EECS 225D
Audio Signal Processing in Humans and Machines

Lecture 16 – Perceptual Audio Coding

2012-3-14

Professor Nelson Morgan

today's lecture by John Lazzaro

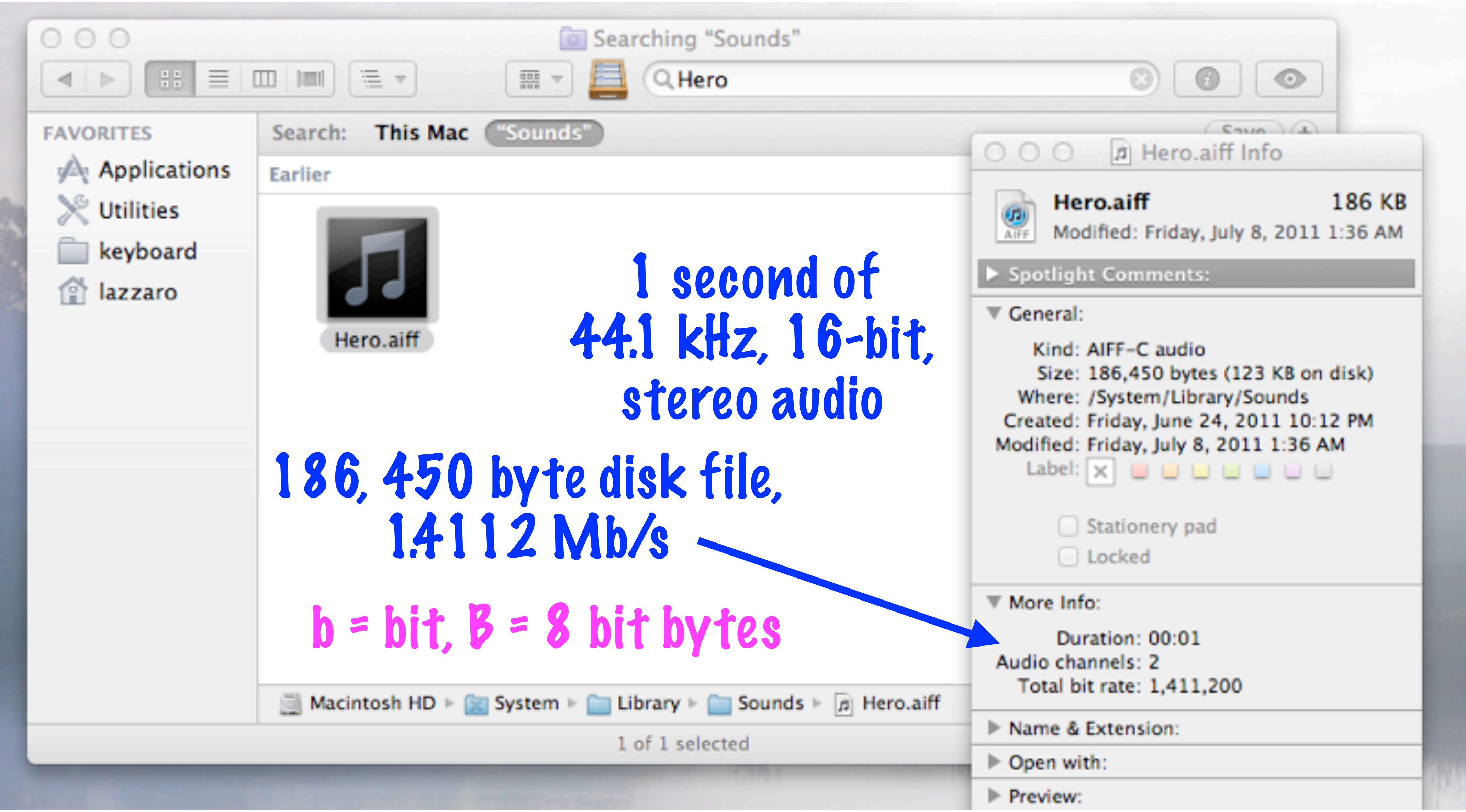
www.icsi.berkeley.edu/eecs225d/spr12/

"Hero" sound Play

Today's lecture: Audio Coding

- **Compression: Lossless and Lossy**
- Quantization and Noise
- Psychoacoustic Masking
- Time-Frequency Tradeoffs
- Research Topics

OS X System Sound: Hero.aiff



Play

How well does gzip work on audio files?

File size reduced by a factor of 1.56

(measured in units of 4KB disk blocks)

```
% gzip -c Hero.aif > Hero.aif.gz
% ls -1 -s -S  Hero.aif.gz Hero.aif
376 Hero.aif
240 Hero.aif.gz
% gunzip -c Hero.aif.gz > Hero-Check.aif
% diff -s Hero.aif Hero-Check.aif
Files Hero.aif and Hero-Check.aif are identical
% _
```

"Lossless" compression.

(decompression is bit-accurate).

Lossless algorithms remove redundant bits -- bits that are not needed to exactly reconstruct the original file.

"Shorten": Tony Robinson, Cambridge, 1992.

Redundancy removal can be improved if the algorithm can be specialized for audio waveforms.

Apple Lossless (after shorten, FLAC, ...)

File size reduced by a factor of 3.1

(double the performance of gzip on the same file)

Lossless, just like gzip.

```
% afconvert -f 'm4af' -d 'alac' Hero.aif Hero-Loss
% ls -1 -s -S Hero-Lossless.m4a Hero.aif
376 Hero.aif
120 Hero-Lossless.m4a
% afconvert -f "AIFC" -d 'BEI16' Hero-Lossless.m4a
% diff -s Hero.aif Hero-Check.aif
Files Hero.aif and Hero-Check.aif are identical
% _
```

To reduce file size by larger factors, we need to go beyond removing redundancy.

One approach: Remove information that is irrelevant information for a particular use case.

Example: Remove audio information whose loss a human listener cannot perceive.

Lossy perceptual audio coding.

MPEG 4 Advanced Audio Codec (AAC)

Request 128 kb/s

Encoder adjusts quality to meet request

File size reduced by a factor of 9.4

```
% afconvert -f 'm4af' -d 'aac ' -b 128000 Hero.aif
% ls -1 -s -S Hero-AAC.m4a Hero.aif
376 Hero.aif
 40 Hero-AAC.m4a
% afconvert -f "AIFC" -d 'BEI16' Hero-AAC.m4a Hero-
% diff -s Hero.aif Hero-Check.aif
Binary files Hero.aif and Hero-Check.aif differ
% _
```

Lossy: Decompression does not restore original file.

Today's Lecture
How it works

Listening Test: Original: Play

128 kb/s (9.4X): Play

16 kb/s (23.5X): Play

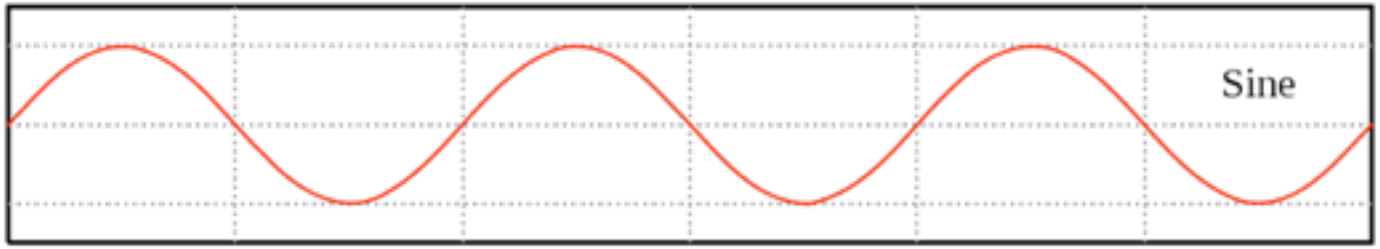
Today's lecture: Audio Coding

- Compression: Lossless and Lossy
- **Quantization and Noise**
- Psychoacoustic Masking
- Time-Frequency Tradeoffs
- Research Topics

Quantization, noise, and compression ...

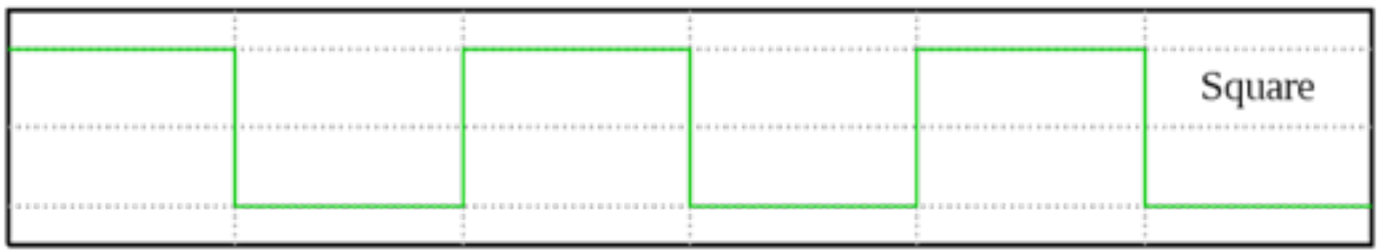
To compress a real-valued discrete-time waveform, quantize the samples to reduce bits/sample, and then apply lossless compression.

$s(t)$



Play

$s(t) + e(t)$



Play

Quantization corrupts the signal $s(t)$ with noise term $e(t)$.

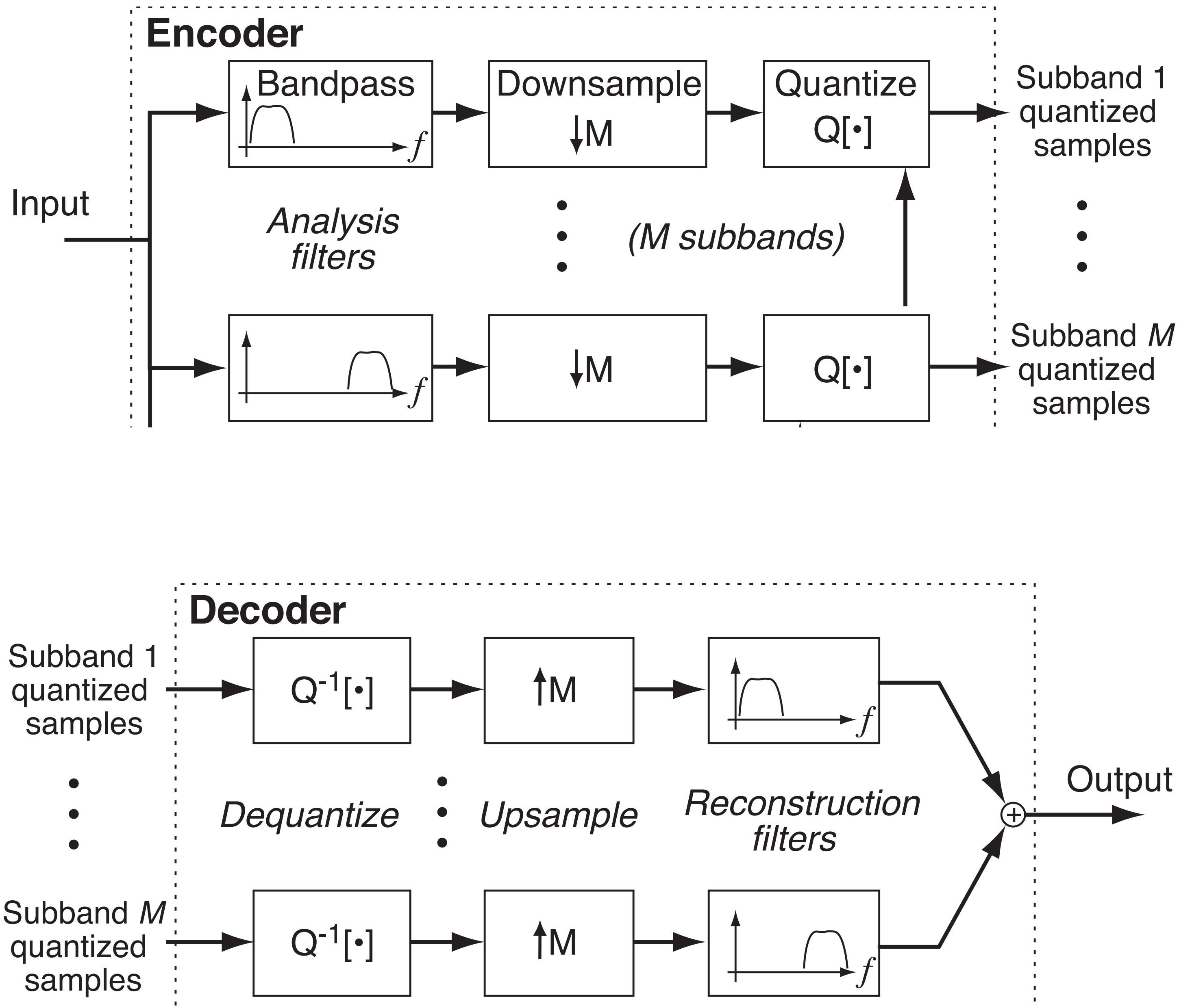
In this example, quantizing to 1 bit is clearly objectionable.
However, a 40 dB reduction in $e(t)$ yield a better result.

Quantizing with more bits acts to reduce $e(t)$.

$s(t) + 0.01 * e(t)$ Play

$s(t) + 0.1 * e(t)$ Play

Which leads to this architecture ...

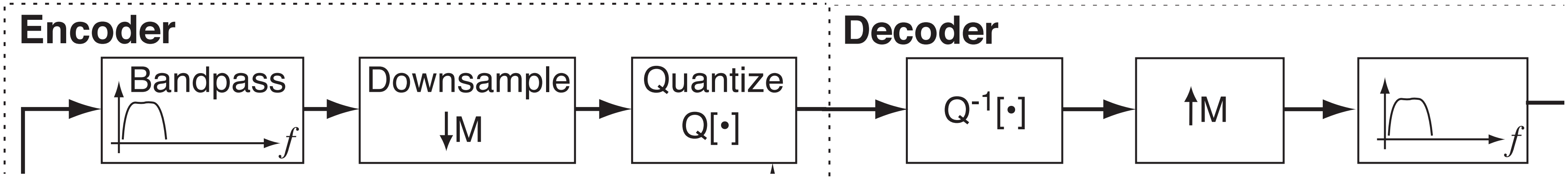


Filter bank splits audio input into **M sub-bands**.

Quantize to **minimize** the number of **bits** needed across **all M channels**.

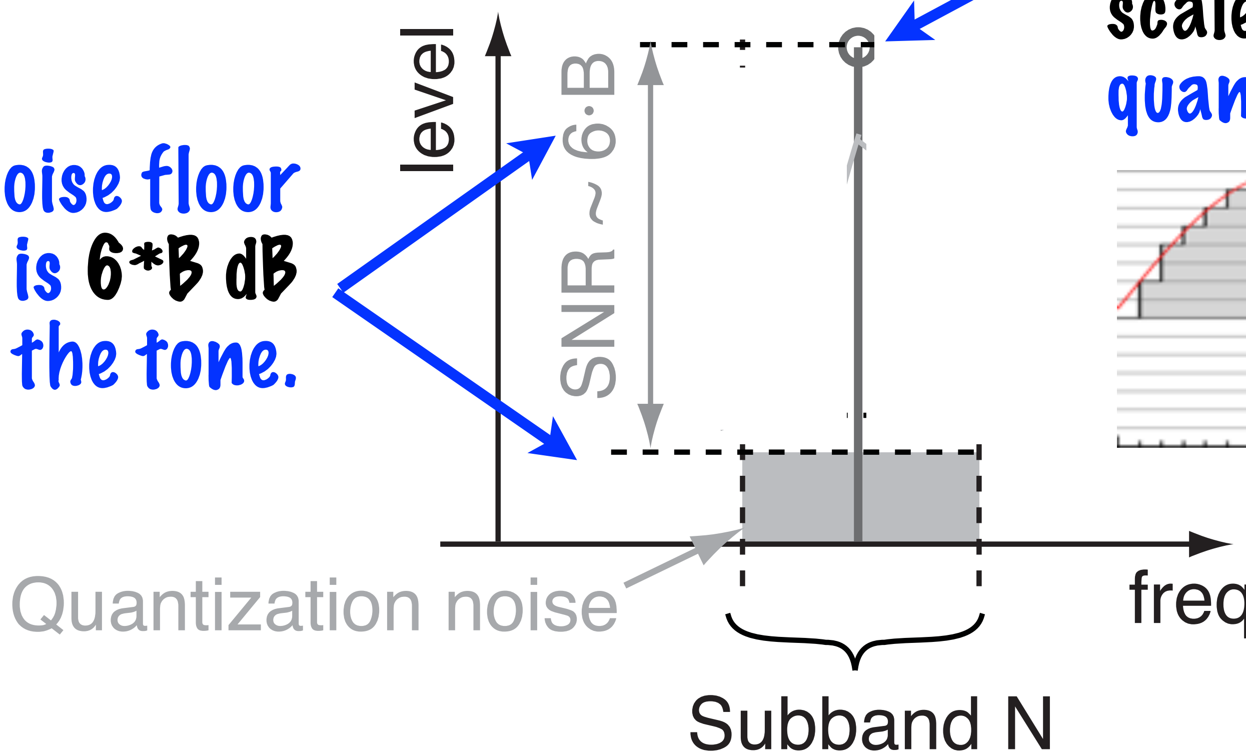
Constraint: Human **imperceptibility** of the **encode -> decode** process.

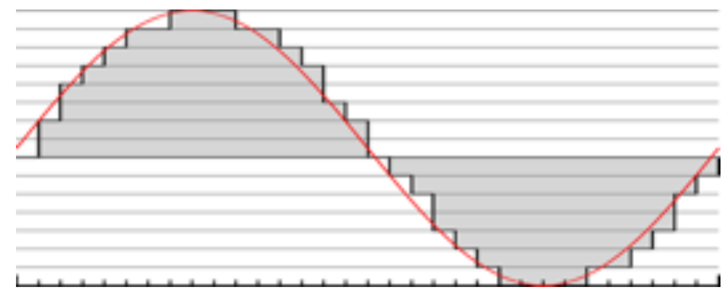
Quantization noise in a sub-band ...



A tone in the sub-band, scaled to fully use B quantize bits.

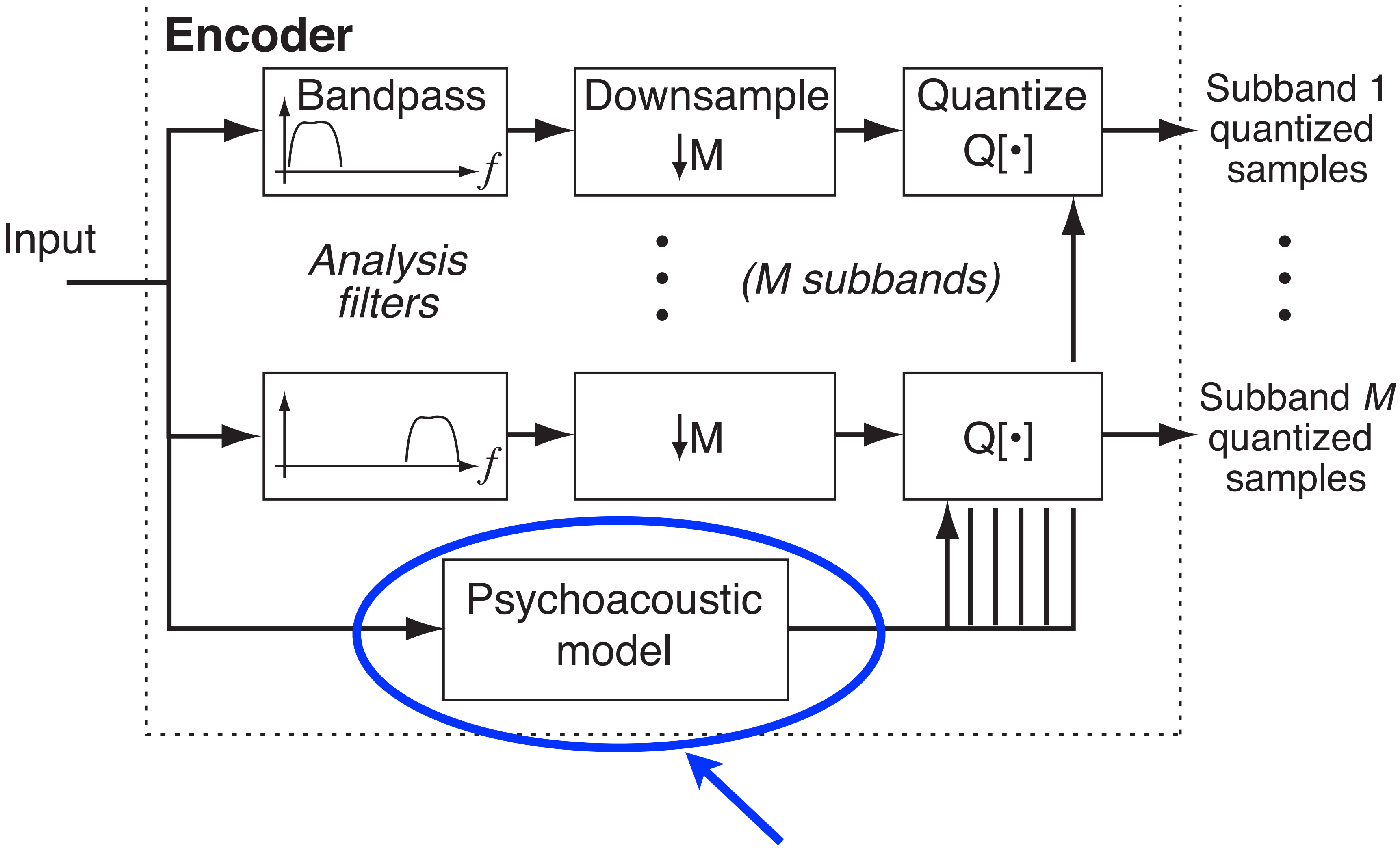
The noise floor is 6*B dB below the tone.





(Approximate result. See the book for the fine print)

If a B is too small, noise may be audible

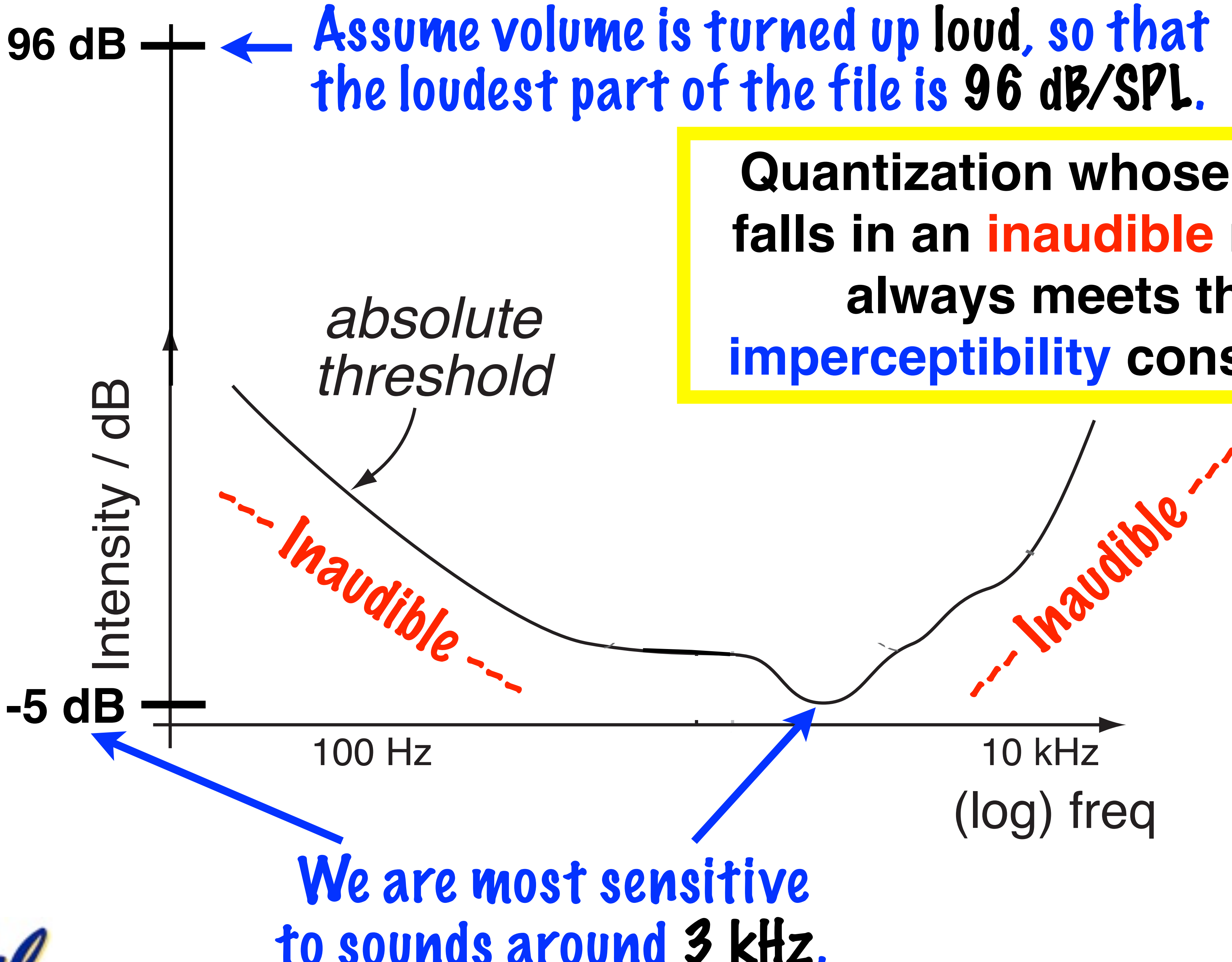


Encoder includes a model of human perception.
Candidate sets of M quantizations are tested against the model to check imperceptibility.

Today's lecture: Audio Coding

- Compression: Lossless and Lossy
- Quantization and Noise
- **Psychoacoustic Masking**
- Time-Frequency Tradeoffs
- Research Topics

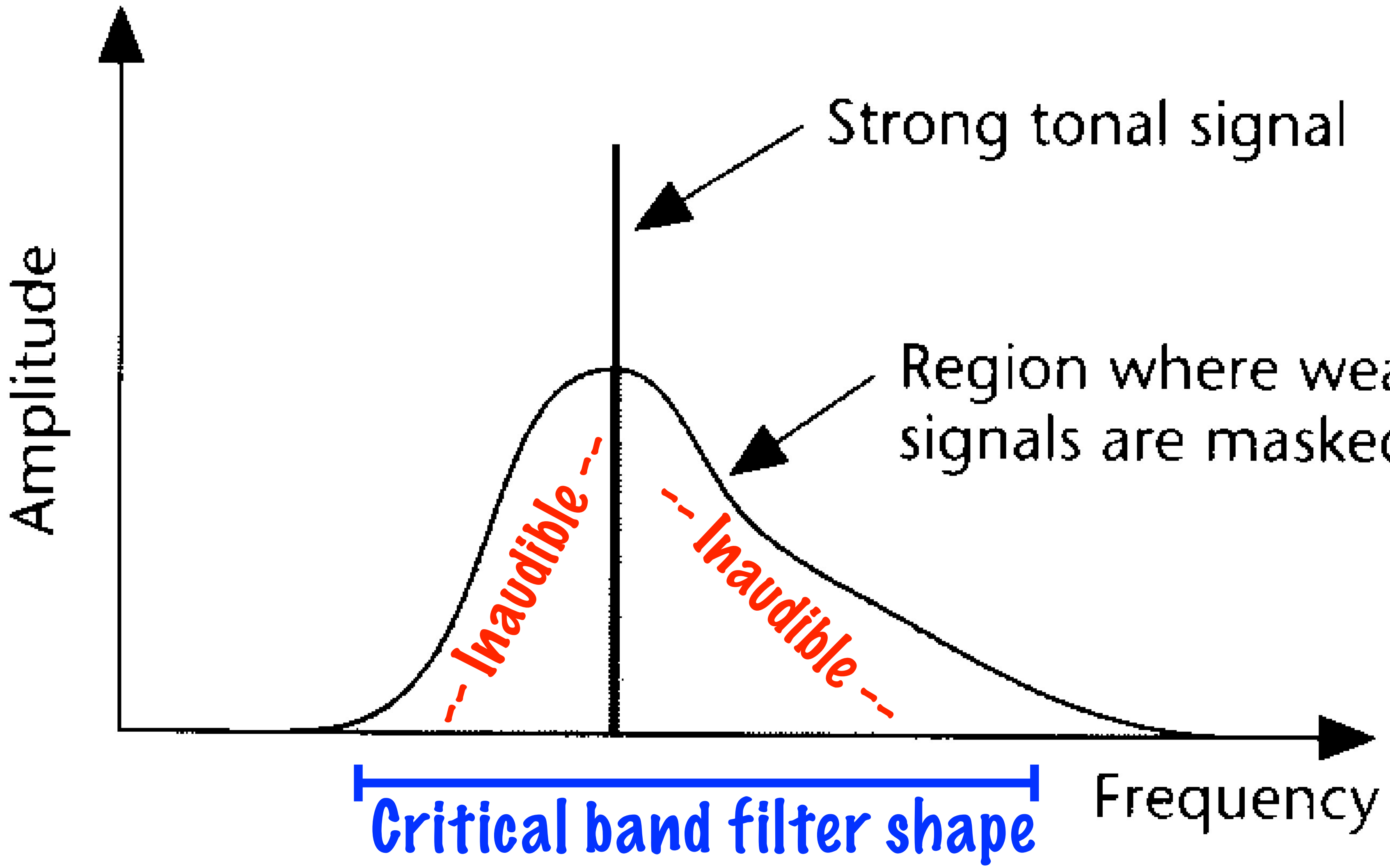
The absolute threshold of hearing ...



Quantization whose noise falls in an inaudible region always meets the imperceptibility constraint.

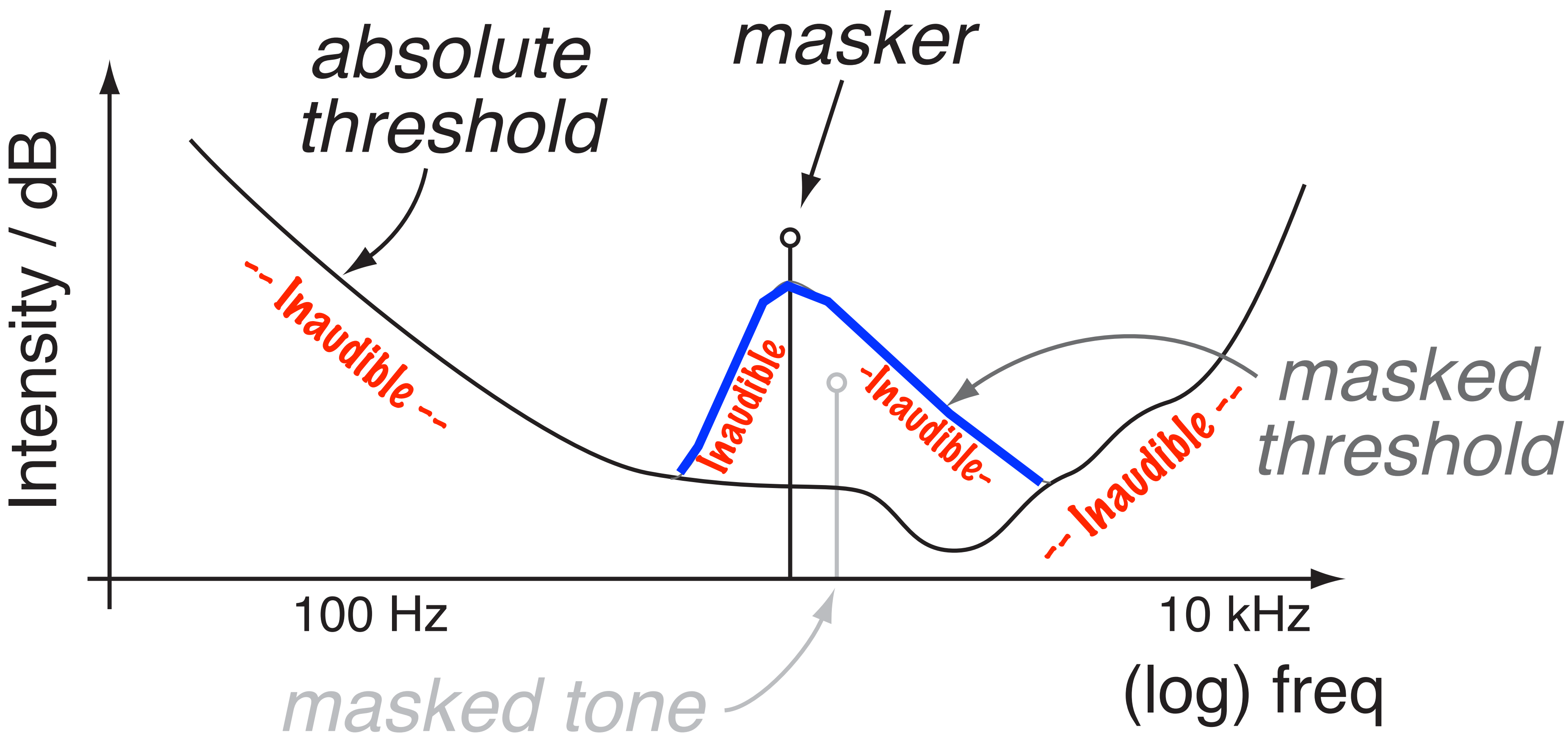
Tonal masking ....

Quantization noise will be imperceptible if it falls in the inaudible skirt surrounding a tonal signal.



Maskers compose using `max()` function

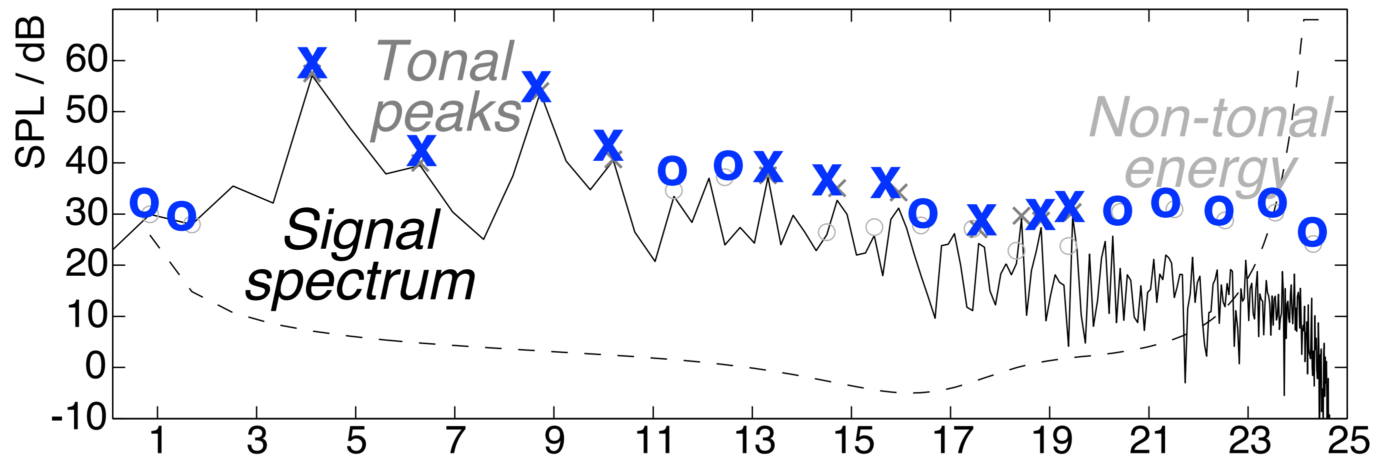
Given a short segment of wide-band audio, we can identify narrow-band maskers and compute a composite masking function for the audio signal.



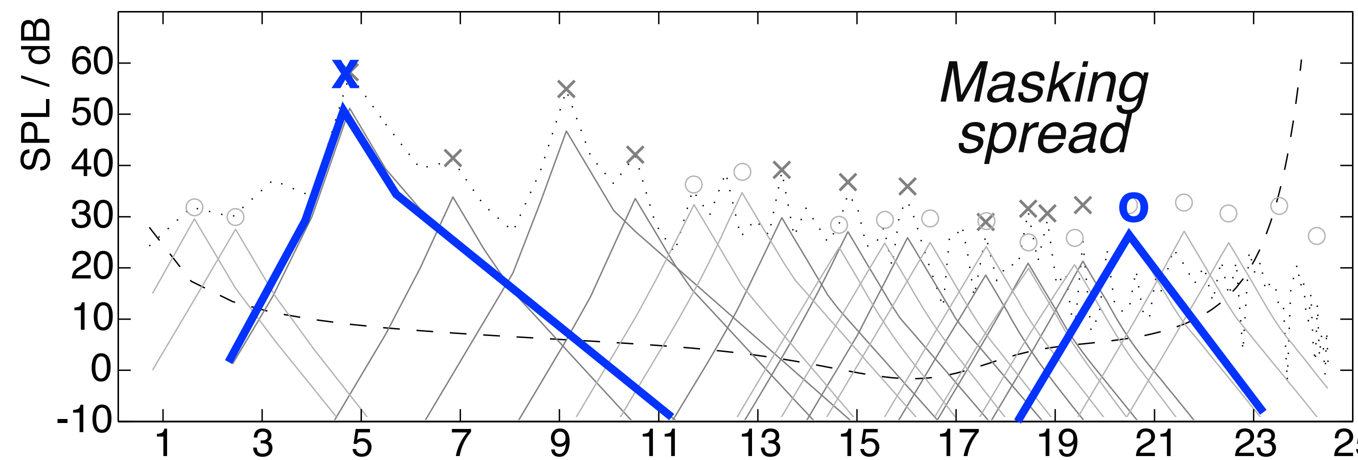
Effectively, tonal maskers locally raise the absolute threshold.

Computing a mask.

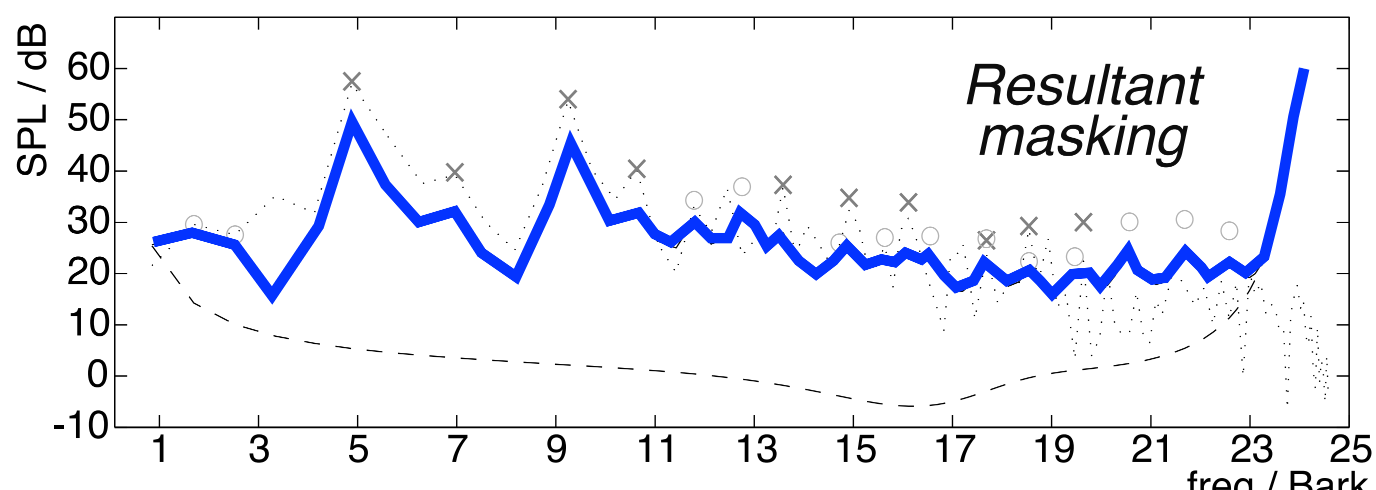
(analysis of a 26 ms audio "frame")



[1] Identify tonal (x) and non-tonal (o) energy peaks.



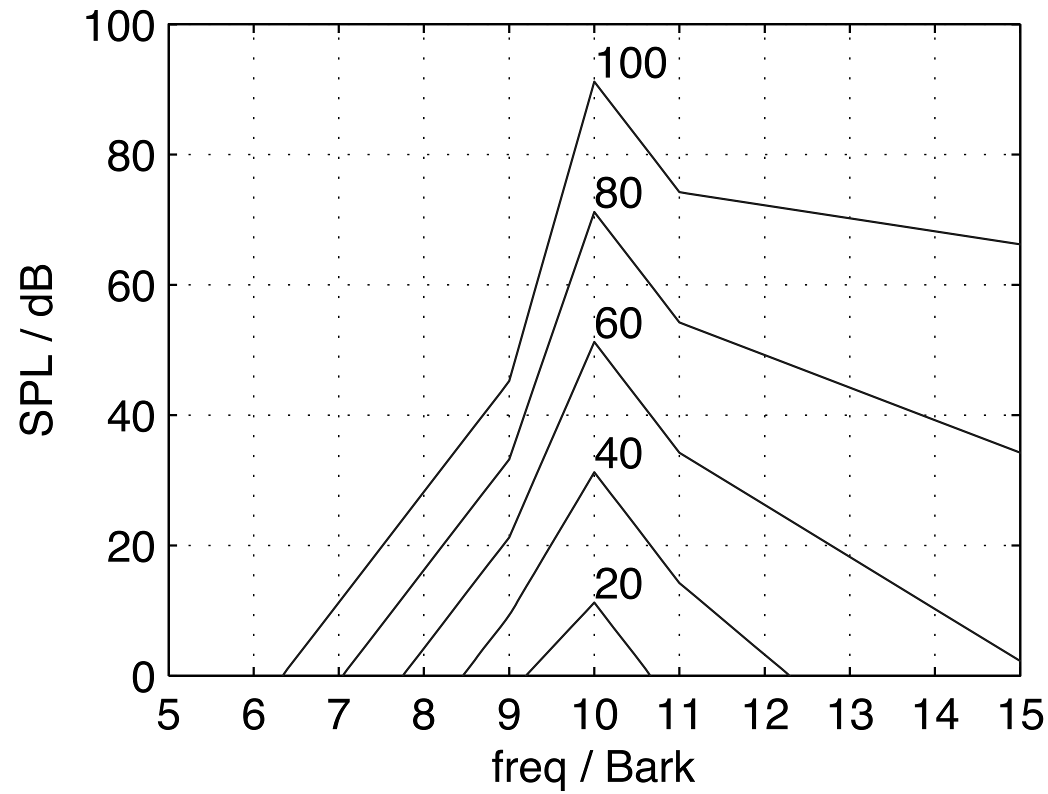
[2] Place a local masking function for each peak.



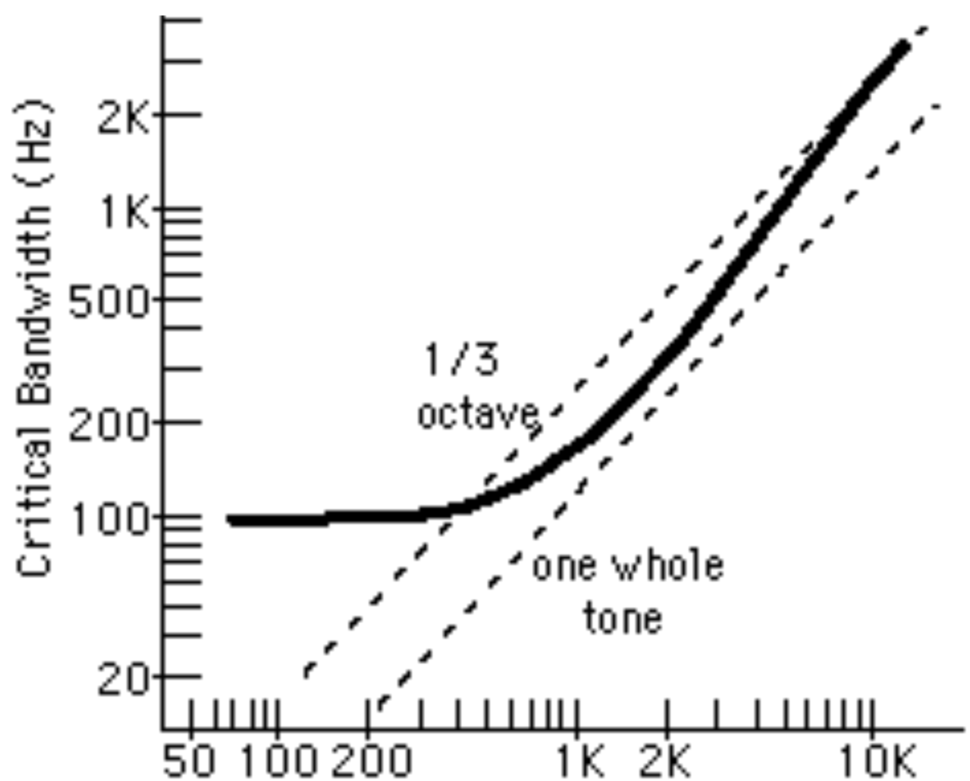
[3] Apply `max()` over frequency to compute the composite masker.

Masking functions

Masking function **widens** with masker level, following **cochlear filter** response shapes.



Masking function **widen** at higher channels, following **critical bandwidth**. The Bark scale **warping** handles this effect.

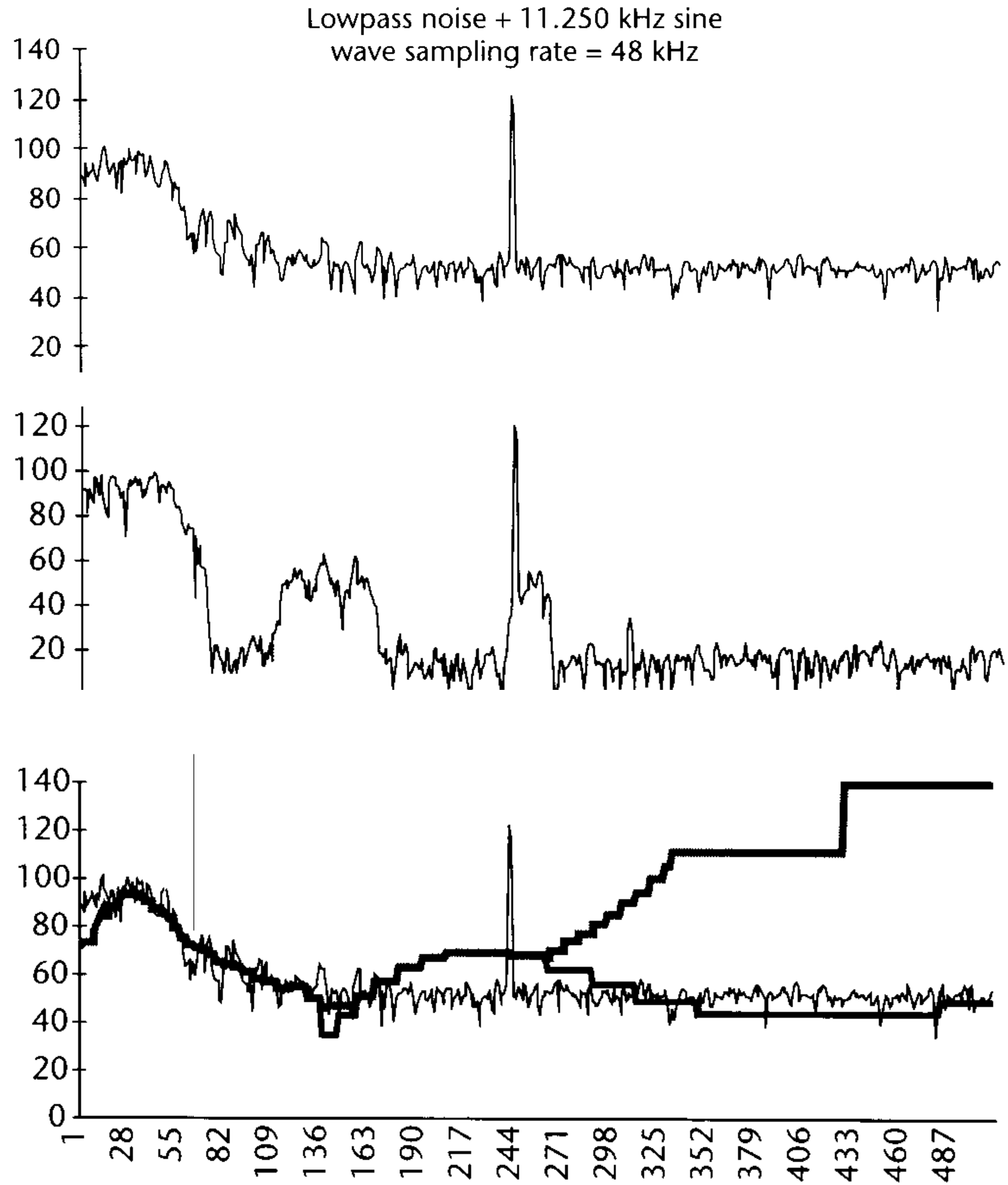


What is "lost"?

Input spectrum a mono audio frame.

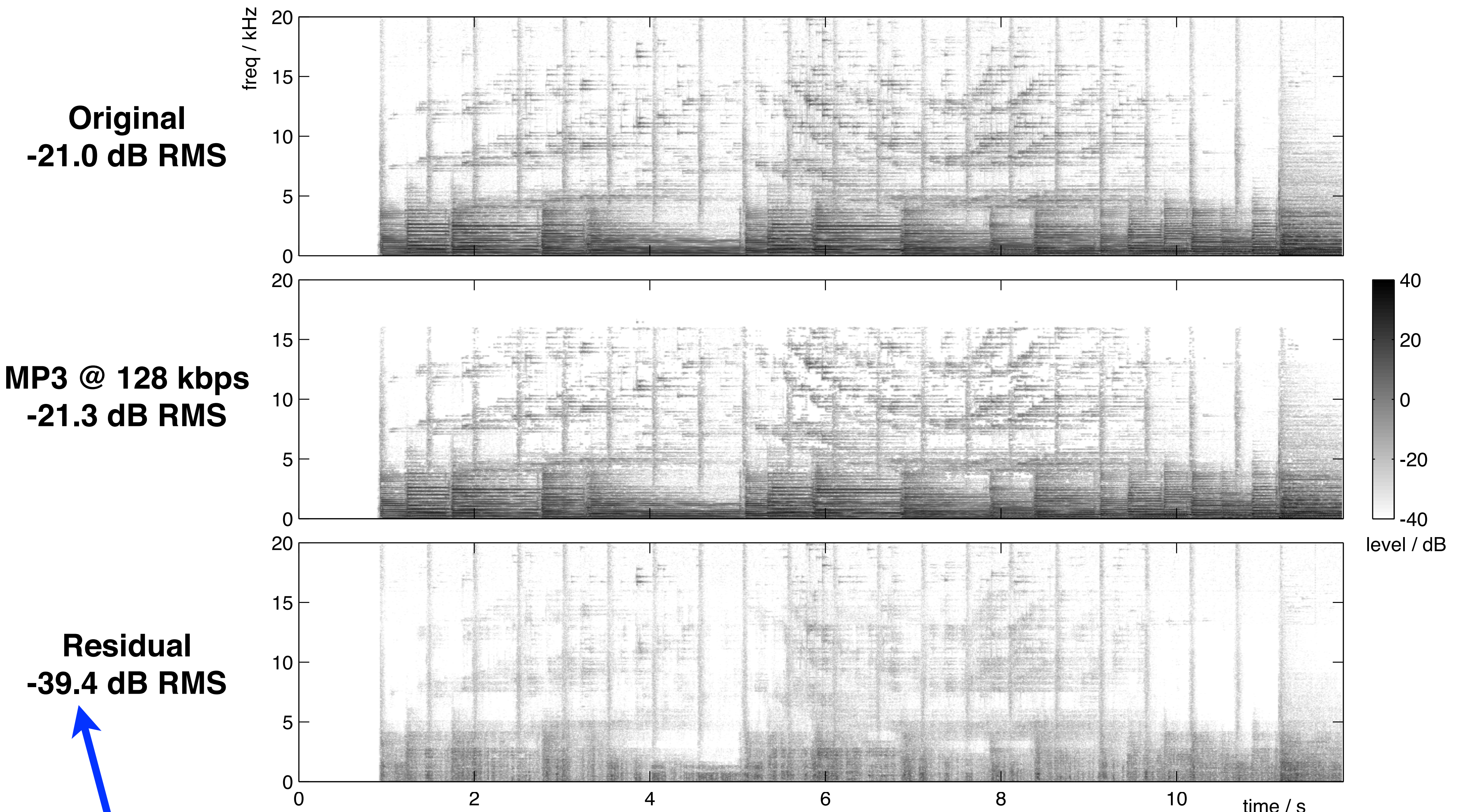
Spectrum of encoded audio (64 kb/s).

Masking profile that guided the quantization.



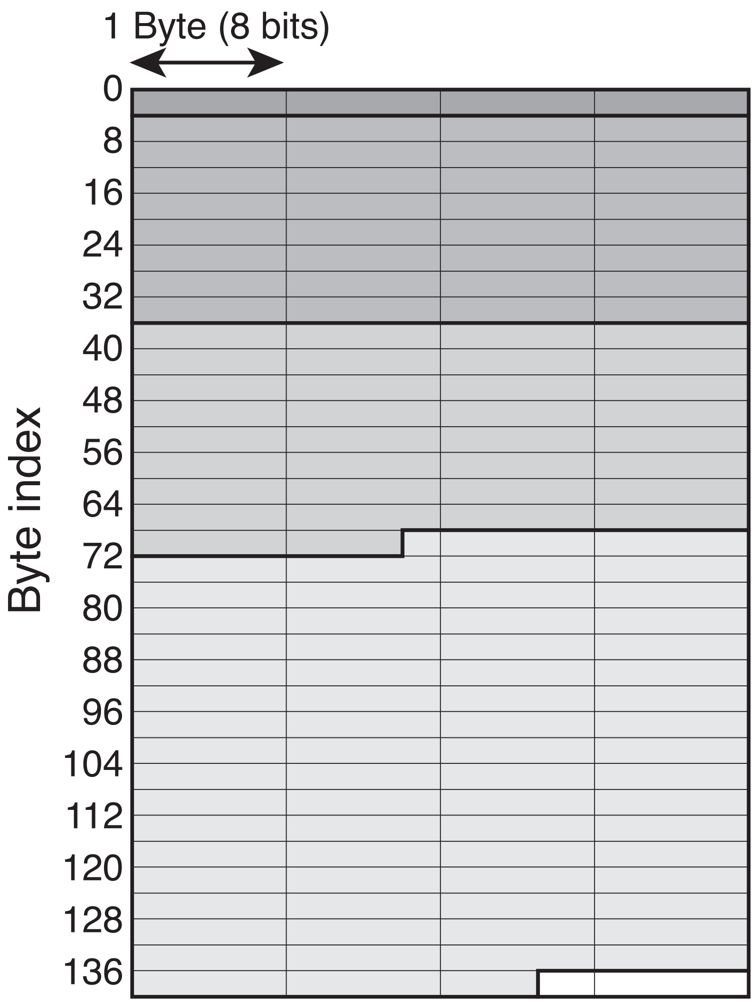
What is "lost"?

10 seconds of pop music content encoded using MP3 @ 128 kb/s.



Noise is only 10 to 20 dB below signal ... but carefully placed!

The bit level: An encoded frame in a file



Header: defines layer, bitrate, channels, etc. (4 bytes)

Subband bit allocation indices:
32 subbands x 2 channels x 4 bits
= 32 bytes

Subband scale factor indices:
32 subbands x 2 channels x 6 bits
(only for subbands with nonzero bit allocation)
≤ 48 bytes

Quantized subband samples:
32 subbands x 2 channels x 12 samples
x 2-15 bits / sample (as per bit allocation,
only for subbands with nonzero bit allocation)

Padding to make frame an integer number of 4 byte blocks

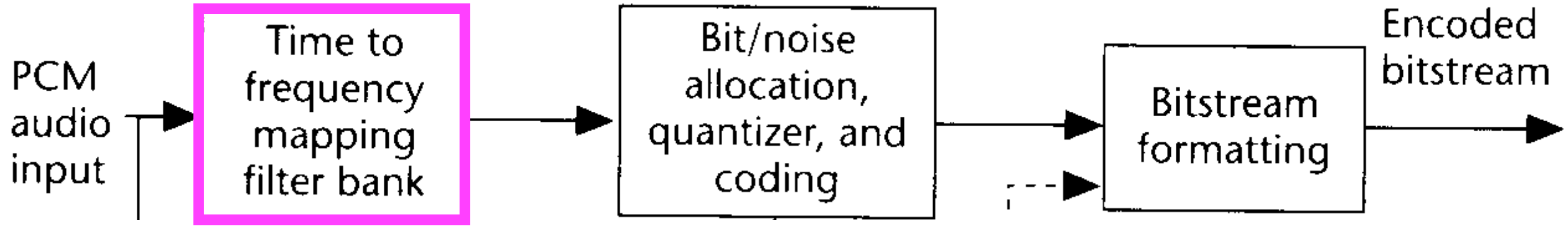
FIGURE 35.12 Bit usage layout in an example MPEG-1 Audio Layer I frame encoding 384 stereo samples in 140 bytes, for a bit rate of 128 kbps.

MP3: Lossless (Huffman) encoding used on "sample" field.

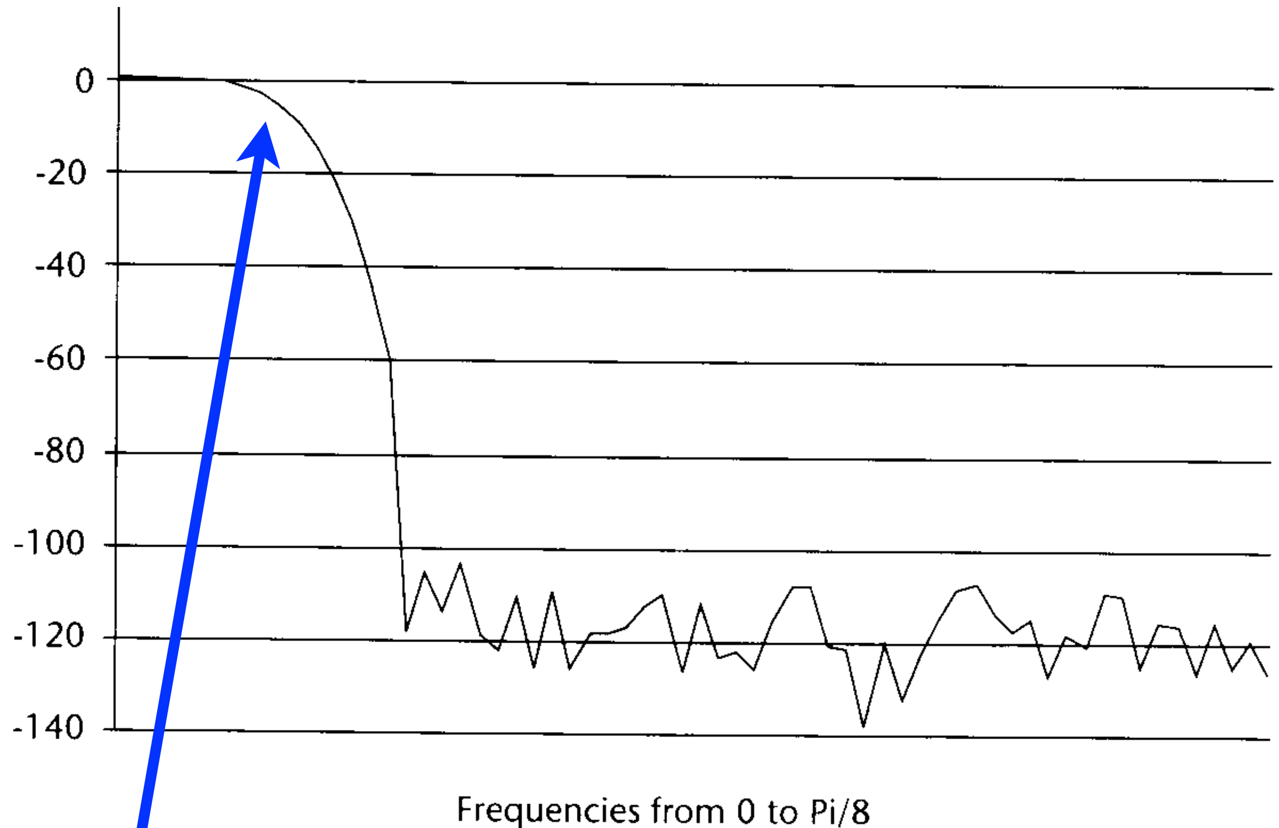
Today's lecture: Audio Coding

- Compression: Lossless and Lossy
- Quantization and Noise
- Psychoacoustic Masking
- **Time-Frequency Tradeoffs**
- Research Topics

Time-Frequency Tradeoff

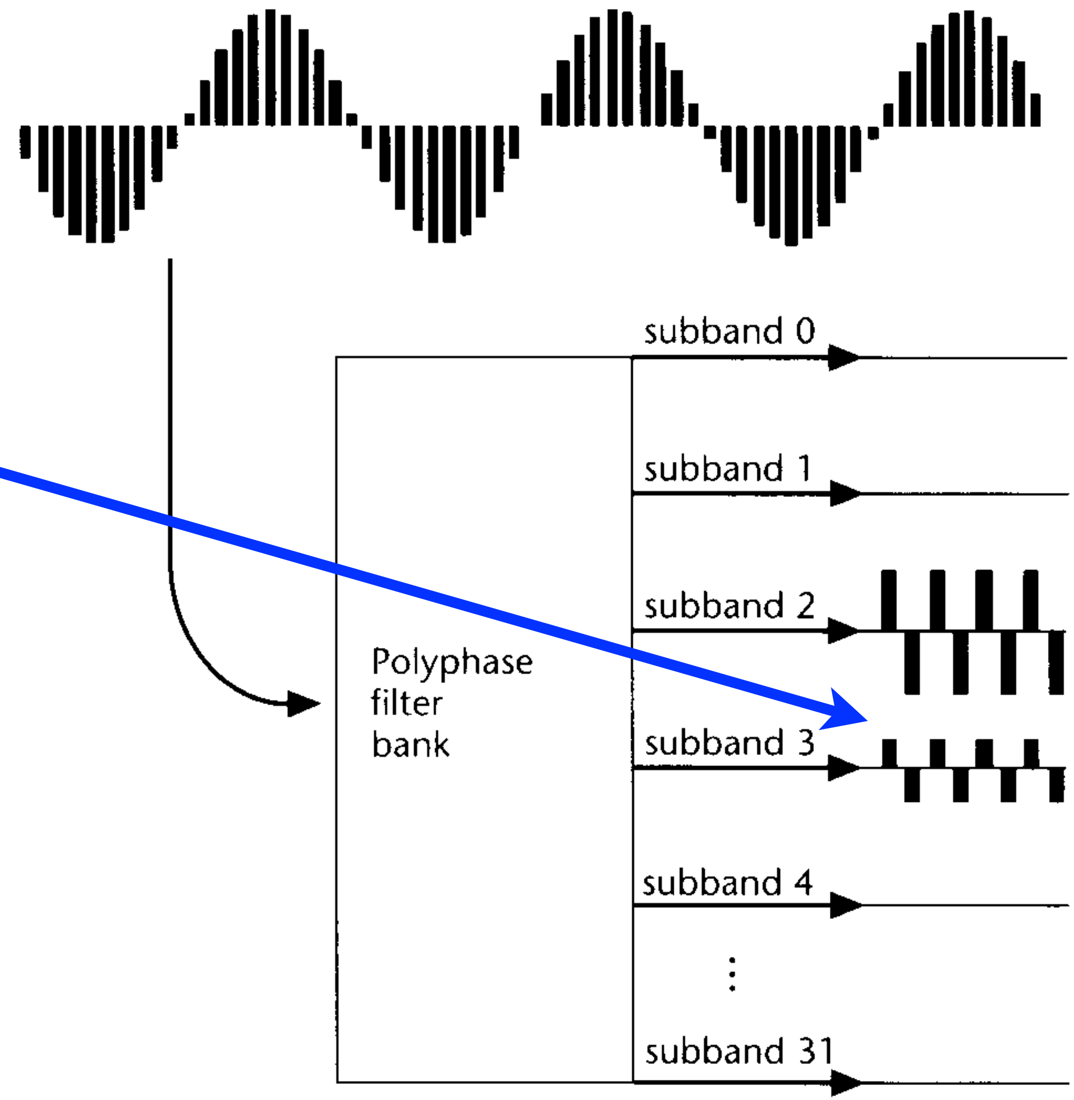


Good time resolution is required in the filter bank ...



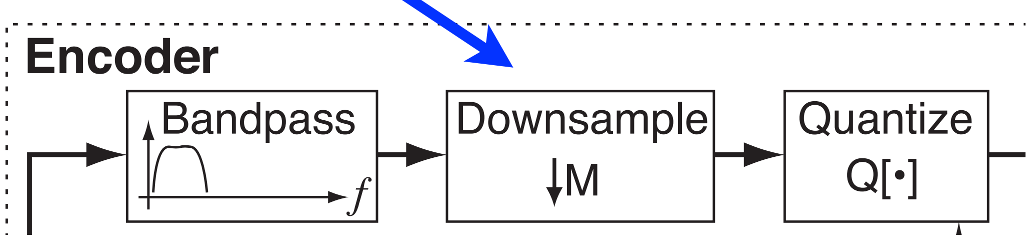
... which implies a gentle rolloff in frequency.

Time-Frequency Tradeoff



... which results in **aliases** appearing in sub-band outputs ...

... which **fold over** as we move the sub-band to baseband by **downsampling**.



Solution: Quadrature-mirror filter banks

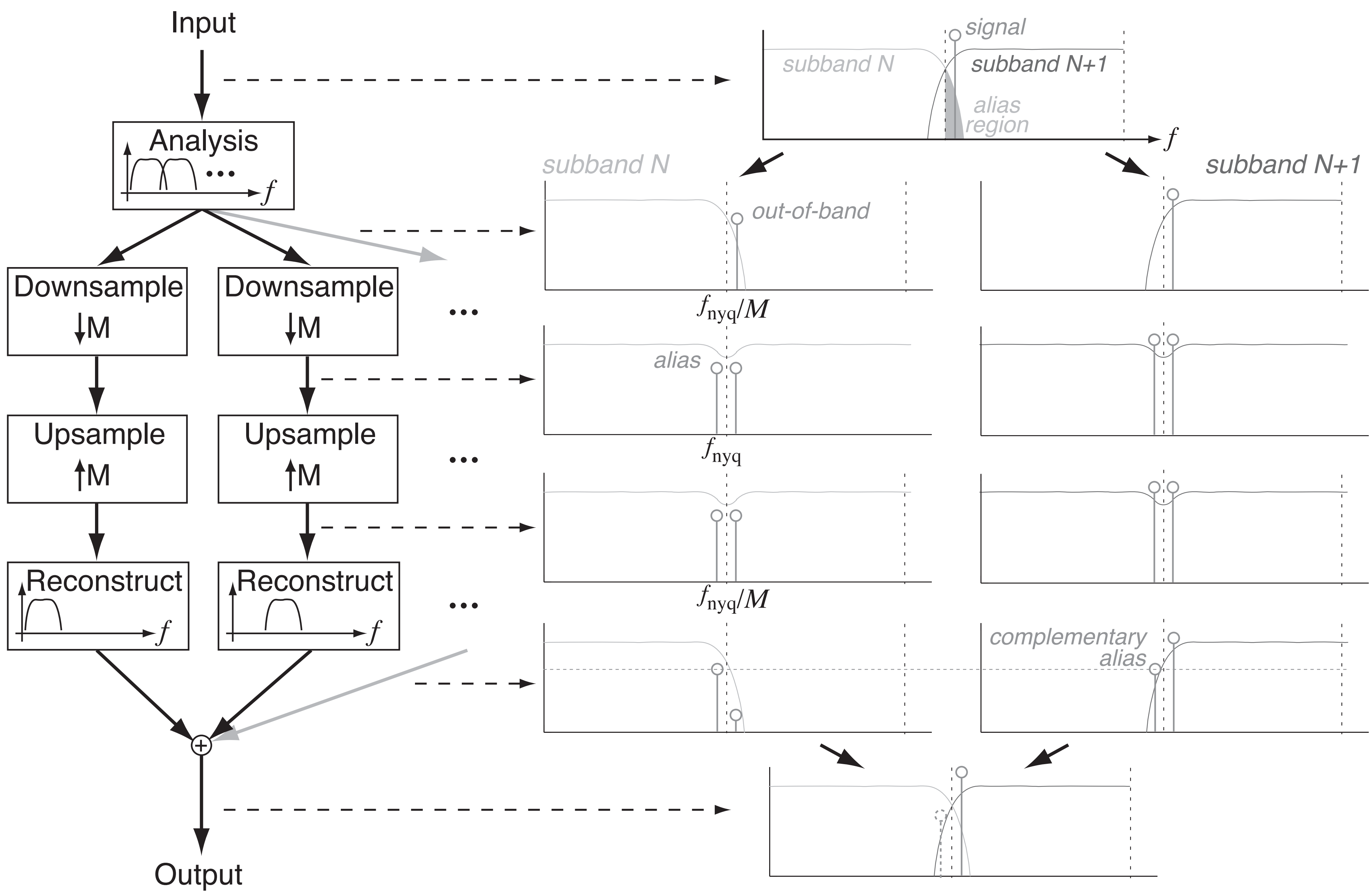
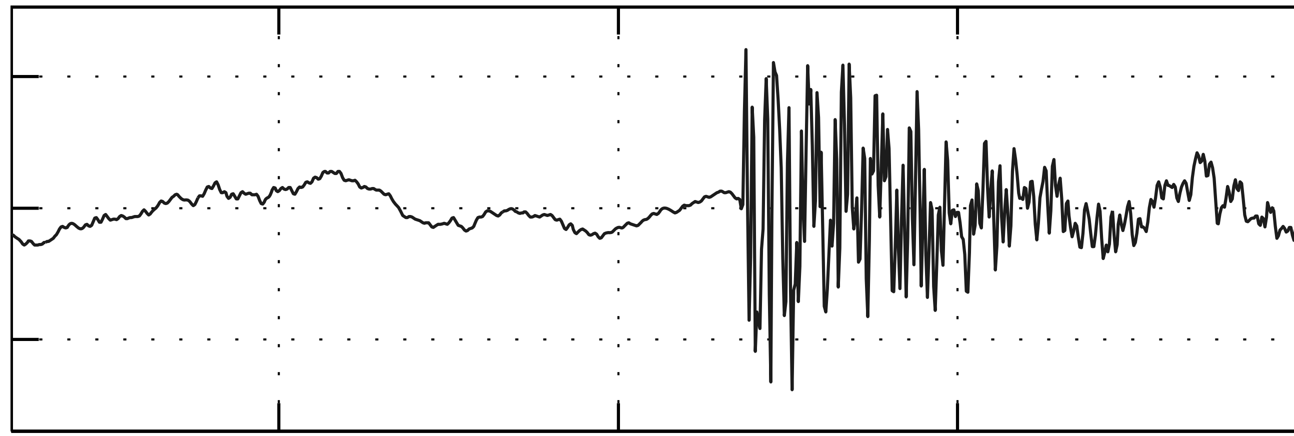


FIGURE 35.8 Alias cancellation in quadrature-mirror filterbanks.

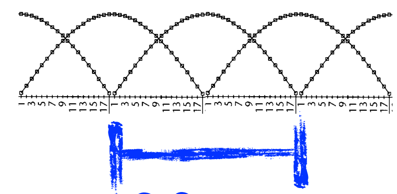
Frame Windows

Calculated mask yields imperceptible noise once the hit begins, but not during the silence before the click.

Input: castanet hit.

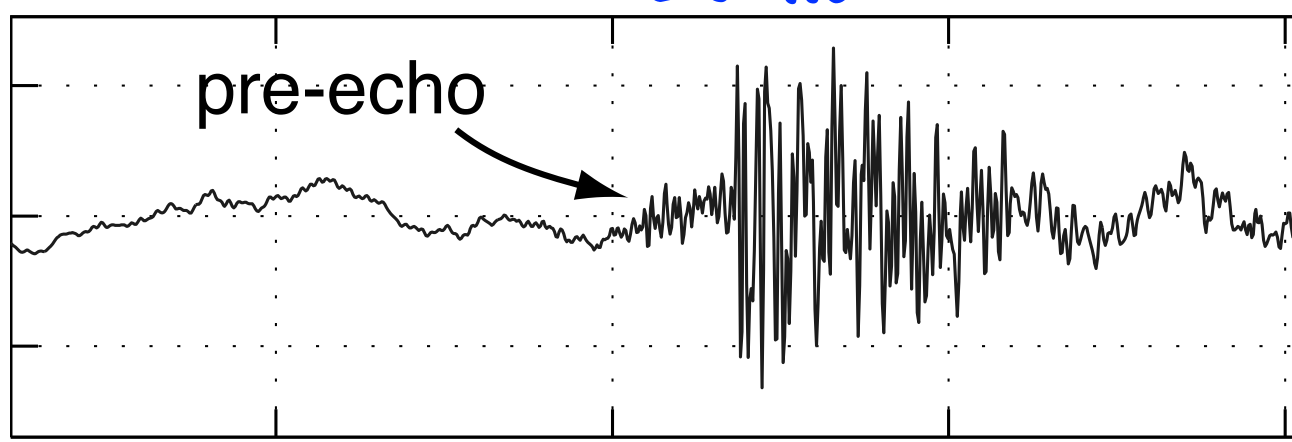


MPEG-2 (Level II)
Frame Window



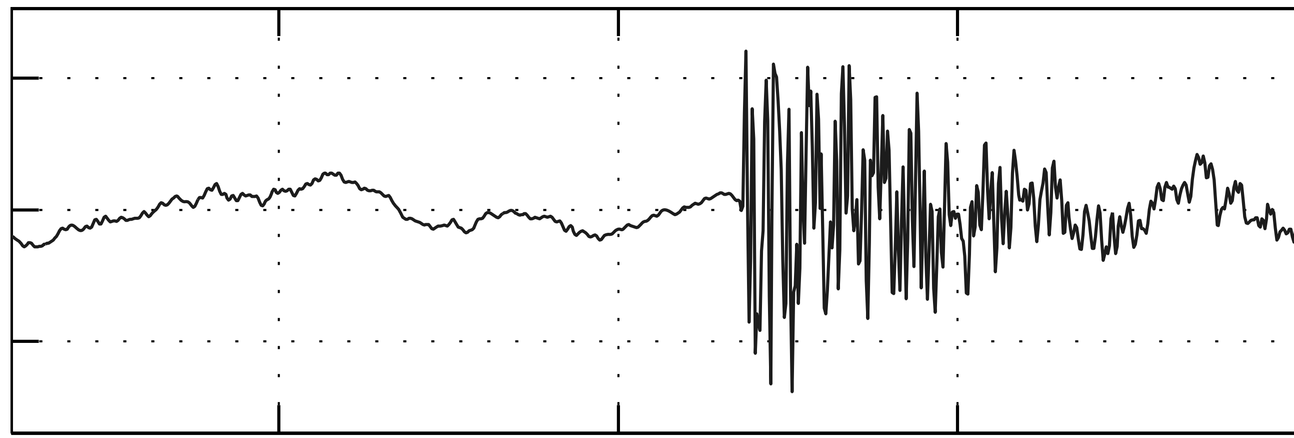
26 ms

Decoder output, with artifacts.

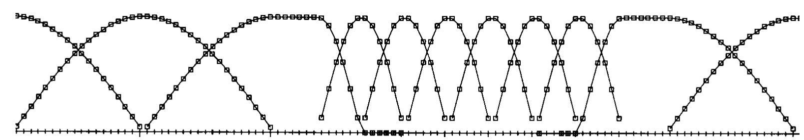


"MP3" solution: Variable-length frames ...

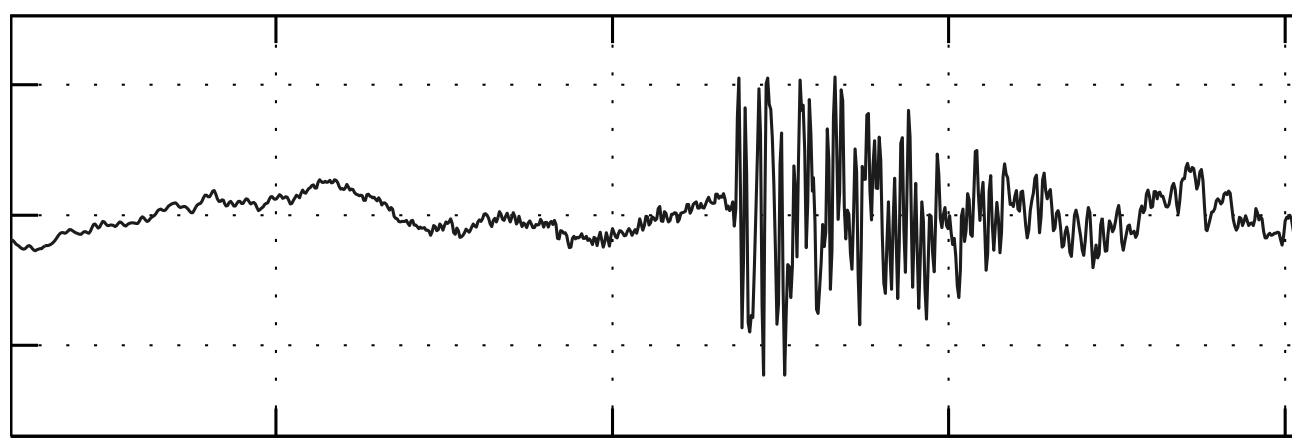
Input: castanet hit.



"MP3"
Frames

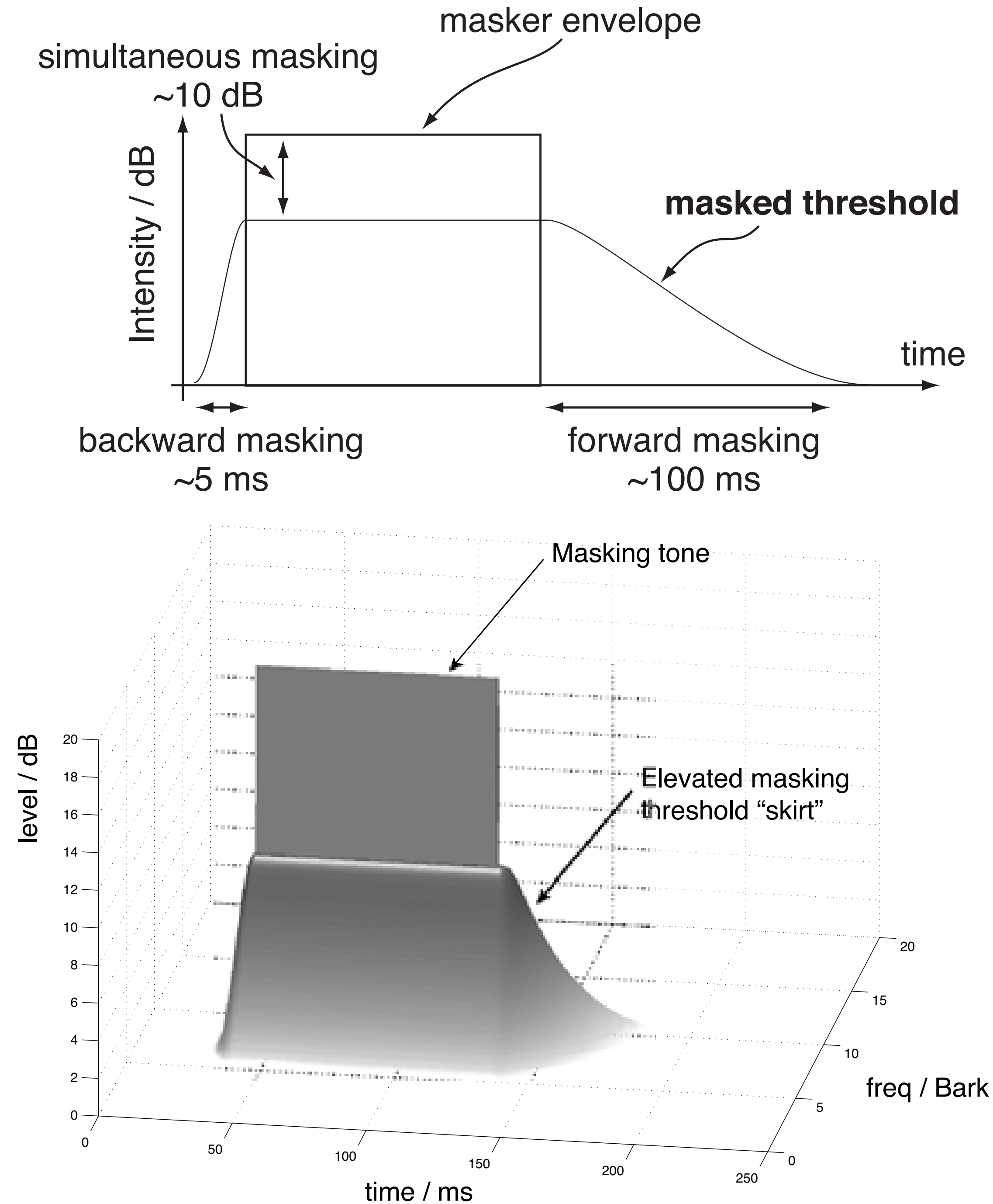


Decoder output, with reduced artifacts.



Temporal masking effects ...

Masking phenomena have temporal properties which must be considered when encoding examples like “castanets”



Today's lecture: Audio Coding

- Compression: Lossless and Lossy
- Quantization and Noise
- Psychoacoustic Masking
- Time-Frequency Tradeoffs
- **Research Topics**