Audio compression

Prof. Dr. Uluğ Bayazıt

Applications for audio coding

- High fidelity audio storage: CD, DAT, MD…
- Interpersonal communication: video conferencing, multimedia applications...
- HDTV standards (mainly MPEG).
- Digital sound broadcasting

Outline

- Sampling techniques (PCM)
  - Linear (Uniform quantization)
  - Non-linear (companding: **comp**ress-uniform quantize- exp**and**)
    - μ-Law
    - A-Law
- Generic Coding Techniques
  - DPCM
  - ADPCM
- Psychoacoustic Coding
  - MPEG

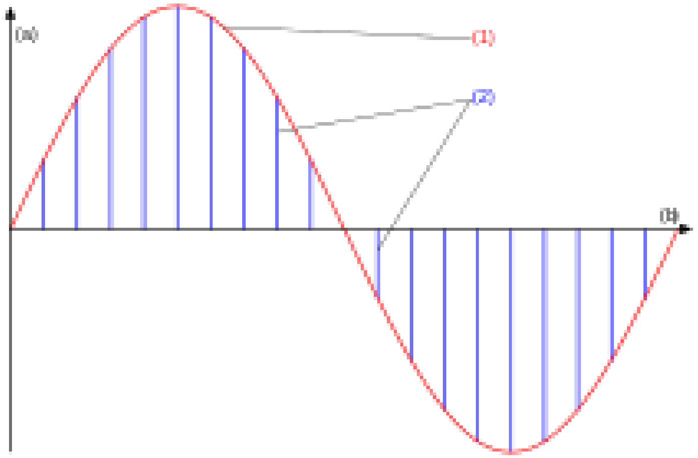
Existent standards

Standard	BW (Hz)	Comp-ression	Sample Rate (KHz)	Precision (bits)	Bit rate (kbps)	Quality
IMA-ADPCM	200-20000	ADPCM	8 - 44.1	4	32-350	Telephone, CD
G.711	200-3200	µ-law PCM	8	8	64	Telephone
G.722	50-7000	DPCM	16	4	64	AM Radio
G.728	200-3200	low-delay CELP	8	2	16	Telephone
G.723	200- 3200	ADPCM	8	8	5.3 or 6.3	H.323
Audio CD	20-20000	Linear PCM	44.1	16	1411.2 (stereo)	CD
MPEG-1	20-20000	sub-band coding	32-48	2-15	256-384	near-CD

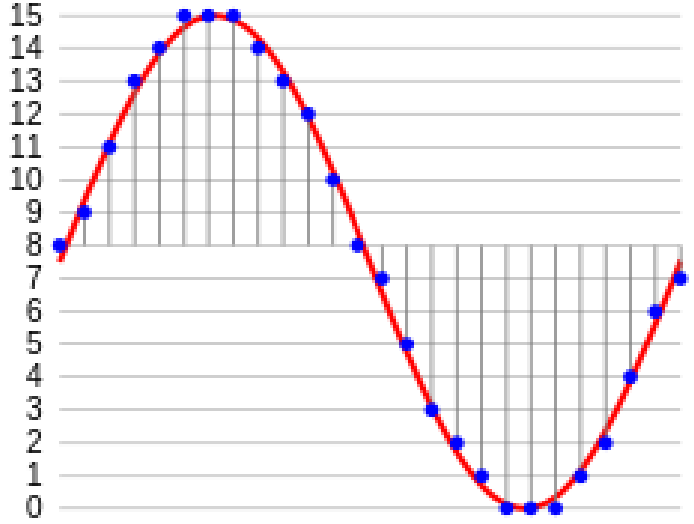
IMA: Interactive Multimedia Association

CELP: Code excited linear prediction

PAM (Pulse amplitude modulation) vs. PCM (Pulse code modulation)



Principle of PAM: (1) original signal,
(2) PAM signal, (a) amplitude of signal, (b) time



Principle of PCM: Red – Original signal,
Blue: Quantized samples for 4-bit PCM

Nyquist says: Sampling frequency must be at least twice the highest frequency in the signal!
Then can obtain original signal by interpolating the samples via low-pass filtering.

- Rect filtering ⇔sinc fitting

PCM approximation

- Greater the number of quantization levels, lower the quantization noise
  - Each bit of resolution adds 6 dB of dynamic range i.e. Number of bits required depends on the amount of noise that is tolerated
    - SNR ≈ $6.02N$

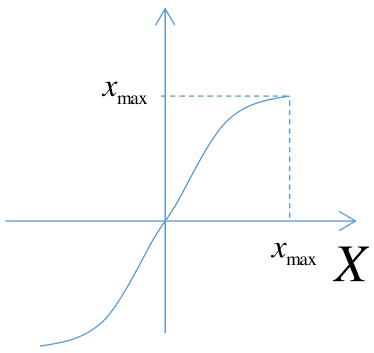
Uniform or Linear Quantization

- Quantization levels are evenly spaced.
- 16-bit samples provide plenty of dynamic range.
  - CDs do this.
- Compression factor 1:1.
  - i.e. does not count as compression, but counts as digital representation
- File formats
  - .WAV (MS)
  - .AIFF (Unix and Mac)
- …but SNR is worse for low-level signals than for high-level signals.

Nonuniform Quantization

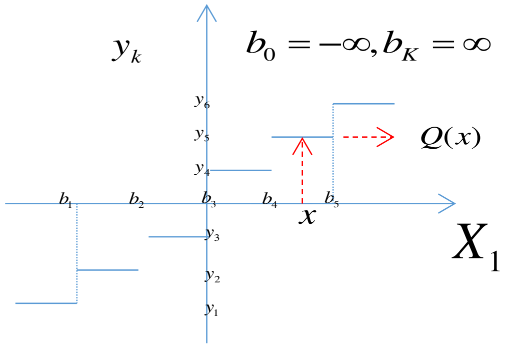
- Optimal nonuniform quantization
  - Decision boundaries optimized for quantization levels (nearest neighbor assignment of input points)
  - Quantization levels optimized for decision boundaries (centroid (bin conditional expected value) reconstruction
  - Not interested in this course!
- Companding (**comp**ress-uniform quantize- exp**and)**

$$X_1 = c(X)$$



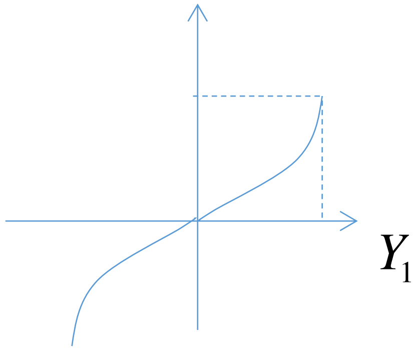
Compress (encoder)

$$Y_1 = Q(X_1)$$



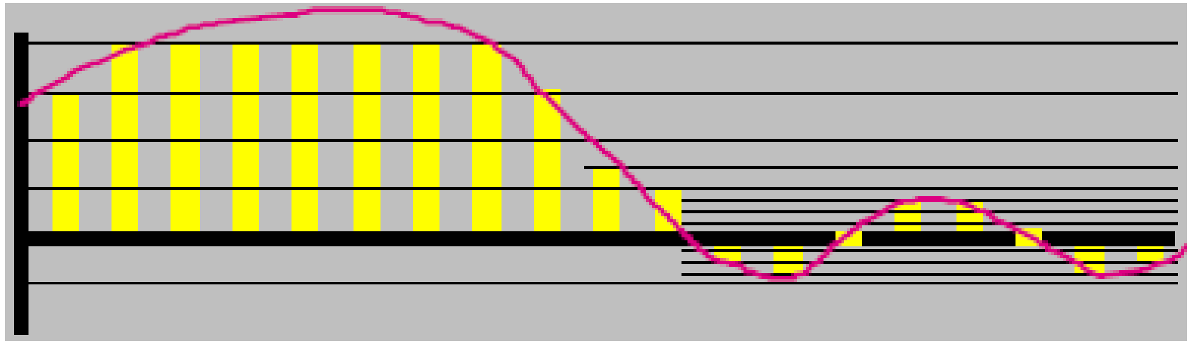
Uniform
Quantize

$$Y = c^{-1}(Y_1)$$



Expand (decoder)

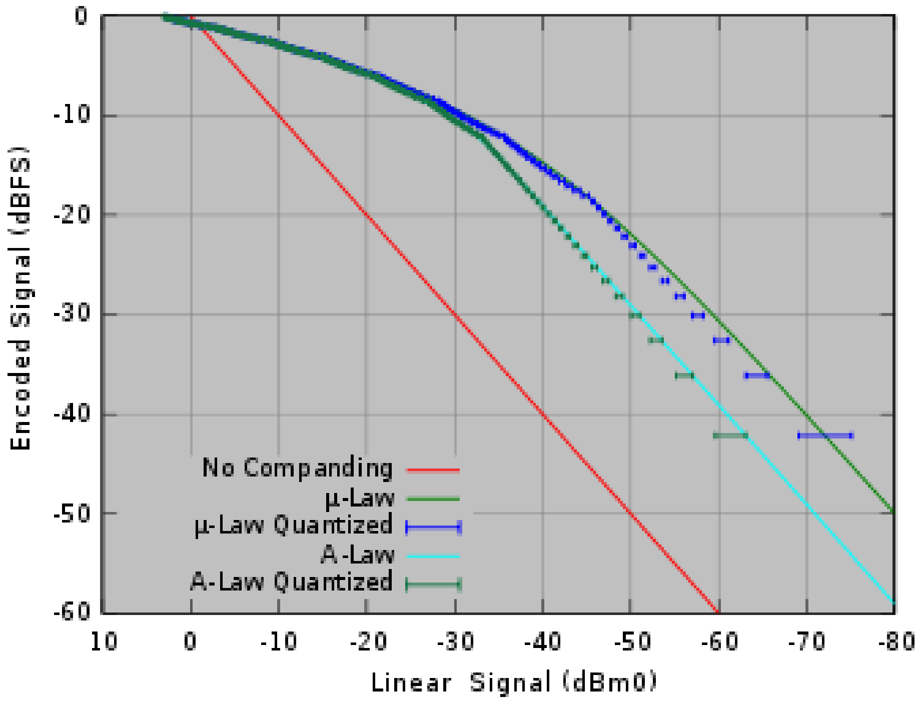
Companding



- Humans are less sensitive to changes in “loud” sounds than “quiet” sounds
  - Low-amplitude samples need to be represented with greater accuracy than high amplitude samples.
- A different kind of nonuniform quantization of the signal’s amplitude
  - Quantization step-size decreases logarithmically with signal level
  - Stretch the high probability regions around the origin
  - Compress the low probability regions away from the origin
  - As a result the input to the uniform quantizer is approximately uniformly distributed

Companding

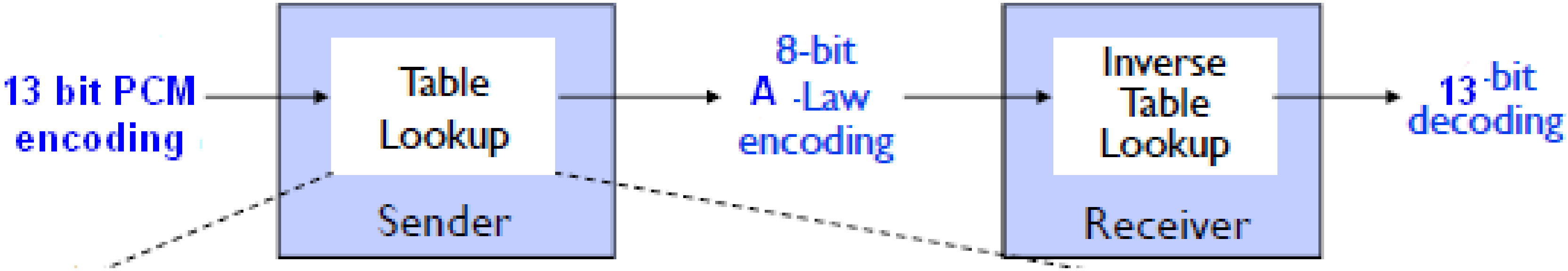
- Two commonly used characteristics in audio transmission:
  - µ-law: $$c(x) = x_{\max} \frac{\ln\left(1+\mu \frac{|x|}{x_{\max}}\right)}{\ln(1+\mu)} \operatorname{sgn}(x)$$
  - A-law: $$c(x) = \begin{cases} \frac{A|x|}{1+\ln A} \operatorname{sgn}(x), & 0 \le \frac{|x|}{x_{\max}} \le \frac{1}{A} \\ x_{\max} \frac{1+\ln A \frac{|x|}{x_{\max}}}{1+\ln A} \operatorname{sgn}(x) & \text{else} \end{cases}$$



A-law

- Provides 13-bit quality (dynamic range) with an 8-bit encoding
- Used in European PCM digital communication systems
- Simple to compute encoding.
- Used in G.711 waveform voice coding standard
  - G.711 is a standard in
    - H.320 (Multimedia Audio/Video/Data over ISDN)
    - H.323 (Audio-Visual communication sessions on packet networks such as IP/X.25) specifications

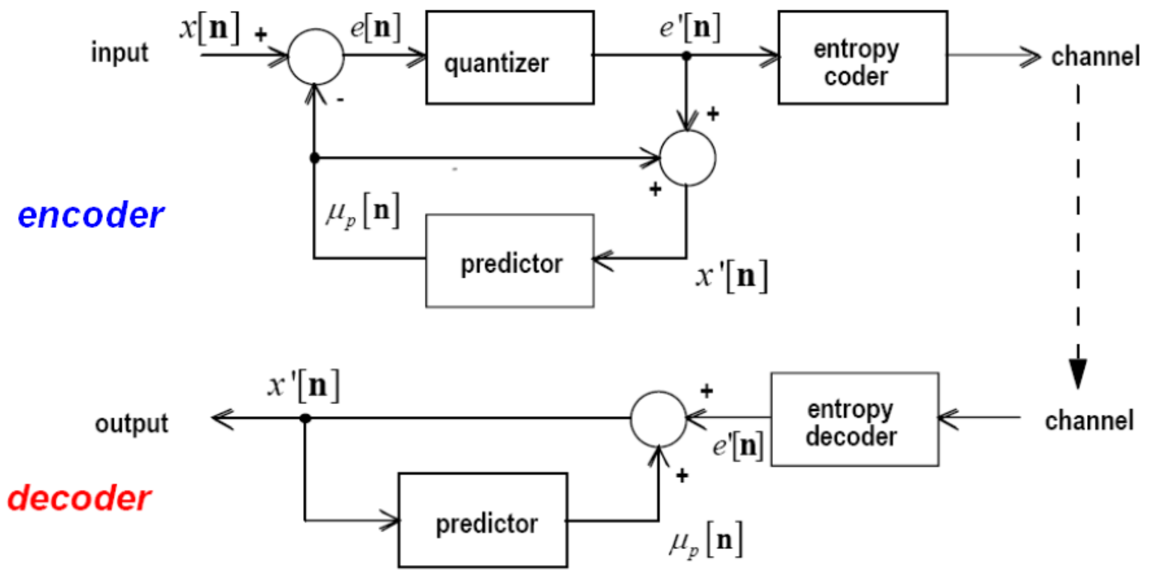
A-law encoding



Segment Number	Input Amplitude Range	Normalized Amplitude Range	Normalized Step Size	A-Law Code
1	0..31	0..1/64	1/2048	0..31
2	32..63	1/64..1/32	1/1024	32..47
3	64..127	1/32..1/16	1/512	48..63
4	128..255	1/16..1/8	1/256	64..79
5	256..511	1/8..1/4	1/128	80..95
6	512..1023	1/4..1/2	1/64	96..111
7	1023..2047	1/2..1	1/32	112..127

Difference Encoding

- Exploit temporal redundancy in samples.
- Difference between 2 *x-bit* samples can be represented with significantly fewer than *x-bits* **if they are fairly correlated**
- Transmit the difference (rather than the sample).
- Differential-PCM (DPCM)



Reconstruction error = quantization error

$$\boxed{x' - x = e' - e = q}$$

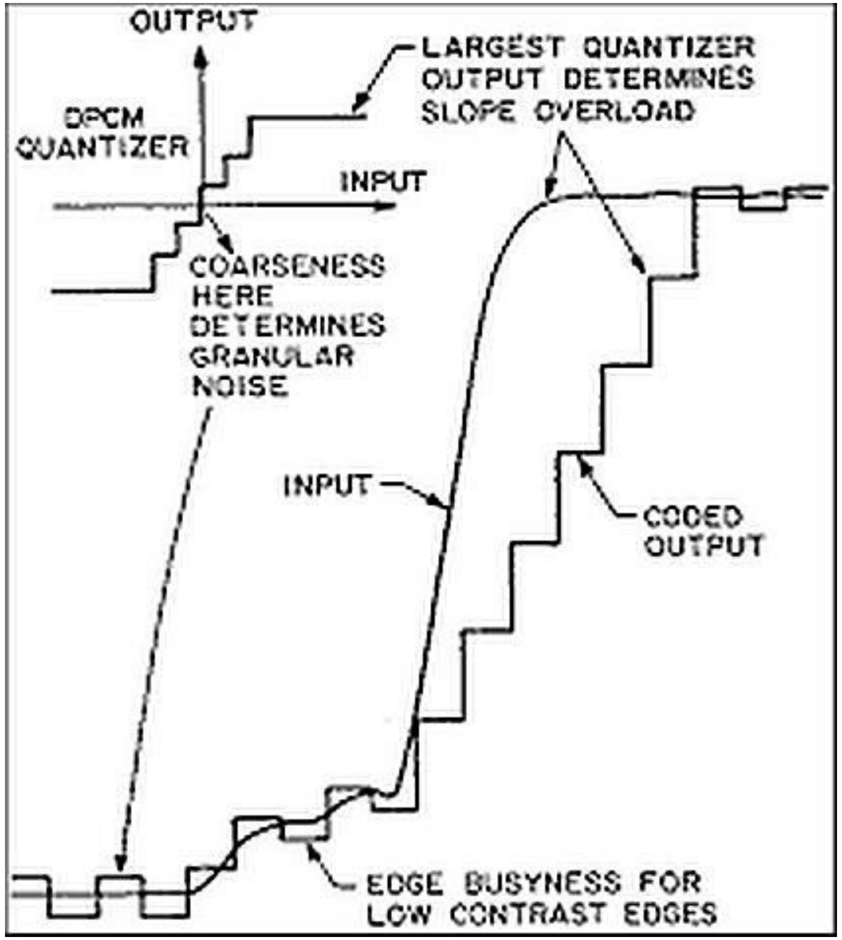
$\mu_p[n] = x'[n-1]$: Simple difference

$\mu_p[n] = \sum_{k=1}^{N} h[k]x'[n-k]$: Linear predictor

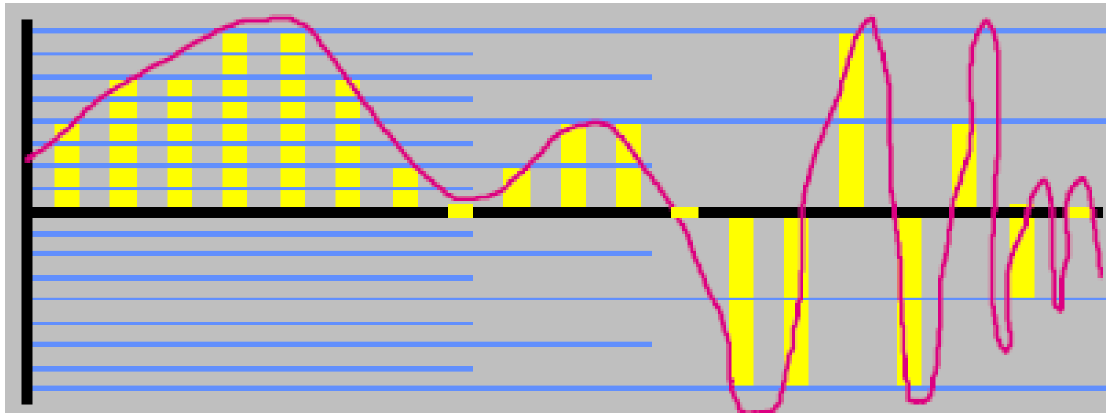
$\mu_p[n] = 0$: PCM

Two types of distortion in DPCM

- Slope overload for high dynamic range input
- Granular noise for low dynamic range input



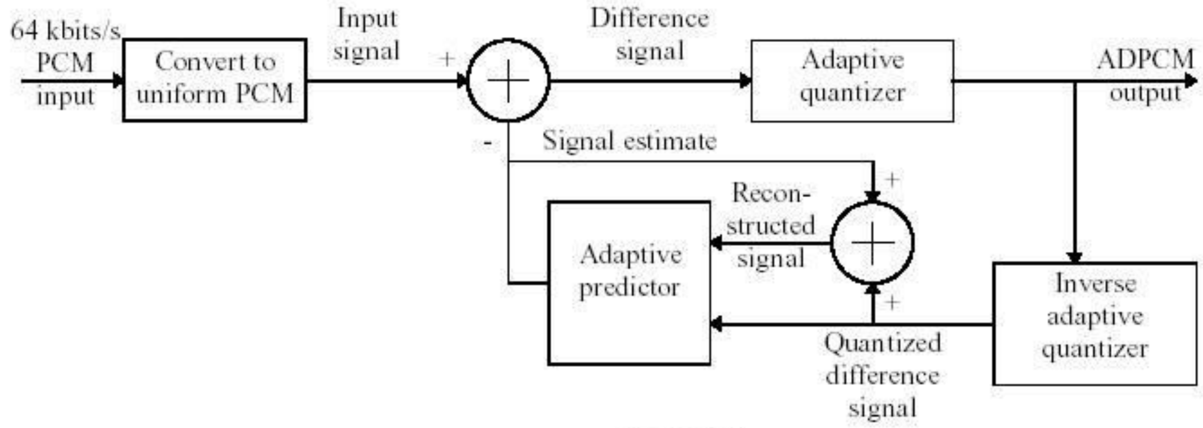
Adaptive DPCM (ADPCM)



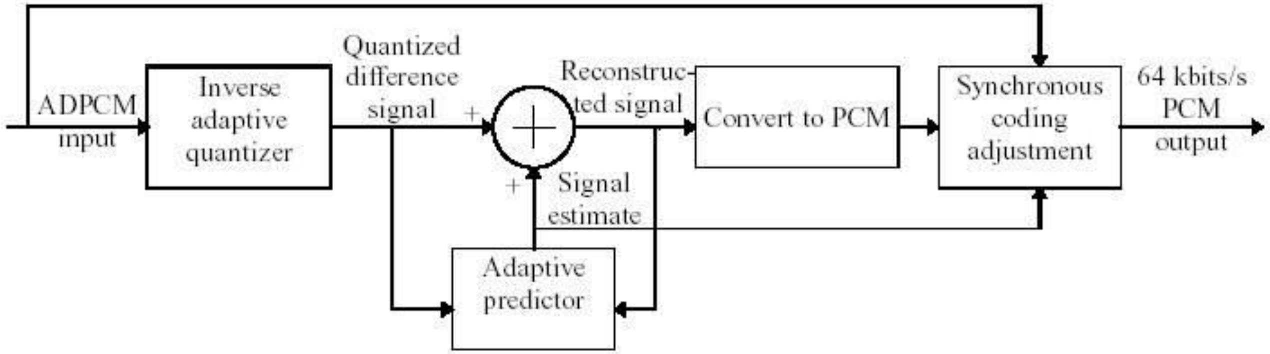
- Use a larger step-size to encode differences between high-frequency samples & a smaller step-size for differences between low-frequency samples.
- Use previous sample values to estimate changes in the signal in the near future.

ADPCM

- To ensure differences are always small:
  - Adaptively change the step-size (quantization bin size).
  - Adaptively attempt to predict next sample value
    - Need to update filter coefficients



a) encoder

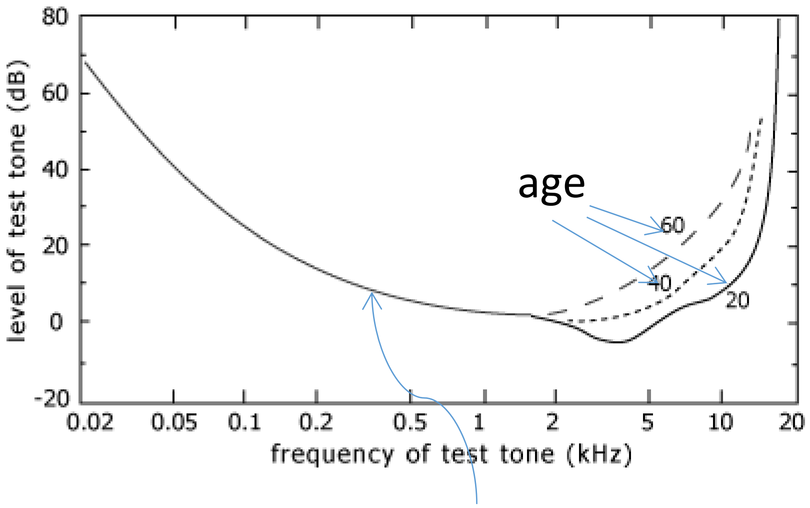


b) decoder

MPEG Psychoacustic Coding

- **Psychoacoustic Principles**
  - Based on extensive studies of human perception.
  - Average human does not hear all frequencies the same way.
  - Limitations of the human sensory system leads to facts that can be used to discard unnecessary data in an audio signal.
  - Two main properties of the human auditory system:
    - *Absolute threshold of hearing*
    - *Auditory masking*

Absolute threshold of hearing

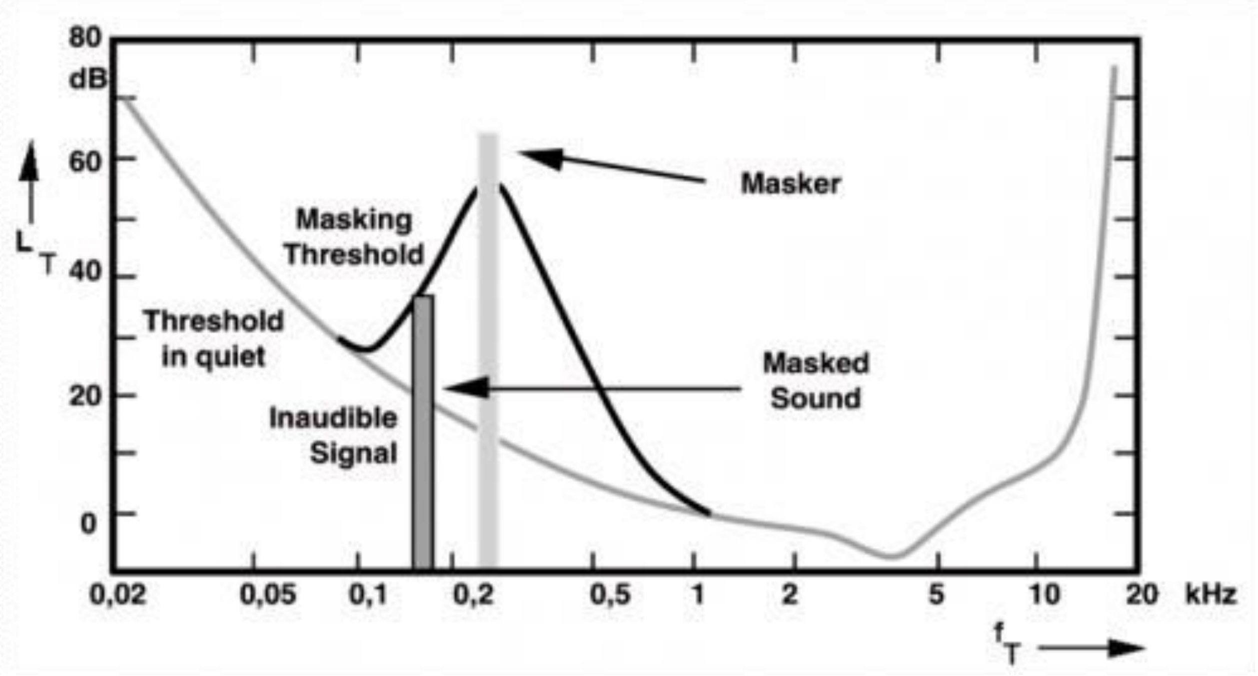


$$T_q(f) = 3.64(f/1000)^{-0.8} - 6.5e^{-0.6(f/1000-3.3)^2} + 10^{-3}(f/100)^4$$

- Minimum sound level of a pure tone that an average human ear with normal hearing can hear with no other sound present
- Frequency, age and gender dependent
- Range is about 20 Hz - 20 kHz, most sensitive at 2 kHz to 4 kHz.
- Dynamic range (quietest to loudest is about 96 dB)
- Normal voice range is about 500 Hz - 2 kHz

Simultaneous Masking

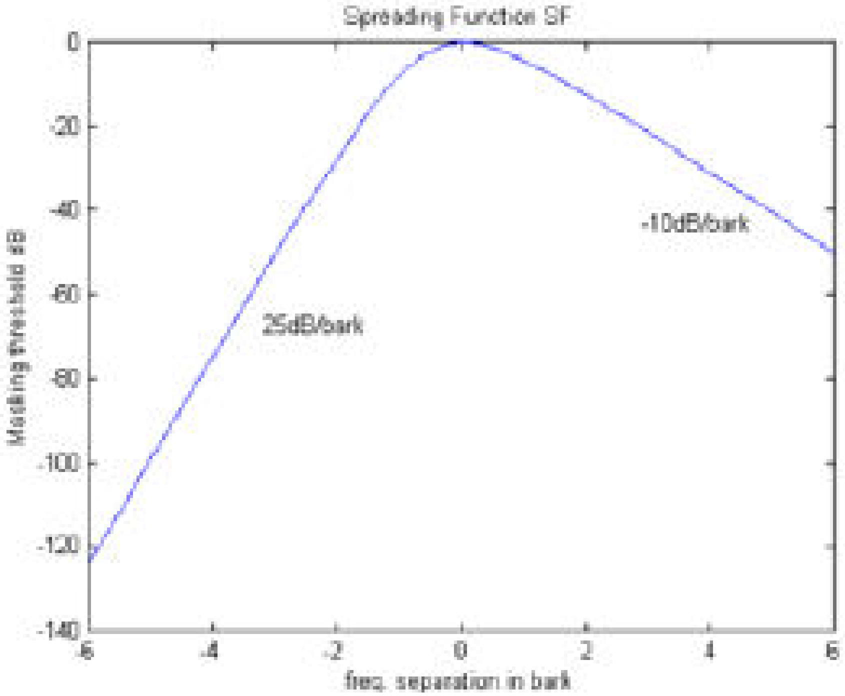
- The presence of tones at certain frequencies makes us unable to perceive tones at other "nearby" frequencies
- Absolute threshold modified by masker signal
  - modified range called masking threshold
- Humans cannot distinguish between tones within
  - 100 Hz at low frequencies
  - 4 kHz at high frequencies



Masking Threshold

- Approximates a triangular function modeled by spreading function, SF(x) (db), where x has units of bark

$$SF(x) = 15.81 + 7.5(x + 0.474) - 17.5\sqrt{1 + (x + 0.474)^2}$$



Example:https://www.youtube.com/watch?v=2HDka1hYiCk

Critical bands

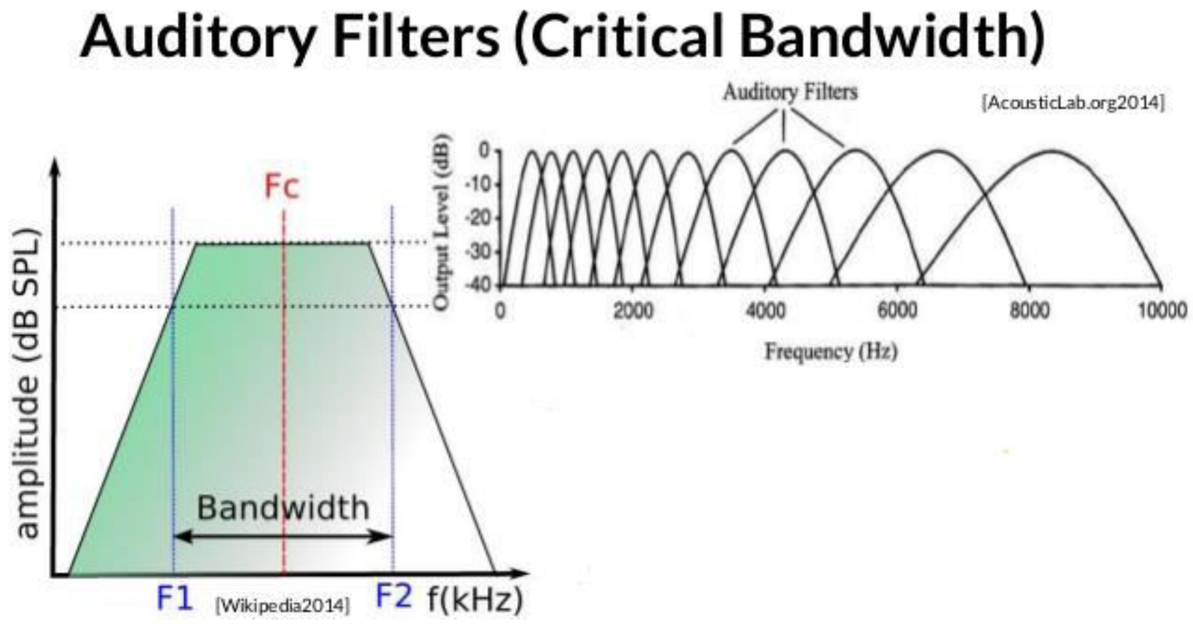
- the frequency bandwidth of the "auditory filter" created by the cochlea, the sense organ of hearing within the inner ear
- the band of audio frequencies within which a second tone will interfere with the perception of the first tone by auditory masking
- Human auditory system has limited and frequency dependent resolution
  - Critical band number (Bark) for a given frequency, *z(f):*
    - $f < 500Hz \Rightarrow z(f) \approx f/100$
    - $f > 500Hz \Rightarrow z(f) \approx 9 + 4 \log_2(f/1000)$

    $z(f) = 13.0 \arctan(0.00076 f) + 3.5 \arctan((f/7500)^2)$
  - One bark = width of one critical band (variable!)

Bark scale critical bands

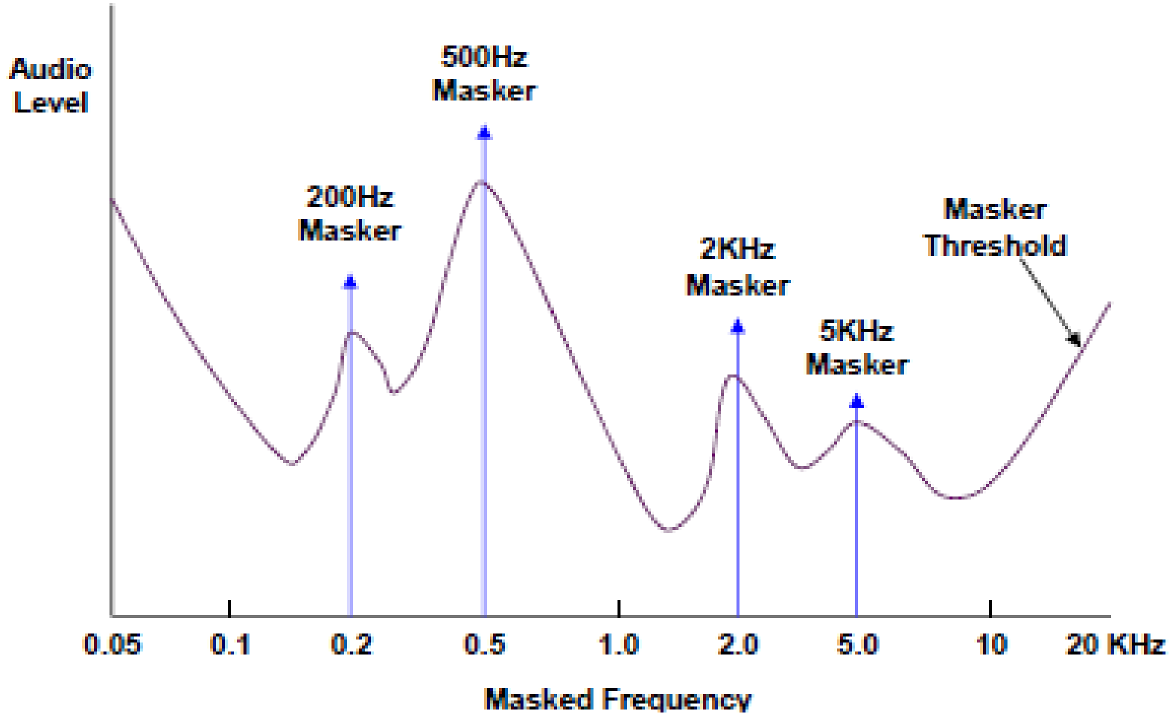
TABLE I.

Number	Center frequencies Hz	Cut-off frequencies Hz	Bandwidth Hz
		20	
1	50	100	80
2	150	200	100
3	250	300	100
4	350	400	100
5	450	510	110
6	570	630	120
7	700	770	140
8	840	920	150
9	1000	1080	160
10	1170	1270	190
11	1370	1480	210
12	1600	1720	240
13	1850	2000	280
14	2150	2320	320
15	2500	2700	380
16	2900	3150	450
17	3400	3700	550
18	4000	4400	700
19	4800	5300	900
20	5800	6400	1100
21	7000	7700	1300
22	8500	9500	1800
23	10 500	12 000	2500
24	13 500	15 500	3500



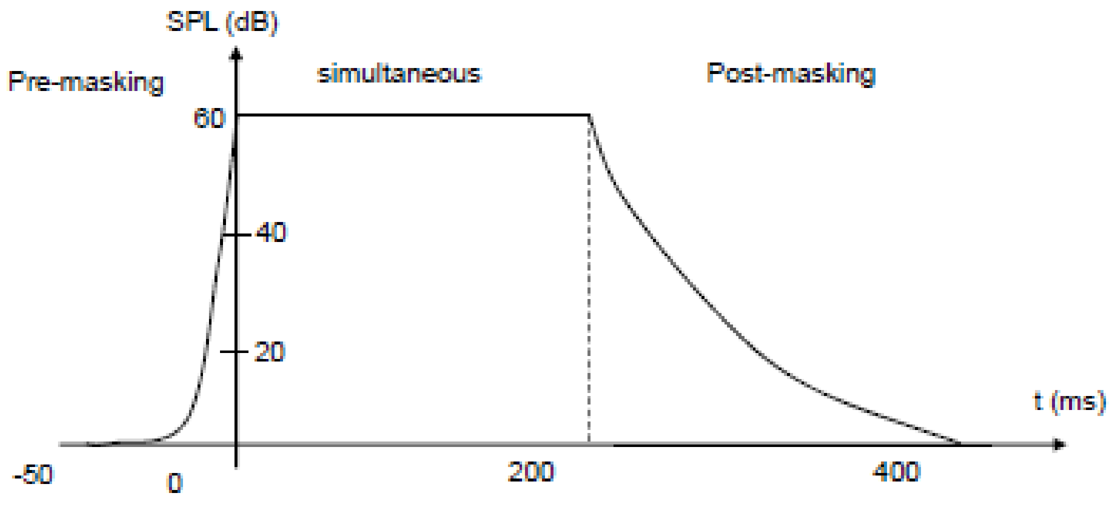
Global Masking Threshold

- Additive process
  - The minimum level of audibility in the presence of masking noise
  - Time varying function of frequency at each frequency at a given time



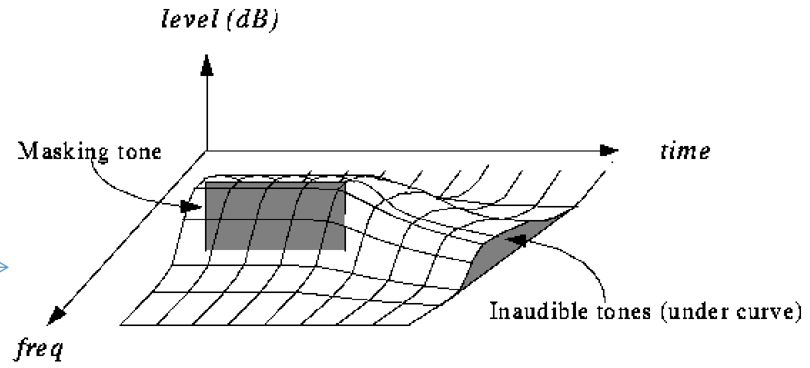
Temporal Masking

- After we hear a loud sound, it takes a little while until we can hear a soft tone nearby.
  - As much as 50 ms before and 200 ms after.
- Example;
  - Play 1 kHz masking tone at 60 dB and 1.1 kHz test tone at 40dB.



Recap

- Auditory masking
- Range of masking is 1 Bark (or critical band)
- Two factors for masking
  - frequency masking
  - temporal masking.



- How can we take advantage of masking for compression?

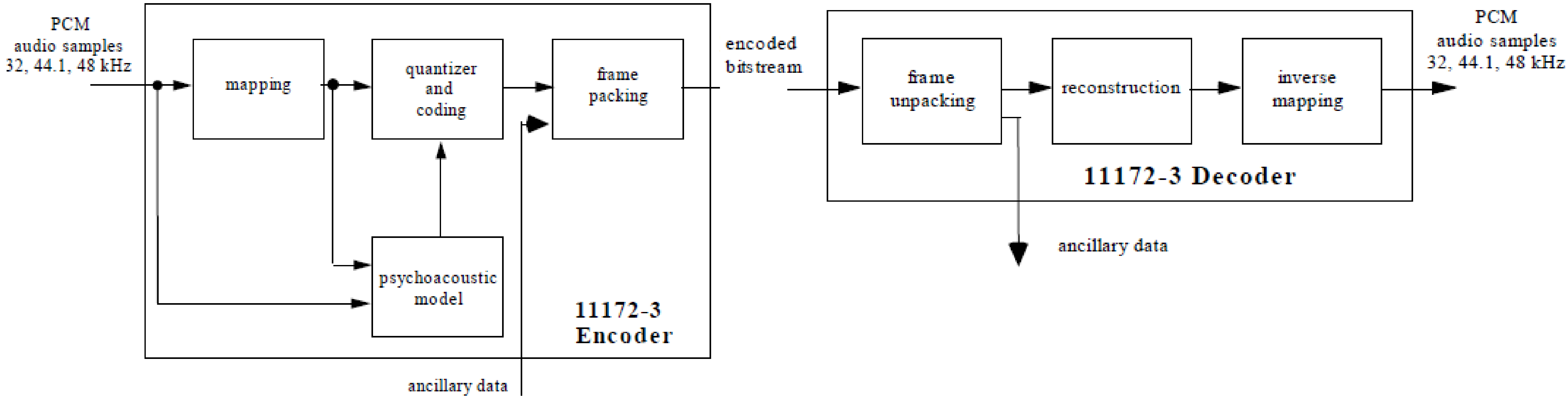
MPEG-1 Audio

- Founded in 1988 (ISO/IEC/ JTC 1/SC29/WG11)
  - Audio and video standard for encoding and decoding.
  - Specify general functionality of application.
- MPEG-1: 1.5 Mbps for audio and video
  - 1.2 Mbps for video
  - 0.3 Mbps for audio
    - Uncompressed CD audio - 44,100 samples/sec * 16 bits/sample * 2 ch > 1.4 Mbps
- Compression ratio from 2.7:1 to 42:1
  - 16 bit stereo sampled at 48 kHz is reduced to 256 kbps
- Sampling frequencies
  - 32, 44.1 and 48 kHz
- One or two audio channels
  - Monophonic, Dual-monophonic, Stereo, Joint Stereo

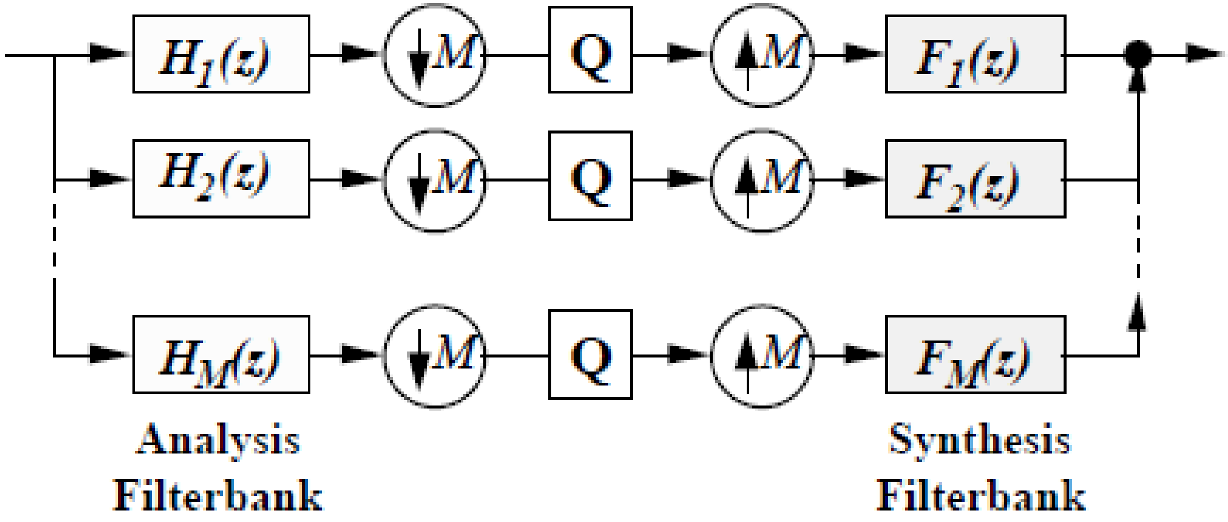
MPEG-1 Audio Layers

- Layer 1
  - DCT type filter with one frame and equal frequency spread per band.
  - Psychoacoustic model only uses frequency masking.
  - ~384 kbits/s for perceptually lossless quality (4:1)
- Layer 2
  - Use three frames in filter (before, current, next, a total of 1152 samples).
  - This models a little bit of the temporal masking.
  - ~192 kbits/s for perceptually lossless quality (8:1)
- Layer 3 (*MP3)*
  - Better critical band filter is used (non-equal frequencies)
  - Psychoacoustic model includes temporal masking effects, and takes into account stereo redundancy.
  - Huffman coder.
  - ~128 kbits/s for perceptually lossless quality (12:1)

MPEG 1 Generic Encoder/Decoder

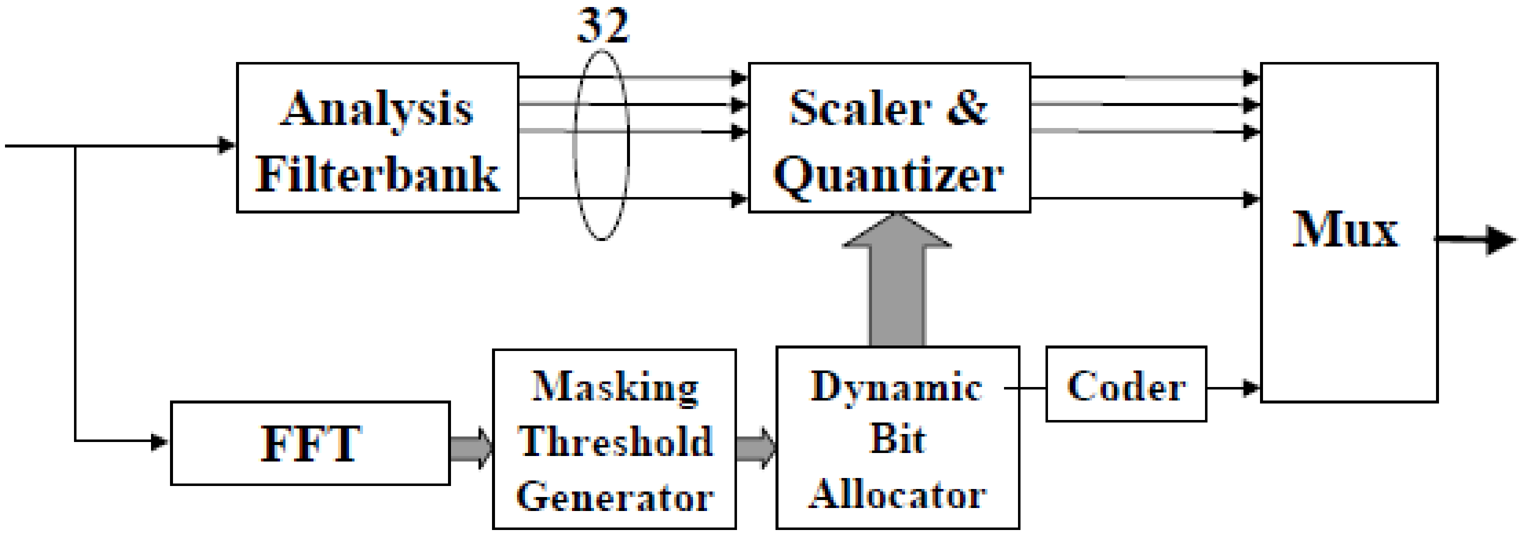


Subband Filterbanks (Mapping/Inverse Mapping)



- Critical downsampling
- Quantizer Q resolution (bit allocation) should be based on signal-to-masking ratio (SMR)
- Ear’s critical bands are not uniform, but logarithmic

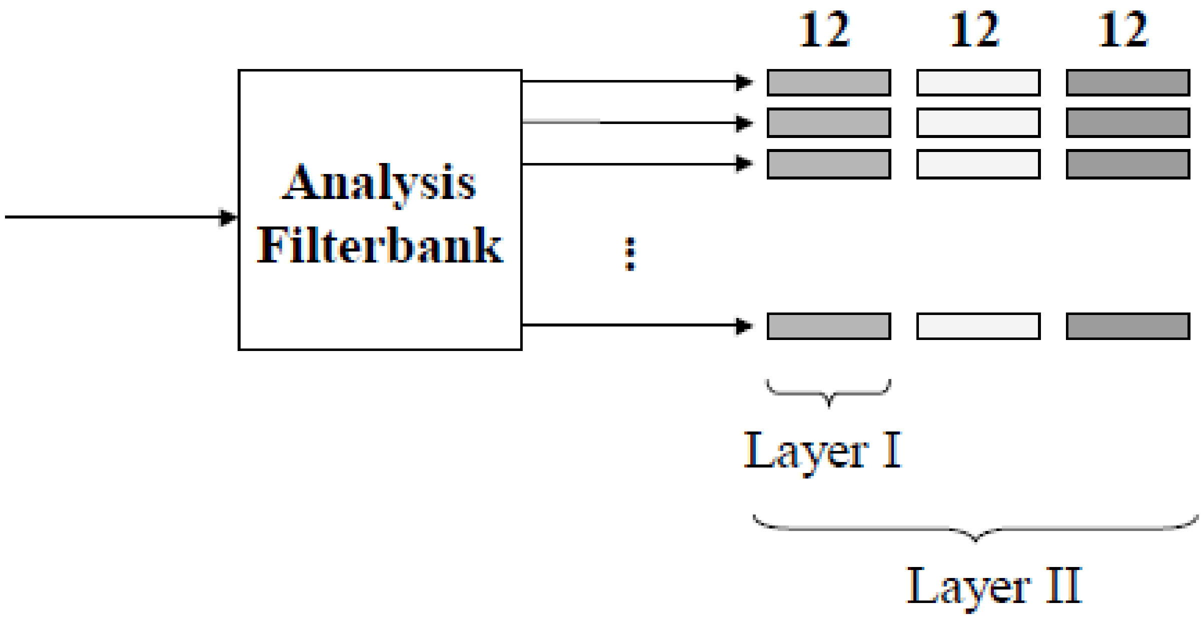
Layer I and II Encoder



- 32 PCM samples yields 32 subband samples.
  - Each sub-band corresponds to a frequency band evenly spaced from 0 to Nyquist frequency.
  - For example, @48 kHz sampling rate, each sub-band is 750 Hz wide.

Framing

- Samples out of each filter are grouped into blocks, called *frames.*
  - Blocks of 12 for Layer 1 (384 samples).
  - Blocks of 36 for Layers 2 and 3 (1152 samples)
  - @48 kHz, 32x12 represents 8 ms of audio.
- Block companding: Each block normalized by *scalefactor*
  - For Layer II, up to 3 scalefactors, with 2-bit scalefactor select
  - Each block is allocated bits



Psychoacoustic Model

- *Apply FFT to get detailed spectral information about the signal.*
  - 512-point FFT for Layer 1
  - 1024-point for Layer 2 and 3
- From each subband's samples, separate *tonal (sinusoidal)* and *nontonal (noise)* maskers in the signal.
- Determine *masking threshold* for each sub-band.
- Determine the *global masking threshold.*
- Determine the *minimum masking threshold* in each subband.
- Calculate the *signal-to-mask ratio (SMR)* in each sub-band, which is the ratio of *signal energy* to *minimum masking threshold.*

Spectral Analysis: DFT/FFT

- Transforms PCM samples from time to frequency domain.
- Needs finer frequency resolution for an accurate calculation of the masking thresholds.
- Incoming audio sample *s(n) normalized* according to FFT length *N.*

$$x(n) = \frac{s(n)}{N 2^{b-1}}$$

  - *N - FFT length*
  - b - number of bits per sample

- Segment the signal into 512 samples => 12 ms frames @44.1kHz

Discrete Fourier Transform (DFT)

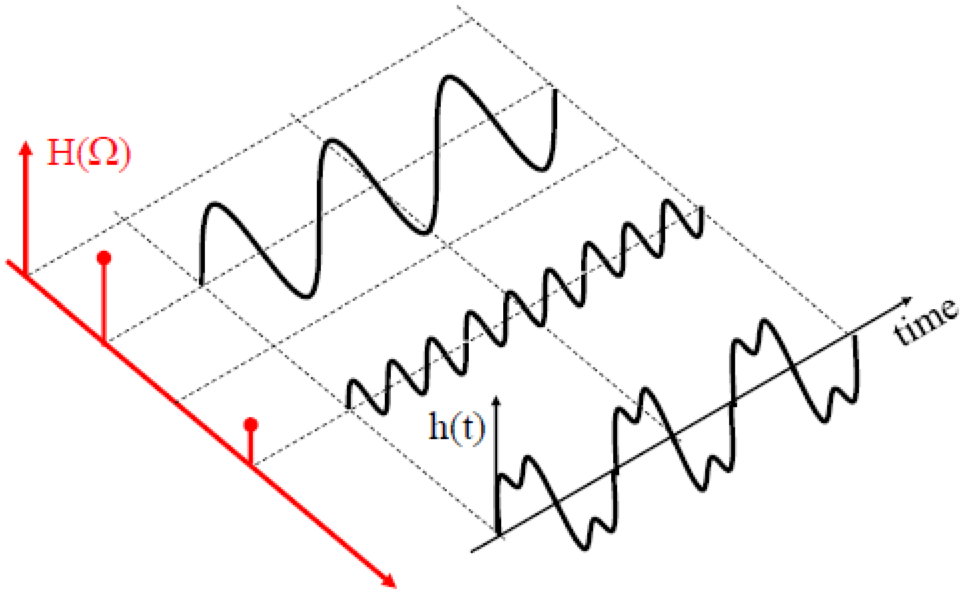
- Spectral Analysis requires Time ⇔ Frequency transform.
- Discrete Fourier Transform (DFT) provides uniformly spaced samples of the Discrete-Time Fourier Transform (DTFT).
  - Need to have a time limited signal
- DFT and Inverse-DFT

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j\frac{2\pi nk}{N}} \quad \mathrm{k} = 0,\ldots,\mathrm{N}-1$$

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] e^{j\frac{2\pi nk}{N}} \quad \mathrm{n} = 0,\ldots,\mathrm{N}-1$$

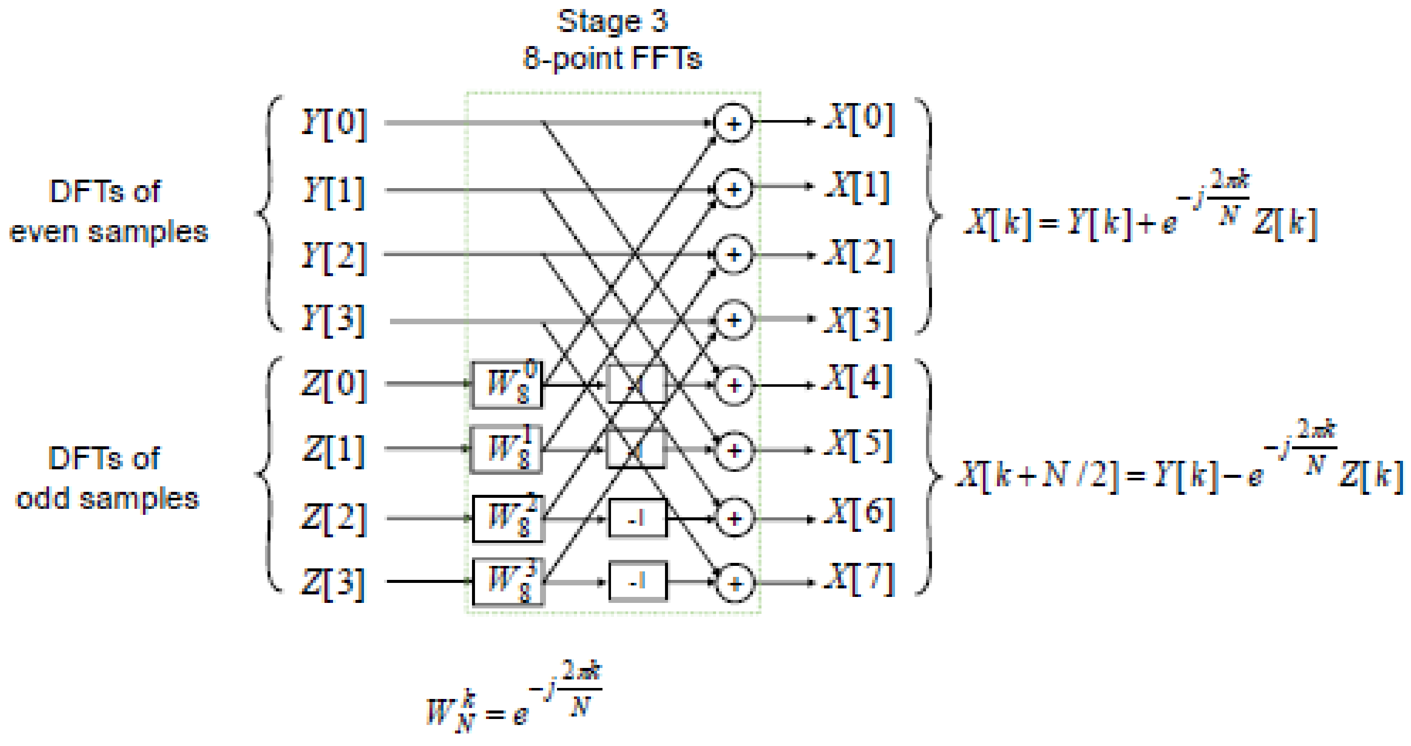
Time⇔Frequency transformation

- Analyze different tones in the signal

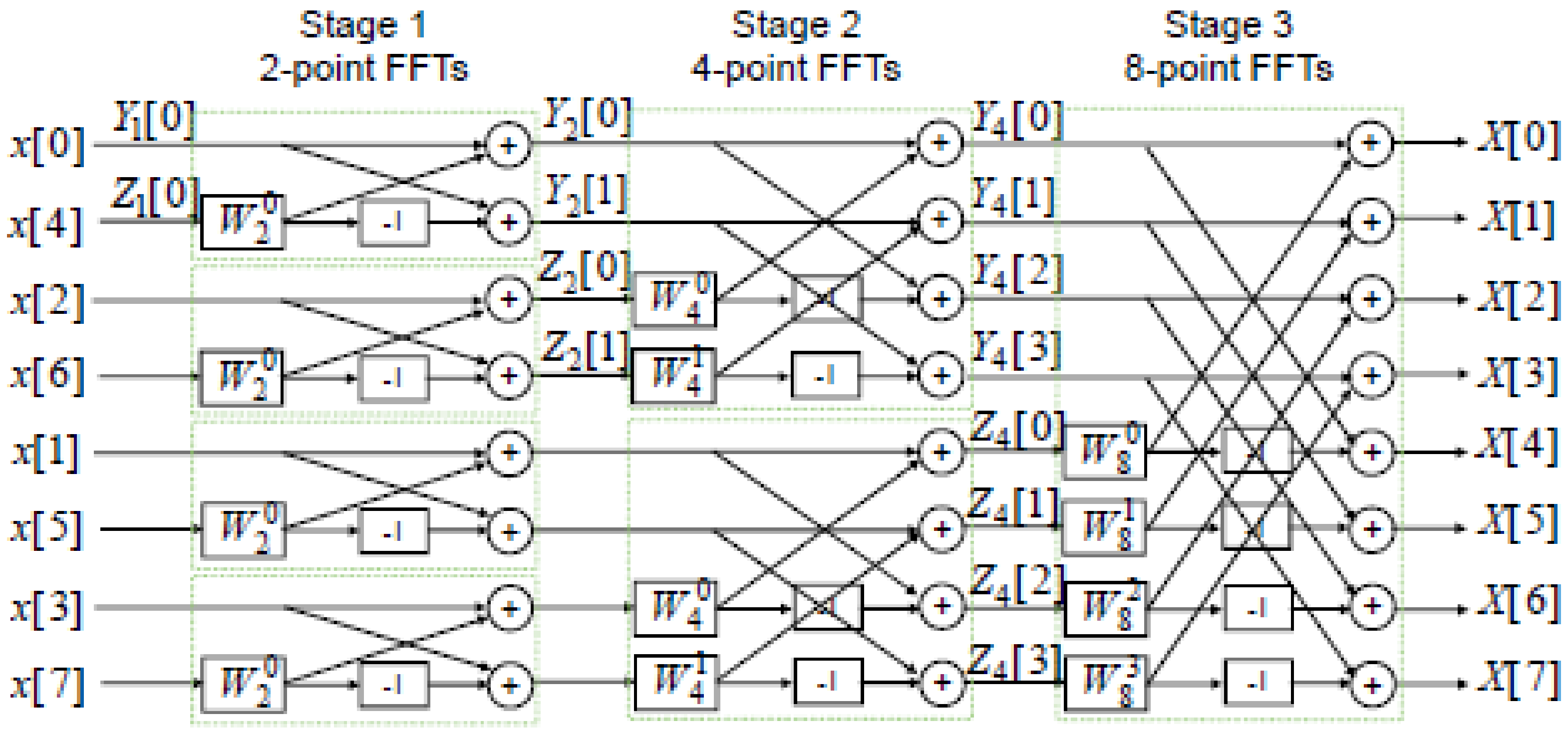


Fast Fourier Transform

- Fast implementation of DFT
- Based on splitting the signal x[n] into two parts: even samples Y[n], odd samples Z[n]



Fast Fourier Transform



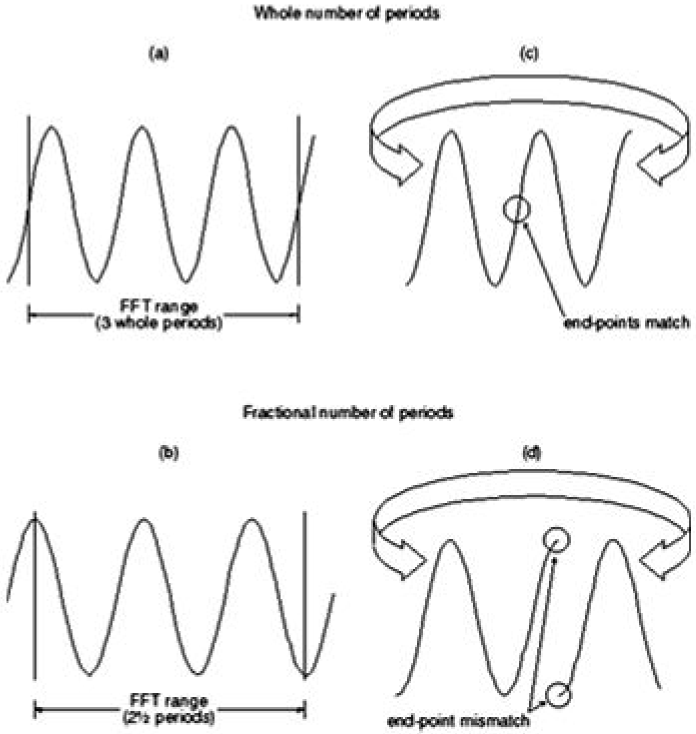
$$W_N^k = e^{-j\frac{2\pi k}{N}}$$

Spectral Analysis:PSD

- Power Spectral Density, P(k), using 512-point FFT (N=512)

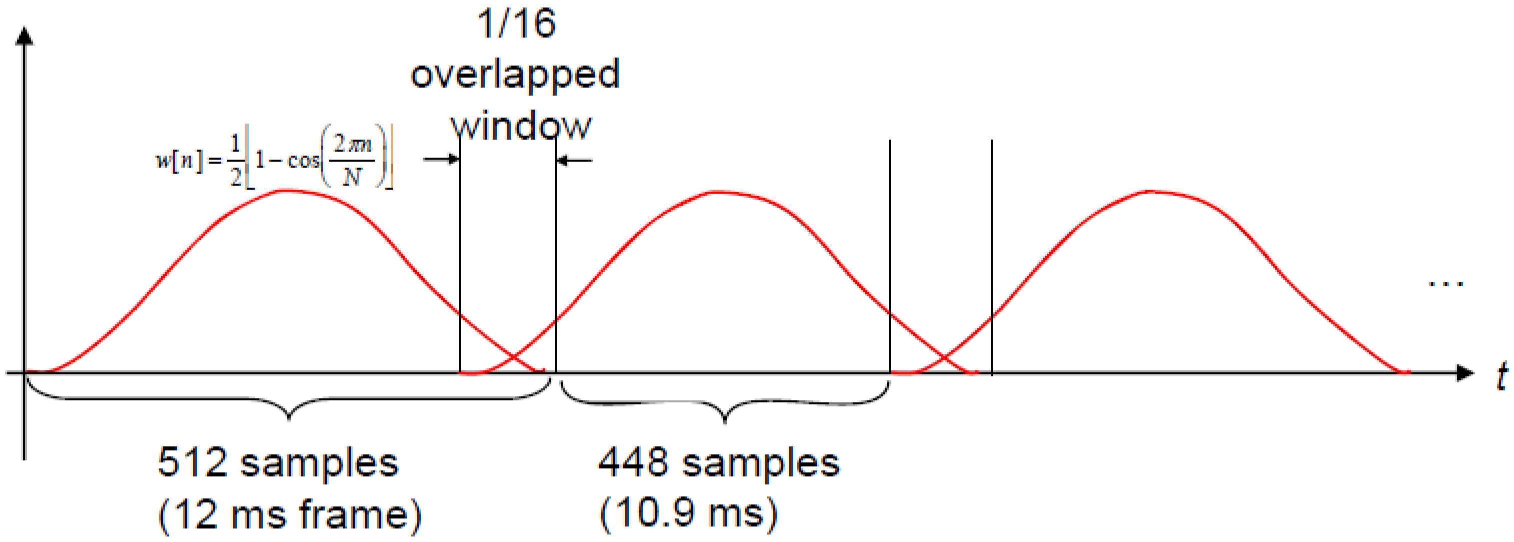
$$\mathrm{P(k)} = 90.302\,\mathrm{dB} + 10\log\left|\sum_{\mathrm{n}=0}^{N-1} w[n]x[n]e^{-j\frac{2\pi kn}{N}}\right| dB \quad 0 \le k \le \frac{N}{2}$$

- Generates N/2 frequency components or bins.
- w[n]: Hanning window $w[n] = \frac{1}{2}\left[1 - \cos\left(\frac{2\pi n}{N}\right)\right]$
  - DFT inherently assumes that data is a single period of a periodically repeating waveform!
  - window with ends dying off at zero ensures continuity of the periodic extension



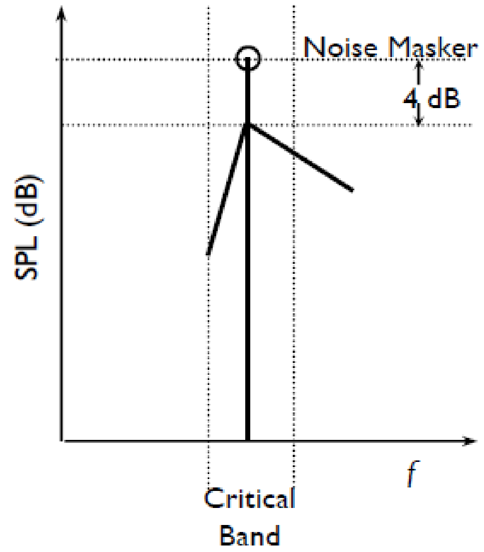
Implementation of Hanning window

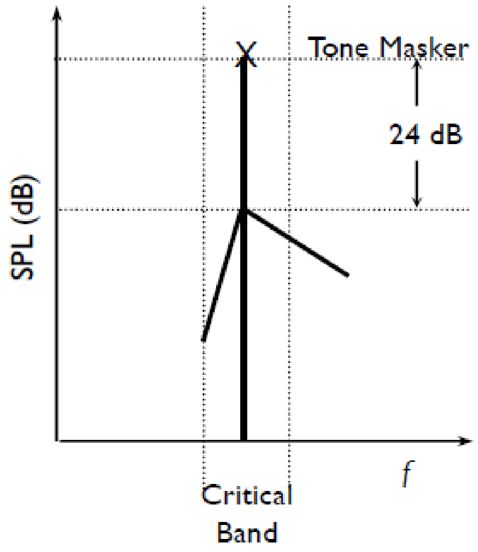
- Overlap and add



Identification of Tonal and Noise Maskers

- Tonal maskers are signals that generate pure tone, i.e., harmonically rich. (also narrowband noise)
- Noise masker has no single dominant freq., i.e., more noise like.
- Tonal and noise maskers have different masking characteristics.
- Component considered tonal if *P(k)-P(k±Δk)≥7 dB, where Δk is Bark* distance.





Tonal and Noise Maskers

- *P(k) - Output of FFT*
- *Tonal maskers, $P_{TM}(k)$*
  - Energies from 3 adjacent spectral components centered around the peak are combined.

$$P_{TM}(k) = 10\log\sum_{j=-1}^{1} 10^{0.1P(k+j)}$$

- *Noise maskers, $P_{NM}(k)$*
  - Energies from all spectral components within each critical band are combined.

$$P_{NM}(k) = 10\log\sum_{j} 10^{0.1P(k+j)}$$

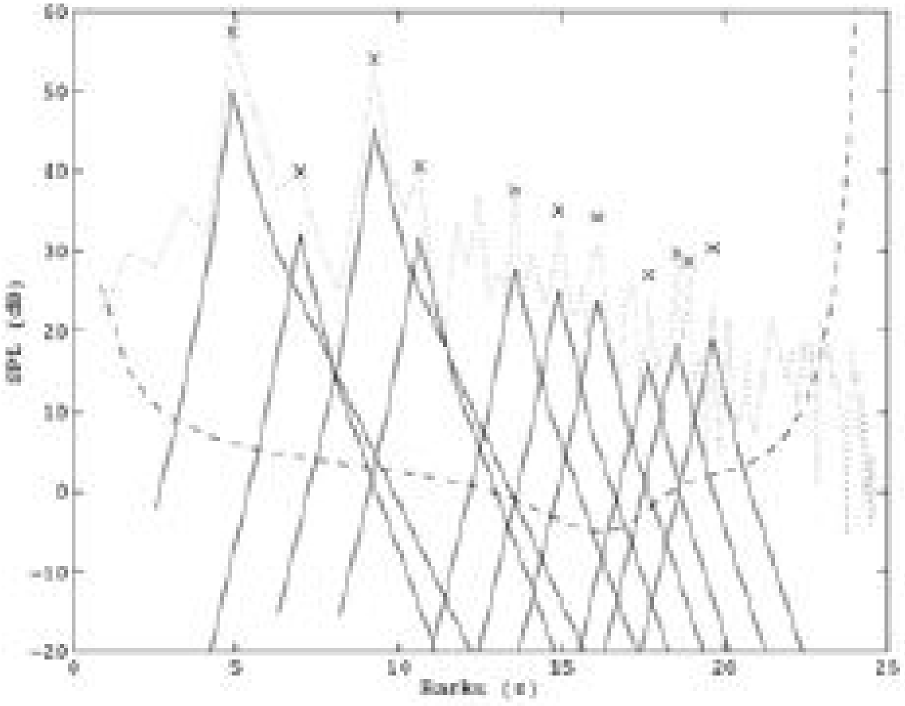
- Only retain maskers that satisfy $P_{TM,\ NM}(k) \geq T_q(k)$
- Sliding 0.5 Bark-window is used to replace any pair of maskers occurring within a distance 0.5 Bark by the stronger of the two.

Calculation of Individual Masking Threshold

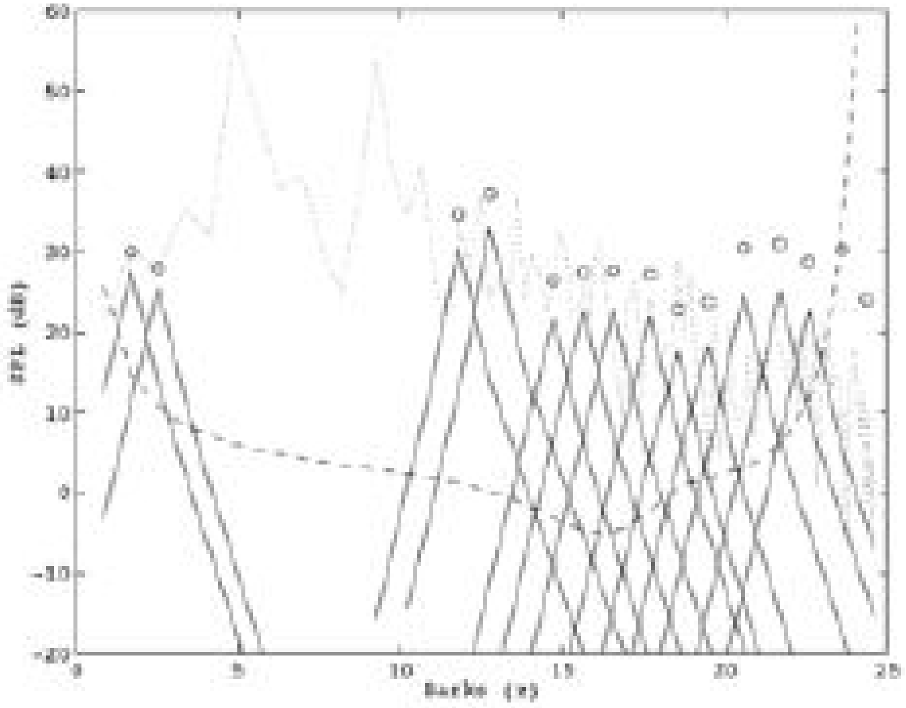
- Masking does not only occur within the critical band, but also spreads to other bands
  - $SF(x) = 15.81+7.5(x+0.474) - 17.5[1+(x+0.474)^2]^{0.5}$
- *Tonal masking threshold:* $T_{TM}(i, j) = P_{TM}(j) - 0.275z(j) + SF(i, j) - 6.025$
- *Noise masking threshold:* $T_{NM}(i, j) = P_{NM}(j) - 0.175z(j) + SF(i, j) - 2.025$
  - $SF(i, j)$ - *spread of masking from bin j to maskee bin i*
  - $z(j)$ - *Bark frequency of frequency bin j*
  - $P_{TM,NM}(j)$ - *SPL of the tonal/noise masker in frequency bin j*

Individual masking thresholds

- Tonal and noise masking thresholds separately examined



Tonal maskers

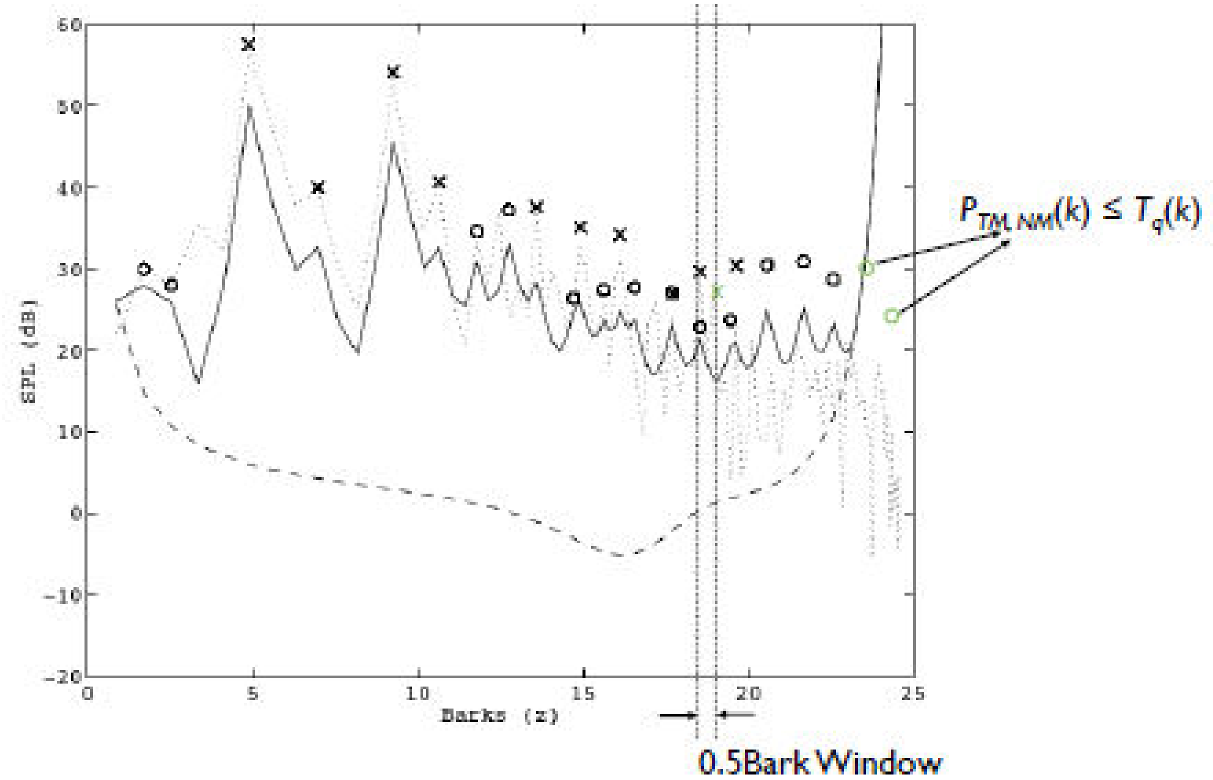


Noise maskers

Calculation of Global Masking Threshold

$$T_g(i) = 10\log\left(10^{0.1T_q(i)} + \sum_{l=1}^{L} 10^{0.1T_{TM}(i,l)} + \sum_{m=1}^{M} 10^{0.1T_{NM}(i,m)}\right)$$

L :# tonal maskers
M :# noise maskers



Masking Model Example

- Suppose the levels of the first 16 of the 32 sub-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

Tone occurs here

- The pre-computed masking model specifies a masking of 12 dB in the 7th band and 15 dB in the 9th.
  - The signal level in 7th band is 10 (< 12 dB), so ignore it.
  - The signal level in 9th band is 35 (< 15 dB), so send it.
- Only the signals above the masking level needs to be sent.

Scaling

- Block of 12 samples for each subband is scaled to normalise the peak signal level within a subband.
  - Largest signal quantized using *6-bit scale-factor.*
- The receiver needs to know the scale factor and quantisation levels used.
  - Fairly small side information included along with the samples

Bit allocation to subbands

- Goal: Minimize the Noise-to-Mask ratio (NMR) over all bands
- For each subband
  - determine SMR=20$\log_{10}$(Signal energy/Minimum masking threshold)
  - Look up estimate of SNR assuming a given number of quantizer levels
  - NMR=SMR-SNR
- Allocate bits to the subband with highest NMR
- Lookup new estimate of SNR for the subband allocated more bits, and recalculate NMR

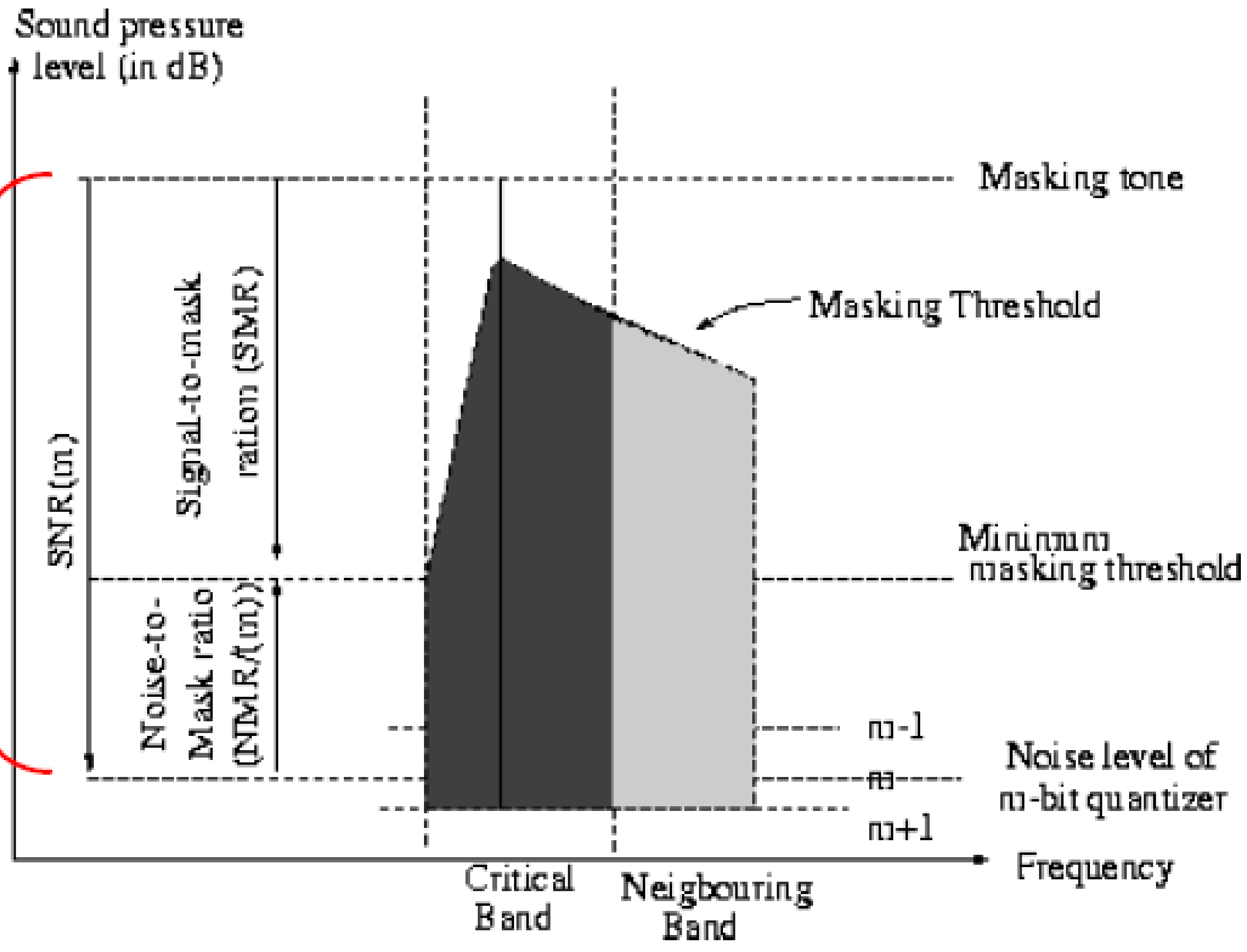
Bit allocation

Each bit increases resolution by ~6dB

$$SNR = 20\log\frac{V}{V_{QN}}$$

$$= 20\log\frac{2^{N-1}}{1/2}$$

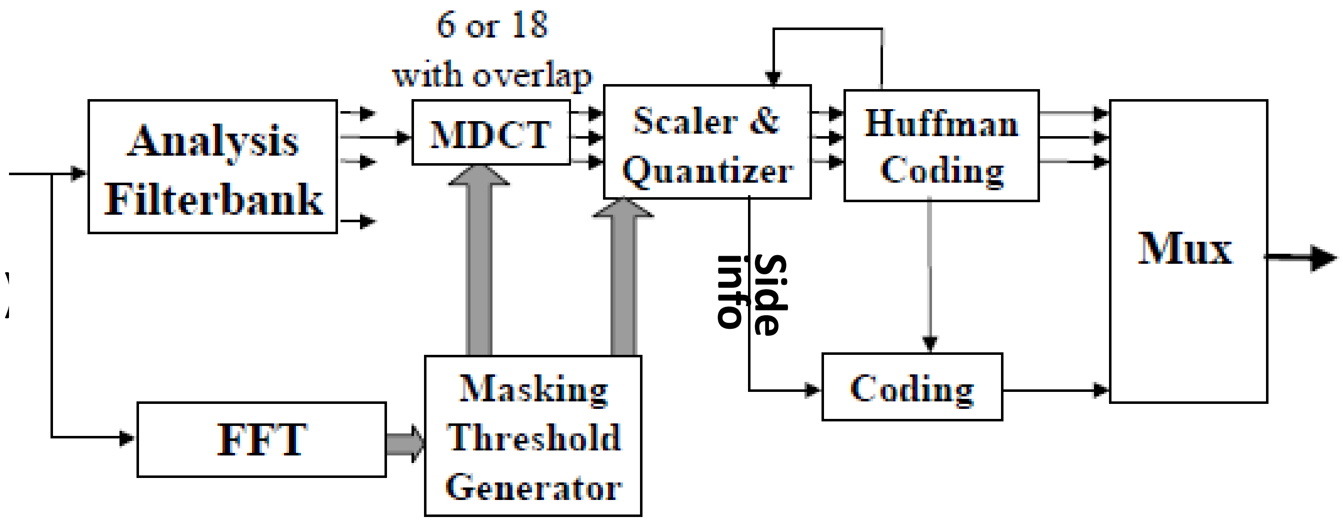
$$= 6.02N(\text{dB})$$



Determined by Tonal or Noise masking threshold

Layer III Encoder

- New Features
- Modified DCT (MDCT)
  - DCT with overlap
  - Long/short window switching
    - Short for better temporal resolution (to prevent pre-echoes)
    - Long for better frequency resolution
- Nonuniform quantization
- Entropy coding
  - Run-length and Huffman coding
- Bit reservoir (buffer)



Frame Structure

Header Info	Side Info	Subband Sanples	Aux Data

- Header info: Sync bits, system info, CRC (cyclic redundancy code)
- Side info: bit allocation, scalefactor, (and scalefactor select for Layer II and III)
- Subband samples:
  - 32 × 12 for Layer I,
  - 32 × 36 for Layer II,
  - 32x18 (hybrid subband/MDCT) for Layer III
- Packetization: 4-byte header, 184-byte payload

Stereo Redundancy Coding

- Four modes: mono, stereo, dual with two separate channel, joint stereo
- In joint stereo mode
  - Human stereo perception > 2kHz is based on envelope
  - Intensity stereo coding > 2kHz
    - Encode (L + R)
    - Assign independent left- and right- scalefactors
- Layer III supports (L+R) and (L–R) coding

MPEG Audio

Performance

Effectiveness of MPEG audio

Layer	Target bit rate	Ratio	Quality @ 64 kbits	Quality @ 128 kbits	Theoretical Min. Delay
I	192 kbps	4:1	-	-	19 ms
II	128 kbps	6:1	2.1 to 2.6	4+	35 ms
III	64 kbps	12:1	3.6 to 3.8	4+	59 ms

5: perfect, 4:just noticeable, 3 : slightly annoying, 2: annoying, 1: very annoying

Complexity

Complexity

Layer	Complexities	
	Encoder	Decoder
I	1.5	1.0
II	2…4	1.25
III	> 7.5	2.5

MPEG-2 Advanced Audio Coding

- **Improvements over MP3**
- More sample frequencies (from 8 to 96 kHz) than MP3 (16 to 48 kHz)
- Up to 48 channels
  - MP3 supports up to two channels in MPEG-1 mode and up to 5.1 channels in MPEG-2 mode)
- Arbitrary bit-rates and variable frame length. Standardized constant bit rate with bit reservoir.
  - Low bit rate speech coding to high quality audio coding
- Higher efficiency at low rates and simpler filter bank
  - rather than MP3's hybrid coding, AAC uses a pure MDCT
- Higher coding efficiency for stationary signals
  - AAC uses a blocksize of 1024 or 960 samples, allowing more efficient coding than MP3's 576 sample blocks especially at low rates

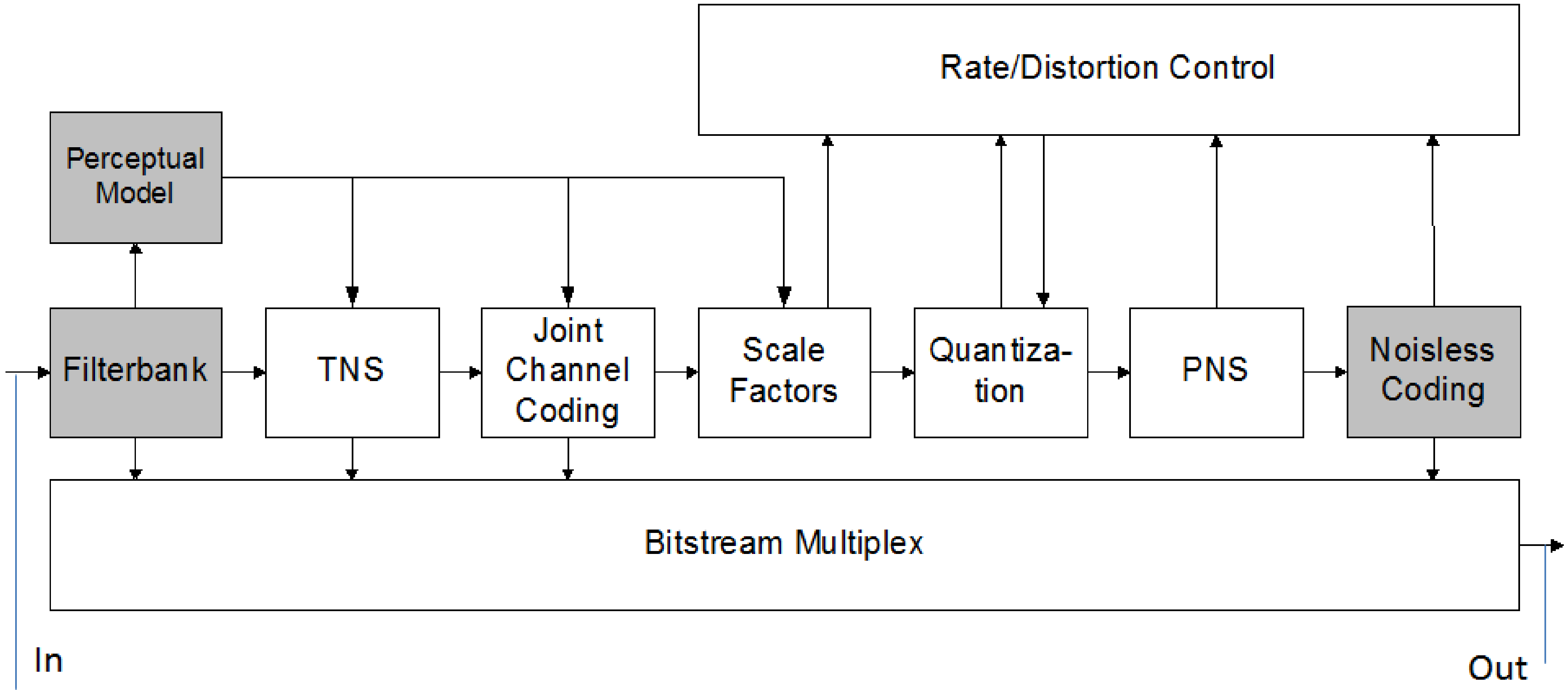
MPEG-2 Advanced Audio Coding

- Higher coding accuracy for transient signals
  - AAC uses a blocksize of 128 or 120 samples, allowing more accurate coding than MP3's 192 sample blocks
- Can use Kaiser-Bessel derived window function to eliminate spectral leakage at the expense of widening the main lobe
- Much better handling of audio frequencies above 16 kHz
- More flexible joint stereo
  - different methods can be used in different frequency ranges
- Additional modules (tools) to increase compression efficiency: Temporal Noise Shaping, Backwards Prediction, Perceptual Noise Substitution etc... These modules can be combined to constitute different encoding profiles.

Joint stereo

- Intensity stereo
  - Joint frequency encoding
    - merge a given frequency range of multiple sound channels together
    - the resulting encoding will preserve the sound information of that range not as a bundle of separate channels but as one homogeneous data stream
- M/S Stereo
  - Haar transform (L+R,L-R)
- Can switch between/combine these techniques per frame

AAC Coding Stages



AAC Coding Features

- Perceptually irrelavant signal components discarded
  - A psychoacoustic model is used in quantization
- Redundancies in the signal exploited
  - Time domain to frequency domain conversion by MDCT
    - Can switch between a single MDCT block of length 1024 points or 8 blocks of 128 points
      - Stationary vs. Non-stationary signals (do changes in statistics occur over time?)
- Internal error correction codes added
  - Luhn-mod-N algorithm

MPEG-2 AAC Profiles

- **Low Complexity (LC)** – the simplest and most widely used and supported;
- **Main Profile (Main)** – like the LC profile, with the addition of backwards prediction;
- **Scalable Sample Rate (SSR)** (MPEG-4 AAC-SSR) – a.k.a. Sample-Rate Scalable (SRS);

MPEG-4

- MPEG-4 Part 3 defined various new compression tools (a.k.a. Audio Object Types)

Audio Profile	Audio Object Types	Date
Synthetic Audio	TTSI, Main synthesis	1999
Speech Audio	CELP, HVXC, TTSI	1999
Scalable Audio	AAC LC, AAC LTP, AAC Scalable, TwinVQ, CELP, HVXC, TTSI	1999
Natural Audio	AAC Main, AAC LC, AAC SSR, AAC LTP, AAC Scalable, TwinVQ, CELP, HVXC, TTSI, ER AAC LC, ER AAC LTP, ER AAC Scalable, ER TwinVQ, ER BSAC, ER AAC LD, ER CELP, ER HVXC, ER HILN, ER Parametric	2000
Mobile Audio Internetworking	ER AAC LC, ER AAC Scalable, ER TwinVQ, ER BSAC, ER AAC LD	2000
Main Audio	AAC Main, AAC LC, AAC SSR, AAC LTP, AAC Scalable, TwinVQ, CELP, HVXC, TTSI, Main synthesis	1999
Low Delay Audio	CELP, HVXC, TTSI, ER AAC LD, ER CELP, ER HVXC	2000
High Quality Audio	AAC LC, AAC LTP, AAC Scalable, CELP, ER AAC LC, ER AAC LTP, ER AAC Scalable, ER CELP	2000
High Efficiency AAC	AAC LC, SBR	2003
HE-AAC v2	AAC LC, SBR, PS	2006
HD-AAC	AAC LC, SLS	2009
ALS Simple	ALS	2010
AAC	AAC LC	2003