Loudness level
  - The level of a 1 KHz tone that is as loud as the sound
- Unit
  - phon

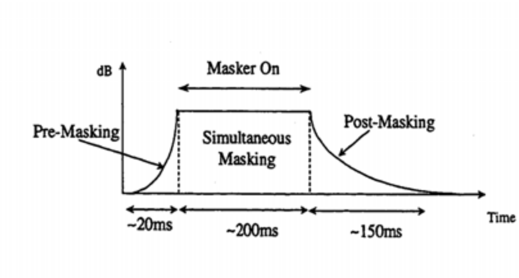
Fletcher-Munson Equal Loudness Curves



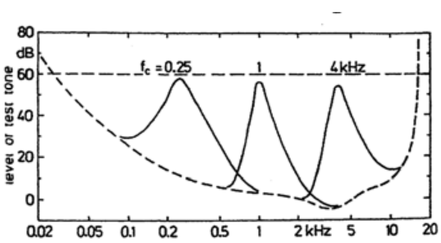
Frequency Masking



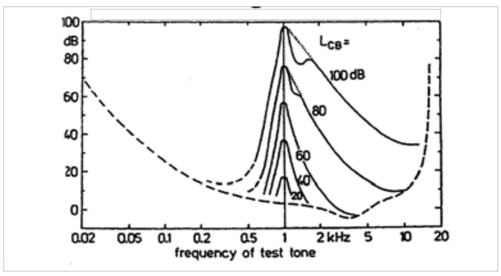
Temporal Masking



Narrow-Band Noise Masking Tones



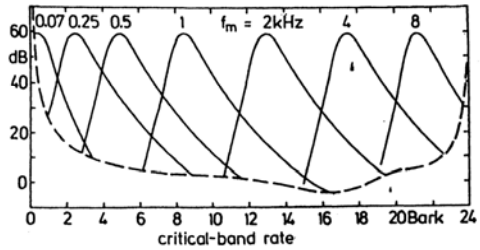
Masking Thresholds at Different Masking Levels



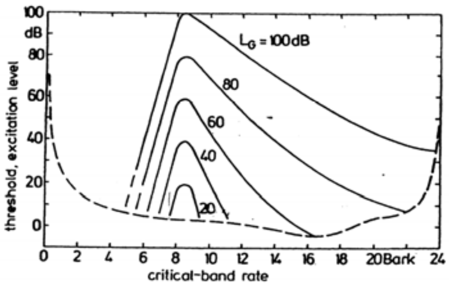
Bark Scale

Bark	f_l/f_u (Hz)	f_c	Δf
0	0	50	100
1	100	150	100
2	200	250	100
3	300	350	100
4	400	450	110
5	510	570	120
6	630	700	140
7	770	840	150
8	920	1000	160
9	1080	1170	190
10	1270	1370	210
11	1480	1600	240
12	1720	1850	280
13	2000	2150	320
14	2320	2500	380
15	2700	2900	450
16	3150	3400	550
17	3700	4000	700
18	4400	4800	900
19	5300	5800	1100
20	6400	7000	1300
21	7700	8500	1800
22	9500	10500	2500
23	12000	13500	3500
24	15500		

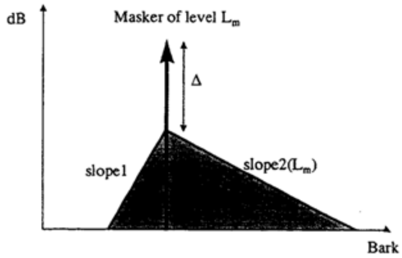
Threshold vs. Critical-Band Rate



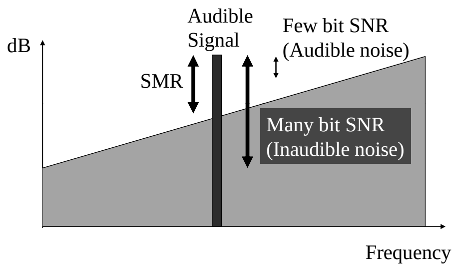
Threshold vs. Critical-Band Rate



Simple Masking Model



Bit Allocation Using Masking Thresholds



Transform Coding Data Rates

- Encoding in frequency domain
  - N equally spaced frequency bands
  - Encode each band with R_k bits
- Data rate of a critically sampling system

$$I = F_s \langle R_k \rangle \quad b/s$$

$$\langle R_k \rangle = (1/N) \sum_{k=0}^{N-1} R_k$$

- Typical data rate
  - from 64 kb/s/ch to 128 kb/s/ch

Example: TDAC Transform

- Sampling frequency $F_s = 48\ KHz$
- Window length 1024
- Bit rate 128 kb/s/ch
- Average bits per sample

$$\langle R_k \rangle = 128/48 = 2.6667$$

- Number of bits for each new block of data

$$\sum_{k=0}^{N-1} R_k = \langle R_k \rangle * 512 = 1365\ b/Block$$

53

Floating Point Quantization

- Effect of the scale factor
  - Scale $x_{\max}$ to the order of the signal so that the error in terms of the number of mantissa bits

$$\langle q^2 \rangle = \frac{\langle x^2 \rangle}{3 \cdot 2^{2R}}$$

- Get coding gain if R_k can reduce the error

$$\langle q^2 \rangle = \frac{1}{N} \sum_{k=0}^{N-1} \left(\frac{X_k^2}{3 \cdot 2^{2R_k}} \right)$$

54

Optimal Bit Allocation

- Optimization problem

$$\underset{\{R_k\}}{Minimize} \left\{ \frac{1}{N} \sum_{k=0}^{N-1} \left(\frac{X_k^2}{3 \cdot 2^{2R_k}} \right) \right\} \quad under \quad \frac{1}{N} \sum_{k=0}^{N-1} R_k = R$$

- Solution
  - Lagrange multiplier
  - Take derivative
  - Solve for R_k

55

Optimal Bit Allocation (cont.)

$$\frac{\partial}{\partial R_l} \left\{ \frac{1}{N} \sum_k \left(\frac{X_k^2}{3 \cdot 2^{2R_k}} \right) + \frac{\lambda}{N} \left(\sum_k R_k - NR \right) \right\} = 0$$

$$\sum_k R_k = NR$$

$$\lambda = \frac{X_k^2}{3/2\ln 2} 2^{-2R_k}$$

$$R_k = \frac{1}{2} \log_2 \frac{X_k^2}{\lambda} - \frac{1}{2} \log_2 (3/2\ln 2)$$

$$R_k = R + \frac{1}{2} \log_2 (X_k^2) - \frac{1}{2N} \log_2 \left(\prod_j X_j^2 \right)$$

56

Application to Perceptual Coding

- Not to minimize the average error power
- To get the quantization noise below the masking curve
- To maximize SNR-SMR for signals above the masking curve

$$\text{SNR} \sim \log(x^2/q^2) \quad \text{SMR} \sim \log(x^2/M^2)$$

$$\text{SNR} - \text{SMR} \sim -\log(q^2/M^2)$$

57

Application to Perceptual Coding (cont.)

- New problem

$$Minimize \left\{ \frac{1}{N} \sum_k \frac{(X_k/M_k)^2}{3 \cdot 2^{2R_k}} \right\} \quad under \quad \frac{1}{N} \sum_k R_k = R$$

- New solution

$$R_k = R + \frac{1}{2} \log_2 \left((X_k/M_k)^2 \right) - \frac{1}{2N} \log_2 \left(\prod_j (X_j/M_j)^2 \right)$$

58

A Caveat…

- The above algorithm sometimes gives negative R_k when X_k / M_k is much below its geometric mean
  - Rounds those R_k to zero
  - Take bits away from other parts of the spectrum
  - Use approximate solution allocating bits one by one locally

History

- Moving Picture Expert Group (MPEG)
  - Established in 1988
  - Joint Technical Committee (JTC1): ISO, IEC
  - Develop standards for coded representation of moving pictures and associated audio
- Original work items
  - MPEG-1, up to 1.5 Mb/s (ISO/IEC 11172)
  - MPEG-2, up to 10 Mb/s (ISO/IEC 13818)
  - MPEG-3, up to 40 Mb/s
- MPEG-3 was dropped in July '92

History (cont.)

- MPEG-4
  - First proposed in 1991
  - Approved in July 1993
  - Targets audiovisual coding at very low bit rates
  - Scalability, 3-D, etc.
  - ISO/IEC FDIS in 1999 (ISO/IEC 14496)
- MPEG-7
  - Started in the Fall of 1996
  - Standardize the description of multimedia contents of multimedia data base search
  - Scheduled to become ISO/IEC standard in 2001

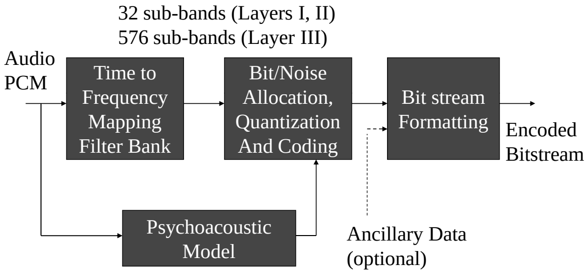
MPEG-1 Audio Layers

- Layer I
  - Simplest configuration, 32 to 224 kb/s/ch
  - Best for data rates above 128 kb/s/ch
  - Used in Philips's DCC at 192 kb/s/ch
- Layer II
  - Intermediate complexity, 32 to 384 kb/s/ch
  - Best for data rates of 128 kb/s/ch
  - Used in DAB, CD-Interactive, etc.
- Layer III
  - Highest quality and complexity, 32 to 160 kb/s/ch
  - Best for data rates below 128 kb/s/ch
  - Used for transmission over ISDN, Internet, etc.

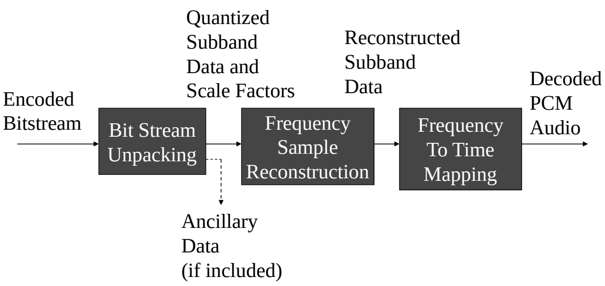
MPEG-1 Audio Layers (cont.)

- Single-chip, real-time decoders exist for all three layers
- Layers II and III
  - Perceptually lossless at 128 kb/s/ch (compression ratio of 6:1, 16 bits per sample, 48 KHz sampling rate)
  - Selected by ITU-R TG 10/2 for broadcast applications

MPEG-1 Encoder Building Blocks



MPEG-1 Decoder Building Blocks



65