Audio Coding

C.M. Liu
Perceptual Signal Processing Lab
College of Computer Science
National Chiao-Tung University

http://www.csie.nctu.edu.tw/~cmliu/Courses/Compression/

Office: EC538
(03)5731877
cmliu@cs.nctu.edu.tw
Overview

- **Speech coding**
  - Based on a model of speech *production*

- **Audio coding**
  - Based on a psychoacoustic model of audio *perception*

- **General idea**
  - Analyze the signal and eliminate inaudible sounds

- **Psychoacoustic model captures**
  - Physiological perception limits (sensor limitations)
  - Psychological perception limits (signal processing limitations)
Threshold of Audibility
dB Sound Pressure Level (SPL) Table

<table>
<thead>
<tr>
<th>Situation</th>
<th>SPL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colt .45 Pistol (25 feet)</td>
<td>140</td>
</tr>
<tr>
<td>Threshold of Pain</td>
<td>130</td>
</tr>
<tr>
<td>Underground Train</td>
<td>120</td>
</tr>
<tr>
<td>Average Home Hi-Fi Level</td>
<td>110</td>
</tr>
<tr>
<td>Average Factory</td>
<td>100</td>
</tr>
<tr>
<td>Average Conversation</td>
<td>90</td>
</tr>
<tr>
<td>Average Office</td>
<td>80</td>
</tr>
<tr>
<td>Residential Ambient Noise</td>
<td>70</td>
</tr>
<tr>
<td>Quiet Whisper (5 feet)</td>
<td>60</td>
</tr>
<tr>
<td>Threshold of Hearing</td>
<td>50</td>
</tr>
<tr>
<td>0.0002 Dyne/Sq. cm</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>0</td>
</tr>
</tbody>
</table>
Age-adjusted SPL Table
Spectral Masking

![Graph showing spectral masking with SPL (dB) on the y-axis and frequency (Hz) on the x-axis. The graph includes various regions such as original threshold of audibility, audible region, and raised threshold of audibility.]

PSP LAB
Psychoacoustic

- Introduction
- Masking Effect
- Critical Band
Psychoacoustic - Introduction

- **Sound Pressure**
  - Sounds are easily described by means of the time-varying sound pressure $p(t)$.
  - The unit of sound pressure is the PASCAL (Pa).
  - In psychoacoustics, values of the sound pressure between $10^{-5}$ Pa and $10^2$ Pa are relevant.

- **Sound Pressure Levels**
  - Normally used to cope with the broad range of sound pressure.
    \[ L = 20 \log_{10}(p/p_0) \text{ dB} \]
  - The reference value of the sound pressure $p_0$ is standardized to $p_0 = 20 \, \mu\text{Pa}$. 

PSP LAB
Psychoacoustic - Introduction (c.1)

- Sound Intensity and Sound Intensity Levels
  \[ L = 20 \log\left(\frac{p}{p_0}\right) dB = 10 \log\left(\frac{I}{I_0}\right) dB \]
  - The reference value \( I_0 \) is defined as \( 10^{-12} \text{ W/m}^2 \)

- Noise Density
  - When dealing with noises, it is advantageous to use density instead of sound intensity
    - e.g., the sound intensity within a bandwidth of 1 Hz.
  - The noise power density, although not quite correct, is also used.
  - The logarithmic correlate of the density of sound intensity is called sound intensity density level, usually shortened to density level, \( \ell \).
  - For white noise, \( \ell \) and \( L \) are related by the equation
    \[ L = [\ell + 10 \log(\Delta f / \text{Hz})] dB \]
    where \( \Delta f \) represents the bandwidth of the sound.
Psychoacoustic - Introduction(c.2)

- **Normative Elements on Human Ear**
  - **Threshold in quiet**
    - A function of frequency that the sound pressure level of a pure tone that is just audible
  - **Masking**
    - Property of the human auditory system by which an audio signal cannot be perceived in the presence of another audio signal.
  - **Masking threshold**
    - A function in frequency and time below which an audio signal cannot be perceived by the human auditory system.
  - **Critical band**
    - Loosely speaking, the perception of a particular frequency, say $\Omega_0$, by the auditory system is influenced by energy in a critical band of frequencies around $\Omega_0$.
    - The ear acts as a multichannel real-time analyzer with varying sensitivity and bandwidth throughout the audio range.
Psychoacoustic - Masking effect

- **Masking of pure tones by noise**
  - Pure tones masked by broad-band noise
  - The masked thresholds rise with increasing frequency.
  - The slope of this increase corresponds to about 10 dB.
  - At low frequencies, the masked thresholds lie about 17 dB above the given density level.
2. Psychoacoustic - Masking effect (c.1)

- Masking of pure tones by narrow-band noise
  - The bandwidth is about 100 Hz below and 0.2 f above 500 Hz.
  - The level of each masking noise is 60 dB and the corresponding bandwidths of the noise are 100, 160, and 700 Hz, respectively.
Psychoacoustic-- Masking Effects (c.2)

- Masking of pure tones by narrow-band noise (c.1)

Fig. 4.4. Level of test tone just masked by critical-band wide noise with centre frequency of 1 kHz and different levels as a function of the frequency of the test tone.
Psychoacoustic - Masking effect (c.3)

- Pure tones masked by Low-pass or High-pass noise
- The cut-off frequencies is 0.9 KHz and 1.1 kHz,
Psychoacoustic - Masking effect (c.4)

- Masking of pure tones by pure tones
  - Masking tone -- 1 kHz, 80 dB.
Psychoacoustic - Masking effect (c.5)

- Pure tones masked by pure tones
Pure tones masked by complex tones
Psychoacoustic - Masking effect (c.7)

- **Temporal effect**
  - **Simultaneous masking**
    - When two signal presence simultaneously, the phenomenon of the weaker signal become inaudible are called simultaneous masking
  - **Premasking**
    - The test sound has to be a short burst or sound impulse which can be presented before the masker stimulus is switched on
  - **Postmasking**
    - The test sound is presented after the masker is switched off, then quite pronounced effects occur
Psychoacoustic - Masking effect (c.8)

- Premasking & Postmasking do not offer much efforts than simultaneous masking

![Diagram showing the effect of masking on sensation level over time after masker onset and delay time.](image-url)
Temporal Masking

![Diagram showing temporal masking with SPL (dB) vs. Time (msec)]

- Premasking
- Masking sound
- Postmasking
MPEG Audio Coding

- Video coding
  - MPEG-1 ➔ VCD (~VHS)
  - MPEG-2 ➔ DVD

- Audio coding in MPEG
  - Layers: I, II, III common to both standards
  - Commonly: layer III == mp3

- Standards:
  - Normative (mandatory) sections
    - Required for compliance, generally output format
  - Informative (optional) sections
    - Not mandatory—e.g., encoding algorithms
MPEG Audio Coding (2)

- **Layers**
  - Backward compatible
  - Upwardly more sophisticated

- **Block-diagram view**
  - Input: 16-bit PCM audio; output: fixed bitrate encoded audio
Polyphase Filter Bank
Polyphase Filter Bank (2)

MPEG/Audio Filter Bank Bands

Critical Band Widths

Increasing Frequency
Example frequency response:
Psychoacoustic Model

- **Model 1:**
  - Less computationally expensive
  - Makes some serious compromises in what it assumes a listener cannot hear

- **Model 2:**
  - Provides more features suited for Layer III coding
  - Assumes more CPU capacity
Psychoacoustic Model (2)

- Time-to-frequency domain conversion
  - Polyphase filter bank + DFT

- Model 1:
  - 512 samples for Layer I
  - 1024 samples for Layers II

- Model 2:
  - 1024 samples and two calculations per frame (MDCT)
Need to separate sound into “tones” and “noise” components

Model 1:
- Local peaks are tones
- Remaining spectrum per critical band lumped into noise at a representative frequency.

Model 2:
- Calculates “tonality” index to determine likelihood of each spectral point being a tone based on previous two analysis windows
“Smear” each signal within its critical band
- Model 1: masking function
- Model 2: spreading function

Adjust calculated threshold by using a “quiet” mask:
- A masking threshold for each frequency when no other frequencies are present.

Calculate signal-to-mask (SMR) per band
Pass results on to coding/framing unit
Psychoacoustic Model Example

Input

lowpass noise + 11.250 kHz Sinewave
sampling rate = 48 kHz
Psychoacoustic Model Example: Transformation to Perceptual Domain
Psychoacoustic Model Example: Masking Thresholds
Psychoacoustic Model Example: Signal-to-Mask Ratios
Psychoacoustic Model Example:
Model Output

lowpass noise + 11.250 kHz Sinewave
sampling rate = 48 kHz
Layer 1 Coding: Overview

- Compression—4:1
- Time frequency mapping
  - 32 subband filters
  - Filter output down-sampled @ 1/32
- Grouping (framing)
  - 12 samples x 32 subbands = 384 samples/frame
  - Each group coded w/ 0-15 bits/sample
- Scale factors
  - Used to optimize quantizer performance
  - Determined based on the 12 samples in the frame
  - 63 scale factors ➔ 6 bits
Layer-Specific Framing

Note: Each subband filter produces 1 sample out for every 32 samples in.
Layer I Coding: Bit Allocation

- **Task:**
  - Determine number of bits to allot for each subband given SMR from psychoacoustic model.

- **Algorithm**
  - Calculate mask-to-noise ratio:
    \[ \text{MNR} = \text{SNR} - \text{SMR} \text{ (in dB)} \]
    - SNR given by MPEG-I standard (as function of quantization levels)
  - Repeat until no bits to allocate left:
    - Allocate bit to subband with lowest MNR.
    - Re-calculate MNR for subband w/ allocated bits.
Layer I Coding: Framing

Frame 1
- Header
- CRC
- Bit allocation
- Scale factors
- Subband data

Frame 2

Frame 3
Layer I Coding: Modes

- **Stereo**
  - Two independently coded channels, synchronized

- **Joint stereo**
  - Two channels coded together
    - ‘Mid’ channel: $M = (L + R)/2$
    - ‘Side’ channel: $S = (L - R)/2$

- **Dual channel**
  - Two independent channel, unsynchronized

- **Single channel (mono)**
  - One channel
Layer II Coding: Overview

- Generally, Layer I methodology w/ some improvements
- Compression—up to 8:1
- Framing
  - $3 \times 12\ \text{samples} \times 32\ \text{bands} = 1152\ \text{samples/frame}$
  - Less overhead per frame
- Scale factor
  - Layer I: 1 per 12 samples
  - Layer II: 1 per 24/36 samples
Layer II Coding: Quantization

- **Layer I:**
  - 1 of 14 possible quantizers per band

- **Layer II**
  - Quantization depends on sampling & bit rates
    - Some band may get 0 bits
  - “Granules” optimization
    - Granule = 3 samples
    - Quantization level based on granules not individual samples
    - E.g.
      - 3 samples @ 5 q.levels = $3^5 = 243$ possible values $\rightarrow$ 8 bits
    - Alternative
      - 1 sample @ 5 q.levels = 3 bits x 3 samples $\rightarrow$ 9 bits
Layer II Coding: Framing

Frame 1
- Header
- CRC
- Bit allocation
- Scalefactor selection index
- Scale factors
- Subband data

Frame 2
- Scalefactor selection index
- Scale factors

Frame 3
- Scalefactor selection index
- Scale factors

42
Layer III Coding (mp3)

- **Problem**
  - For low frequencies, Layer I/II bands are significantly wider than critical bands

- **Obvious solution**
  - Increase spectrual resolution (more bands)

- **The catch**
  - Maintain backward compatibility

- **Solution**
  - First, use the 32-band decomposition
  - Then, use MDCT w/ 50% overlap to subdivide into 6/18
Layer III Coding (2)

- Note that increased spectral resolution ➔ lower temporal resolution
- QE is spread over entire block
  - Larger block ➔ more error spreading
- Normally, backward temporal masking is short (~20ms)
  - QE would appear as a pre-echo
  - Example:
Layer III Coding (3):
MDCT Coefficients
Layer III Coding (4):
Reconstructed Signal
Layer III Coding (5)

- Observation
  - For sharp, short sounds, we need small blocks to control the spread of QE

- mp3 controls the size of the window
  - Long window: 36 samples
  - Short window: 12 samples
  - Start/stop window: 30 samples
  - Example
Layer III Coding (6)
Window Transition Diagram
Layer III Coding (7)

- Long windows
  - $32 \times 18 = 576$ frequencies

- Short windows
  - $32 \times 6 = 192$ frequencies

- Mixed mode
  - Two lowest subbands w/ long windows, the rest—short

- Note
  - Number of samples/frame is always 1152
Layer III Coding (8): Coding & Quantization

- **Outer loop—distortion control loop**
  - Scale factors assigned in bands of coefficients
  - 21 factors for long blocks & 12 for short ones

- **Inner loop—rate control loop**
  - Scaled MDCT coeff are quantized
    - Quantization is nonuniform & companded
  - High-frequency coeff are usually zeroes
    - Grouped into one region and RL is Huffman coded
  - Preceding coeff of the -1, 0, 1 variety are grouped in quadruplets & Huffman coded
  - Remaining coeff are split into 2-3 groups and appropriately Huffman coded
Control loop:
- Runs until target rate is achieved
- Modifies quantization/Huffman codes

Distortion loop:
- Checks psychoacoustic model for allowable distortion
- Modifies scale factors

Bit reservoir
- Coder runs under target rate
- Main data may precede frame header
- However, it cannot split into a following frame
Layer III Coding (10):
Coding & Quantization /3/
MPEG Audio Data Format
## MPEG Audio Frame Header Format

### Example

<table>
<thead>
<tr>
<th>Bits</th>
<th>Binary</th>
<th>Hex</th>
<th>Meaning</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>11111111111111</td>
<td>FFB</td>
<td>MP3 Sync Word</td>
<td>Sync Word</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>F</td>
<td>Version</td>
<td>1 = MPEG</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>F</td>
<td>Layer</td>
<td>01 = Layer 3</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>B</td>
<td>Error Protection</td>
<td>1 = No</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>A</td>
<td>Bit Rate</td>
<td>1010 = 160</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bits</th>
<th>Frequency</th>
<th>Pad. Bit</th>
<th>Priv. Bit</th>
<th>Mode</th>
<th>Mode Extension</th>
<th>Copy</th>
<th>Original</th>
<th>Emphasis</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>22</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>23</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>24</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>25</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>26</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>27</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>28</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>29</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>30</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>31</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>32</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

00 = 44100 Hz
0 = Frame is not padded
Unknown
01 = Joint Stereo

0 = Intensity Stereo Off
0 = MS Stereo Off
0 = Not Copyrighted
0 = Copy Of Original Media
0 = None
A. MPEG1 Audio Coding

- Features
- Generic Concepts
- Features of Each Layers
- Coded Bit-Stream
- Layer I, II CODEC
- Layer III CODEC
- Psychoacoustic Model 1
- Psychoacoustic Model 2
- Stereo Control
- Concluding Remarks
A. MPEG1 -- Features

- Sampling Rate
  - 32, 44.1, 48 kHz

- Input Resolution
  - 16 bits uniform PCM

- Modes
  - Stereo, Joint Stereo, Dual Channel, and Single

- Layers
  - Layer 1: 32 - 448 kbps/channel
  - Layer 2: 32 - 384 kbps/channel
  - Layer 3: 32 - 320 kbps/channel
A. MPEG1 -- Generic Concepts

- **Layer 1**: 32 - 448 kbps/channel
  - Simplified of MUSICAM
  - Consumer applications where very low data rates are not mandatory.
    - Digital home recording on taps, Winchester discs, or digital optical disks

- **Layer 2**: 32 - 384 kbps/channel
  - Nearly identical to MUSICAM except the frame header
  - Consumer & professional audio
    - Audio broadcasting, television, recording, telecommunication, and multimedia

- **Layer 3**: 32 - 320 kbps/channel
  - Most effective modules in MUSICAM and ASPEC
  - Most telecommunication, narrowband ISDN, Professional audio with high weights on very low bit rate.
# A. MPEG1 -- Features of Each Layers

<table>
<thead>
<tr>
<th></th>
<th>Layer I</th>
<th>Layer II</th>
<th>Layer III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis/Synthesis</td>
<td>32 subbands</td>
<td>32 subbands</td>
<td>Hybrid (subband + MDCT)</td>
</tr>
<tr>
<td>Bit allocation representation</td>
<td>explicit</td>
<td>indexing</td>
<td>Indexing</td>
</tr>
<tr>
<td>Suggested psychacoustic model</td>
<td>Model 1</td>
<td>Model 1</td>
<td>Model 2</td>
</tr>
<tr>
<td>Output bit-rate</td>
<td>32 - 448 Kbps</td>
<td>32 - 384 Kbps</td>
<td>32 - 320 Kbps</td>
</tr>
<tr>
<td>Efficient bit-rate</td>
<td>160 - 224 Kbps</td>
<td>96 - 128 Kbps</td>
<td>64 - 96 Kbps</td>
</tr>
<tr>
<td>Sampling frequency</td>
<td>32, 44.1, 48 Khz</td>
<td>32, 44.1, 48 KHz</td>
<td>32, 44.1, 48 KHz</td>
</tr>
<tr>
<td>Intensity stereo</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Quantization</td>
<td>Uniform</td>
<td>Uniform</td>
<td>non-uniform</td>
</tr>
<tr>
<td>Segmentation</td>
<td>Fixed</td>
<td>Fixeded</td>
<td>dynamic</td>
</tr>
<tr>
<td>Segmentation</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Slot Size</td>
<td>4 bytes</td>
<td>1 bytes</td>
<td>1 bytes</td>
</tr>
<tr>
<td>Frame Size</td>
<td>384 samples</td>
<td>1152 samples</td>
<td>1152 samples</td>
</tr>
<tr>
<td>Frame-self decodability</td>
<td>Yes</td>
<td>Yes</td>
<td>needs previous frames</td>
</tr>
<tr>
<td>Technical originality</td>
<td>simplified MUSICAM</td>
<td>Refined MUSICAM</td>
<td>Hybrid from MUSICAM and ASPEC</td>
</tr>
</tbody>
</table>
A. MPEG1 -- Audio Coded Bitstream Syntax
A. MPEG1 -- Audio Coded Bitstream

Syntax (c.1)

- **Frame in Layer 1 & 2**
  - Part of the bitstream that is decodable by itself
  - In Layer 1 it contains information for 384 samples and Layer 2 for 1152 samples
  - Starts with a syncword and ends just before the next syncword.
  - Consists of an integer number of slots (four bytes in Layer 1 and one in Layer 2).

- **Frame in Layer 3**
  - Part of the bit stream that is decodable with the use of previously acquired side and main information.
  - Contains information for 1152 samples
  - The distance between the start of consecutive syncwords is an integer number of slots (one byte in Layer 3).
  - The audio information belonging to one frame is generally contained between two successive syncwords.
A.MPEG1-- Audio Coded Bitstream
Syntax (c.1)

- **Audio Data for Layer 1 Single Channel**

  - Scalefactor \((i+1) = \frac{\text{Scalefactor (i)}}{1.25992104989487}\)

  - The number of bits per samples in the band

  - Normalization Factor of the samples in the band

  - Bank 0
    - (12 samples for 384 PCMs)

  - Bank 1

  - Bank 31
A.MPEG1 -- Audio Coded Bitstream Syntax (c.2)

- Audio Data for Layer 2 Single Channel
  - Three successive subband samples are grouped to a granule and coded with one code word for quantization steps-- 3, 5, 9.

![Diagram showing the structure of audio data and bit allocation](image-url)

- Band 0
  - Bit-Allocation (hi-2; mi-3; li-4)
- Band i
  - SCFSI (2)
  - Scalefactor (6-18)
  - Samples (?)
- Band 31
  - (36 samples for 1152 PCMs)
A.MPEG1 -- Audio Coded Bitstream Syntax (c.2)

- Audio Data for Layer 3
  - Side Information 17 or 32 bytes one or two channel
    - window type, the Huffman Table numbers, the region table apply, scalefactor descriptors, a pointer to the end of the main data.
  - Main Data
    - The scalefactor and Huffman data

![Diagram of Audio Coded Bitstream Syntax]
A.MPEGI—Layer I, II Codec

- The Codec Process

- Analysis Filter Bank
- FFT
- Masking Threshold
- Dynamic Bit Allocation
- Signal to Mask Ratio
- Quantization Samples
- Scaler and Quantizer
- Frame Packing
- Encoded Bitstream
- Digital Channel
- Synthesis Filter Bank
- Dequantizer & Descaler
- Frame Unpacking
A. MPEGI-- Layer I

- **Bit Rates**
  - 32, 64, 96, ... 448 kb/s

- **Sampling Frequency**
  - 44.1, 48, 32 kHz and one reserved code.

- **Frame**
  - 384 16-bit PCM samples

- **Encoding Method**
  - The analysis filter bank and psychoacoustic model execute in parallel
  - **Filter Bank**
    - 32 equal bandwidth polyphase pseudo-QMF
A.MPEGI -- Layer I (c.1)

- **Scalefactor**
  - A normalization factor before quantization.
  - For each subband, using the max(abs(12 subband samples)) to obtain scalefactor by search table (Table 3-B.1.).
  - Each scale factor is represented by 6 bits.

- **Psychoacoustic Model**
  - Calculated in 512 pt. FFT with shift length 384 samples.
  - Provide more sufficient frequency resolution than filter bank.
  - Produces Signal-to-Mask Ratio for each subband.

- **Dynamic Bit Allocation**
  - Use iterative procedure to determine the quantization level for each subband.
  - Let quantization noise is under masking threshold calculated by psychoacoustic model.
Initially, allocate the quantizer of each subband with zero bit, and we know the SMR of each subband.

- Calculate available bit number in this frame.
- From the quantizer step size find the SNR of each subband (Table 3-C.2).
- Calculate Mask-to-Noise Ratio (MNR) by \( MNR = SNR - SMR \).
- Pick the minimum MNR, allocate one more bit of each sample in this subband.
- Repeat the above four steps until there is no more bits to allocate.
Quantization and encoding of subband samples (4 bits)

- From the step size - quantizer table (Table 3C.3), using A, B, and scalefactor to quantize the subband samples:
  - $X_1 = X / \text{scalefactor}$
  - $X_2 = A \cdot X_1 + B$
  - Take the N MSB bits
  - Invert the MSB bit (avoid confusing with sync. word 1111..1)

Minimum Rate Distortion Curve:
Minimum MNR versus number of bits required to encode a layer 1 frame determined for a particular frame.
A.MPEGI-- Layer II

- **Encoding Method**
  - **Filter bank**
    - Use the same analysis filter bank with layer I
    - Since the number of input samples become 3*384, the output samples become 3*12*32
  - **Scalefactor**
    - Like Layer I, but has 3 scalefactors in one subband
    - Coding with the 3 scalefactors 6-18 bits

1. Look up the Table 3-B.1
2. Calculate two successive differences and label the differences into 5 classes
   
   \[
   \text{dscf1} = \text{scf1} - \text{scf2} \\
   \text{dscf2} = \text{scf2} - \text{scf3}
   \]
3. Lookup Table 3-C.4 for 6-18 bits
Psychoacoustic model
- Use 1024 pt. FFT with shiftlength 1152 pt.
- Produce Signal-to-Mask Ratio (SMR), like layer I

Dynamic bit allocation
- Performs like layer I

Quantization and encoding of subband samples
- Like Layer I, but has finer quantization with up to 16 b amplitude.
- The number of available quantizers decreases with increased subband index.
- If quantization level is 3, 5, or 9, 3 consecutive samples are coded into one codeword
  - eg. v3=9z+3y+x (3 based)
  - eg. v5=25z+5y+x (5 based)
  - eg. v9=81z+9y+x (3 based)

Packing
A. MPEGI-- Layer I, II Decoder

- Layer I, II Decoding
  - No Psychacoustic Model is needed.
  - The computation power for Encoder and Decoder for Layer 1 is about 2:1 and Layer 2 is 3:1.

```
Begin
Input Encoded Bit Stream
Decoding of Bit Allocation
Decoding of Scalafactor
Output PCM Samples
Synthesis Subband Filters
Requantization of Samples
End
```
A.MPEG1 -- Synthesis Subband Filter Flow Chart

Begin

Input 32 New Subband Samples
Si = 0, ..., 31

For i = 0 to 63 do
V(i) = \sum_{k=0}^{31} N_{ik} S_k

For i = 1023 down to 64 do
V(i) = V(i-64)

Build a 512 value vector U for i = 0 to 7 do for j = 0 to 31 do
U(64i+j) = V[128i+j];
U[64i+32+j] = V[128i+96+j]

Window by 512 coefficients
Produce vector W for i = 0 to 511 do
Wi = Ui * Di

Calculate 32 samples for j = 0 to 31 do
S_j = \sum_{k=0}^{15} W_{j+32i}

Output 32 reconstructed PCM Samples

End
A.MPEGI-- Layer III Codec

- Analysis Filter Bank
- MDCT with dynamic windowing
- Scaler and Quantizer
- Huffman Coding
- Coding of Side Information
- Packing

- FFT
- Masking Threshold
- Dequantizer & Descaler
- Huffman Coding
- Decoding of Side Information
- Unpacking

- Synthesis Filter Bank
- MDCT with dynamic windowing
- PSP LAB
A. MPEGI-- Layer III Codec

The coder process:

Layer I Analysis Filter

- Buffer for 36 subband 0 samples
- Buffer for 36 subband 31 samples

Psychoacoustic Model: decide block type and allow distortion

- Block: long: 36 short: 3 * 12
- Block: long: 36 short: 3 * 12
- MDCT output: long: 18 short: 6

Scaler and quantizer

Huffman coding

coding of side information
Features

- Hybrid polyphase/MDCT filter bank.
- Different frequency resolution for the attributes of samples.
- Nonuniform quantization and Huffman coding are used to increase coding efficiency.
- A buffer technique called bit reservoir is used to maintain coding efficiency and to keep the quantization noise below the masking threshold.
- The scalefactors of gr1, gr2 can be grouped.
- Long blocks and short blocks are switched for controlling pre-echo effects.
- Long block: the 576 frequency lines can be divided to 21 scalefactor bands
- Short block: the 192 frequency lines can be divided to 12 scalefactor bands
A. MPEGI-- Layer III Codec (c.2)

- **Encoding Method**
  - **Filter Bank**
    - Using the same filter bank with layer I to get the subband samples
    - According to the block type of psychoacoustic model to do MDCT with dynamic windowing.
  - **Psychoacoustic Model (c-29)**
    - The model is calculated twice in parallel
      - 1024 pt. FFT with shiftlength 576 pt. for long block
      - 256 pt. FFT with shiftlength 192 pt. for short block
    - According to the perceptual entropy of signal determine using short or long block
    - Produce the block type, threshold, perceptual entropy, time signal
A.MPEGI-- Layer III Codec (c.3)

- Quantization
  - Nonuniform quantization
    
    \[ is(i) = n \text{int}(((xr(i) / quant)^{0.75}) - 0.0946) \]
    
    - \( xr(i) \): absolute value of frequency line at index \( i \)
    - \( quant \): actual quantizer step size
    - \( n \text{int} \): nearest integer function
    - \( \text{quantized absolute value at index } i \).
  
  - Bigger values are quantized less accurately than smaller values.

- Huffman Coding
  - A series of zero at high frequencies is coded by run-length coding.
  - All the contiguous quadruples consisting of values 0, -1, 1 are assigned to the count1 section (two tables).
  - The remaining pairs whose absolute values form the last section (32 tables).
The choice of Human table depends largely on the dynamic range of the values.

Allocation
- Using two iteration loops to control the distortion be under the masking threshold
- Inner loop: quantize the input vector and increases the quantizer step size until output vector can be coded with the available amount of bits.
- Outer loop: Check the output from inner loop, if the allow distortion is exceeded, amplifies the scalefactor band and call the inner loop again (Noise allocation)

Bit Reservoir Technique
- A buffer technique
- The amount of bits corresponding to a frame is no longer constant, but varies with a constant long term average.
Window switching logic

- Normal
- Start
- Stop
- Short

Transitions:
- Normal to Start: attack
- Start to Stop: attack
- Stop to Start: attack
- Normal to Stop: no attack
- Start to Short: attack
- Stop to Short: no attack
- Short to Start: attack

States:
- Normal
- Start
- Stop
- Short

Events:
- Attack
- No attack
Four Types of Window Functions
A.MPEG-I -- Layer III Codec (c.7)

- Illustration of the window switching decision

![Diagram of Input Sound Signal with time (t) and blocks: normal, start, short, stop, normal block]

**Input Sound Signal**

- t

- Normal block
- Start block
- Short block
- Stop block
- Normal block
A.MPEGI-- Layer III Codec (c.8)

- **Scalefactor Bands**
  - 576 spectral lines for the long MDCT are grouped into 21 scalefactor bands.
  - 192 spectral lines for short MDCT are grouped into 12 scalefactor bands.
  - Scalefactors are sent either for each granule in a frame or for both granules together, depending upon the contents of the scalefactor selection information (sfsi) variable.
  - The 21 scalefactor bands are assigned to four groups.
    - For each group, a scalefactor selector information (1 bit) for one factor for a granule or for both granules.
  - The number of bits of the scalefactor is specified by a four bit variable called scalefac_compress.
  - The 21 bands are divided into two groups 0-10 and 11-20. The scalefac_compress variable indexes a table which returns two numbers called slen1 and slen2, the number of bits assigned to the bands in groups 1 and 2.
A. MPEGI-- Layer III Codec (c.8)

- **Allocation Function**
  - Outer loop adjusts the scale factor to shape the noise spectrum.
  - Inner loop sets the `global_gain` to the value which brings the number of bits to encode the granule to the value closest but not exceeding `maxbits`.
  - During the initial stage, the maximum allowable quantization noise, `xmin`, is determined for each scale factor band.
  - The maximum number of bits for encoding the granule, `maxbits`, is determined and all scalfactors are initialized to their lowest values.
  - The iteration terminates when the scalfactors are increased in small increments until the quantization noise is below `xmin` or until the scalfactors cannot be increased any more.

The main bottleneck to the iteration occurs in the inner loop where the number of bits needed to encode the granule is determined. This involves quantizing the 576 values and counting the number of bits needed to Huffman encode the quadruples and `big_value` pair.
A.MPEGI-- Layer III Codec (c.9)

- Global gain
- 576 spectral lines
- SNR
- Xmin

PSP LAB
A.MPEGI-- Layer III Codec (c.9)

- Decoding Method
  - Huffman decoding
  - Requantization and all scaling
  - Synthesis filter bank [ preprocess alias reduction ]
  - IMDCT
  - Overlapping and add with previous block
A. MPEGI-- Synthesis Subband Filter Flow Chart

Begin

Get Bit Stream, Find Header

Decode Scalefactor

Decode Samples

Inverse Quantize Samples

Reorder if (blocksplit_flag) and (block_type == 2)

Alias Cancellation

Synthesize via MDCT & Overlap-Add (Using either 18 or 666 depending on blocksplit_flag and block_type [gr])

Synthesize via Polyphase MDCT

Output Samples

End
A. MPEG-1 -- Psychoacoustic Model I

- **Input**
  - Layer I: 384 PCM samples frame; Layer II: 3*384 PCM samples

- **Output**
  - Signal-to-Mask Ratio (SMR) for each subband

- **Steps**

- **FFT Analysis**
  - Find the corresponding data segment with length \( N = 512 \) for layer I, and 1024 for layers II and III.
    - \( s(l) \quad 0 \leq l \leq N-1 \)
  - **Hanning window**
    
    \[
    h(i) = \sqrt{8/3} \times 0.5 \times \left\{1 - \cos\left[2\pi(i)/(N-1)\right]\right\}; 0 \leq i \leq N-1
    \]
**Power Density Spectrum**

\[ X(k) = 10 \log \left( \frac{1}{N} \sum_{l=0}^{N-1} h(l) s(l) e^{-2 j k l \pi / N} \right)^2 \, dB \quad k = 0, 1, \ldots, N / 2 \]

**Determination of Sound Pressure Level**

- Determine the sound pressure level in each subband, \( L_{sb}(n) \) by choosing one largest energy to represent this subband.

\[ L_{sb} = \max_{X(k) \text{ in subband } n} \left[ X(k), 20 \log(scf_{max}(n) \times 32768) - 10 \right] dB \]

- \( X(k) \) is the sound pressure level of the spectral line with index \( k \) of the FFT with the maximum amplitude in the frequency range corresponding to subband \( n \).
- The -10dB term corrects for the difference between peak and RMS level.
Determine the Threshold in Quiet

- The threshold in quiet $LT_q(k)$ is called **absolute threshold**.
- These tables depend on the sampling rate of the input PCM signal (see Table 3-D.1).
- An offset depending on the overall bit rate is used for the absolute threshold. This offset is -12 dB for bit rates $\geq 96$ kbits/s and 0 dB for bit rates $< 96$ kbits/s per channel.

Finding of Tonal and Nontonal Components

**Concepts:** The different masking effect between tonal and nontonal sound

- Labelling of local maximum.
  - $X(k) > X(k-1)$ and $X(k) \geq X(k+1)$
  - $X(k) - X(k+j) \geq 7$ dB; the range of $j$ depends on Layers and frequency.
A. MPEGI -- Psychoacoustic Model I (c.3)

- Listing of tonal components and calculation of the sound pressure level.
  - Index number $k$
  - Sound pressure level $X_{tm}(k) = X(k-1) + X(k) + X(k+1)$
  - Tonal Flag.

- Listing of non-tonal components and calculation of the power.
  - The power of the spectral lines are summed to form the sound pressure level of the new non-tonal component corresponding to that critical band.
  - Index number $k$ of the spectral line nearest to the geometric mean of the critical band. (see Table 3-D.2, Layer 1: 23 critical bands for 32kHz, 24 bands for 44.1k; 25 for 48k; Layer 2: 25 bands for 32 KHz, 26 bands for 44.1 and 48 kHz)
  - Sound pressure level $X_{nm}(k)$ in dB
  - Non-tonal flag
A. MPEGI -- Psychoacoustic Model I (c.4)

- FFT Spectrum
- Magnified Portion
- Power Spectrum
- Decimated Spectrum
Decimation of the Masking Components.
- Concepts: Reduce the number of Maskers
- Eliminate the tonal $X_{tm}(k)$ and non-tonal $X_{nm}(k)$ components below absolute threshold
- Eliminate the tonal components within a distance of 0.5 Bark, remove the smaller components from tonal lists

Calculation of the Individual Masking Thresholds
- Of the original domain samples indexed by $k$, only a subset of the samples indexed by $i$ are considered for threshold calculation.
- Calculate individual tonal and non-tonal masking threshold to predetermined frequency
  \[
  LT_{tm}(z(j),z(i)) = X_{tm}(z(j)) + av_{tm}(z(j)) + vf(z(j),z(i)) \text{ dB}
  \]
  \[
  LT_{nm}(z(j),z(i)) = X_{nm}(z(j)) + av_{nm}(z(j)) + vf(z(j),z(i)) \text{ dB}
  \]
  \[
  av_{tm} = -1.525 - 0.275*z(j) - 4.5 \text{ dB}
  \]
  \[
  av_{nm} = -1.525 - 0.175*z(j) - 0.5 \text{ dB}
  \]
A.MPEGI -- Psychoacoustic Model I (c.6)

- Non-Tonal Components
- Decimated Non-Tonal Components
- Non-Tonal Masking
- Tonal Masking
A. MPEGI -- Psychoacoustic Model I

(c.7)

- $vf = 17(dz + 1) - (0.4 \times X(z(j)) + 6) \, \text{dB}$ for $-3 \leq dz < 0.1$ Bark
- $vf = (0.4 \times X[z(j)] + 6) \times dz \, \text{dB}$ for $-1 \leq dz < 0$ Bark
- $vf = -17 \, \text{dB}$ for $0 \leq dz < 0$ Bark
- $vf = -(dz-1) \times (17-0.15 \times X[z(j)]) - 17 \, \text{dB}$ for $1 \leq dz < 8$ Bark

.dequeue

Determine the Global Masking Threshold

- Calculate total masking energy for every frequency component (normal square amplitude scale).

$$LT_g(i) = 10 \log(10^{LT_q(i)/10} + \sum_{j=1}^{m} 10^{LT_m(j,i)/10} + \sum_{j=1}^{n} 10^{LT_m(j,i)/10})$$

- The summation can be reduced to +8 to -3 Barks
A. MPEGI -- Psychoacoustic Model I (c.8)

Masking Function for a 40dB

Masking Function for a 80dB

Absolute Threshold for Layer 1

Absolute Threshold for Layer 2
Determine the Minimum Masking Threshold

\[
LT_{\text{min}}(n) = \text{MIN}_{f(i) \text{in subband } n} [LT_g(i)] dB
\]

- Determine the minimum masking level in each subband.

Calculate the signal-to-mask ratio

![Graphs showing global and minimum masking thresholds across subbands and spectral lines.](image)
A. MPEGI -- Psychoacoustic Model II

- Input
  - Short block: 192 new sample with 256 pt. FFT
  - Long block: 576 new sample with 1024 pt. FFT

- Output
  - Block type (Layer III)
  - Signal-to-Mask Ratio (SMR) for each scalefactor band

- Calculation of block type and SMR
  - FFT: Using Hanning window, get the polar representation of the transform \( r_w \) and \( f_w \)
  - Use a polynomial predictor to predict magnitude and phase
    \[
    r^A_w = 2.0 r_w(t-1) - r_w(t-2) \\
    f^A_w = 2.0 f_w(t-1) - f_w(t-2)
    \]
Calculate Euclidian distance between predict value $c_w$ for each line.

$$c_w = \left( \frac{\left( r_w \cos \tilde{f}_w - \tilde{f}_w \cos \tilde{A}_w \right)^2 + \left( r_w \sin \tilde{f}_w - \tilde{f}_w \sin \tilde{A}_w \right)^2}{r_w + a_b(\tilde{A}_w)} \right)^{0.5}$$

Unpredictivity Measure

Calculate energy & unpredictability in each partition. (Table 3-D.3)

- A resolution of approximately one FFT line or 1/3 critical band.
- The Table Contents
  - Index of the partition, $b$
  - Lowest frequency line, $w_{lowb}$
  - Highest frequency line, $w_{highb}$
  - The median Bark value, $b_{valb}$
  - A lower limit for the SNR in the partition that controls stereo unmasking effects, $b_{minvalb}$
  - The value for tonal masking noise for that partition, $b_{TMNb}$

\[ e_b = \sum_{w=lowb}^{highb} r_w^2 \]

\[ c_b = \sum_{w=lowb}^{highb} r_w^2 c_w \]
A.MPEGI -- Psychoacoustic Model II

(c.2)

Partition number - critical band rate

Conversion from spectral line to partition number

Fourier Power Spectrum

Unpredictability

Partition energy & Partitioned Unpredictability

Convolved partitioned energy & unpredictability
A.MPEGI -- Psychoacoustic Model II (c.3)

- Calculate energy($ecb_b$) & noise($cb_b$) masking effects and normalize them($cb_b$, $en_b$)

$$\begin{align*}
  cb_b &= \frac{ct_b}{ecb_b} \\
  en_b &= ecb_b \times rnorm_b \\
  rnorm_b &= \frac{1}{\max_b \sum_{bb=0}^{b_{\max}} \text{sprdngf}(bval_{bb}, bval_b)}
\end{align*}$$

- Convert chaos measure $cb_b$ to tonality $tbb$
  
  - $0.05 < cb_b < 0.5 \Rightarrow 0 < t_b < 1$
  
  $$tbb = -0.299 - 0.43 \log_e(cb_b)$$
According to the tonality $(t_{b_b})$ to calculate spread threshold $(\text{SNR}_b)$

$\text{SNR}_b = \max(\minval_{b_b}, t_{b_b} \times \text{TMN}_b + (1-t_{b_b}) \times 5.5)$

Calculate the power ratio of $\text{SNR}_b$

$bc_b = 10^{\frac{-\text{SNR}_b}{10}}$

$n_{b_b} = e_{n_{b}} \times bc_b$

Renormalization $(n_{b_w})$

$n_{b_w} = \frac{n_{b_{b}}}{whight_{b} - w - wlow_{b} + 1}$

Include the absolute threshold for final Energy Threshold of audibility

$\text{th}_{w_{b}} = \max(n_{b_w}, absth_{w})$
A.MPEGI-- Psychoacoustic Model II (c.5)

- Pre-Echo Control (C-30)
  \[ \text{thr}(b) = \max(q\text{thr}(b), nbb(b), nbb_I(b), nbb_{II}(b)) \]
  \[ nbb_I(b) = 2 \times nbb(b) \text{ from the last block} \]
  \[ nbb_{II}(b) = 16 \times nbb(b) \text{ from the block before the last block} \]

- Calculate SMR in each scalefactor band

\[ n_{\text{part}}^n = \sum_{w_{\text{low}}}^{w_{\text{high}}} \text{thr}_w \text{ for narrow scalefactor band} \]
\[ n_{\text{part}}^n = \min(\text{thr}_{w_{\text{low}}}, \ldots, \text{thr}_{w_{\text{high}}}) \times (w_{\text{high}_n} - w_{\text{low}_n} + 1) \text{ for wide scalefactor band} \]

\[ e_{\text{part}}^n = \sum_{w_{\text{low}}}^{w_{\text{high}}} r_w^2 \]

\[ SMR_n = 10 \log_{10} \left( \frac{e_{\text{part}}^n}{n_{\text{part}}^n} \right) \]
A.MPEGI -- Psychoacoustic Model II

(c.6)

1. Masking Threshold
2. Masking Threshold mapped to Fourier spectral domain
3. Signal-to-masker ratio
4. Spreading Function
5. Tone Masking Function
6. The minval function

PSP LAB
A. MPEGI -- Adaptation of the Psychoacoustic Model 2 for Layer 3

- Tone masking ratio is now fixed at 29dB and the noise masking tone is 6 dB.
- Mainval function is specified as a function of partition number.
- A perceptual entropy is returned.
  - calculated from the weighted average of the logarithm of the ratio of the masking threshold over the energy for all partitions.
  - The ratio is then transformed to the scale factor bands.
- The unpredictability measure is now computed using long FFTs (1024) for the first 6 lines and short FFTs (256) for the next 200 lines, for the remaining lines, the unpredictivity assumes a value of 0.4.
A. MPEGI-- Adaptation of the Psychoacoustic Model 2 for Layer 3

- Preecho control is incorporated by computing a new threshold based on the current threshold and the thresholds calculated for the previous two blocks.

![Spreading Function](image1.png)

![Minval Function](image2.png)
A.MPEGI-- Intensity Stereo

- Concepts
  - At high frequency, the location of the stereophonic image within a critical band is determined by the temporal envelope and not by the temporal fine structure.

- Technique (adopted in Layer 1&2)
  - For some subbands, instead of transmitting separate left and right subband samples only the sum-signal is transmitted, but with the scalefactors for both the left and right channels.
  - The quantization, coding, bit allocation are preferred in the same way as in independent coding.
A. MPEGI-- Intensity Stereo and MS Stereo

- **MS_Stereo Switching**
  - Switch is on when

  $$\sum_{i=0}^{511} [rl_i^2 + rr_i^2] < 0.8 \sum_{i=0}^{511} [rl_i^2 + rr_i^2]$$

  - The values rli and rri correspond to the energies of the FFT line spectrum of the left and right channel calculated within the psychoacoustic model.

- **MS_Matrix**

  $$M_i = \frac{R_i + L_i}{\sqrt{2}}; S_i = \frac{R_i - L_i}{\sqrt{2}}$$
### Concluding Remarks

<table>
<thead>
<tr>
<th>Features</th>
<th>Layer I</th>
<th>Layer II</th>
<th>Layer III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mapping into frequency coefficients</td>
<td>subband filtering into 32 equally spaced bands</td>
<td>subband + overlap transform with dynamic windowing.</td>
<td></td>
</tr>
<tr>
<td>Scaling factor</td>
<td>one scalefactor for 12 consecutive output</td>
<td>1 to 3 for 36 consecutive each band</td>
<td>by clustering of frequency into bands, 1 per band output each band</td>
</tr>
<tr>
<td>Bit allocation</td>
<td>found by iteration based on MNR (bit allocation )</td>
<td>based on SMR (noise allocation )</td>
<td></td>
</tr>
</tbody>
</table>
B. MPEG II & IV

- **MPEG II**
  - The syntax, semantics, coding techniques are maintained.
  - Multichannel and multilingual audio.
  - Lower sampling frequencies (16, 22.05, 24 kHz) and lower bit rates.
  - Multichannel Features: 3/2, 3/1, 3/0, 2/2, 2/1, 2/0, 1/0)
  - Downward compatibility
  - NBC (Non backward compatible) syntax will be defined.

- **MPEG IV**
  - Address at low and very low bit rates
    - A stereo audio signal over a 64 kb/s rate channel
  - In the field of digital multichannel surround systems
    - The use of interchannel correlations and interchannel masking effects
Remark A: Derivation of the Analysis/Synthesis Filters for MPEGI-- Audio

The Modulated Filter Banks

The Analysis Filters $h_k(n)$

$$h_k(n) = p(n) \cos \left[ \left( k + \frac{1}{2} \right) \left( n - \frac{L-1}{2} \right) \frac{\pi}{M} - \phi_k \right]$$

where $L$ and $M$ are the filter length and filterbank number.

The Synthesis Filters $f_k(n)$

$$f_k(n) = p(n) \cos \left[ \left( k + \frac{1}{2} \right) \left( n - \frac{L-1}{2} \right) \frac{\pi}{M} + \phi_k \right]$$
A.1 Requirements for PR

- The phase restriction
  \[ \phi_k - \phi_{k-1} = (2p + 1) \frac{\pi}{2} \]
  where \( p \) is a constant.

- The selection for the \( h(n) \)
  \[ h(n) = h(L - n - 1); \quad \sum_{i=0}^{L-M-2s-1} h(m+iM)h(m+(i+2)M) = \delta(s) \]
A.2 The parameters in MPEG 1-- Layer 1&2

- \( L=513, \ p=512/2, \ M=8 \)
- The phase
  \[
  \phi_k = (k + \frac{1}{2})(2p + 1) \frac{\pi}{2} = (k + \frac{1}{2})(16 + 1) \frac{\pi}{2}
  \]
- The Analysis & Synthesis Filters
  \[
  h_k(n) = p(n) \cos \left[ (k + \frac{1}{2})(n - \frac{M}{2}) \frac{1}{M} \right]
  \]
  \[
  f_k(n) = p(n) \cos \left[ (k + \frac{1}{2})(n + \frac{M}{2}) \frac{1}{M} \right]
  \]
A.3 Derivation of the Analysis Filter Banks

The Analysis Filters

\[ X_k(m) = h_k(n) * x(n) \bigg|_{n=mM} = \sum_{r=0}^{L-1} h_k(r) x(mM - r) \]

\[ = \sum_{r=0}^{L-1} x(mM - r) p(r) \cos \left( (k + \frac{1}{2})(r - 16) \frac{\pi}{32} \right) \]

Let \( z(r) = x(mM - r) p(r) \)

So

\[ X_k(m) = \sum_{r=0}^{L-1} z(r) \cos \left( (k + \frac{1}{2})(r - 16) \frac{\pi}{32} \right) \]

Let \( r = 64l + s \), where \( l = 0,1,\ldots,7 \); \( s = 0,1,\ldots,63 \)

So

\[ X_k(m) = \sum_{s=0}^{63} \sum_{l=0}^{7} z(64l + s)(-1)^l \cos \left( (k + \frac{1}{2})(s - 16) \frac{\pi}{32} \right) \]
A.4 The Synthesis Filters

\[ y(n) = \sum_{k=0}^{31} v_k(n) = \sum_{k=0}^{31} \sum_{r=-\infty}^{\infty} X_k(r) f(n-32r) = \sum_{k=0}^{31} \sum_{r=-\infty}^{\infty} X_k(r) p(n-32r) \cos \left[ (k + \frac{1}{2})(n-32r + 16) \frac{\pi}{32} \right] \]

Let \( n = 32l + s; s = 0,1, \ldots, 31 \); block number \( l = 0,1, \ldots \).

\[ y(32l + s) = \sum_{r=-\infty}^{\infty} p(32(l-r) + s) \sum_{k=0}^{31} X_k(r) \cos \left[ (k + \frac{1}{2})(32(l-r) + s + 16) \frac{\pi}{32} \right] \]

Let \( q = l - r \)

\[ y(32l + s) = \sum_{q=0}^{15} p(32q + s) \sum_{k=0}^{31} X_k(l-q) \cos \left[ (k + \frac{1}{2})(32q + s + 16) \frac{\pi}{32} \right] = \sum_{q=0}^{15} p(32q + s)U(l;32q + s) \]

where \( U(l;32q + s) = \sum_{k=0}^{31} X_k(l-q) \cos \left[ (k + \frac{1}{2})(32q + s + 16) \frac{\pi}{32} \right] \)

Define \( V(l; j + 64i) = \sum_{k=0}^{31} X_k(l-i) \cos \left[ (k + \frac{1}{2})(j + 16) \frac{\pi}{32} \right] \) where \( j = 0,1, \ldots, 63 \)

So if \( q = \text{even} = 2i \):

\[ U(l;32q + s) = U(l;64i + s) = \sum_{k=0}^{31} X_k(l-q) \cos \left[ (k + \frac{1}{2})(64i + s + 16) \frac{\pi}{32} \right] \]

\[ = (-1)^i \sum_{k=0}^{31} X_k(l-2i) \cos \left[ (k + \frac{1}{2})(s + 16) \frac{\pi}{32} \right] = (-1)^i V(l,128i + s) \]

If \( q = \text{odd} = 2i + 1 \):

\[ U(l;32q + s) = U(l;64i + 32 + s) = \sum_{k=0}^{31} X_k(l-q) \cos \left[ (k + \frac{1}{2})(s + 32 + 16) \frac{\pi}{32} \right] \]

\[ = (-1)^i V(l,128i + 96 + s) \]
Remark B Windows

- Short-Time Spectral Analysis

\[ S_k(e^{j\omega}) = \sum_{n=-\infty}^{\infty} w(k - n)s(n)e^{-j\omega n} \]

- Role of Windows

  - The window, \( w(n) \), determines the portion of the signals that is to be processed by zeroing out the signal outside the region of interest.
  
  - The ideal window frequency response has a very narrow main lobe which increases the resolution, and no side lobes (or frequency leakage).
### B.1 Windowing Functions--Some Examples

<table>
<thead>
<tr>
<th>Name of Window</th>
<th>Time-Domain Sequence, $h(n)$, $0 \leq n \leq M - 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bartlett (triangular)</td>
<td>$\frac{2 \left(n - \frac{M - 1}{2}\right)}{M - 1}$</td>
</tr>
<tr>
<td>Blackman</td>
<td>$0.42 - 0.5 \cos\frac{2\pi n}{M - 1} + 0.08 \cos\frac{4\pi n}{M - 1}$</td>
</tr>
<tr>
<td>Hamming</td>
<td>$0.54 - 0.46 \cos\frac{2\pi n}{M - 1}$</td>
</tr>
<tr>
<td>Hanning</td>
<td>$\frac{1}{2} \left(1 - \cos\frac{2\pi n}{M - 1}\right)$</td>
</tr>
<tr>
<td>Kaiser</td>
<td>$I_0 \left[ \alpha \sqrt{\left(\frac{M - 1}{2}\right)^2 - \left(n - \frac{M - 1}{2}\right)^2} \right]$</td>
</tr>
</tbody>
</table>
| Lanczos                      | $\left\{ \begin{array}{ll} 
\frac{\sin\left[2\pi\left(n - \frac{M - 1}{2}\right)/\left(M - 1\right)\right]}{2\pi\left(n - \frac{M - 1}{2}\right)/\left(M - 1\right)} & L > 0 \\
1, \frac{n - \frac{M - 1}{2} - \alpha \frac{M - 1}{2}}{\alpha} & 0 < \alpha < 1 
\end{array} \right.$ |
| Tukey                        | $\frac{1}{2} \left[1 + \cos\left(\frac{n - (1 + \alpha)(M - 1)/2}{1 - \alpha}(M - 1)/2\right)\right]$ |
B.2 Windowing Functions-- Time Shape

![Graph of windowing functions showing magnitude vs. M-1 for various types including Rectangular, Hamming, Hanning, Blackman, Tukey, Bartlett, and Lanczos.](image)
### B.3 Windowing Functions--Performance

**TABLE 8.2 Important Frequency-Domain Characteristics of Some Window Functions**

<table>
<thead>
<tr>
<th>Type of Window</th>
<th>Approximate Transition Width of Main Lobe</th>
<th>Peak Sidelobe (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rectangular</td>
<td>$4\pi / M$</td>
<td>$-13$</td>
</tr>
<tr>
<td>Bartlett</td>
<td>$8\pi / M$</td>
<td>$-27$</td>
</tr>
<tr>
<td>Hanning</td>
<td>$8\pi / M$</td>
<td>$-32$</td>
</tr>
<tr>
<td>Hamming</td>
<td>$8\pi / M$</td>
<td>$-43$</td>
</tr>
<tr>
<td>Blackman</td>
<td>$12\pi / M$</td>
<td>$-58$</td>
</tr>
</tbody>
</table>
B.4 Windowing Functions -- Low-Pass
FIR Filter Design with Windowing

Rectangular Window (M=61)

Hamming Window (M=61)

Blackman Window (M=61)

Kaiser Window (M=61)
Remark C: QMF/DCT/MDCT

- **QMF**

\[
\begin{align*}
    h_l(n) &= h_u(n) = 0 \quad \text{for } n < 0 \text{ and } n \geq N \\
    h_l(n) &= h_l(N - 1 - n) \quad n = 0, 1, 2, N / 2 - 1 \\
    h_u(n) &= -h_u(N - 1 - n) \quad n = 0, 1, \ldots, N / 2 - 1 \\
    h_u(n) &= (-1)^n h_l(n) \\
    \left| H_l(e^{j\omega}) \right|^2 + \left| H_u(e^{j\omega}) \right|^2 &= 1
\end{align*}
\]
C.2 DCT

- **DCT**
  \[ F_u(u) = \sqrt{\frac{2}{N}} \alpha(u) \sum_{n=0}^{N-1} x(n) \cos\left(\frac{(2n + 1)u\pi}{2N}\right) \quad u = 0, 1, \ldots, N - 1 \]

  \[ \alpha(0) = 1/\sqrt{2}; \quad \alpha(u) = 1 \text{ for } u \neq 0 \]

- **IDCT**
  \[ x(n) = \sqrt{\frac{2}{N}} \sum_{u=0}^{N-1} \alpha(u) F(u) \cos\left(\frac{(2n + 1)u\pi}{2N}\right) \quad n = 0, 1, \ldots, N - 1 \]

- **2D DCT & 2D IDCT**
  \[ F(u, v) = \frac{2}{N} \alpha(u)\alpha(v) \left[ \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} x(n) \cos\left(\frac{(2n + 1)u\pi}{2N}\right) \cos\left(\frac{(2m + 1)v\pi}{2N}\right) \right] \]

  \[ \alpha(0) = \alpha(0) = 1/\sqrt{2}; \quad \alpha(u) = \alpha(v) = 1 \text{ if } u \neq 0, v \neq 0 \]

  \[ x(m, n) = \frac{2}{N} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} \alpha(u)\alpha(v) F(u, v) \cos\left(\frac{(2m + 1)u\pi}{2N}\right) \cos\left(\frac{(2n + 1)v\pi}{2N}\right) \]
C.3 MDCT

**MDCT**

\[
F(u) = \sum_{n=0}^{2N-1} h(n)x(n) \cos \left( \frac{1}{2N}(2u + 1)(2n + 1 + N) \right) \quad u = 0, 1,
\]

\[
h^2(N - 1 - n) = h^2(n) = 2
\]

\[
h^2(N + n) + h^2(2N - 1 - n)^2 = 2 \quad \text{for } 0 \leq n < N.
\]

\[
h(n) = \pm \sqrt{2} \sin \left( (n + \frac{1}{2}) \frac{\pi}{2N} \right)
\]