

マルチメディア配信システム

- No.7 音響圧縮技術 -

渡辺 裕

Multimedia Distribution System

- No.7 Audio Compression Technology -

Hiroshi Watanabe

音声信号方式

- アナログ電話
 - 音声帯域4kHz -> もしデジタル化すれば8kHz sampling
- FMラジオ
 - 周波数帯域15kHz -> もしデジタル化すれば30kHz sampling
- オーディオCD
 - 周波数帯域22.05kHz, 16bit -> 44.1 kHz sampling
- DVD オーディオ
 - 周波数帯域192kHz, 24bit (2チャンネル時のみ), 96kHz, 24bitマルチチャンネルステレオ

Audio Signal Formats

- Analog telephone
 - Speech bandwidth 4kHz -> If digitized, 8kHz sampling
- FM radio
 - Bandwidth 15kHz -> If digitized, 30kHz sampling
- Audio CD
 - Bandwidth 22.05kHz, 16bit -> 44.1 kHz sampling
- DVD audio
 - Bandwidth 192kHz, 24bit (2 channel), 96kHz, 24bit multi-channel stereo

音響符号化

- 低ビットレート(<32kbps)
 - LPC(線形予測符号化)ボコーダ
 - マルチパス線形符号化
 - CELP(符号励振線形予測)
- 中ビットレート(32kbps-64kbps)
 - ADPCM(適応差分パルス符号変調)
 - サブバンドADPCM
- 高ビットレート(>64kbps)
 - サブバンド符号化
 - ATC(適応変換符号化) + 心理聴覚重みづけビット割り当て

Audio Coding

- Low bitrate (<32kbps)
 - LPC(Linear Predictive Coding) Vocoder
 - Multipass Linear Predictive Coding
 - CELP(Code Exited Linear Prediction)
- Middle bitrate (32kbps-64kbps)
 - ADPCM(Adaptive Differential Pulse Coded Modulation)
 - Subband ADPCM
- High bitrate (>64kbps)
 - Subband Coding
 - MDCT(Modified DCT)+Psycho-acoustic weighted bit allocation

符号化標準

■ 規格とアルゴリズム

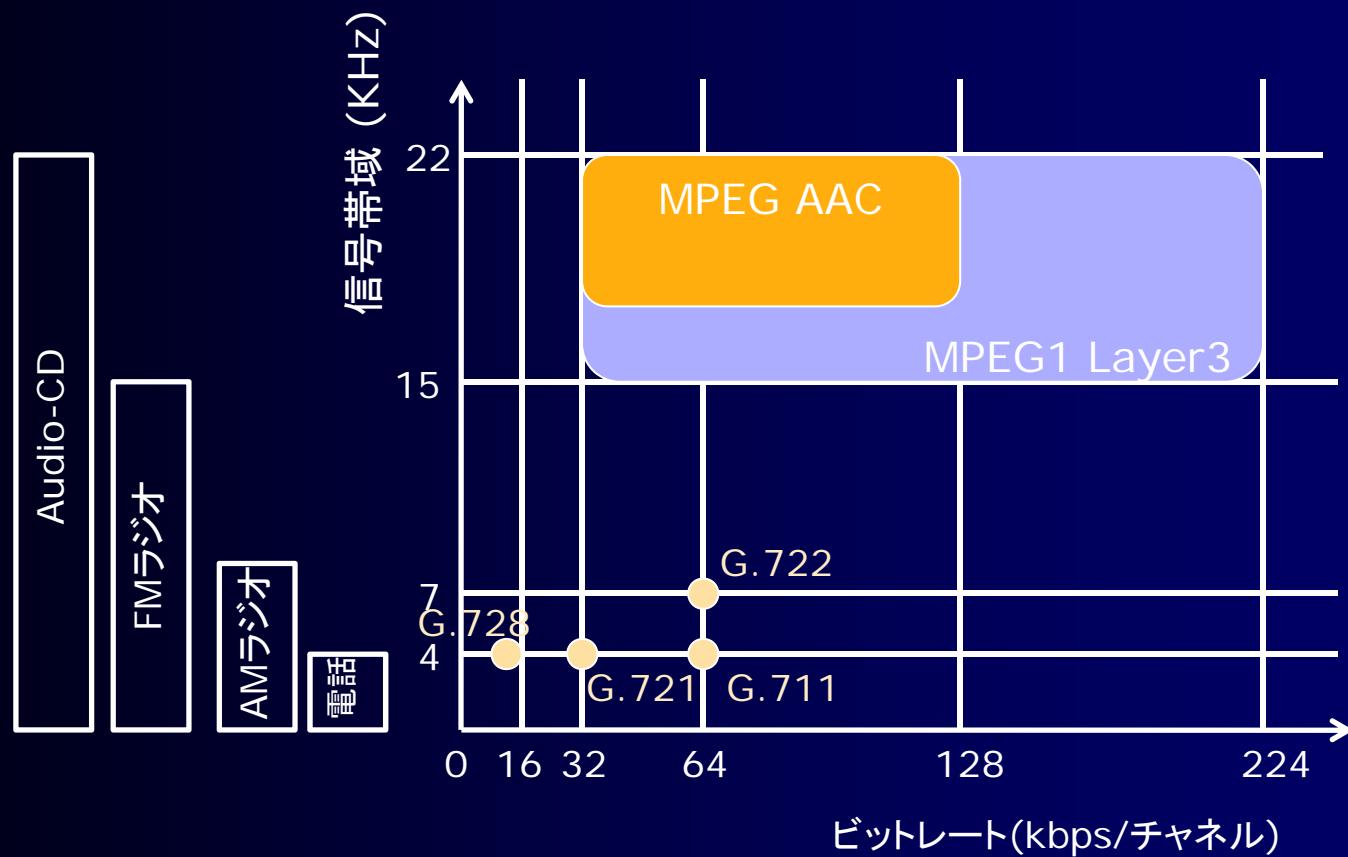
時期	名称	ビットレート	アルゴリズム
1972	CCITT G.711	64 kbps	非線形量子化
1984	CCITT G.721	32 kbps	ADPCM
1986	CCITT G.722	64 kbps	サブバンドADPCM
1991	CCITT G.728	16 kbps	低遅延CELP符号化
1993	ISO MPEG	32k-224kbps	サブバンド+MDCT
1997	ISO MPEG AAC	32k-128kbps	MDCT+ステレオ処理

Standard

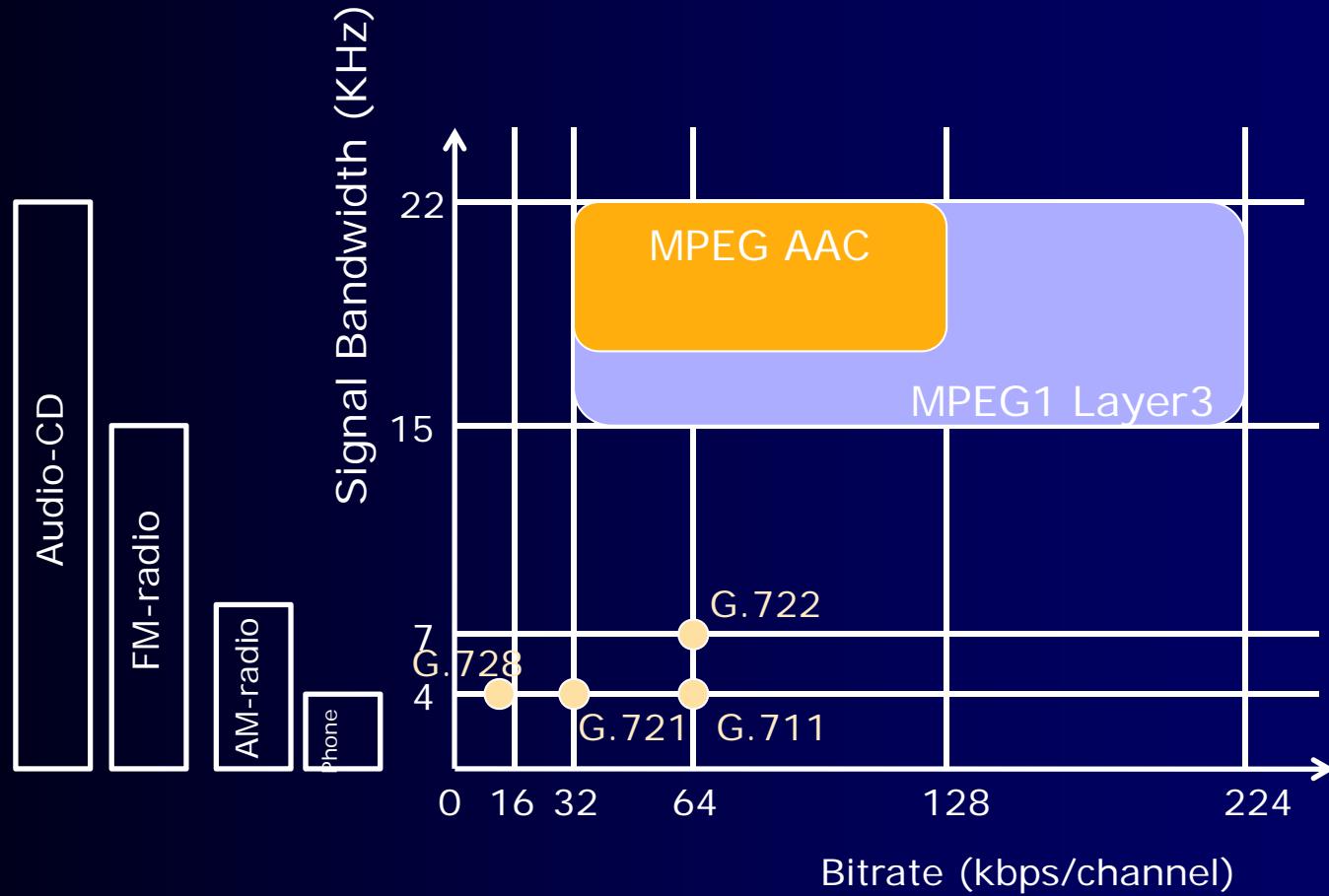
■ Standard and Algorithm

Year	Name	Bitrate	Algorithm
1972	CCITT G.711	64 kbps	Non-linear Quantization
1984	CCITT G.721	32 kbps	ADPCM
1986	CCITT G.722	64 kbps	Subband-ADPCM
1991	CCITT G.728	16 kbps	Low-delay CELP
1993	ISO MPEG	32k-224kbps	Subband+MDCT
1997	ISO MPEG AAC	32k-128kbps	MDCT+Stereo

対象信号

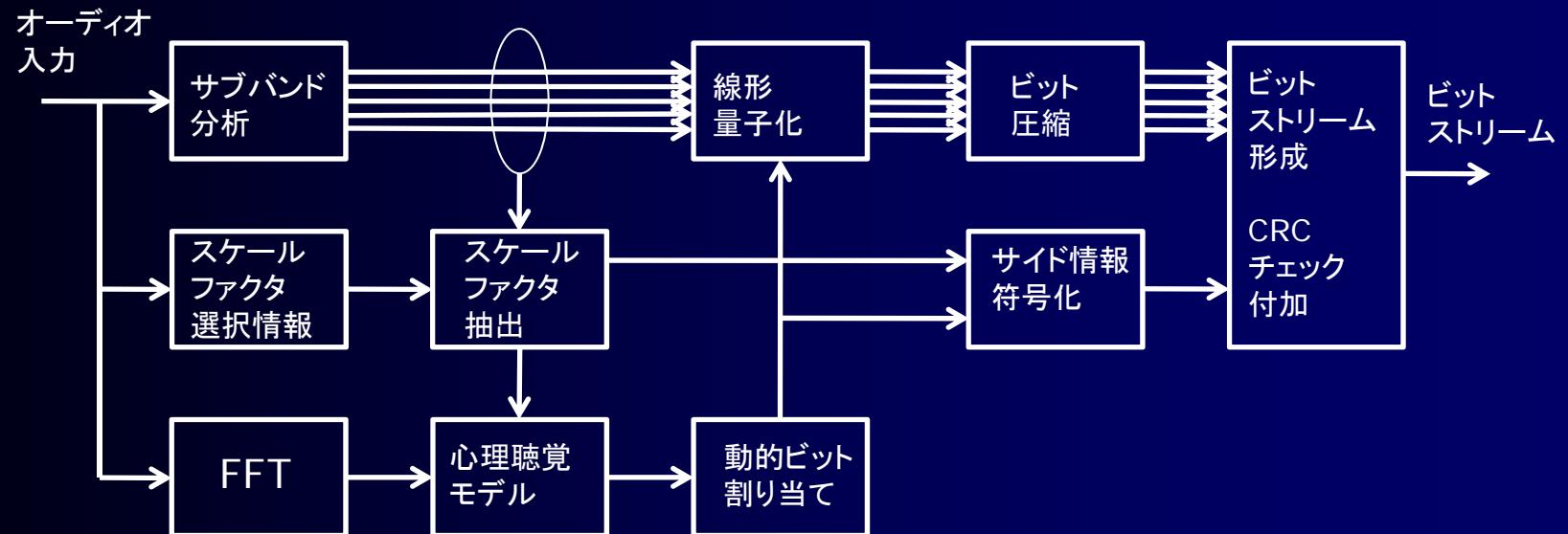


Target Signal



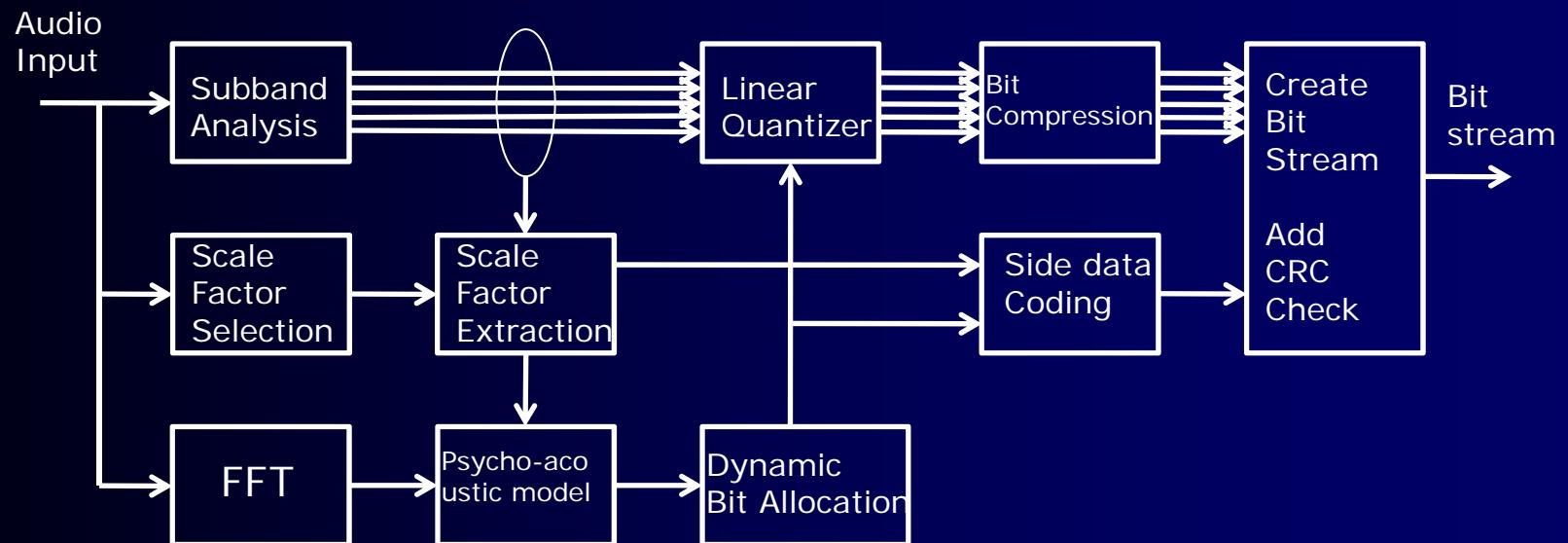
MPEG Audio 基本構成

■ MPEG-1 Layer I, II 符号化器の構成



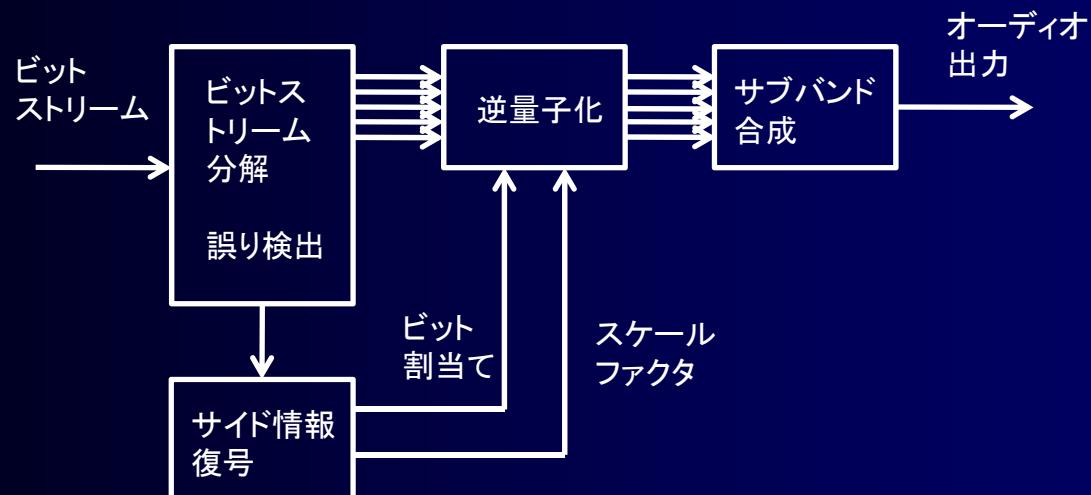
MPEG Audio Basic Scheme

■ MPEG-1 Layer I, II Encoder Structure



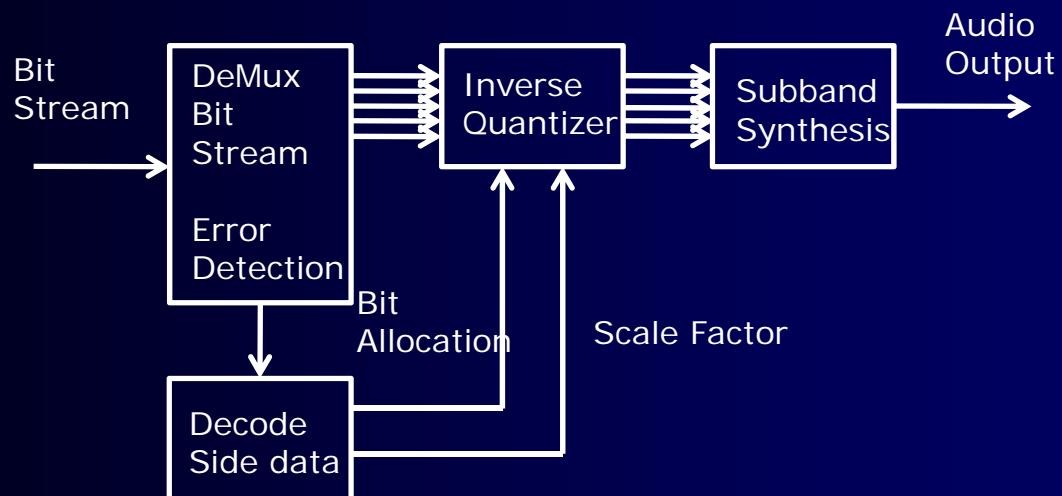
MPEG Audio 基本構成 (2)

■ MPEG-1 Layer I, II 復号器の構成



MPEG Audio Basic Scheme (2)

■ MPEG-1 Layer I, II Decoder Structure



MPEG Audio 基本構成 (3)

■ 符号化器における主要処理

- 16bit PCM 入力
- ポリフェーズフィルタバンク(512tap)による32帯域分割
- 量子化誤差マスキングレベル計算
- 心理聴覚モデルによるビット割当て
- アンシラリデータの組込み

MPEG Audio Basic Scheme (3)

- Main Processes at an encoder
 - Input: 16bit PCM
 - Subband Analysis (32 bands) by Polyphase Filter Bank (512tap)
 - Calculation of masking level for quantization error
 - Bit allocation based on psycho-acoustic model
 - Embed ancillary data

サブバンド分析

- 入力信号512点のデータ $x(i)$ にLPFインパルス応答 $c(i)$ をかける

$$y(i) = c(i)x(i) \quad (i = 0, 1, \dots, 511)$$

- 周期加算信号の計算

$$z(i) = \sum_{j=0}^7 y(64j + i)$$

- サブバンド信号の導出

$$s(i) = \sum_{i=0}^{63} z(i) \cos \frac{(2i+1)(k-16)\pi}{64}$$

Subband Analysis

- Multiply LPF impulse response $c(i)$ to input 512 point data $x(i)$

$$y(i) = c(i)x(i) \quad (i = 0, 1, \dots, 511)$$

- Calculation of periodic added signal

$$z(i) = \sum_{j=0}^7 y(64j + i)$$

- Derive subband signal

$$s(i) = \sum_{i=0}^{63} z(i) \cos \frac{(2i+1)(k-16)\pi}{64}$$

スケールファクタ

■ 各サブバンドのダイナミックレンジを揃える

- レイヤIでは各バンド12サンプル毎, 原信号では384サンプル
(12×32 バンド)
- レイヤIIでは1152サンプル毎, レイヤIの3倍のデータ量となるため, トリプレッツ(3つ組のパターン)を構成
- 隣接する二つのスケールファクタの差分を計算
- 差分の連続パターンが1bitのスケールファクタ選択情報と1-3bitの伝送パターンで符号化

Scale Factor

- Align dynamic range of each subband
 - レイヤIでは各バンド12サンプル毎, 原信号では384サンプル(12*32バンド)
 - レイヤIIでは1152サンプル毎, レイヤIの3倍のデータ量となるため, トリプレッツ(3つ組のパターン)を構成
 - 隣接する二つのスケールファクタの差分を計算
 - 差分の連続パターンが1bitのスケールファクタ選択情報と1-3bitの伝送パターンで符号化

心理聴覚分析（1）

- FFT分析
 - ブロック長はレイヤIで512, レイヤIIは1024
 - ポリフェーズフィルタバンクの遅延量に中心を設定
- 音圧計算
 - FFT 出力とFFT分析の1ブロック内で最大となるスケールファクタのうち, 大きい方の値を音圧と定義
- 純音成分と非順音成分の選別
 - 両隣のスペクトルラインより大きく, 左右 j サンプル離れたすべてのスペクトルラインより, 7dB大きなスペクトルラインを純音成分とする

Psycho-acoustic Analysis (1)

- FFT Analysis
 - Block length are 512 for Layer I, 1024 for Layer II
 - Set center to the delay of polyphase filter bank
- Calculation of acoustic pressure
 - Define acoustic pressure as the larger one either FFT output or the maximum scale factor in one block of FFT analysis
- Selection of pure tone and non-pure tone
 - Set pure tone that is 7 dB larger than spectrum lines far from j-sample as well as larger than the neighborhood spectrum lines

心理聴覚分析 (2)

■ 純音決定処理

- $X(k) > X(k-1)$ および $X(k) \geq X(k+1)$ を同時に満足するスペクトルラインを検出
- $X(k) - X(k+j) > 7 \text{ dB}$ を満足するスペクトルラインを選択
- 幅 j の値はレイヤI, レイヤIIそれぞれに対し, 周波数帯域ごとに与えられる
- 純音成分に隣接するスペクトルラインの音圧を, 純音成分の音圧に加算
 - $X(k) \leftarrow X(k) + X(k-1) + X(k+1)$

Psycho-acoustic Analysis (2)

- Decision process for pure tone
 - Detect spectrum line that satisfies $X(k) > X(k-1)$ and $X(k) \geq X(k+1)$ simultaneously
 - Select spectrum line that satisfies $X(k) - X(k+j) > 7 \text{ dB}$
 - Width value j is given depends on each subband for Layer I and II
 - Add acoustic pressure of neighborhood spectrum to the one of pure tone
 - $X(k) \leftarrow X(k) + X(k-1) + X(k+1)$

心理聴覚分析 (3)

- 幅 j で与えられる範囲内に存在する $X(k)$ 以外のスペクトルラインをすべてゼロに再設定
 - 以上の処理で得られた純音以外のスペクトルラインを、各帯域内ですべて加算し、非純音成分の音圧として定義
 - 非純音成分の音圧を、各帯域の中心に最も近い位置に再配置
- 純音成分と非純音成分の間引き
- 求められた純音成分と非純音成分のうち、しきい値未満のものは棄却。さらに、0.5バーク以内の距離に複数の純音成分がある場合には、最大の純音成分以外を棄却

Psycho-acoustic Analysis (3)

- Set values of all spectrum lines to 0 within the width j except for $X(k)$
 - Add all non-pure tone spectrum lines obtained the above process within each subband, and define it as a non-pure tone pressure
 - Replace non-pure tone pressure to the center of each subband
-
- Sub-sampling of pure tone and non-pure tone
 - For the obtained pure and non-pure tone, reject the one less than a threshold. Further, only the maximum pure tone is kept if there are several pure tones in the distance of 0.5 bark

バーク

- 臨界帯域を示す単位で、人間の聴覚特性性能に近似した非線形周波数写像のスケール
 - 1 バークは低域周波数では狭い周波数帯域に相当し、高域周波数ではより広い周波数帯域に相当する
 - 尺度の範囲は1から24で、聴覚の24の臨界帯域 (critical band) に対応している。帯域の境界は順に、20、100、200、300、400、510、630、770、920、1080、1270、1480、1720、2000、2320、2700、3150、3700、4400、5300、6400、7700、9500、12000、15500(Hz)である。

Bark

- A unit to specify critical bandwidth. It is a scale of nonlinear frequency mapping that approximates human acoustic characteristics
 - 1 bark corresponds narrow frequency band at low frequency range, and wider frequency band at high frequency range
 - Range of measure is from 1 to 24, corresponding 24 critical bands for psycho-acoustics. Borders are 20, 100, 200, 300, 400, 510, 630, 770, 920, 1080, 1270, 1480, 1720, 2000, 2320, 2700, 3150, 3700, 4400, 5300, 6400, 7700, 9500, 12000, 15500 (Hz)

心理聴覚分析 (4)

■ 個別マスキングしきい値の計算

- 高域周波数ほど大きな間引きを実施
- 新しいサンプル点に、純音と非純音を再配置
- 純音および非純音マスキングしきい値の導出
 - $F(X(z(j)), av(z(j)), vf(z(j), z(i)))$

$X(z(j))$: j番目のマスカ一音圧
 $av(z(j))$: マスキング指数
 $vf(z(j), z(i))$: マスキング関数

■ 全体マスキングレベルの決定

- 低音側3バーク、高音側8バーク以内に含まれる純音成分と非純音成分によるマスキングしきい値を加算

Psycho-Acoustic Analysis (4)

- Calculation of threshold of individual masking value
 - Large sub-sampling applied to high frequency
 - Replace pure/non-pure tone to new samples
 - Derive thresholds for pure/non-pure tone masking
 - $F(X(z(j)), av(z(j)), vf(z(j), z(i)))$
- X(z(j)): j-th masker pressure
- av(z(j)): masking indicator
- vf(z(j), z(i)): masking function
- Decide overall masking level
 - Add masking threshold of pure/non-pure tone contained in 3 bark for low frequency side and 8 bark for high frequency side

心理聴覚分析 (5)

- 最小マスキングレベルの決定
 - 各サブバンドに対応する周波数軸上の範囲内で最小の全体マスキングしきい値を抽出
- 信号対マスク比の計算
 - 信号対マスク比(SMR)を各サブバンドの音圧と最小マスキングレベルの比として与える

Psycho-Acoustic Analysis (5)

- Decide the minimum masking level
 - Extract overall masking threshold in each subband
- Calculation of signal to mask ratio
 - Signal to mask ratio (SMR) is given by the ratio of acoustic pressure and the minimum masking level

ビット割当て

■ SMRに基づき各サブバンドへのビット割当て決定

- 総ビット数からヘッダ、CRCチェック、アンシラリデータを差し引く
- 最小マスク対雑音比(MNR)を有するサブバンドを探索
- 該当サブバンドの量子化ステップを1段小さくする
- 新しい量子化ステップに対応する信号対雑音比(SMR)を表から選択し新たなMNRを求める
- 割当てビット数から量子化ビット数を差し引き、再度ビット割当てを繰り返す

Bit Allocation

- Decide bit allocation to each subband based on SMR
 - Subtract bits of header, CRC check and ancillary data from total bits
 - Search subband having the minimum mask to noise ratio (MNR)
 - Decrease quantization step of the above subband
 - Obtain new MNR by selecting SMR that corresponds to the new quantization step from the table
 - Subtract quantization bit from allocated bit, and perform the bit allocation again

量子化処理

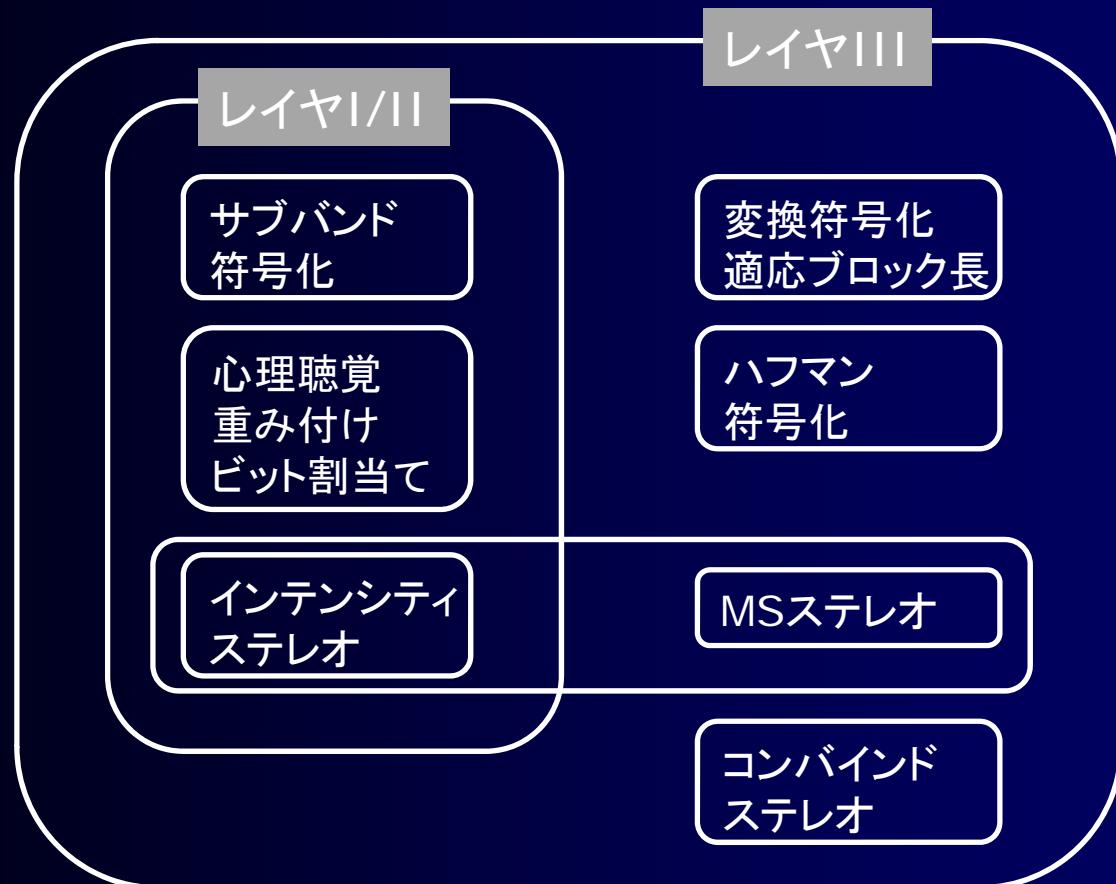
- 各サブバンドをスケールファクタで正規化 $X(n)$
 - サブバンドごとに割当てられたビット数に対応した値 $A(n)$,
 $B(n)$ により $A(n)X(n) + B(n)$ に従って量子化
 - 上位Nビットを取り最上位ビットを反転
- レイヤIIでは効率化のために、連続する3サンプル x, y, z に対して
量子化ステップ数が3,5,9の場合の符号化サンプル v_3, v_5, v_9 を
与える
 - $v_3 = 9z + 3y + x$
 - $v_5 = 25z + 5y + x$
 - $v_9 = 81z + 9y + x$

Quantization Process

- Normalize each subband by scale factor
 - To each subband, $A(n)$, $B(n)$ that correspond to allocated bits, quantize using $A(n)X(n)+B(n)$
 - Take upper N bit, MSB is reversed
- In Layer II, for efficiency, give coding sample v_3 , v_5 , v_9 that are the case when quantization steps are 3,5,9 for consecutive 3 samples x,y,z
 - $v_3 = 9z + 3y + x$
 - $v_5 = 25z + 5y + x$
 - $v_9 = 81z + 9y + x$

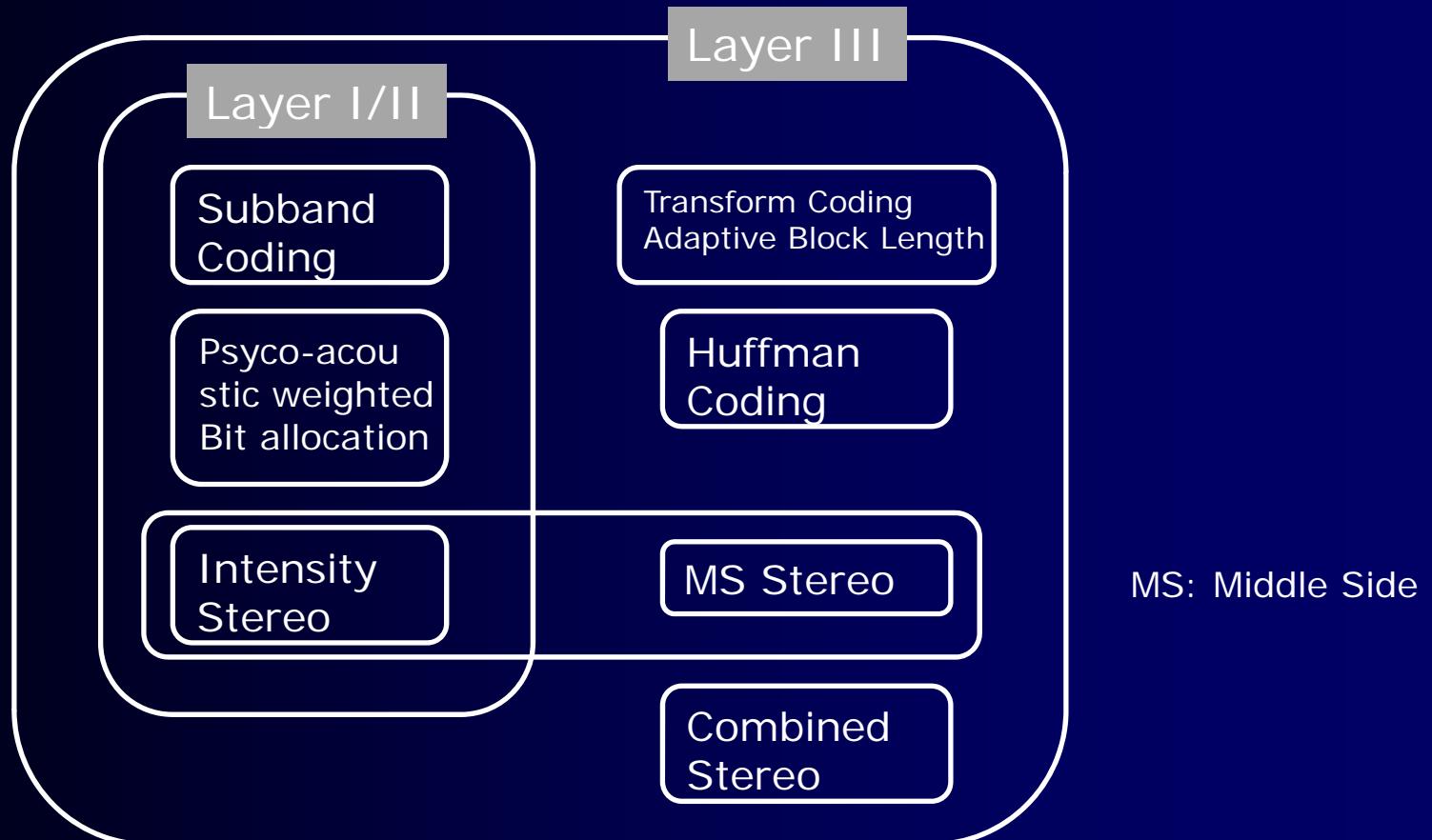
符号化技術 (2)

■ レイヤ間の要素技術の関係



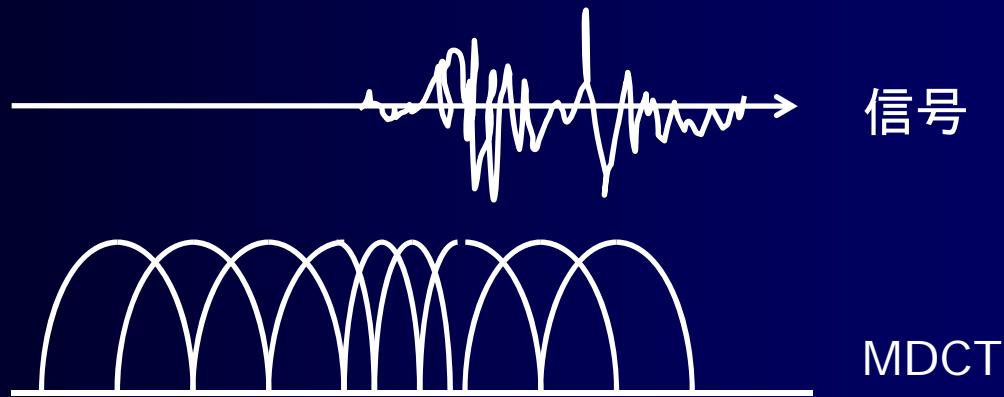
Compression Technology (2)

■ Relation of technology elements between Layers



ハイブリッドフィルタバンク

- レイヤIIIではハイブリッドフィルタバンクを採用
 - 32PFBに加えてMDCT(LOT)
 - $21 * 18 = 576$ 帯域に分割
 - MDCTはプリエローに対応，窓長可変(18, 36)



Hybrid Filter Bank

- Hybrid filter bank is adopted in Layer III
 - 32PFB with MDCT(LOT)
 - $21 * 18 = 576$ subbands
 - MDCT can respond to pre-echo, variable window(18, 36)

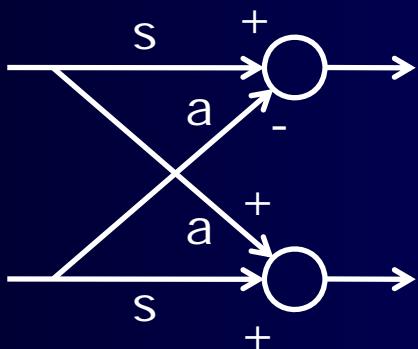


折り返し歪除去

- MDCT出力に対して帯域境界に近い8サンプル、8段にわたって処理

$$s = \frac{1}{\sqrt{1+c^2}}$$

$$a = \frac{c}{\sqrt{1+c^2}}$$

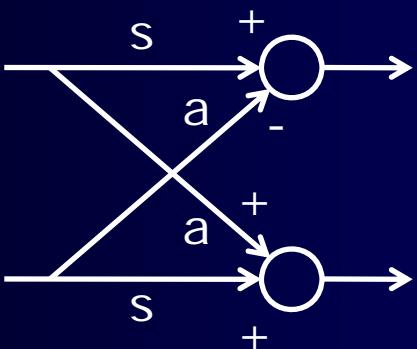


Aliasing Noise Removal

- Process butterfly operation to MDCT output, boundary 8 sample, 8 steps

$$s = \frac{1}{\sqrt{1+c^2}}$$

$$a = \frac{c}{\sqrt{1+c^2}}$$



ステレオ符号化

■ ジョイントステレオ符号化

- インテンシティステレオ

- 左右両チャネルの和信号と比率を2チャネルとして処理
- 低域周波数は左右独立に符号化

- MS(middle side)ステレオ

- 左右両チャネルの和信号と差信号を2チャネルとして処理
- 最も簡単な2点直交変換
- 両チャネルの相関が大きい時には、和信号と差信号の情報差が大きくなり、偏在によるデータ圧縮効果が得られる

- コンバインドステレオ

- FFTにより得られた左右のスペクトル差が、しきい値より多い場合にインテンシティステレオとMSステレオを切り替える

Stereo Coding

- Joint Stereo Coding
 - Intensity Stereo
 - Sum of two channel and ratio is processed
 - Low frequency is independently coded
 - MS(middle side) Stereo
 - Sum and difference of two channel is processed
 - Simplest two point orthogonal transform
 - Compression efficiency can be obtained when correlation of two channel is large because of bias effect
 - Combined Stereo
 - Switch Intensity and MS stereo when spectrum ratio between two channel obtained by FFT is large

AAC符号化での新技術

■ 時間領域量子化雑音整形

- 量子化雑音を信号振幅の大きさにより整形
- 符号化時にMDCT係数の一部を線形予測分析
- 線形予測係数を用いたトランスバーサルフィルタ処理
- 復号時に巡回型フィルタ処理
- 量子化雑音は信号振幅の大きな部分に集中
- しきい値を超えた場合にのみ適用

■ 予測

- 直前ブロックの量子化データにより、現ブロックの時間領域量子化雑音整形出力を予測、予測誤差のみを符号化
- 各スペクトルに対して2次のラティス型バックワード適応予測

New Technology at AAC

- Temporal Noise Shaping
 - Quantization noise shaping based on signal amplitude
 - Linear prediction for MDCT coefficients at encoder
 - Transversal filtering using linear prediction
 - Feedback filtering employed at decoder
 - Quantization noise concentrates to large amplitude
 - Threshold control
- Prediction
 - Current TNS output is predicted by the previous one, and only the difference is coded
 - 2nd order lattice backward adaptive prediction

他の符号化技術

■ MPEG-4で導入された方式

- Twin VQ (transform domain weighted interleaved vector quantization)
 - MDCTで周波数領域信号に変換しLPC分析で正規化
 - バーク尺度領域に変換して求めた包絡により正規化
 - 電力で正規化後重みづけインターリーブベクトル量子化
- CELP符号化
 - LPC分析後量子化係数を得てLPC合成フィルタを音源コードブックの出力で駆動
 - 長期予測係数を求め、LPC合成フィルタと入力信号の差が聴感的に最小化されるよう音源コードブックの出力を探索
 - LPC係数の量子化はベクトル量子化、音源はマルチパルス

Other Technologies

- Coding method introduced in MPEG-4
 - Twin VQ (transform domain weighted interleave vector quantization)
 - Normalize by LPC analysis to frequency domain signal obtained by MDCT
 - Normalize by envelope converted to Bark domain
 - Weighted interleave VQ after normalize by power
 - CELP Coding
 - Excite LPC analysis filter by source codebook output after LPC analysis
 - Long term prediction coefficient is obtained, search codebook to minimize input and LPC analysis output
 - LPC coefficient is vector quantized, source is multi-pulse

他の符号化技術 (2)

– ハーモニックベクトル励振符号化

- LPC分析後量子化された係数をLPC逆フィルタに供給し入力信号の残差を求める
- 予測残差を周波数領域でピッチ分析および包絡線分析
- 有声音部では包絡線を重みづけベクトル量子化
- 無声音部では励振ベクトルの閉ループ探索を実行

Other Technologies (2)

- Harmonic Vector Excited Coding (HEVC)
 - Calculate difference between input and inverse LPC filter after performing LPC analysis and quantization
 - Prediction error is fed into pitch analysis and envelope analysis in frequency domain
 - In voice part, envelope is weighted vector quantized
 - Otherwise, excited vector's closed loop search is performed