5. 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
5. 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
5. 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.
### 5. 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><strong>Analysis/Synthesis</strong></td>
<td>32 subbands</td>
<td>32 subbands</td>
<td>Hybrid (subband + MDCT)</td>
</tr>
<tr>
<td><strong>Bit allocation representation</strong></td>
<td>explicit</td>
<td>indexing</td>
<td>Indexing</td>
</tr>
<tr>
<td><strong>Suggested psychacoustic model</strong></td>
<td>Model 1</td>
<td>Model 1</td>
<td>Model 2</td>
</tr>
<tr>
<td><strong>Output bit-rate</strong></td>
<td>32 - 448 Kbps</td>
<td>32 - 384 Kbps</td>
<td>32 - 320 Kbps</td>
</tr>
<tr>
<td><strong>Efficient bit-rate</strong></td>
<td>160 - 224 Kbps</td>
<td>96 - 128 Kbps</td>
<td>64 - 96 Kbps</td>
</tr>
<tr>
<td><strong>Sampling frequency</strong></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><strong>Intensity stereo</strong></td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Quantization</strong></td>
<td>Uniform</td>
<td>Uniform</td>
<td>non-uniform</td>
</tr>
<tr>
<td><strong>Segmentation</strong></td>
<td>Fixed</td>
<td>Fixeded</td>
<td>dynamic</td>
</tr>
<tr>
<td><strong>Entropy coding</strong></td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Slot Size</strong></td>
<td>4 bytes</td>
<td>1 bytes</td>
<td>1 bytes</td>
</tr>
<tr>
<td><strong>Frame Size</strong></td>
<td>384 samples</td>
<td>1152 samples</td>
<td>1152 samples</td>
</tr>
<tr>
<td><strong>Frame-self decodability</strong></td>
<td>Yes</td>
<td>Yes</td>
<td>needs previous frames</td>
</tr>
<tr>
<td><strong>Technical originality</strong></td>
<td>simplified MUSICAM</td>
<td>Refined MUSICAM</td>
<td>Hybrid from MUSICA and ASPEC</td>
</tr>
</tbody>
</table>
5. MPEG1-- Audio Coded Bitstream Syntax

- Frame
  - Header (32)
    - Syn. Word (12)
    - ID (1)
  - Error-Check (16)
    - Layer (2)
    - Protection Bit (1)
    - Bit-Rate Index (4)
    - Sampling Freq. (2)
  - Audio-Data
    - Padding bit (12)
    - Private-bit (1)
    - Mode (2)
    - Mode-Extension (2)
    - Copy right (1)
    - Emphasis (2)
  - Ancillary Data
5. 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.
5. MPEG1-- Audio Coded Bitstream Syntax (c.1)

Audio Data for Layer 1 Single Channel

\[
\text{Scalefactor (i+1)} = \text{Scalefactor (i)} / 1.25992104989487
\]

- **Bit-Allocation (4)**
- **Scalefactor (6)**
- **Samples (?)**

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
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.
5. 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 appy, scalefactor descriptors, a pointer to the end of the main data.
- Main Data
  - The scalefactor and Huffman data
5. MPEGI-- Layer I, II Codec

The Codec Process

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

NCTU/CSIE/DSP LAB
Audio Processing Group

Backgrounds47
5. 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
5. 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.
5. MPEGI-- Layer I (c.2)

Allocation Procedure

- 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
5. MPEGI-- Layer I (c.2)

• 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)

• Packing

Minimum Rate Distortion Curve: Minimum MNR versus number of bits required to encode a layer 1 frame determined for a particular frame.
5. 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
   - dscf1 = scf1 - scf2
   - dscf2 = scf2 - scf3
3. Lookup Table 3-C.4 for 6-18 bits
5. MPEGI-- Layer II (c.1)

- 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
5. 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.
5. MPEGI--Synthesis Subband Filter Flow Chart

Begin

Input 32 New subband Samples $S_i$ $i = 0, ..., 31$

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

For $i = 0$ to $63$ do

$$V_i = \sum_{k=0}^{31} N_{ik} S_k$$

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 $W_i = U_i * D_i$

Calculate 32 samples for $j = 0$ to $31$ do

$$S_j = \sum_{k=0}^{15} W_{j+32i}$$

Output 32 reconstructed PCM Samples

End
5. MPEGI-- Layer III Codec

**Analysis Filter Bank** → **MDCT with dynamic windowing** → **Scaler and Quantizer** → **Huffman Coding** → **Packing**

**FFT** → **Masking Threshold**

**Synthesis Filter Bank** → **MDCT with dynamic windowing** → **Dequantizer & Descaler** → **Huffman Coding** → **Unpacking**

**Coding of Side Information**

**Decoding of Side Information**
5. MPEGI-- Layer III Codec

**The coder process:**

- **Input sample**
  - Layer I Analysis Filter
  - Buffer for 36 subband 0 samples
  - Buffer for 36 subband 31 samples

- **Analysis Filter**
  - Buffer for 36 subband 0 samples
  - Buffer for 36 subband 31 samples

- **Psychoacoustic Model:**
  - Decide block type and allow distortion

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

- **Aliasing**

- **Scaler and quantizer**

- **Huffman coding**
  - Coding of side information
  - Packing
5. MPEGI-- Layer III Codec (c.1)

- **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
5. 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
5. MPEGI-- Layer III Codec (c.3)

- Quantization
  - Nonuniform quantization

\[
is(i) = n \text{ int}(((x_r(i) / \text{quant})^{0.75}) - 0.0946)
\]

- \(x_r(i):\) absolute value of frequency line at index \(i\)
- \(\text{quant}:\) actual quantizer step size
- \(\text{nint}:\) nearest integer function
- 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).
5. MPEGI-- Layer III Codec (c.4)

- 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.
5. MPEGI-- Layer III Codec (c.5)

- **Window switching logic**

![Diagram showing window switching logic](image)
5. MPEGI-- Layer III Codec (c.6)

- Four Types of Window Functions

![Graphs showing different window functions for MPEGI Layer III Codec]
5. MPEGI-- Layer III Codec (c.7)

Illustration of the window switching decision
5. 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.
5. 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 scalefactors are increased in small increments until the quantization noise is below xmin or until the scalefactors 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.
5. MPEGI-- Layer III Codec (c.9)

- global_gain
- 576 spectral lines
- SNR
- Xmin

NCTU/CSIE/DSP LAB
Audio Processing Group
5. 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
5. MPEG1-- Synthesis Subband Filter Flow Chart

Begin

Get Bit Stream, Find Header

Decode Scalefactor

Decode Samples

Inverse Quantize Samples

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

Alias Cancellation

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

Synthesize via Polyphase MDCT

Output Samples

End
5. MPEGI-- 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

\[
 h(i) = \sqrt{8 / 3} \times 0.5 \times \{1 - \cos[2\pi(i) / (N - 1)]\}; \quad 0 \leq i \leq N - 1
\]
5. MPEGI -- Psychoacoustic Model I(c.1)

- **Power Density Spectrum**

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

- **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

\[
L_{sb} = \text{MAX}_{X(k) \text{ in subband } n} \left[ X(k), 20 \log(s_{\text{max}}(n) * 32768) - 10 \right] \text{ dB}
\]

- \( X(k) \) 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.
5. MPEGI -- Psychoacoustic Model I (c.2)

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 non-tonal 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.
5. 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, Layer1: 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
5. MPEGI -- Psychoacoustic Model I (c.4)

- FFT Spectrum
- Magnified Portion
- Power Spectrum
- Decimated Spectrum
5. MPEGI -- Psychoacoustic Model I (c.5)

- **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
    
    \[
    LT_{tm}[z(j),z(i)] = X_{tm}[z(j)] + av_{tm}[z(j)] + vf[z(j),z(i)] dB
    \]
    \[
    LT_{nm}[z(j),z(i)] = X_{nm}[z(j)] + av_{nm}[z(j)] + vf[z(j),z(i)] dB
    \]
    
    $av_{tm} = -1.525 - 0.275*z(j) - 4.5 dB$
    $av_{nm} = -1.525 - 0.175*z(j) - 0.5 dB$
5. MPEGI -- Psychoacoustic Model I (c.6)
5. MPEGI -- Psychoacoustic Model I (c.7)

- \( v_f = 17(dz + 1) - (0.4 \times X(z(j)] + 6) \) dB for \(-3 \leq dz < 0.1\) Bark
- \( v_f = (0.4 \times X[z(j)] + 6) \times dz \) dB for \(-1 \leq dz < 0\) Bark
- \( v_f = -17 \) dB for \(0 \leq dz < 0\) Bark
- \( v_f = -(dz - 1) \times (17 - 0.15X[z(j)]) - 17 \) dB for \(1 \leq dz < 8\) Bark

 поверхностного взаимодействия

### 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
5. 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
5. MPEGI-- Psychoacoustic Model I (c.8)

- **Determine the Minimum Masking Threshold**

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

- Determine the minimum masking level in each subband.

- **Calculate the signal-to-mask ratio**

\[
\text{SMR}_{sb}(n) = L_{sb}(n) - LT_{\text{min}}(n) \text{ dB}
\]
5. 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

\[
\begin{align*}
r_w &= 2.0r_w(t-1) - r_w(t-2) \\
\hat{f}_w &= 2.0\hat{f}_w(t-1) - \hat{f}_w(t-2)
\end{align*}
\]
5. **MPEGI -- Psychoacoustic Model II (c.1)**

- Calculate Euclidian distance between predict value $c_w$ for each line.

$$c_w = \frac{\left( (r_w \cos f_w - r_{w'} \cos f_{w'})^2 + (r_w \sin f_w - r_{w'} \sin f_{w'})^2 \right)^{0.5}}{r_w + \text{abs}(r_w)}$$

**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, $w_{\text{low}_b}$
    - Highest frequency line, $w_{\text{high}_b}$
    - The median Bark value, $b_{\text{val}_b}$
    - A lower limit for the SNR in the partition that controls stereo unmasking effects, $b_{\text{minval}_b}$
    - The value for tonal masking noise for that partition, $b_{\text{TMN}_b}$
5. MPEGI -- Psychoacoustic Model II (c.2)

- Critical band rate vs. partition number
- Conversion from spectral line to partition number
- Fourier Power Spectrum
- Unpredictability
- Partition energy & Partitioned Unpredictability
- Convolved partitioned energy & unpredictability
5. MPEGI -- Psychoacoustic Model II (c.3)

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

$$cb_b = \frac{ct_b}{ecb_b}$$

$$en_b = ecb_b * rnorm_b$$

$$rnorm_b = \frac{1}{\sum_{bb=0}^{b_{\text{max}}} \text{sprdngf}(bval_{bb}, bval_b)}$$

- Convert chaos measure $cb_b$ to tonality $tb_b$
  - $0.05 < cb_b < 0.5 \Rightarrow 0 < tb_b < 1$
  - $tb_b = -0.299 - 0.43 \log(e, cb_b)$

$$ecb_b = \sum_{bb=1}^{b_{\text{max}}} e_{bb} \text{sprdngf}(bval_{bb}, bval_b)$$

$$ct_b = \sum_{bb=1}^{b_{\text{max}}} c_{bb} \text{sprdngf}(bval_{bb}, bval_b)$$
5. MPEGI -- Psychoacoustic Model II (c.4)

- According to the tonality \( tb_b \) to calculate spread threshold \( SNR_b \) (TMN, NMT)

\[
SNR_b = \max(\minval_b, tb_b \times TMN_b + (1-tb_b) \times 5.5)
\]

- Calculate the power ratio of \( SNR_b \)

\[
b_c_b = 10^{-\frac{SNR_b}{10}}
\]

\[
nb_b = \text{en}_b \times b_c_b
\]

- Renormalization ( \( nb_w \))

\[
nb_w = \frac{nb_b}{\text{weight}_b - w - \text{low}_b + 1}
\]

- Include the absolute threshold for final Energy Threshold of audibility

\[
\text{thr}_w = \max(nb_w, \text{absthr}_w)
\]
5. MPEGI-- Psychoacoustic Model II (c.5)

- **Pre-Echo Control (C-30)**
  \[
  \text{thr}(b) = \max(q\text{thr}(b), n\text{bb}(b), n\text{bb}_I(b), n\text{bb}_{II}(b))
  \]
  \[
  n\text{bb}_I(b) = 2 \times n\text{bb}(b) \text{ from the last block}
  \]
  \[
  n\text{bb}_{II}(b) = 16 \times n\text{bb}(b) \text{ from the block before the last block}
  \]
- **Calculate SMR in each scalefactor band**

\[
\text{npart}_n = \sum_{\text{w} \text{lower}}^{\text{w} \text{higher}} \text{thr}_w \text{ for narrow scalefactor band}
\]

\[
\text{npart}_n = \min(\text{thr}_\text{low}_n, \ldots, \text{thr}_\text{high}_n) \times (\text{w} \text{high}_n - \text{w} \text{low}_n + 1)
\text{ for wide scalefactor band}
\]

\[
\text{epart}_n = \sum_{\text{w} \text{lower}}^{\text{w} \text{higher}} r_w^2
\]

\[
\text{SMR}_n = 10 \log_{10}\left(\frac{\text{epart}_n}{\text{npart}_n}\right)
\]
5. MPEGI -- Psychoacoustic Model II (c.6)

- **Masking Threshold**: The masking threshold is shown in the first graph, with values plotted against partition number.

- **Masking Threshold mapped to Fourier spectral domain**: The second graph displays the masking threshold in the Fourier spectral domain, with values plotted against spectral line number.

- **Signal-to-masker ratio**: The third graph illustrates the signal-to-masker ratio, with values plotted against subband number.

- **Spreading Function**: The spreading function is depicted in the fourth graph, with values plotted against partition number.

- **Tone Masking Function**: The fifth graph shows the tone masking function, with values plotted against partition number.

- **The minval function**: The sixth graph presents the minval function, with values plotted against critical band rate.
5. 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.
5. 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.
5. 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.
5. MPEGI-- Intensity Stereo and MS Stereo

- **MS_Stereo Switching**
  - Switch is on when
    \[
    \sum_{i=0}^{511} \left[ r_{li}^2 + r_{ri}^2 \right] < 0.8 \sum_{i=0}^{511} \left[ r_{li}^2 + r_{ri}^2 \right]
    \]
  - The values \( r_{li} \) and \( r_{ri} \) correspond to the energies of the FFT line spectrum of the left and right channel calculated within the psychoacoustic model.

- **MS_Matrix**
  \[
  M_i = \frac{R_i + L_i}{\sqrt{2}}; \quad S_i = \frac{R_i - L_i}{\sqrt{2}}
  \]
### 5. MPEGI-- 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></td>
<td>subband + overlap transform with dynamic windowing.</td>
</tr>
<tr>
<td>Scaling factor</td>
<td>one scalefactor for 12 consecutive output each band</td>
<td>1 to 3 for 36 consecutive output each band</td>
<td>by clustering of frequency into bands, 1 per band</td>
</tr>
<tr>
<td>Bit allocation</td>
<td>found by iteration based on MNR (bit allocation )</td>
<td></td>
<td>base on SMR (noise allocation )</td>
</tr>
</tbody>
</table>
6. 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 MPEG-1 Audio

The Modulated Filter Banks

- The Analysis Filters $h_k(n)$

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

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

- The Synthesis Filters $f_k(n)$

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

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

- \( L=513, \ p=512/2M=8 \)
- **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

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

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

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

So

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

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

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

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

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

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

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 \( j = 0, 1, ......., 63 \)

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>$2 \left(1 - \frac{n}{M-1}\right)$</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>
<tr>
<td></td>
<td>$I_0 \left[ \alpha \left(\frac{M-1}{2}\right) \right]$</td>
</tr>
<tr>
<td>Lanczos</td>
<td>$\left{ \frac{\sin \left[ 2\pi \left(n - \frac{M-1}{2}\right) / (M-1) \right]}{2\pi \left(n - \frac{M-1}{2}\right) / (M-1)} \right}^L$</td>
</tr>
<tr>
<td></td>
<td>$L &gt; 0$</td>
</tr>
<tr>
<td>Tukey</td>
<td>$\frac{1}{2} \left[ 1 + \cos \left( \frac{n - (1 + \alpha)(M-1)/2}{(1 - \alpha)(M-1)/2} \pi \right) \right]$</td>
</tr>
<tr>
<td></td>
<td>$\alpha(M - 1)/2 \leq \left</td>
</tr>
</tbody>
</table>
B.2 Windowing Functions-- Time Shape
### 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**

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

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

**IDCT**

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

**2D DCT & 2D IDCT**

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

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

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

**MDCT**

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

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

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

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