Introduction to MP3 and psychoacoustics

Material from website by Mark S. Drew

http://www.cs.sfu.ca/CourseCentral/365/li/material/notes/Chap4/Chap4.4/Chap4.4.html

Human hearing and voice

* Frequency range is about 20 Hz to 20 kHz, most sensitive at 2 to 4 KHz.
* Dynamic range (quietest to loudest) is about 96 dB
* Normal voice range is about 500 Hz to 2 kHz
* Low frequencies are vowels and bass
* High frequencies are consonants

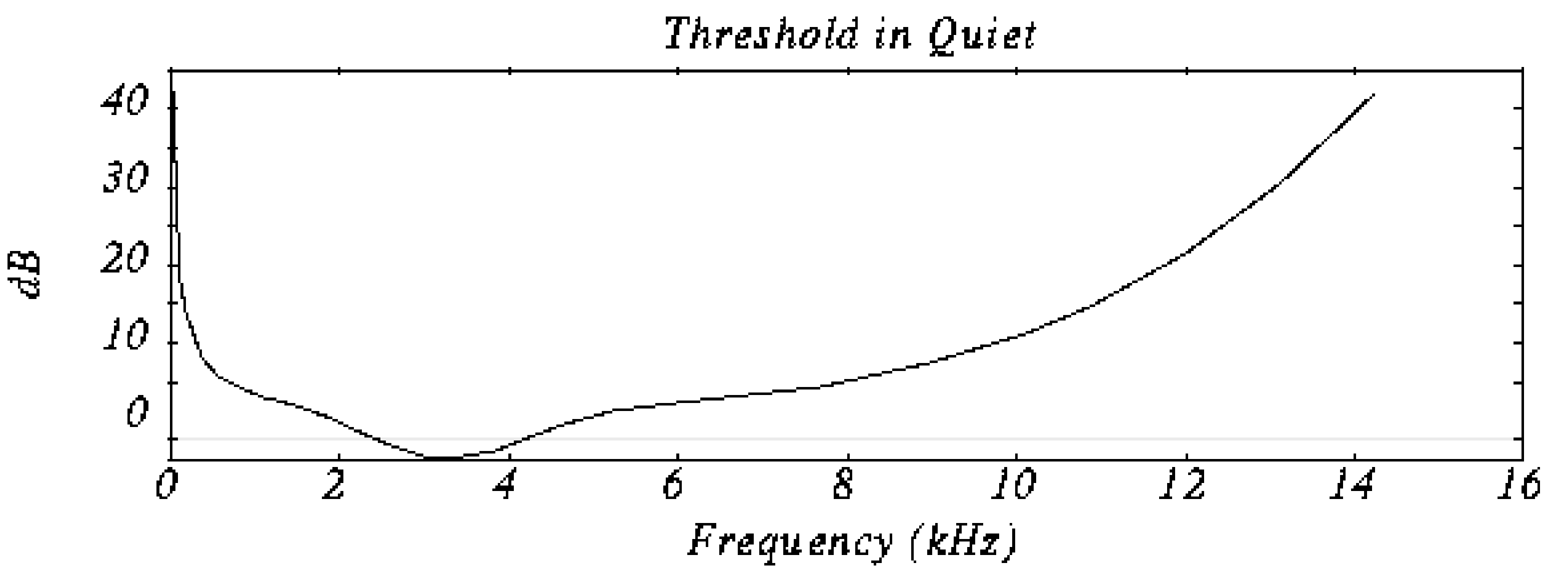
Efficient coding

Send what is audible; throw away what's not.

Or:
Only send what is needed (e.g. telephone cut-off for speech)

Sensitivity of human hearing in relation to frequency

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



Critical Bands

* Human auditory system has a limited, frequency-dependent resolution. The perceptually uniform measure of frequency can be expressed in terms of the width of the *Critical Bands*.
It is less than 100 Hz at the lowest audible frequencies, and more than 4 kHz at the high end. Altogether, the audio frequency range can be partitioned into 25 critical bands.

* A new unit for frequency *bark* (after Barkhausen) is introduced:

1 Bark = width of one critical band

For frequency < 500 Hz, it converts to $freq / 100$ Bark,

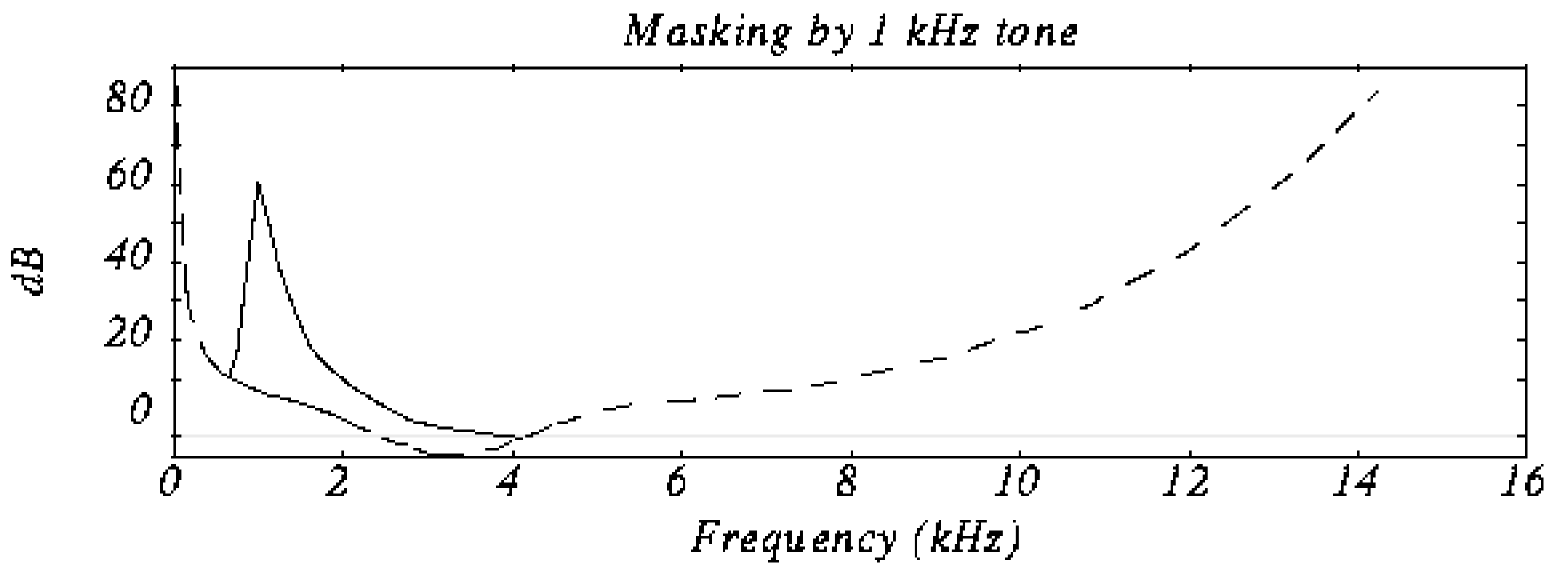
For frequency > 500 Hz, it is $9 + 4 \cdot \log_2(freq/1000)$ Bark.

Frequency Masking

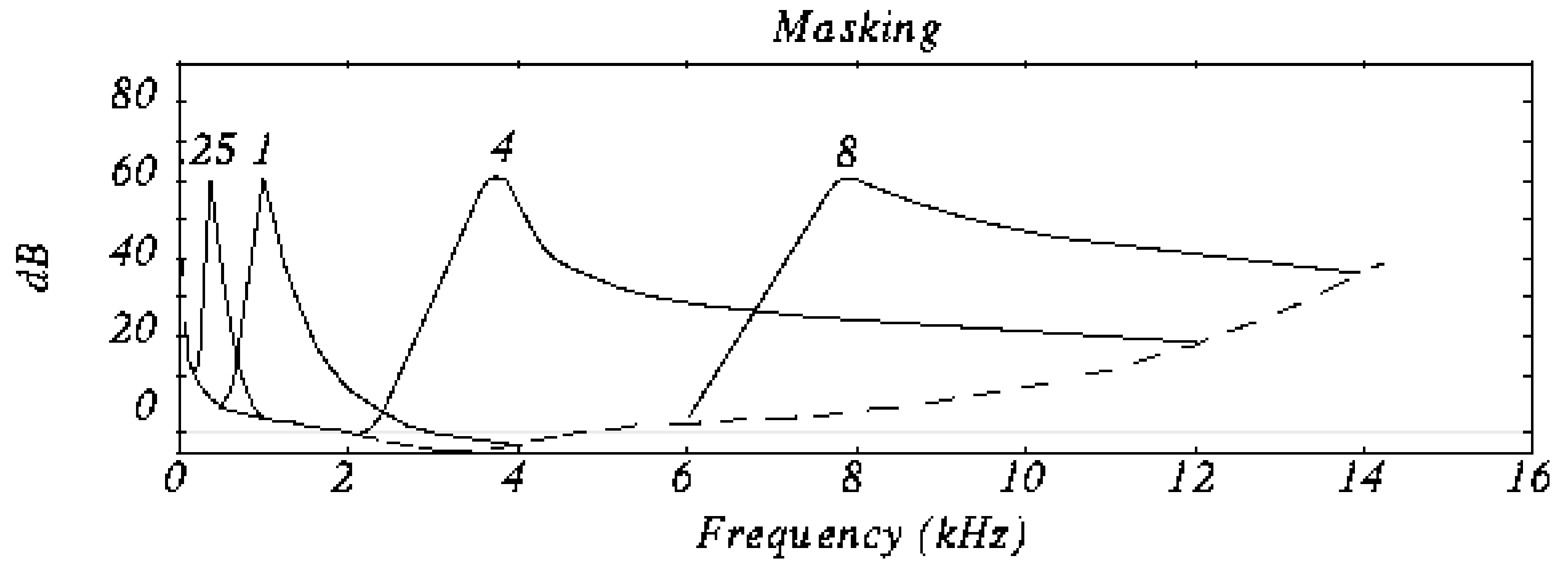
Question: Do receptors interfere with each other?

* 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.

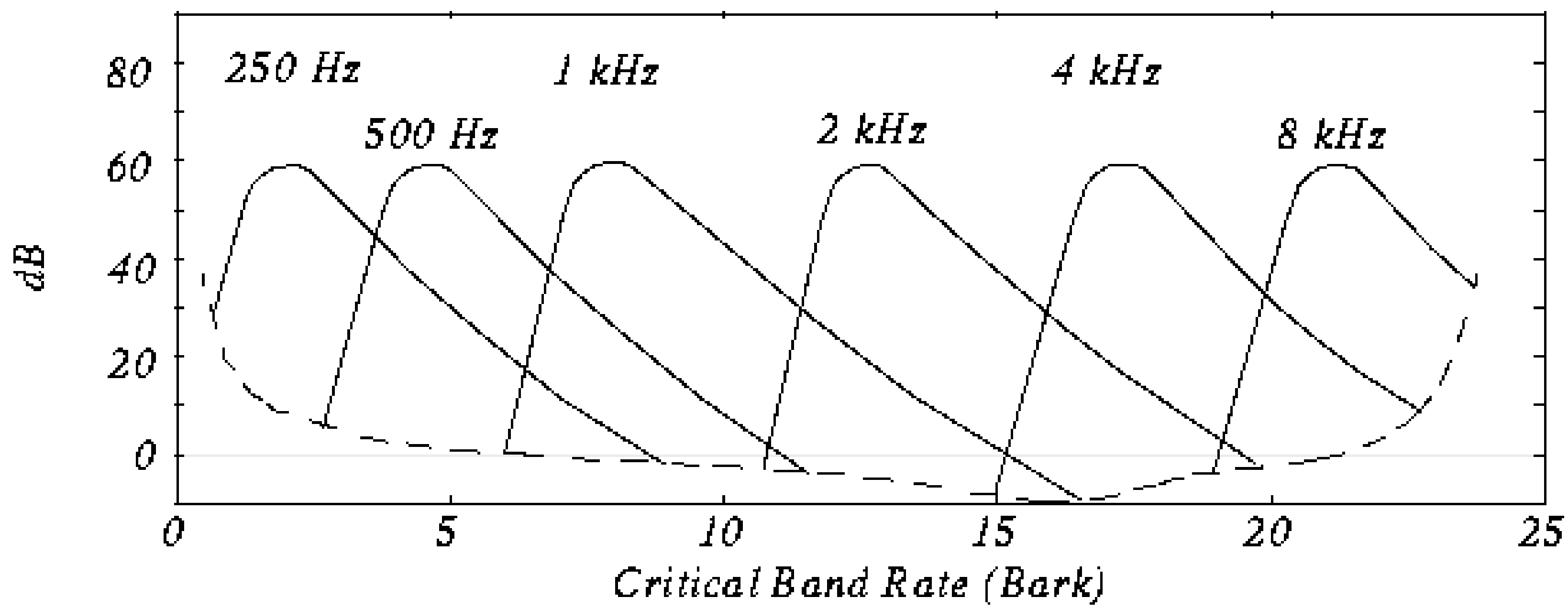
* Vary the frequency of the test tone and plot the threshold when it becomes audible:



Masking with various frequency masking tones.



Frequency Masking shown on critical band scale:



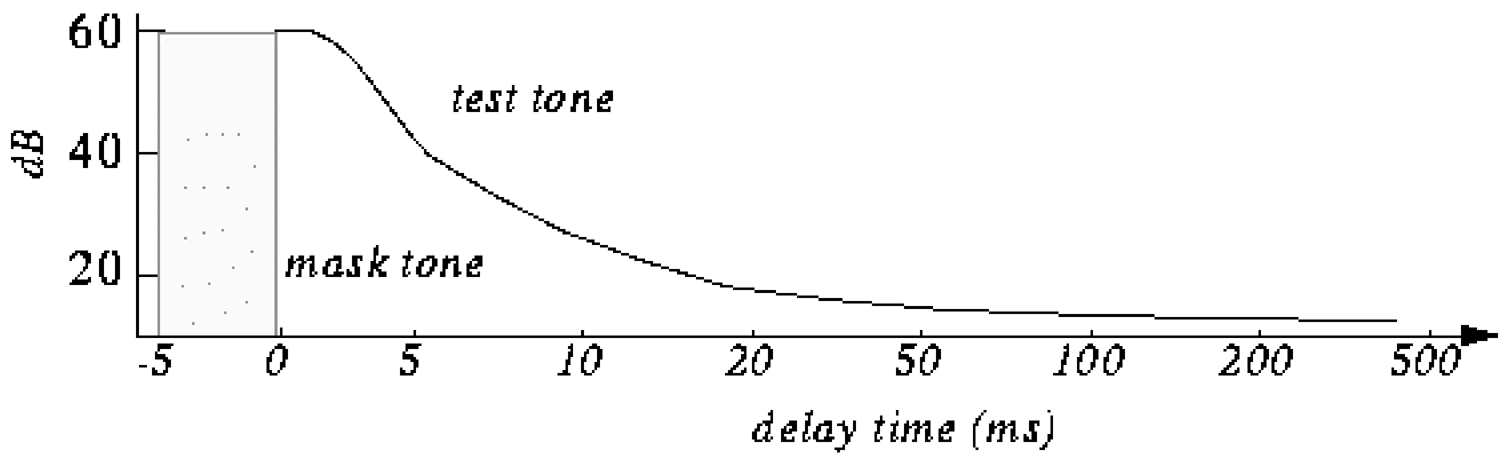
Temporal masking

* If we hear a loud sound, then it stops, it takes a little while until we can hear a soft tone nearby.
* 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.

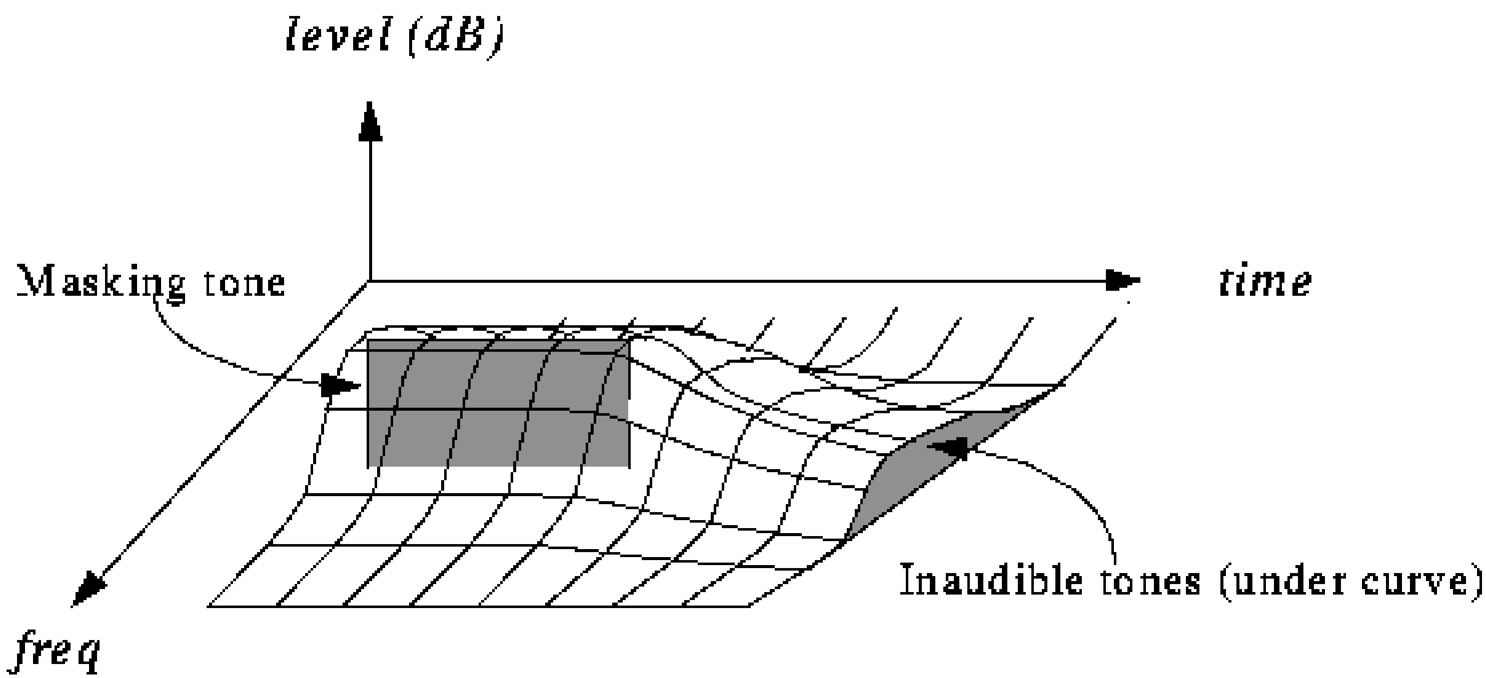
Adjust delay time to the shortest time when test tone can be heard (e.g., 5 ms).

Repeat with different level of the test tone and plot:



Total effect of both frequency and temporal maskings

:

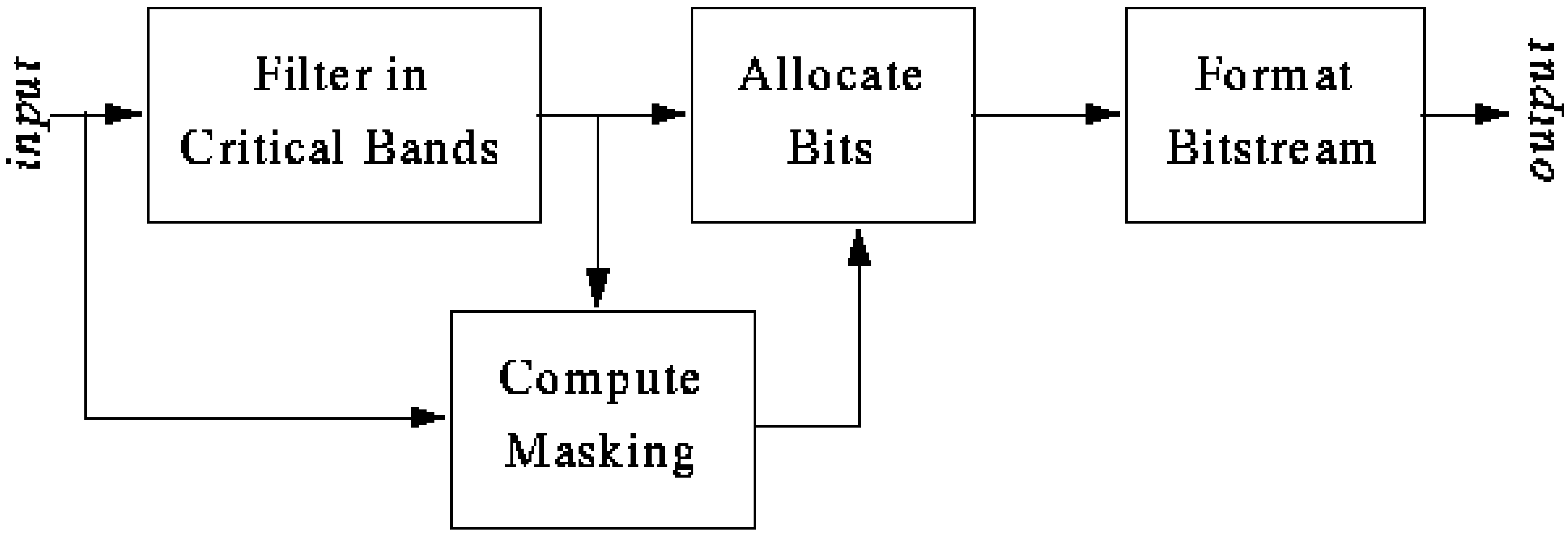


Steps in algorithm:

1. Use convolution filters to divide the audio signal (e.g., 48 kHz sound) into 32 frequency subbands --> *subband filtering*.

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

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 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



Example of running algorithm

Example:

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

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 -- a saving of 2 bits (= 12 dB).]