MP3 (aka MPEG1-LayerIII) の要素技術

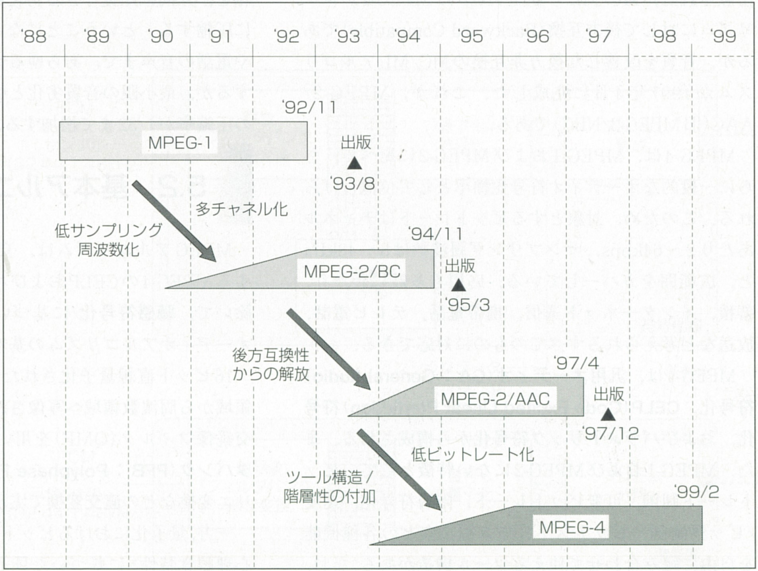
2024.3-

あらすじ

1. MP3概要
   - MPEG/Audioの歴史
   - コーデック構造
2. ハイブリッドフィルタバンク
   - フィルタバンク
   - MDCT (Modified Discrete Cosine Transform)
   - エイリアス削減バタフライ演算
3. 量子化
4. 符号化
5. 聴覚心理モデルII

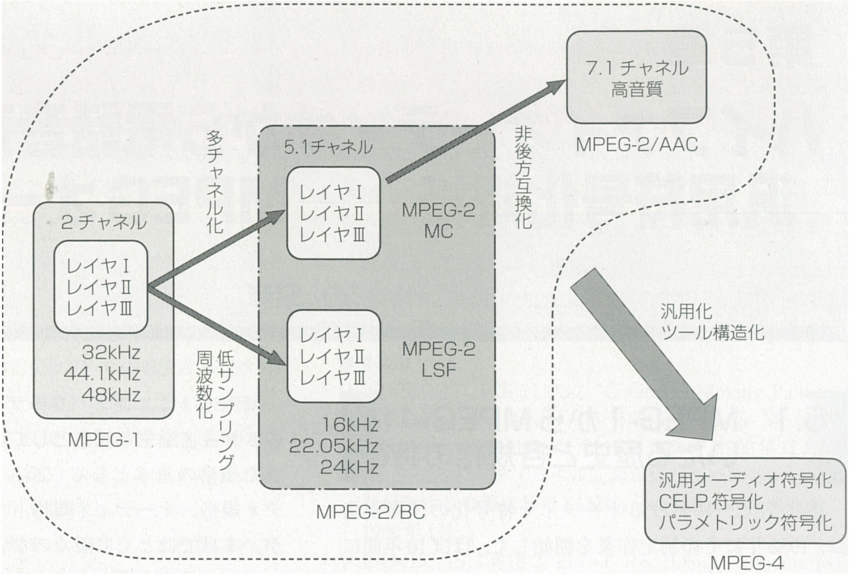
1. MP3 概要
   - MPEG/Audio の歴史
   - コーデック構造
2. ハイブリッドフィルタバンク
   - フィルタバンク
   - MDCT (Modified Discrete Cosine Transform)
   - エイリアス削減バタフライ演算
3. 量子化
4. 符号化
5. 聴覚心理モデル II

MPEG/Audioの歴史



MPEG/Audio の系譜．[1] より引用

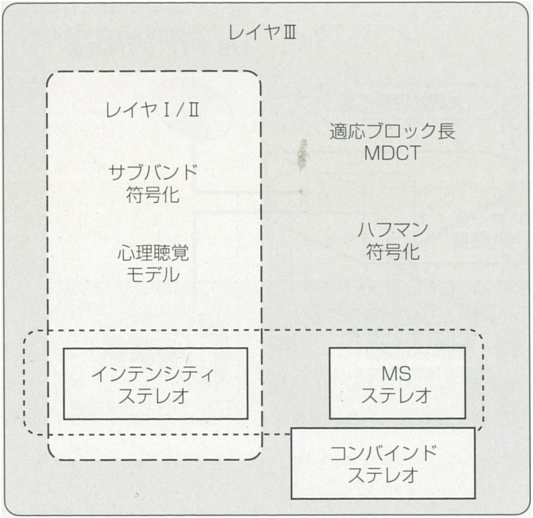
MPEG/Audioの歴史



MPEG/Audio の関係．[1] より引用

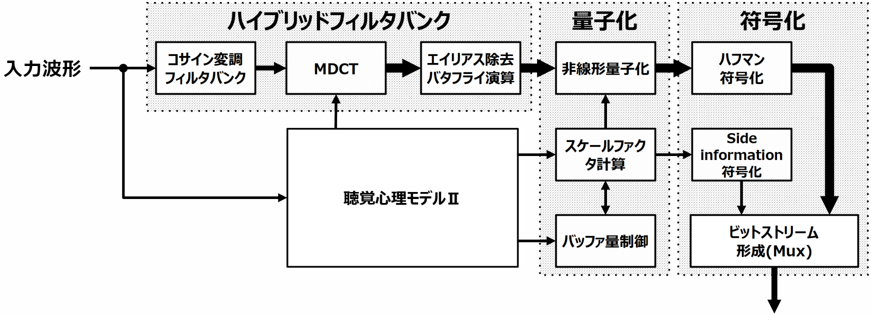
MPEG1の要素技術

- レイヤI,II,IIIの順に圧縮率向上
- レイヤI,IIはサブバンド符号化がメイン
- 聴覚心理モデルはI,IIとIIIで異なる



[1] より引用

MP3のエンコーダ構造

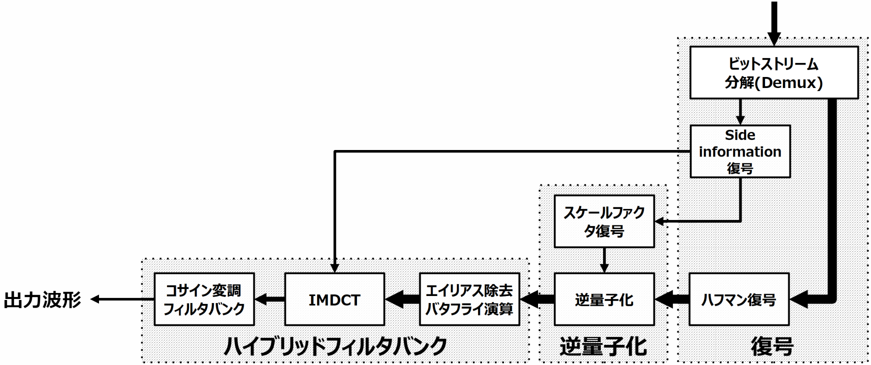


ハイブリッドフィルタバンク 32バンドのフィルタバンクの後，18点のMDCT → 576点のスペクトルを計算

量子化 臨界帯域・マスキングの情報を元にスペクトルを量子化

符号化 低域を精密・高域を荒く符号化

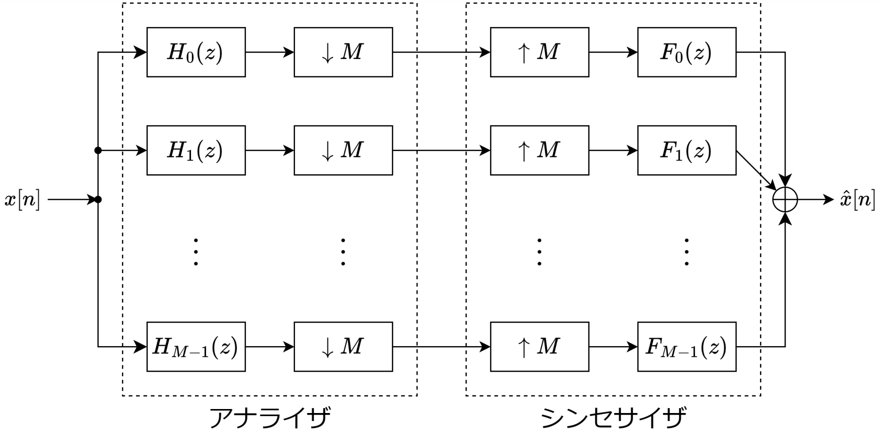
MP3のデコーダ構造



エンコーダの逆の操作

1. MP3概要
   - MPEG/Audioの歴史
   - コーデック構造
2. ハイブリッドフィルタバンク
   - フィルタバンク
   - MDCT (Modified Discrete Cosine Transform)
   - エイリアス削減バタフライ演算
3. 量子化
4. 符号化
5. 聴覚心理モデルII

M 分割フィルタバンク [2]



- 信号を M 個の帯域に分割
- M 個の分析フィルタ h_k・合成フィルタ f_k を使用

コサイン変調フィルタバンク [2, 3]

コサイン変調フィルタバンク

1 つの実係数・直線位相プロトタイプフィルタ $p_0[n]$ から，分析フィルタ h_k と合成フィルタ f_k を次で設定：

$$h_k[n] = 2p_0[n] \cos\left[\frac{\pi}{M}\left(k+\frac{1}{2}\right)\left(n-\frac{L-1}{2}\right)+\theta_k\right] \tag{1}$$

$$f_k[n] = h_k[L-1-n] \tag{2}$$

M：分割帯域数，L：タップ長，$\theta_k = (-1)^k \frac{\pi}{4}$

詳細は補足 1 節に記載

MP3のフィルタバンク

$M = 32, L = 33$ としたコサイン変調フィルタバンクに近いが異なる！（θ_k がない！）

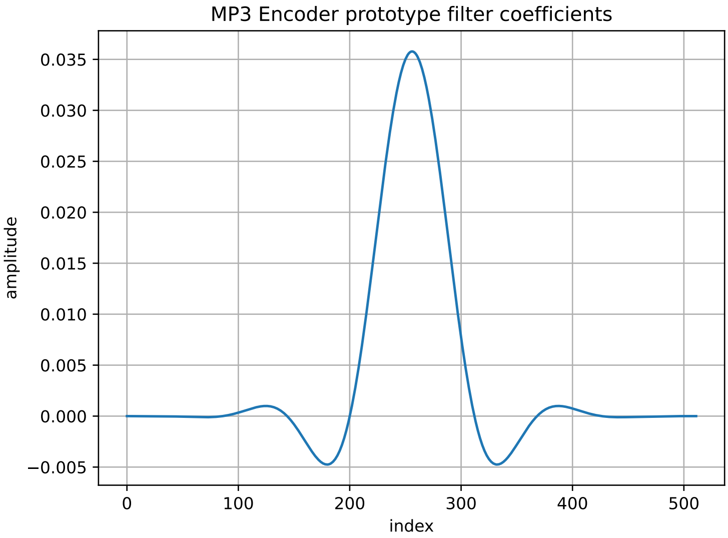
$$h_k[n] = p_0[n] \cos\left[\frac{\pi}{32}\left(k + \frac{1}{2}\right)(n - 16)\right] \tag{3}$$

$$f_k[n] = 32 p_0[n] \cos\left[\frac{\pi}{32}\left(k + \frac{1}{2}\right)(n + 16)\right] \tag{4}$$

$$p_0[n] = \begin{cases} -C_n & \lfloor n/64 \rfloor \text{ が偶数} \\ C_n & \lfloor n/64 \rfloor \text{ が奇数} \end{cases} \tag{5}$$

$C_n \ (n = 0, ..., 511)$ は規格で設定

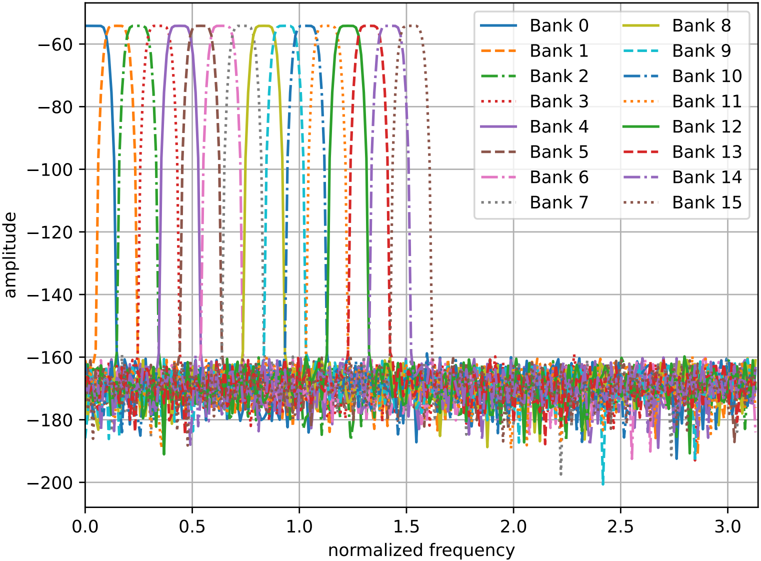
フィルタバンクの特性



- $p_0[n]$ の形状．対称（= 直線位相特性をもつ）．

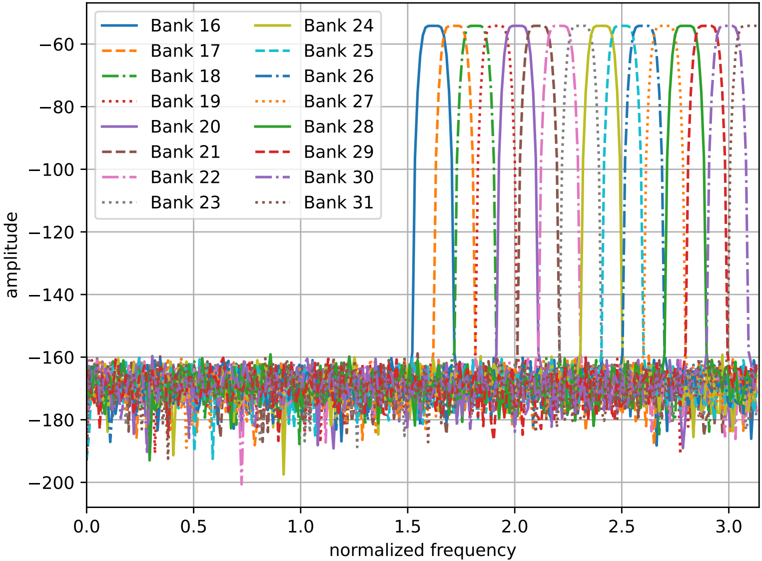
フィルタバンクの周波数特性

バンク $k = 0, \ldots, 15$ の周波数特性



フィルタバンクの周波数特性

バンク $k = 16, ..., 31$ の周波数特性



フィルタバンクの実装

プログラムでは，入力 $x[t]$ からバンド k の出力 $y_k[t]$ を

$$y_k[t] = \sum_{s=0}^{63} t_{k,s} \sum_{u=0}^{7} x[t-s-64u]C_{s+64u}$$

$$t_{k,s} := \cos\left[\frac{\pi}{32}\left(k+\frac{1}{2}\right)(s-16)\right]$$

で計算．この式が FIR フィルタ出力計算式

$$y_k[t] = \sum_{n=0}^{511} x[t-n]h_k[n] = \sum_{n=0}^{511} x[t-n]p_0[n]t_{k,n} \tag{6}$$

から導かれることを示す．

フィルタバンクの実装

(6) 式を変形していくと，

$$y_k[t] = \sum_{n=0}^{511} x[t-n]p_0[n]t_{k,n} = \sum_{u=0}^{7}\sum_{s=0}^{63} x[t-s-64u]p_0[s+64u]t_{k,s+64u}$$
$$= \sum_{u=0}^{7}\sum_{s=0}^{63} x[t-s-64u](-1)^u C_{s+64u}t_{k,s+64u} \tag{7}$$

ここで，

$$t_{k,s+64u} = \cos\left[\frac{\pi}{32}\left(k+\frac{1}{2}\right)(s+64u-16)\right] = \cos\left[\frac{\pi}{32}\left(k+\frac{1}{2}\right)(s-16)+\pi\left(2k+1\right)u\right]$$
$$= \cos\left[\frac{\pi}{32}\left(k+\frac{1}{2}\right)(s-16)\right]\cos\left[\pi(2k+1)u\right]$$
$$-\sin\left[\frac{\pi}{32}\left(k+\frac{1}{2}\right)(s-16)\right]\sin\left[\pi(2k+1)u\right] = (-1)^u t_{k,s}$$

だから，これを式 (7) に代入すれば，

$$y_k[t] = \sum_{u=0}^{7}\sum_{s=0}^{63} x[t-s-64u]C_{s+64u}t_{k,s} = \sum_{s=0}^{63} t_{k,s}\sum_{u=0}^{7} x[t-s-64u]C_{s+64u}$$

プログラムの計算式が導かれた．

フィルタバンクは完全再構成か？

- C_n の導出方法が不明．厳密に完全再構成性を示せない
- 再構成信号 $\hat{x}[n]$ が入力信号の遅延$+$定数倍になるか観察

$$
\begin{cases}
y_k[n] = \sum_{i=0}^{511} h_k[i]x[n-i] & \text{バンク } k \text{ の分析フィルタ出力} \\
z_k[n] = \sum_{i=0}^{511} f_k[i]y_k[n-i] & \text{バンク } k \text{ の合成フィルタ出力} \\
\hat{x}[n] = \sum_{k=0}^{31} z_k[n] & \text{再構成信号}
\end{cases}
$$

フィルタバンクは完全再構成か？

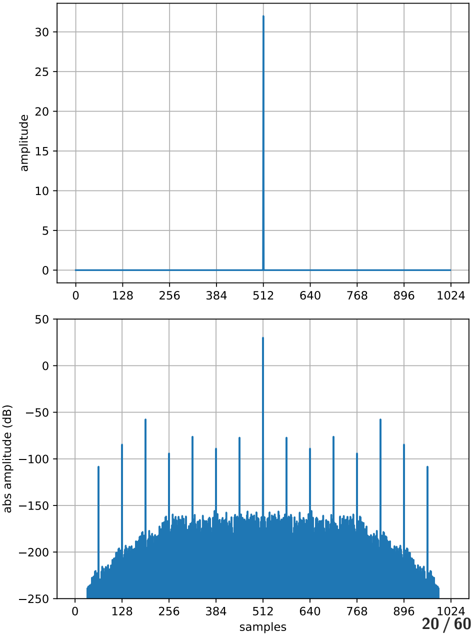
$$
\begin{aligned}
\hat{x}[n] &= \sum_{k=0}^{31} z_k[n] = \sum_{k=0}^{31} \left(\sum_{i=0}^{511} f_k[i] y_k[n-i] \right) \\
&= \sum_{k=0}^{31} \left\{ \sum_{i=0}^{511} f_k[i] \left(\sum_{j=0}^{511} h_k[j] x[n-i-j] \right) \right\} \\
&= \sum_{k=0}^{31} \sum_{i=0}^{511} \sum_{j=0}^{511} f_k[i] h_k[j] x[n-i-j] \\
&= \sum_{k=0}^{31} \sum_{m=0}^{1022} \sum_{i=\max\{0,m-511\}}^{\min\{511,m\}} f_k[i] h_k[m-i] x[n-m] \quad (m := i+j) \\
&= \sum_{m=0}^{1022} x[n-m] \underbrace{\sum_{i=\max\{0,m-511\}}^{\min\{511,m\}} \sum_{k=0}^{31} f_k[i] h_k[m-i]}_{=g[m]} \qquad (8)
\end{aligned}
$$

フィルタバンクは完全再構成か？

- $g[m]$ のグラフ（右図）
- $g[m] \approx 32\delta_{m,512}$ だから，

$$\hat{x}[n] \approx 32x[n-512]$$

近似的に完全再構成



MP3のMDCT (Modified DCT)

- サブバンドフィルタで32帯域に分割した信号に対し，18点MDCTを実行
  - 出力：$32 \times 18 = 576$ 点のスペクトルデータ
- MDCTの前・IMDCTの後で窓関数を適用
  - MP3では4種類の窓関数を使用
  - 窓関数は完全再構成条件を満たす

MDCT と IMDCT

入力信号を $x[n]$，再構成信号を $y[n]$ として

MDCT (Modified Discrete Cosine Transform)

$$X_k = \sum_{n=0}^{2N-1} x[n] \cos\left[\frac{\pi}{N}\left(k+\frac{1}{2}\right)\left(n+\frac{1}{2}+\frac{N}{2}\right)\right] \tag{9}$$

IMDCT (Inverse MDCT)

$$y[n] = \frac{2}{N}\sum_{k=0}^{N-1} X_k \cos\left[\frac{\pi}{N}\left(k+\frac{1}{2}\right)\left(n+\frac{1}{2}+\frac{N}{2}\right)\right] \tag{10}$$

- 時間領域は $2N$ 点，周波数領域は N 点の変換 $(k = 0, \ldots, N-1)$

MDCT と完全再構成条件

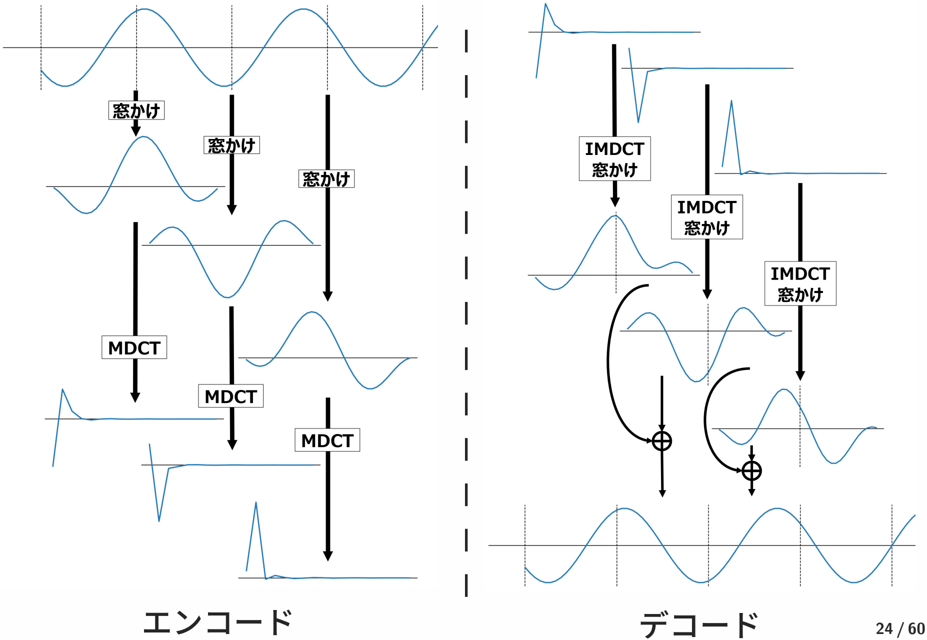
(10) 式に (9) 式を代入して整理すると

$$y[n] = \begin{cases} x[n] - x[N-1-n] & (n = 0, ..., N-1) \\ x[n] + x[3N-1-n] & (n = N, ..., 2N-1) \end{cases} \tag{11}$$

となる（証明は補足）.

ハーフオーバーラップで処理するとき，$x[n], y[n]$ にうまく窓関数を適用すると完全再構成にできる．その条件は？

ハーフオーバーラップアドの手順



MDCT と完全再構成条件

Princen–Bradley 条件（完全再構成条件）[4]

長さ $2N$ の分析窓を w_a，合成窓を w_s としたとき，

$$w_a[n]w_s[n] + w_a[n+N]w_s[n+N] = 1 \quad (12)$$

$$w_a[n+N]w_s[2N-1-n] = w_a[n]w_s[N-1-n] \quad (13)$$

$$n = 0, \ldots, N-1$$

ならば，MDCT・IMDCT によるハーフオーバーラップアドは完全再構成

Princen–Bradley 条件の導出

m フレーム目の n 時刻の入力 $x_m[n]$ は，フレームあたり N サンプルでスライドしており

$$x_m[n] = x_{m-1}[n+N] \tag{14}$$

が成り立つとする．窓かけした信号 $g_m[n]$ を

$$g_m[n] := w_a[n] x_m[n] \quad (n = 0, ..., 2N-1) \tag{15}$$

と書く．$g_m[n]$ を MDCT・IMDCT して再構成した信号 $z_m[n]$ は，(11) 式より，

$$z_m[n] = \begin{cases} g_m[n] - g_m[N-1-n] & (n = 0, ..., N-1) \\ g_m[n] + g_m[3N-1-n] & (n = N, ..., 2N-1) \end{cases} \tag{16}$$

Princen–Bradley 条件の導出

ハーフオーバーラップアドした結果を $\hat{x}_m[n]$ と書くと，(16) 式より，

$$
\begin{aligned}
\hat{x}_m[n] &= w_s[n]z_m[n] + w_s[n+N]z_{m-1}[n+N] \\
&= w_s[n](g_m[n] - g_m[N-1-n]) \\
&\quad + w_s[n+N](g_{m-1}[n+N] + g_{m-1}[3N-1-(n+N)]) \\
&= w_s[n](w_a[n]x_m[n] - w_a[N-1-n]x_m[N-1-n]) \\
&\quad + w_s[n+N](w_a[n+N]x_{m-1}[n+N] + w_a[2N-1-n]x_{m-1}[2N-1-n]) \\
&= w_s[n](w_a[n]x_m[n] - w_a[N-1-n]x_m[N-1-n]) \\
&\quad + w_s[n+N](w_a[n+N]x_m[n] + w_a[2N-1-n]x_m[N-1-n]) \\
&= x_m[n](w_a[n]w_s[n] + w_a[n+N]w_s[n+N]) \\
&\quad + x_m[N-1-n](w_a[n+N]w_s[2N-1-n] - w_a[n]w_s[N-1-n])
\end{aligned}
$$

この結果を $x_m[n] = \hat{x}_m[n]$ として両辺比較することで条件が得られる

MP3 と Princen–Bradley 条件

Princen–Bradley 条件（分析窓と合成窓が同一）

分析・合成窓が同じ $w[n] = w_a[n] = w_s[n]$ とき，

$$
\begin{aligned}
w[n]^2 + w[n+N]^2 &= 1 &\quad (17)\\
w[n+N]w[2N-1-n] &= w[n]w[N-1-n] &\quad (18)\\
n &= 0, \ldots, N-1
\end{aligned}
$$

とくに**窓関数が対称** $w[n] = w[2N-1-n]$ であれば，

$$w[n+N]w[2N-1-n] = w[N-1-n]w[2N-1-n] = w[N-1-n]w[n]$$

となり式 (18) が満たされる [*1]

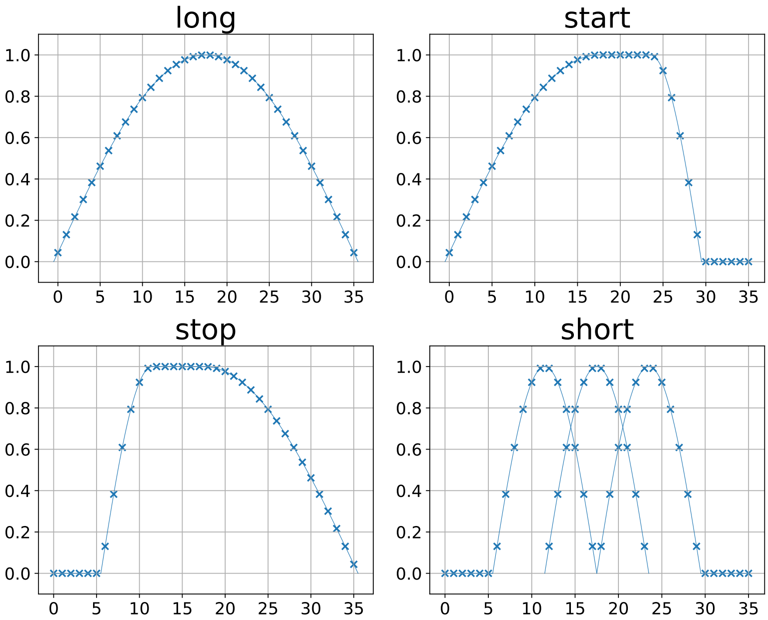
[*1] 逆に，式 (18) を満たしても対称とは限らない．$w[n] = \frac{w[n+N]w[2N-1-n]}{w[N-1-n]}$ だが，一般に $w[n+N] \neq w[N-1-n]$ だから $w[n] \neq w[2N-1-n]$．

MP3 と MDCT – 4 種類の窓関数

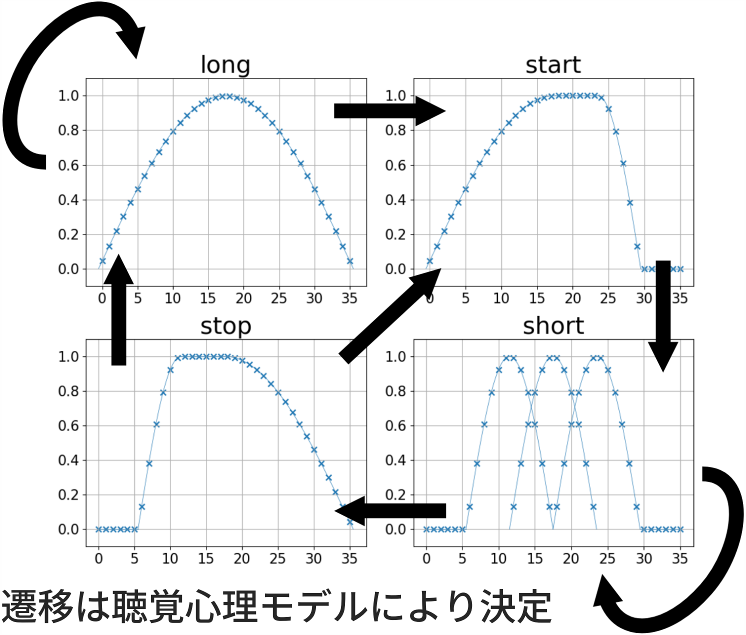
種類	定義
long	$w[n] = \sin\left[\frac{\pi}{36}\left(n+\frac{1}{2}\right)\right] \quad (n=0,...,35)$
short	$w[n] = \sin\left[\frac{\pi}{12}\left(n-6k+\frac{1}{2}\right)\right] \quad (k=1,2,3,\ n=6k,...,6k+11)$
start	$w[n] = \begin{cases} \sin\left[\frac{\pi}{36}\left(n+\frac{1}{2}\right)\right] & (n=0,...,17) \\ 1 & (n=18,...,23) \\ \sin\left[\frac{\pi}{12}\left(n-18+\frac{1}{2}\right)\right] & (n=24,...,29) \\ 0 & (n=30,...,35) \end{cases}$
stop	$w[n] = \begin{cases} 0 & (n=0,...,5) \\ \sin\left[\frac{\pi}{12}\left(n-6+\frac{1}{2}\right)\right] & (n=6,...,11) \\ 1 & (n=12,...,17) \\ \sin\left[\frac{\pi}{36}\left(n+\frac{1}{2}\right)\right] & (n=18,...,35) \end{cases}$

- short の窓長は 12，それ以外は 36

MP3とMDCT – 4種類の窓関数



MP3とMDCT – 窓関数の状態遷移



状態遷移は聴覚心理モデルにより決定

MP3 と Princen–Bradley 条件

サイン窓

$$w[n] = \sin\left[\frac{\pi}{2N}\left(n+\frac{1}{2}\right)\right] \quad (n = 0, ..., 2N-1) \quad (19)$$

は Princen–Bradley 条件を満たす．

（証明）式 (17) は：

$$w[n]^2 + w[n+N]^2 = \sin^2\left[\frac{\pi}{2N}\left(n+\frac{1}{2}\right)\right] + \sin^2\left[\frac{\pi}{2N}\left(n+N+\frac{1}{2}\right)\right]$$
$$= \sin^2\left[\frac{\pi}{2N}\left(n+\frac{1}{2}\right)\right] + \cos^2\left[\frac{\pi}{2N}\left(n+\frac{1}{2}\right)\right] = 1$$

式 (18) は，サイン窓が対称であることより示される：

$$w[2N-1-n] = \sin\left[\frac{\pi}{2N}\left(2N-1-n+\frac{1}{2}\right)\right] = \sin\left[\pi + \frac{\pi}{2N}\left(-n-\frac{1}{2}\right)\right]$$
$$= \sin\left[-\frac{\pi}{2N}\left(n+\frac{1}{2}\right)+\pi\right] = \sin\left[\frac{\pi}{2N}\left(n+\frac{1}{2}\right)\right] = w[n]$$

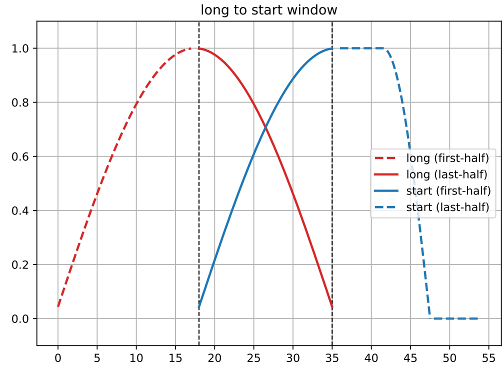
MP3 と Princen–Bradley 条件

- long, short 窓はサイン窓そのものなので完全再構成
- 状態遷移時に完全再構成になるか？
  1. long $\to$ start
  2. stop $\to$ long：1. の対称ケース
  3. start $\to$ short
  4. short $\to$ stop：3. の対称ケース

  1. と 3. だけ確認

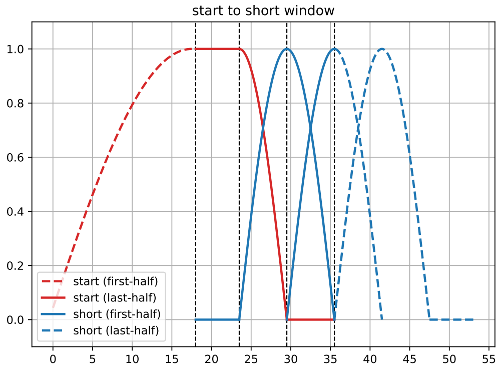
MP3 と Princen–Bradley 条件

1. long $\rightarrow$ start
   - ハーフオーバーラップアドする区間でサイン窓になっているため，完全再構成



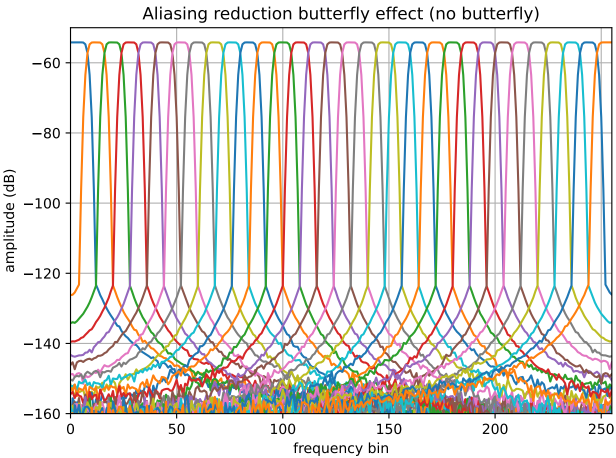
MP3 と Princen–Bradley 条件

3. start $\rightarrow$ short
   - $n = 0, ..., 5$：start 窓が 1, short 窓が 0 なので完全再構成
   - $n = 6, ..., 11$：サイン窓になっているため完全再構成
   - $n = 12, ..., 17$：start 窓が 0, short 窓どうしでサイン窓になっているため完全再構成



エイリアス削減バタフライ演算

- フィルタバンクは，隣接バンクの周波数成分（エイリアス）が混入
- 隣接バンクのスペクトルを使いエイリアス削減（[5]）



エイリアス削減バタフライ演算

- スペクトルを X_k としたとき，

$$X_{18k-i} \leftarrow \mathrm{cs}_i X_{18k-i} - \mathrm{ca}_i X_{18k+i+1} \tag{20}$$

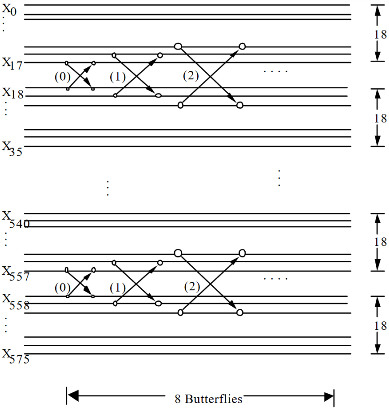
$$X_{18k+i+1} \leftarrow \mathrm{ca}_i X_{18k-i} + \mathrm{cs}_i X_{18k+i+1} \tag{21}$$

$$k = 1, ..., 31,\ \ i = 0, ..., 7$$

- 係数 $\mathrm{cs}_i, \mathrm{ca}_i$ の定義

$$\mathrm{cs}_i := \frac{1}{\sqrt{1 + c_i^2}},\ \ \mathrm{ca}_i := \frac{c_i}{\sqrt{1 + c_i^2}} \tag{22}$$

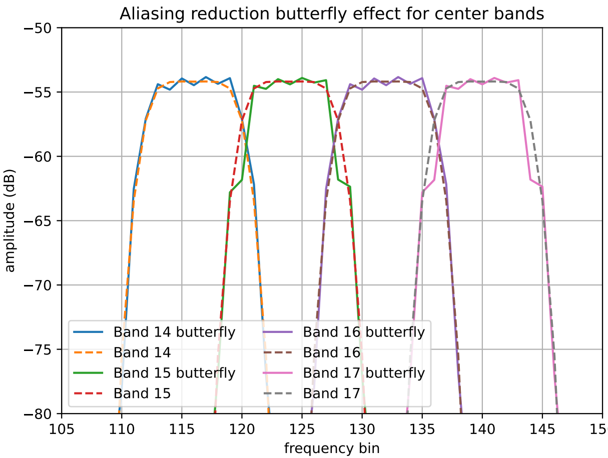
$c_i\ (i = 0, ...7)$ は規格で設定



[6] より引用

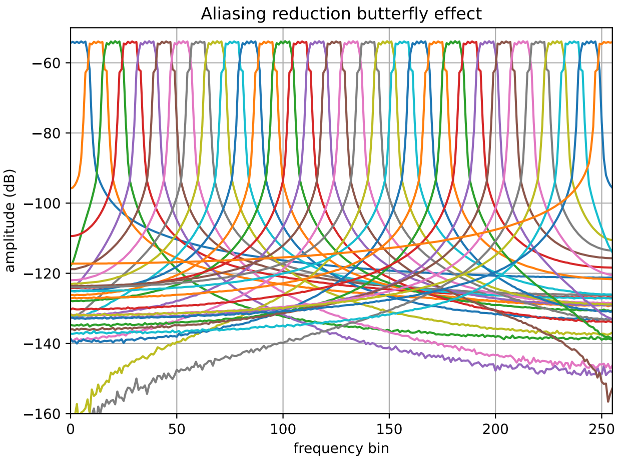
エイリアス削減バタフライ演算

- 14-17 バンドの周波数特性比較
- バタフライ演算により，隣接バンクと交差する振幅が-2dB ほど下に移動（改善）



エイリアス削減バタフライ演算

- 振幅が小さくなると遮断特性は悪化
- 可聴域帯を優先した結果？



1. MP3概要
   - MPEG/Audioの歴史
   - コーデック構造
2. ハイブリッドフィルタバンク
   - フィルタバンク
   - MDCT (Modified Discrete Cosine Transform)
   - エイリアス削減バタフライ演算
3. 量子化
4. 符号化
5. 聴覚心理モデルII

非線形量子化

ブロック長が 36(long, start, stop) のとき，量子化スペクトル X_i^q，逆量子化スペクトル $\hat{X}_i$ は [*2]

$$X_i^q = \mathrm{round}\left[\mathrm{sign}(X_i)\left|X_i(G_i^l)^{-1}\right|^{\frac{3}{4}}\right] \tag{23}$$

$$\hat{X}_i = \mathrm{sign}(X_i^q)|X_i^q|^{\frac{4}{3}}G_i^l \tag{24}$$

$$G_i^l = 2^{\frac{1}{4}g}2^{-\frac{1}{2}(1+\mathrm{scale})(\mathrm{sl}_i+p_i)} \tag{25}$$

- g：グローバルゲイン
- scale：スケールファクタのスケール（dist10 エンコーダーでは常に 0）
- sl_i：スケールファクタ
- p_i：プリエンファシス増幅値（dist10 エンコーダーでは常に 0）

[*2] round を除けば，逆量子により元に戻る

$$\hat{X}_i = \mathrm{sign}(X_i)|X_i^q|^{\frac{4}{3}}G_i^l = \mathrm{sign}(X_i)\left|\left|X_i(G_i^l)^{-1}\right|^{\frac{3}{4}}\right|^{\frac{4}{3}}G_i^l = X_i(G_i^l)^{-1}G_i^l = X_i$$

非線形量子化

ブロック長が 12(short) のとき，

$$X_i^q = \mathrm{round}\left[\mathrm{sign}(X_i)\left|X_i(G_i^s)^{-1}\right|^{\frac{3}{4}}\right] \tag{26}$$

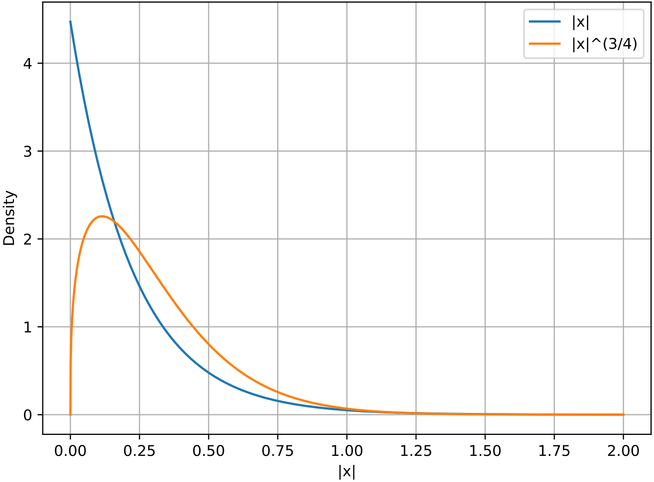
$$\hat{X}_i = \mathrm{sign}(X_i^q)|X_i^q|^{\frac{4}{3}}G_i^s \tag{27}$$

$$G_i^s = 2^{\frac{1}{4}g}2^{2\mathrm{sbgain}_b}2^{-\frac{1}{2}(1+\mathrm{scale})\mathrm{ss}_i} \tag{28}$$

- sbgain_b：サブブロックのゲイン（dist10 エンコーダーでは常に 0）
- ss_i：ショートブロックのスケールファクタ

$3/4$ 乗の効果

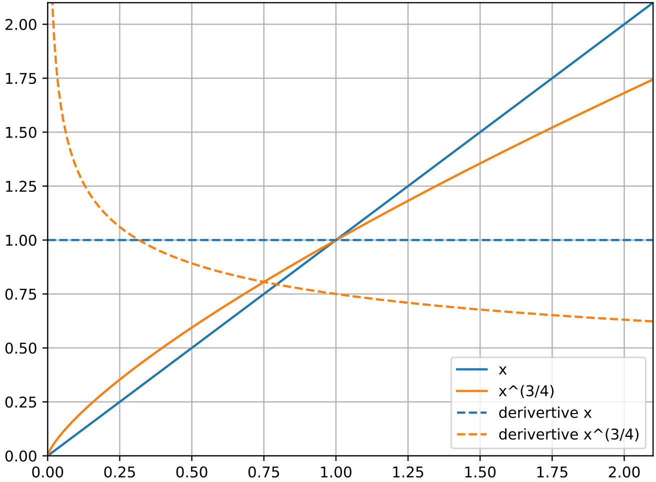
平均 0 分散 0.1 の Laplace 分布の絶対値 $|x|$ と $|x|^{\frac{3}{4}}$ の密度



0 近傍の密度が減少

$3/4$ 乗の効果

$x, x^{\frac{3}{4}}$ とそれらの微分（ステップ幅）

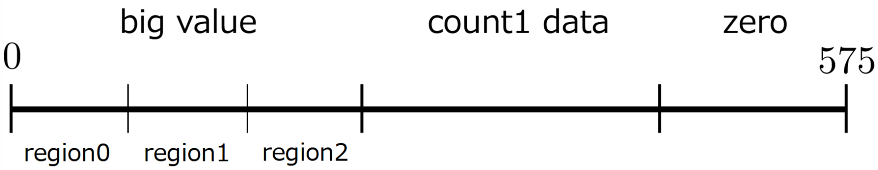


0 近傍は荒く量子化，≈ 0.3 以上では細かく量子化

1. MP3 概要
   - MPEG/Audio の歴史
   - コーデック構造
2. ハイブリッドフィルタバンク
   - フィルタバンク
   - MDCT (Modified Discrete Cosine Transform)
   - エイリアス削減バタフライ演算
3. 量子化
4. 符号化
5. 聴覚心理モデル II

MP3の符号化

量子化スペクトルを区分に分けて符号化



big value 大きい値は線形量子化

- region0,1,2 で異なるハフマンテーブルを使用

count1 data $\{-1, 0, 1\}$ のみで符号化

- 1つのハフマンテーブルを使用

zero 0 のみ（符号化しない）

MP3の符号化（詳細）

big valueの符号化

- 数値2つ組 X_i^q, X_{i+1}^q を

$$X = 16|X_i^q| + |X_{i+1}^q| \tag{29}$$

  として符号化．符号は2bitで出力
- 線形符号化するかの閾値はテーブル毎に設定

count1 dataの符号化

- 数値4つ組 $X_i^q, X_{i+1}^q, X_{i+2}^q, X_{i+3}^q$ を

$$X = 8s(X_{i+2}^q) + 4s(X_{i+3}^q) + 2s(X_i^q) + s(X_{i+1}^q) \tag{30}$$

  として符号化（$s(x)$：$x \neq 0$ なら 1，$x = 0$ なら 0）．符号は4bitで出力

聴覚心理モデル II の概要

- LayerIII の聴覚心理モデル（LayerI, II とは異なる）
- 出力：信号対マスク比 (SMR[*3])，ブロックタイプ
- 処理手順
  1. フレーム切り出し・窓かけ・FFT
  2. Unpredictability 計算
  3. パーティションごとのエネルギー計算
  4. 広がり関数 (Spreading function) の畳み込み
  5. ノイズ許容レベル計算
  6. 聴覚しきい値計算
  7. 知覚エントロピー (Psychoacoustic entropy) 計算
  8. ブロックタイプ判定・信号対マスク比 (SMR) 計算

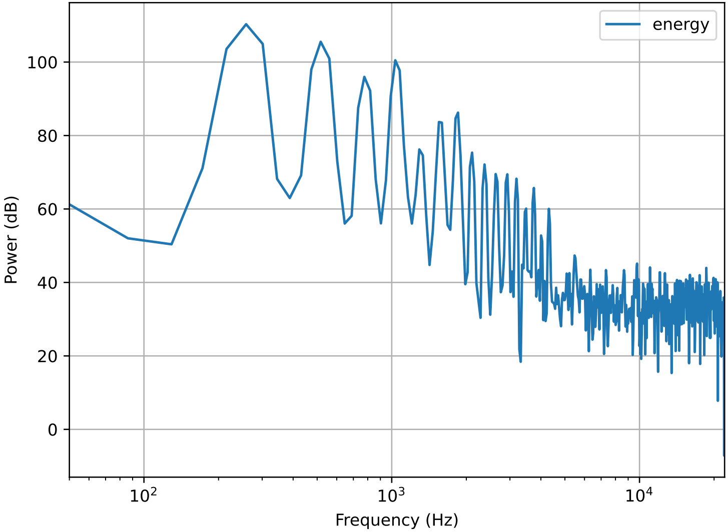
[*3] Signal-to-Masking Ratio

フレーム切り出し・窓かけ・FFT

- long と short の 2 つを計算
  - long はサイズ 1024
  - short はサイズ 256 を 3 つ．信号 $s[n]$ の $s[128b+n]$ $b=1,2,3$ を使用
- Hanning 窓を適用
- スライド (hop) サイズは 576（=DCT スペクトルサイズ）

FFT 係数を w^l(long)，w^{s_b}(short, $b=1,2,3$) と表記

フレーム切り出し・窓かけ・FFT



ピアノ (F0=220) の long のエネルギー（パワースペクトル）

Unpredictability計算

Unpredictability = 予測しずらさの尺度

- 振幅・位相スペクトルを直線予測した複素数を予測とし，誤差絶対値を正規化して評価
- 予測が当っていれば 0，外れていれば 1
- long, short の直線予測（$w^{l'}$ は前フレームの w^{l}）：

$$w^{l*}[n] = (2|w^{l'}[n]| - |w^{l''}[n]|)\exp[j(2\arg(w^{l'}[n]) - \arg(w^{l''}[n]))] \quad (31)$$

$$w^{s*}[n] = (2|w^{s_1}[n]| - |w^{s_3}[n]|)\exp[j(2\arg(w^{s_1}[n]) - \arg(w^{s_3}[n]))] \quad (32)$$

Unpredictability cw の計算式：

$$cw[n] = \begin{cases} \frac{|w^l[n] - w^{l*}[n]|}{|w^l[n]| + |w^{l*}[n]|} & 0 \le n \le 5 \\ \frac{|w^s[j] - w^{s*}[j]|}{|w^s[j]| + |w^{s*}[j]|} & j = \lfloor (n+2)/4 \rfloor,\ 6 \le n \le 205 \\ 0.4 & 206 \le n \le 512 \end{cases} \quad (33)$$

パーティションごとのエネルギー計算

周波数ビンを" パーティション (partition)" 単位に分割

- パーティション内のエネルギーを合算
- 同時に Unpredictability で重みづけしたエネルギーも合算

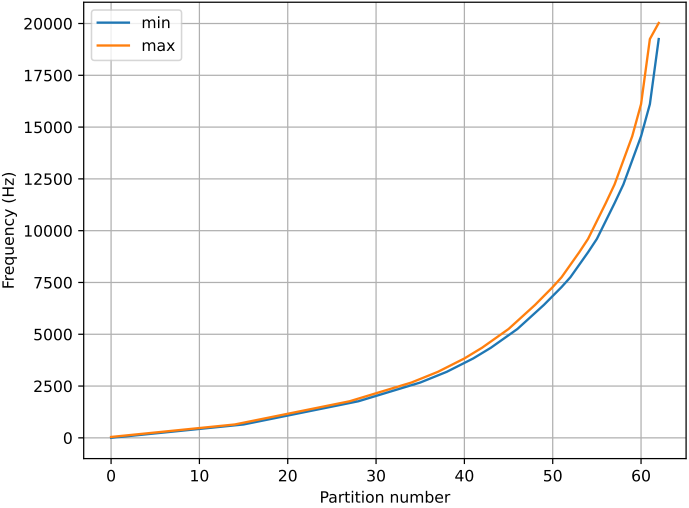
long, short のパーティション b を P_b^l, P_b^s と書くと，

$$eb^l[b] = \sum_{n=\min P_b^l}^{\max P_b^l - 1} |w^l[n]|^2 \tag{34}$$

$$cb^l[b] = \sum_{n=\min P_b^l}^{\max P_b^l - 1} cw[n]|w^l[n]|^2 \tag{35}$$

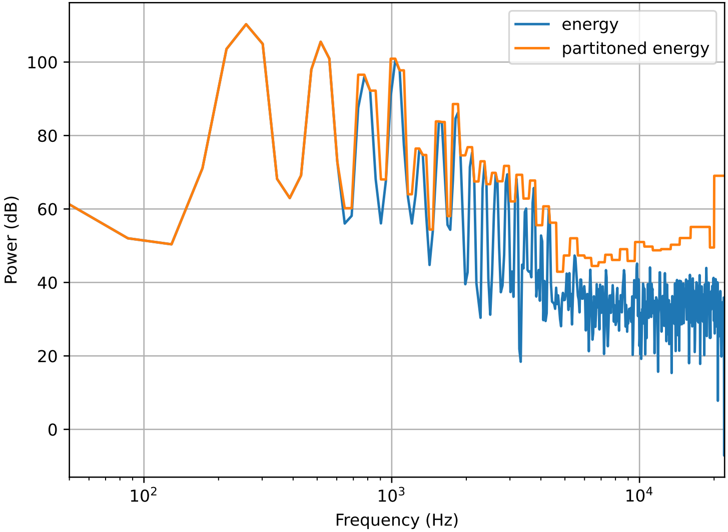
$$eb^s[b] = \sum_{n=\min P_b^s}^{\max P_b^s - 1} |w^s[n]|^2 \tag{36}$$

パーティションごとのエネルギー計算



サンプリング周波数 44.1kHz の long のパーティション．Bark スケールを細かくしたものとみなせる

パーティションごとのエネルギー計算



ピアノ (F0=220) の long パーティション分割後のエネルギー eb^l

広がり関数の畳み込み

ノイズ許容レベル計算

聴覚しきい値計算

知覚エントロピー計算

ブロックタイプ判定・SMR計算

6. 参考文献

7. 証明
- フィルタバンク
- MDCT
- 知覚エントロピー導出

参考文献の紹介

- [5, 7]：エイリアス削減の論文
- [4]：Princen–Bradley の完全再構成条件の導出
- [6]：MP3 の概要説明．
- [1]：MPEG/Audio の他，2000 年代前半の他のコーデックの概要を解説
- [8]：MPEG 標準化に関する書物．技術は概要程度
- [2]：マルチレートに関わる信号処理を広範に解説
- [3]：フィルタバンクに関する詳細書
- [9]：DCT に関する詳しい解説
- [10, 11, 12, 13, 14]：技術解説と簡易 MP3 コーデックの実装例
- [15]：MP3 のソース (dist10) の解説．ただし木を見て森を見ずな印象．実装の補足説明としては優秀だが，コードを数式に直訳した書き方のため，理解しずらい．

参考文献 I

[1] 藤原洋. 画像&音声圧縮技術のすべて : インターネット/ディジタルテレビ/モバイル通信時代の必須技術. 第 6 版. Tech I. CQ 出版社, 2001.

[2] 貴家仁志. マルチレート信号処理. ディジタル信号処理シリーズ. 昭晃堂, 1995.

[3] P. P. Vaidyanathan et al. マルチレート信号処理とフィルタバンク. ディジタル信号処理・画像処理シリーズ. 科学技術出版, 2002.

[4] John Princen and Alan Bradley. "Analysis/synthesis filter bank design based on time domain aliasing cancellation". In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 34.5 (1986), pp. 1153–1161.

[5] Bernd Edler. "Aliasing reduction in sub-bands of cascaded filter banks with decimation". In: *Electronics Letters* 12.28 (1992), pp. 1104–1106.

[6] Rassol Raissi. "The theory behind MP3". In: *MP3' Tech* (2002).

[7] Chi-Min Liu and Wen-Chieh Lee. "The design of a hybrid filter bank for the psychoacoustic model in ISO/MPEG phases 1, 2 audio encoder". In: *IEEE transactions on consumer electronics* 43.3 (1997), pp. 586–592.

[8] 安田浩. *MPEG/*マルチメディア符号化の国際標準. 丸善, 1994.

参考文献 II

[9] 貴家仁志 and 村松正吾. マルチメディア技術の基礎 *DCT(離散コサイン変換) 入門 : JPEG/MPEG* からウェーブレット, 重複直交変換 *(LOT)* まで. I/F essence. CQ 出版, 1997.

[10] 小杉篤史. *Interface Aug.2001* 第 *1* 回音響圧縮技術の基礎 *MP3* と等価的なシステムの構築とサブバンドフィルタバンクの設計. CQ 出版社, 2001.

[11] 小杉篤史 and 城下聡. *Interface Sep.2001* 第 *2* 回音響圧縮技術の基礎 *MP3* と等価的なシステムの構築と *MDCT* の設計. CQ 出版社, 2001.

[12] 小杉篤史 and 城下聡. *Interface Nov.2001* 第 *3* 回音響圧縮技術の基礎 *MP3* と等価的なシステムの構築とハイブリッドフィルタバンクの設計. CQ 出版社, 2001.

[13] 小杉篤史 and 城下聡. *Interface Jan.2002* 第 *4* 回音響圧縮技術の基礎 *MP3* と等価的なシステムの構築と非線形量子化器/符号化器の設計. CQ 出版社, 2002.

[14] 小杉篤史 and 城下聡. *Interface Feb.2002* 第 *4* 回音響圧縮技術の基礎 *MP3* と等価的なシステムの構築と?そのためのビットストリームの設計. CQ 出版社, 2002.

[15] 浦田敏道. 詳細 *MP3* マニュアル. エム研, 1999.

6. 参考文献

7. 証明
- フィルタバンク
- MDCT
- 知覚エントロピー導出

完全再構成

遅延・定数倍を除き入出力が一致すること：

$$\hat{x}[n] = cx[n - n_0], \quad c \neq 0 \tag{37}$$

これは z 領域で，

$$\hat{X}(z) = cz^{-n_0}X(z) \tag{38}$$

となることと等価

ポリフェーズ表現

$H_k(z)$ のインパルス応答を $h_k[n]$ と書くとき，

$$
\begin{aligned}
H_k(z) = \sum_{n=-\infty}^{\infty} h_k[n] z^{-n} &= \sum_{n=-\infty}^{\infty} \sum_{l=0}^{M-1} h_k[nM+l] z^{-(nM+l)} \\
&= \sum_{l=0}^{M-1} z^{-l} \sum_{n=-\infty}^{\infty} h_k[nM+l] z^{-nM} \\
&= \sum_{l=0}^{M-1} E_{k,l}(z^M) z^{-l}
\end{aligned}
$$

h_k の（タイプ I）ポリフェーズ表現

$$
E_{k,l}(z) = \sum_{n=-\infty}^{\infty} h_k[nM+l] z^{-n} \tag{39}
$$

ポリフェーズ表現

$F_k(z)$ のインパルス応答を $f_k[n]$ と書くとき，

$$
\begin{aligned}
F_k(z) &= \sum_{n=-\infty}^{\infty} f_k[n] z^{-n} = \sum_{n=-\infty}^{\infty} \sum_{l=0}^{M-1} f_k[nM+l] z^{-(nM+l)} \\
&= \sum_{n=-\infty}^{\infty} \sum_{l'=0}^{M-1} f_k[nM+M-1-l'] z^{-(nM+M-1-l')} \\
&= \sum_{l'=0}^{M-1} z^{-(M-1-l')} \sum_{n=-\infty}^{\infty} f_k[nM+M-1-l'] z^{-nM} = \sum_{l'=0}^{M-1} z^{-(M-1-l')} R_{k,l}(z^M)
\end{aligned}
$$

f_k の（タイプ II）ポリフェーズ表現

$$
R_{k,l}(z) = \sum_{n=-\infty}^{\infty} f_k[nM+M-1-l] z^{-n} \tag{40}
$$

ポリフェーズ行列表現

(39) 式を l について並べ，行列表現すると

$$\underbrace{\begin{bmatrix} H_0(z) \\ H_1(z) \\ \vdots \\ H_{M-1}(z) \end{bmatrix}}_{\boldsymbol{h}(z)} = \underbrace{\begin{bmatrix} E_{0,0}(z^M) & E_{0,1}(z^M) & \cdots & E_{0,M-1}(z^M) \\ E_{1,0}(z^M) & E_{1,1}(z^M) & \cdots & E_{1,M-1}(z^M) \\ \vdots & \vdots & \ddots & \vdots \\ E_{M-1,0}(z^M) & E_{M-1,1}(z^M) & \cdots & E_{M-1,M-1}(z^M) \end{bmatrix}}_{\boldsymbol{E}(z)} \underbrace{\begin{bmatrix} 1 \\ z^{-1} \\ \vdots \\ z^{-(M-1)} \end{bmatrix}}_{\boldsymbol{e}(z)}$$

(40) 式も同様にして，以下のように書ける

$$\underbrace{\begin{bmatrix} F_0(z) \\ F_1(z) \\ \vdots \\ F_{M-1}(z) \end{bmatrix}}_{\boldsymbol{f}(z)} = \underbrace{\begin{bmatrix} R_{0,0}(z^M) & R_{0,1}(z^M) & \cdots & R_{0,M-1}(z^M) \\ R_{1,0}(z^M) & R_{1,1}(z^M) & \cdots & R_{1,M-1}(z^M) \\ \vdots & \vdots & \ddots & \vdots \\ R_{M-1,0}(z^M) & R_{M-1,1}(z^M) & \cdots & R_{M-1,M-1}(z^M) \end{bmatrix}^{\top}}_{\boldsymbol{R}(z)^{\top}} \underbrace{\begin{bmatrix} z^{-(M-1)} \\ z^{-(M-1)+1} \\ \vdots \\ 1 \end{bmatrix}}_{z^{-(M-1)}\boldsymbol{e}(z^{-1})}$$

ポリフェーズ行列表現

$\tilde{\boldsymbol{e}}(z) = \boldsymbol{e}(z^{-1})^{\mathsf{T}}$ とすると行列表現は

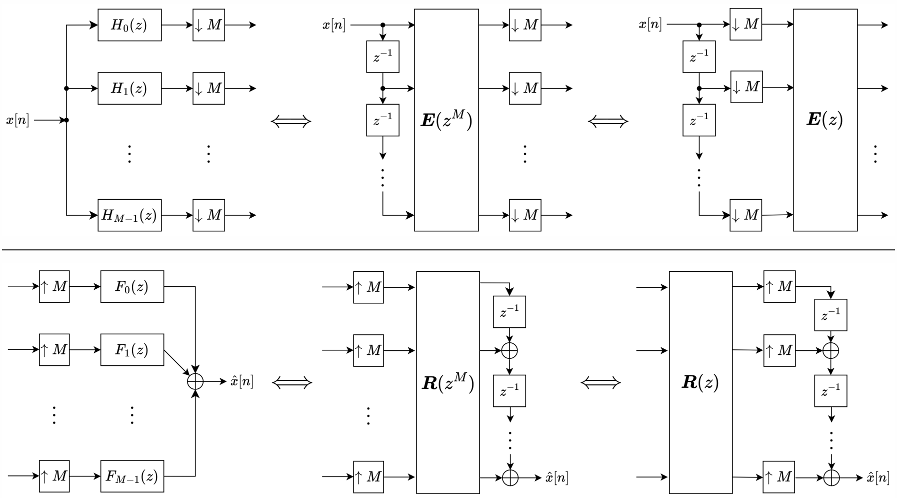
$$\boldsymbol{h}(z) = \boldsymbol{E}(z)\boldsymbol{e}(z) \tag{41}$$

$$\boldsymbol{f}(z)^{\mathsf{T}} = z^{-(M-1)}\tilde{\boldsymbol{e}}(z)\boldsymbol{R}(z) \tag{42}$$

とまとめられる．$\boldsymbol{E}(z), \boldsymbol{R}(z)$ をポリフェーズ行列という

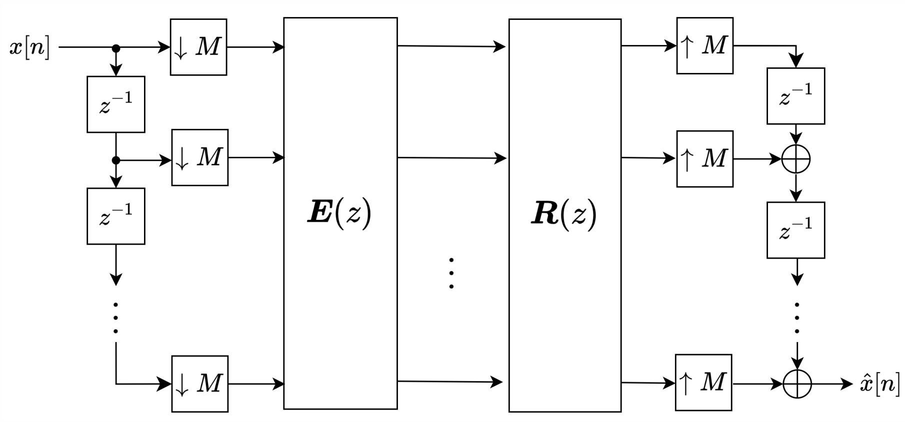
ポリフェーズ行列表現

$\boldsymbol{E}(z), \boldsymbol{R}(z)$ により，アナライザ・シンセサイザは以下のように変形できる



ポリフェーズ行列表現

$\boldsymbol{E}(z), \boldsymbol{R}(z)$ により，M 分割フィルタバンクは以下のように表せる

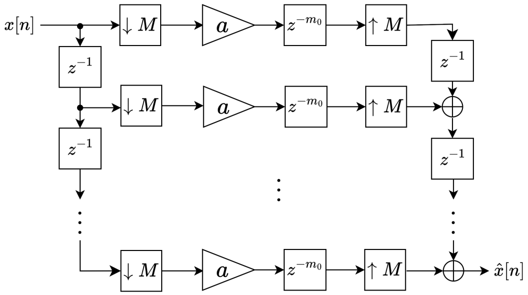


完全再構成 M 分割フィルタバンク

$$\boldsymbol{R}(z)\boldsymbol{E}(z) = az^{-m_0}\boldsymbol{I} \quad (a \neq 0,\ m_0 \in \mathbb{N}) \tag{43}$$

ならば，M 分割フィルタバンクは完全再構成[a]

[a] ∵ 各バンドの遅延が K ならば，$\hat{X}(z) = aMz^{-(M-1+K)}X(z)$



$\boldsymbol{R}(z)\boldsymbol{E}(z) = az^{-m_0}\boldsymbol{I}$ を満たす M 分割フィルタバンク

完全再構成 M 分割フィルタバンク

(43) 式より，$\boldsymbol{R}(z) = az^{-m_0}\boldsymbol{E}(z)^{-1}$ ならば完全再構成．

- しかし，$\boldsymbol{E}(z)^{-1}$ の計算に問題を孕む．

代わりに，$\boldsymbol{E}(z)$ がパラユニタリ [*4]

$$\widetilde{\boldsymbol{E}}(z)\boldsymbol{E}(z) = d\boldsymbol{I}, \quad d \neq 0 \tag{44}$$

ならば，

$$\boldsymbol{R}(z) = az^{-m_0}\widetilde{\boldsymbol{E}}(z)$$

とするとフィルタバンクは完全再構成．

[*4] $\widetilde{\boldsymbol{E}}(z) = \boldsymbol{E}_*(z^{-1})$ で，下付きの $*$ は係数の複素共役

コサイン変調フィルタバンク

式 (2) より，分析合成フィルタ $F_k(z)H_k(z)$ は直線位相特性をもつ

（証明）

$$
\begin{aligned}
F_k(z) &= \sum_{n=-\infty}^{\infty} f_k[n]z^{-n} = \sum_{n=-\infty}^{\infty} h_k[L-1-n]z^{-n} \\
&= \sum_{n'=-\infty}^{\infty} h_k[n']z^{-(L-1-n')} = z^{-(L-1)} \sum_{n'=-\infty}^{\infty} h_k[n']z^{n'} \\
&= z^{-(L-1)} H_k(z^{-1})
\end{aligned}
$$

$H_k(z)$ の周波数特性を（極座標で）$H_k(\omega) = |H_k(\omega)| \exp[j\psi(\omega)]$ と書くと，

$$
\begin{aligned}
F_k(\omega) &= \exp[-j(L-1)\omega]H_k(-\omega) = \exp[-j(L-1)\omega]|H_k(-\omega)| \exp[-j\psi(\omega)] \\
&= \exp[-j(L-1)\omega]|H_k(\omega)| \exp[-j\psi(\omega)] \quad (\because \text{実係数 FIR の振幅特性は偶})
\end{aligned}
$$

だから，$F_k(\omega)H_k(\omega) = \exp[-j(L-1)\omega]|H_k(\omega)|^2$ となって直線位相特性をもつ．

コサイン変調フィルタバンク

$h_k[n]$ の伝達関数を変形する．回転因子 W_{2M} *5 より

$$
\begin{aligned}
&\cos\left[\frac{\pi}{M}\left(k+\frac{1}{2}\right)\left(n-\frac{L-1}{2}\right)+\theta_k\right] \\
&=\frac{1}{2}\left\{\exp\left[j\left\{\frac{\pi}{M}\left(k+\frac{1}{2}\right)\left(n-\frac{L-1}{2}\right)+\theta_k\right\}\right]+\exp\left[-j\left\{\frac{\pi}{M}\left(k+\frac{1}{2}\right)\left(n-\frac{L-1}{2}\right)+\theta_k\right\}\right]\right\} \\
&=\frac{1}{2}\left\{\exp(j\theta_k)W_{2M}^{\left(k+\frac{1}{2}\right)\frac{L-1}{2}}W_{2M}^{-\left(k+\frac{1}{2}\right)n}+\exp(-j\theta_k)W_{2M}^{-\left(k+\frac{1}{2}\right)\frac{L-1}{2}}W_{2M}^{\left(k+\frac{1}{2}\right)n}\right\}
\end{aligned}
$$

これを (1) 式に代入すると，

$$
\begin{aligned}
H_k(z)&=\exp(j\theta_k)W_{2M}^{\left(k+\frac{1}{2}\right)\frac{L-1}{2}}\sum_{n=-\infty}^{\infty}p_0[n]W_{2M}^{-\left(k+\frac{1}{2}\right)n}z^{-n} \\
&\qquad+\exp(-j\theta_k)W_{2M}^{-\left(k+\frac{1}{2}\right)\frac{L-1}{2}}\sum_{n=-\infty}^{\infty}p_0[n]W_{2M}^{\left(k+\frac{1}{2}\right)n}z^{-n} \\
&=\exp(j\theta_k)W_{2M}^{\left(k+\frac{1}{2}\right)\frac{L-1}{2}}P_0\left(W_{2M}^{k+\frac{1}{2}}z\right)+\exp(-j\theta_k)W_{2M}^{-\left(k+\frac{1}{2}\right)\frac{L-1}{2}}P_0\left(W_{2M}^{-\left(k+\frac{1}{2}\right)}z\right)
\end{aligned}
$$

---

*5 $W_{2M}=\exp\left(-j\frac{2\pi}{2M}\right)=\exp\left(-j\frac{\pi}{M}\right)$

コサイン変調フィルタバンク

さらに (52) 式を代入すると [*6]

$$
\begin{aligned}
H_k(z) &= \exp(j\theta_k) W_{2M}^{\left(k+\frac{1}{2}\right)\frac{L-1}{2}} \sum_{l=0}^{2M-1} \left(W_{2M}^{k+\frac{1}{2}} z\right)^{-l} G_l\left(\left(W_{2M}^{k+\frac{1}{2}} z\right)^{2M}\right) \\
&\quad + \exp(-j\theta_k) W_{2M}^{-\left(k+\frac{1}{2}\right)\frac{L-1}{2}} \sum_{l=0}^{2M-1} \left(W_{2M}^{-\left(k+\frac{1}{2}\right)} z\right)^{-l} G_l\left(\left(W_{2M}^{-\left(k+\frac{1}{2}\right)} z\right)^{2M}\right) \\
&= \exp(j\theta_k) W_{2M}^{\left(k+\frac{1}{2}\right)\frac{L-1}{2}} \sum_{l=0}^{2M-1} W_{2M}^{-l\left(k+\frac{1}{2}\right)} z^{-l} G_l(-z^{2M}) \\
&\quad + \exp(-j\theta_k) W_{2M}^{-\left(k+\frac{1}{2}\right)\frac{L-1}{2}} \sum_{l=0}^{2M-1} W_{2M}^{l\left(k+\frac{1}{2}\right)} z^{-l} G_l(-z^{2M}) \\
&= \sum_{l=0}^{2M-1} \left\{ \exp(j\theta_k) W_{2M}^{\left(k+\frac{1}{2}\right)\frac{L-1}{2}} W_{2M}^{-l\left(k+\frac{1}{2}\right)} + \exp(-j\theta_k) W_{2M}^{-\left(k+\frac{1}{2}\right)\frac{L-1}{2}} W_{2M}^{l\left(k+\frac{1}{2}\right)} \right\} z^{-l} G_l(-z^{2M}) \\
&= \sum_{l=0}^{2M-1} 2\cos\left[\frac{\pi}{M}\left(k+\frac{1}{2}\right)\left(l-\frac{L-1}{2}\right)+\theta_k\right] z^{-l} G_l(-z^{2M})
\end{aligned}
$$

[*6] $\left(W_{2M}^{\pm\left(k+\frac{1}{2}\right)}\right)^{2M} = \exp\left[\mp j\frac{\pi}{M}\left(k+\frac{1}{2}\right)2M\right] = \exp[\mp j\pi(2k+1)] = -1$ を使用

コサイン変調フィルタバンク

さらに変形すると

$$H_k(z) = \sum_{l=0}^{M-1} z^{-l} \left\{ t_{k,l} G_l(-z^{2M}) + z^{-M} t_{k,M+l} G_{M+l}(-z^{2M}) \right\} \tag{45}$$

$$t_{k,l} := 2\cos\left[\frac{\pi}{M} \left(k + \frac{1}{2} \right) \left(l - \frac{L-1}{2} \right) + \theta_k \right]$$

(39) 式と (45) 式を見比べると，

$$E_{k,l}(z) = t_{k,l} G_l(-z^2) + z^{-1} t_{k,M+l} G_{M+l}(-z^2) \tag{46}$$

コサイン変調フィルタバンク

(46) 式より，アナライザのポリフェーズ行列は，

$$
\boldsymbol{E}(z) = \underbrace{\begin{bmatrix} t_{0,0} & \cdots\cdots & t_{0,2M-1} \\ t_{1,0} & \cdots\cdots & t_{1,2M-1} \\ \vdots & \cdots & \vdots \\ t_{M-1,0} & \cdots & t_{M-1,2M-1} \end{bmatrix}}_{\boldsymbol{T}} \begin{bmatrix} G_0(-z^2) & & & \\ & \ddots & & \\ & & G_{M-1}(-z^2) & \\ z^{-1}G_M(-z^2) & & & \\ & \ddots & & \\ & & & z^{-1}G_{2M-1}(-z^2) \end{bmatrix}
$$

$$
= \boldsymbol{T} \begin{bmatrix} \boldsymbol{G}_0(z^2) \\ z^{-1}\boldsymbol{G}_1(z^2) \end{bmatrix} \tag{47}
$$

$$
\boldsymbol{G}_i(z) := \mathrm{diag}\begin{bmatrix} G_{Mi}(-z) & G_{Mi+1}(-z) \cdots G_{Mi+M-1}(-z) \end{bmatrix} \tag{48}
$$

と書ける

コサイン変調フィルタバンク

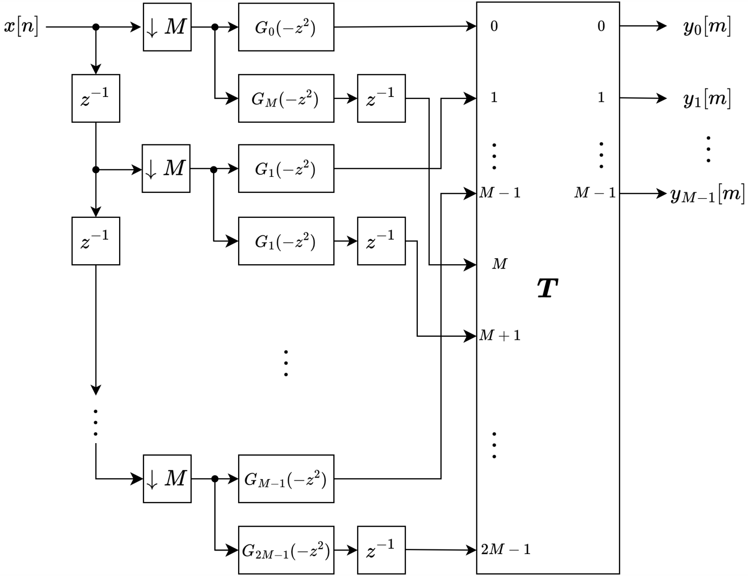
シンセサイザを構成する．フィルタ係数は実だから，

$$\widetilde{\boldsymbol{E}}(z) = \boldsymbol{E}_*(z^{-1})^\mathsf{T} = \begin{bmatrix} \boldsymbol{G}_0(z^{-1}) & z\boldsymbol{G}_1(z^{-1}) \end{bmatrix} \boldsymbol{T}^\mathsf{T} \quad (49)$$

$G_l(z)$ の次数は $2K-1$ で，因果性のためには $2K-1$ の遅延がいるため

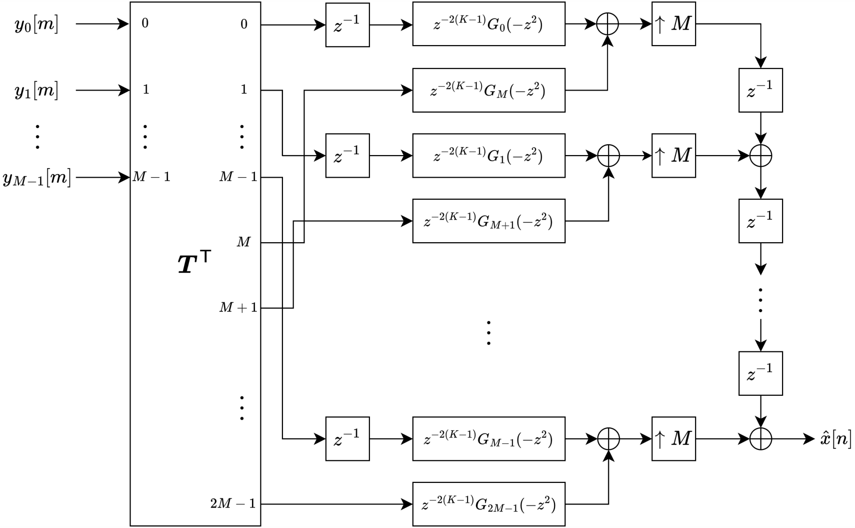
$$\begin{aligned} \boldsymbol{R}(z) &= z^{-(2K-1)} \widetilde{\boldsymbol{E}}(z) \\ &= z^{-(2K-1)} \begin{bmatrix} \boldsymbol{G}_0(z^{-1}) & z\boldsymbol{G}_1(z^{-1}) \end{bmatrix} \boldsymbol{T}^\mathsf{T} \end{aligned} \quad (50)$$

コサイン変調フィルタバンク



アナライザの構成

コサイン変調フィルタバンク



シンセサイザの構成

コサイン変調フィルタバンク

完全再構成条件を導く．(47)，(50) 式より，

$$\begin{aligned}
\boldsymbol{R}(z)\boldsymbol{E}(z) &= z^{-(2K-1)}\widetilde{\boldsymbol{E}}(z)\boldsymbol{E}(z) \\
&= z^{-(2K-1)}\begin{bmatrix} \boldsymbol{G}_0(z^{-1}) & z\boldsymbol{G}_1(z^{-1}) \end{bmatrix}\boldsymbol{T}^\top\boldsymbol{T}\begin{bmatrix} \boldsymbol{G}_0(z) \\ z^{-1}\boldsymbol{G}_1(z) \end{bmatrix} \\
&= 2Mz^{-(2K-1)}\left\{\boldsymbol{G}_0(z^{-1})\boldsymbol{G}_0(z) + \boldsymbol{G}_1(z^{-1})\boldsymbol{G}_1(z)\right\}
\end{aligned}$$

ここで，$\boldsymbol{T}^\top\boldsymbol{T} = 2M\boldsymbol{I}$（後で示す）．$\boldsymbol{G}_0, \boldsymbol{G}_1$ は対角行列だから，$k = 0, ..., M-1$ に対し，

$$G_k(z^{-1})G_k(z) + G_{M+k}(z^{-1})G_{M+k}(z) = \alpha$$

を満たせば完全再構成となる

コサイン変調フィルタバンク

コサイン変調フィルタバンクの完全再構成条件

$k = 0, \ldots, M-1$ に対し，定数 $\alpha \in \mathbb{R}$ があって

$$G_k(z^{-1})G_k(z) + G_{M+k}(z^{-1})G_{M+k}(z) = \alpha \tag{51}$$

となること．ここで，

$$G_k(z) = \sum_{n=-\infty}^{\infty} p_0[2Mn + k] z^{-n} \tag{52}$$

$G_k(z)$ は $p_0[n]$ のポリフェーズ表現

本条件は電力相補条件ともいう

$\boldsymbol{T}^\top\boldsymbol{T} = 2M\boldsymbol{I}$の証明

$\left(\boldsymbol{T}^\top\boldsymbol{T}\right)_{ij} = \sum_{k=0}^{M-1} t_{k,i}t_{k,j}$ であり，

$$
\begin{aligned}
&t_{k,i}t_{k,j} \\
&= 4\cos\left[\frac{\pi}{M}\left(k+\frac{1}{2}\right)\left(i-\frac{L-1}{2}\right)+(-1)^k\frac{\pi}{4}\right]\cos\left[\frac{\pi}{M}\left(k+\frac{1}{2}\right)\left(j-\frac{L-1}{2}\right)+(-1)^k\frac{\pi}{4}\right] \\
&= 2\left\{\cos\left[\frac{\pi}{M}\left(k+\frac{1}{2}\right)\{i+j-(L-1)\}+(-1)^k\frac{\pi}{2}\right]+\cos\left[\frac{\pi}{M}\left(k+\frac{1}{2}\right)(i-j)\right]\right\}
\end{aligned}
$$

$i+j-(L-1) = A$ とおくと，

$$
\begin{aligned}
&\cos\left[\frac{\pi}{M}\left(k+\frac{1}{2}\right)A+(-1)^k\frac{\pi}{2}\right] \\
&= \cos\left[\frac{\pi}{M}\left(k+\frac{1}{2}\right)A\right]\cos\left[(-1)^k\frac{\pi}{2}\right]-\sin\left[\frac{\pi}{M}\left(k+\frac{1}{2}\right)A\right]\sin\left[(-1)^k\frac{\pi}{2}\right] \\
&= -(-1)^k\sin\left[\frac{\pi}{M}\left(k+\frac{1}{2}\right)A\right]
\end{aligned}
$$

$\boldsymbol{T}^\top \boldsymbol{T} = 2M\boldsymbol{I}$の証明

ここで，

$$\sum_{k=0}^{M-1} \cos\left[\frac{\pi}{M}\left(k+\frac{1}{2}\right)(i-j)\right] = M\delta_{ij} \tag{53}$$

$$\sum_{k=0}^{M-1} (-1)^k \sin\left[\frac{\pi}{M}\left(k+\frac{1}{2}\right)A\right] = 0 \tag{54}$$

（δ_{ij}：クロネッカーのデルタ）を示せば，

$$\left(\boldsymbol{T}^\top \boldsymbol{T}\right)_{ij} = 2M\delta_{ij}$$

となり命題が示せる．次ページから計算結果を載せる

$T^\top T = 2MI$ の証明

(53) 式を示す．$i = j$ のとき，$\sum_{k=0}^{M-1} \cos\left[\frac{\pi}{M}\left(k+\frac{1}{2}\right)(i-j)\right] = M$．
$i \neq j$ のとき，

$$
\begin{aligned}
&\sum_{k=0}^{M-1} \cos\left[\frac{\pi}{M}\left(k+\frac{1}{2}\right)(i-j)\right] \\
&= \frac{1}{2}\sum_{k=0}^{M-1}\left\{\exp\left[j\frac{\pi}{M}\left(k+\frac{1}{2}\right)(i-j)\right] + \exp\left[-j\frac{\pi}{M}\left(k+\frac{1}{2}\right)(i-j)\right]\right\} \\
&= \frac{1}{2}\sum_{k=0}^{M-1}\left\{W_{2M}^{-\left(k+\frac{1}{2}\right)(i-j)} + W_{2M}^{\left(k+\frac{1}{2}\right)(i-j)}\right\} \\
&= \frac{1}{2}\left\{W_{2M}^{-\frac{i-j}{2}}\sum_{k=0}^{M-1} W_{2M}^{-(i-j)k} + W_{2M}^{\frac{i-j}{2}}\sum_{k=0}^{M-1} W_{2M}^{(i-j)k}\right\} \\
&= \frac{1}{2}\left[\frac{W_{2M}^{-\frac{i-j}{2}}}{W_{2M}^{-(i-j)}-1}\left\{(-1)^{-(i-j)}-1\right\} + \frac{W_{2M}^{\frac{i-j}{2}}}{W_{2M}^{i-j}-1}\left\{(-1)^{i-j}-1\right\}\right]
\end{aligned}
$$

最後の式変形で等比級数の和 $\sum_{k=0}^{M-1} W_{2M}^{(i-j)k} = \frac{W_{2M}^{(i-j)M}-1}{W_{2M}^{i-j}-1} = \frac{(-1)^{i-j}-1}{W_{2M}^{i-j}-1}$ を使用

$\boldsymbol{T}^\top \boldsymbol{T} = 2M\boldsymbol{I}$ の証明

さらに式変形すると，

$$
\begin{aligned}
&\sum_{k=0}^{M-1} \cos\left[\frac{\pi}{M}\left(k+\frac{1}{2}\right)(i-j)\right] \\
&= \frac{1}{2}\left[\frac{W_{2M}^{-\frac{i-j}{2}}}{W_{2M}^{-(i-j)}-1}\left\{(-1)^{-(i-j)}-1\right\} + \frac{W_{2M}^{\frac{i-j}{2}}}{W_{2M}^{i-j}-1}\left\{(-1)^{i-j}-1\right\}\right] \\
&= \frac{W_{2M}^{-\frac{i-j}{2}}(W_{2M}^{i-j}-1)\left\{(-1)^{-(i-j)}-1\right\} + W_{2M}^{\frac{i-j}{2}}(W_{2M}^{-(i-j)}-1)\left\{(-1)^{i-j}-1\right\}}{2(W_{2M}^{-(i-j)}-1)(W_{2M}^{i-j}-1)} \\
&= \frac{(W_{2M}^{\frac{i-j}{2}} - W_{2M}^{-\frac{i-j}{2}})\left\{(-1)^{-(i-j)}-1\right\} + (W_{2M}^{-\frac{i-j}{2}} - W_{2M}^{\frac{i-j}{2}})\left\{(-1)^{i-j}-1\right\}}{2(W_{2M}^{-(i-j)}-1)(W_{2M}^{i-j}-1)} \\
&= \frac{(W_{2M}^{\frac{i-j}{2}} - W_{2M}^{-\frac{i-j}{2}})\left\{(-1)^{i-j}-1\right\} + (W_{2M}^{-\frac{i-j}{2}} - W_{2M}^{\frac{i-j}{2}})\left\{(-1)^{i-j}-1\right\}}{2(W_{2M}^{-(i-j)}-1)(W_{2M}^{i-j}-1)} \\
&= 0
\end{aligned}
$$

よって (53) 式が示された

$\boldsymbol{T}^\top \boldsymbol{T} = 2M\boldsymbol{I}$の証明

(54)式を示す．

$$
\begin{aligned}
&\sum_{k=0}^{M-1} (-1)^k \sin\left[\frac{\pi}{M}\left(k+\frac{1}{2}\right)A\right] \\
&= \frac{1}{j2}\sum_{k=0}^{M-1} (-1)^k \left\{\exp\left[j\frac{\pi}{M}\left(k+\frac{1}{2}\right)A\right] - \exp\left[-j\frac{\pi}{M}\left(k+\frac{1}{2}\right)A\right]\right\} \\
&= \frac{1}{j2}\sum_{k=0}^{M-1} (-1)^k \left\{W_{2M}^{-A\left(k+\frac{1}{2}\right)} - W_{2M}^{A\left(k+\frac{1}{2}\right)}\right\} \\
&= \frac{1}{j2}\left\{W_{2M}^{-\frac{A}{2}}\sum_{k=0}^{M-1}(-1)^k W_{2M}^{-Ak} - W_{2M}^{\frac{A}{2}}\sum_{k=0}^{M-1}(-1)^k W_{2M}^{Ak}\right\} \\
&= \frac{1}{j2}\left\{W_{2M}^{-\frac{A}{2}}\sum_{k=0}^{M-1} W_{2M}^{-(M+A)k} - W_{2M}^{\frac{A}{2}}\sum_{k=0}^{M-1} W_{2M}^{(M+A)k}\right\} \quad (\because (-1)^k = W_{2M}^{M} = W_{2M}^{-M}) \\
&= \frac{1}{j2}\left\{W_{2M}^{-\frac{A}{2}}\frac{(-1)^{-(M+A)}-1}{W_{2M}^{-(M+A)}-1} - W_{2M}^{\frac{A}{2}}\frac{(-1)^{M+A}-1}{W_{2M}^{M+A}-1}\right\} \quad (\because \text{等比級数の和の公式})
\end{aligned}
$$

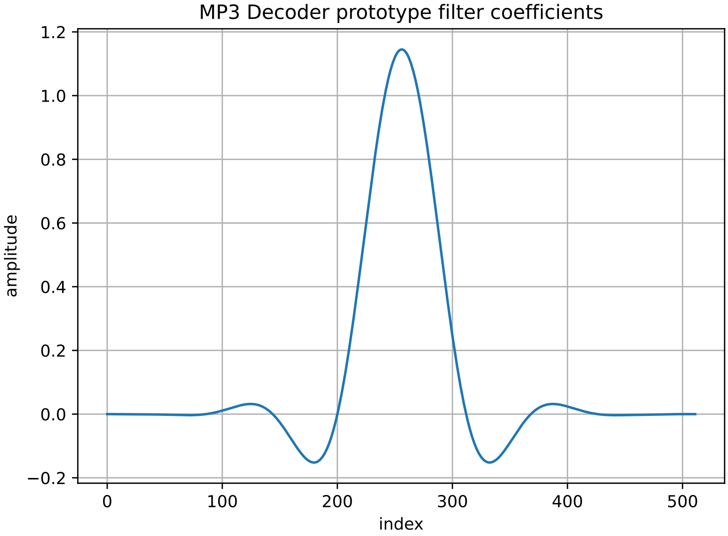
$\boldsymbol{T}^\top\boldsymbol{T} = 2M\boldsymbol{I}$の証明

さらに計算を進めると，

$$
\begin{aligned}
&\sum_{k=0}^{M-1}(-1)^k \sin\left[\frac{\pi}{M}\left(k+\frac{1}{2}\right)A\right] \\
&= \frac{W_{2M}^{-\frac{A}{2}}\left\{(-1)^{-(M+A)}-1\right\}(W_{2M}^{M+A}-1) - W_{2M}^{\frac{A}{2}}\left\{(-1)^{M+A}-1\right\}(W_{2M}^{-(M+A)}-1)}{j2(W_{2M}^{-(M+A)}-1)(W_{2M}^{(M+A)}-1)} \\
&= \frac{\left\{(-1)^{M+A}-1\right\}\left\{W_{2M}^{-\frac{A}{2}}(W_{2M}^{M+A}-1) - W_{2M}^{\frac{A}{2}}(W_{2M}^{-(M+A)}-1)\right\}}{j2(W_{2M}^{-(M+A)}-1)(W_{2M}^{(M+A)}-1)} \\
&= \frac{\left\{(-1)^{M+A}-1\right\}\left(W_{2M}^{M+\frac{A}{2}} - W_{2M}^{-\frac{A}{2}} - W_{2M}^{-M-\frac{A}{2}} + W_{2M}^{\frac{A}{2}}\right)}{j2(W_{2M}^{-(M+A)}-1)(W_{2M}^{(M+A)}-1)} \\
&= \frac{\left\{(-1)^{M+A}-1\right\}\left(-W_{2M}^{\frac{A}{2}} - W_{2M}^{-\frac{A}{2}} + W_{2M}^{-\frac{A}{2}} + W_{2M}^{\frac{A}{2}}\right)}{j2(W_{2M}^{-(M+A)}-1)(W_{2M}^{(M+A)}-1)} \\
&= 0
\end{aligned}
$$

よって，(54) 式が示された．

MP3のフィルタバンクの特性



- デコーダの係数．エンコーダの係数の 32 倍

電力相補条件の確認

MP3 のフィルタバンクでは

$$G_k(z) = \sum_{n=-\infty}^{\infty} p_0[2nM + k]z^{-n} = \sum_{n=0}^{7} p_0[64n + k]z^{-n} = \sum_{n=0}^{7} (-1)^n C_{64+k} z^{-n}$$

$$G_{M+k}(z) = \sum_{n=0}^{7} p_0[64n + 32 + k]z^{-n} = \sum_{n=0}^{7} (-1)^n C_{64n+32+k} z^{-n}$$

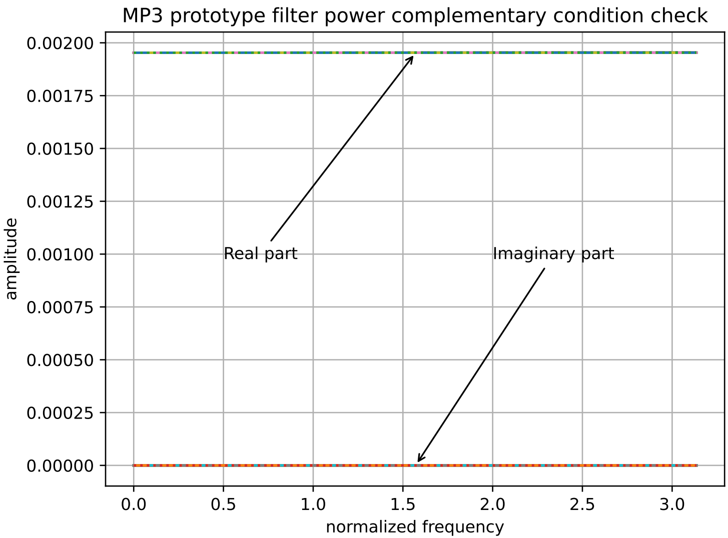
$$G_k(z^{-1}) = \sum_{n=-\infty}^{\infty} p_0[64n + k](z^{-1})^{-n} = \sum_{n=-\infty}^{\infty} p_0[-64n + k]z^{-n} = \sum_{n=-7}^{0} (-1)^n C_{-64n+k} z^{-n}$$

$$G_{M+k}(z^{-1}) = \sum_{n=0}^{7} p_0[-64n + 32 + k]z^{-n} = \sum_{n=0}^{7} (-1)^n C_{-64n+32+k} z^{-n}$$

$k = 0, \ldots, 31$ で実際に計算すると，

$$G_k(z^{-1})G_k(z) + G_{M+k}(z^{-1})G_{M+k}(z) \approx \frac{1}{512}$$

電力相補条件の確認（計算結果）



"ほぼ"完全再構成と言ってよい

MDCT・IMDCTによる再構成信号

式(10)に式(9)を代入すると，

$$
\begin{aligned}
y[n] &= \frac{2}{N}\sum_{k=0}^{N-1}\sum_{m=0}^{2N-1} x[m]\cos\left[\frac{\pi}{N}\left(k+\frac{1}{2}\right)\left(m+\frac{1}{2}+\frac{N}{2}\right)\right]\cos\left[\frac{\pi}{N}\left(k+\frac{1}{2}\right)\left(n+\frac{1}{2}+\frac{N}{2}\right)\right] \\
&= \frac{1}{N}\sum_{k=0}^{N-1}\sum_{m=0}^{2N-1} x[m]\left\{\cos\left[\frac{\pi}{N}\left(k+\frac{1}{2}\right)(n+m+1+N)\right]+\cos\left[\frac{\pi}{N}\left(k+\frac{1}{2}\right)(n-m)\right]\right\} \\
&= \frac{1}{N}\sum_{m=0}^{2N-1} x[m]\sum_{k=0}^{N-1}\left\{\cos\left[\frac{\pi}{N}\left(k+\frac{1}{2}\right)(n+m+1+N)\right]+\cos\left[\frac{\pi}{N}\left(k+\frac{1}{2}\right)(n-m)\right]\right\}
\end{aligned}
$$

ここで，

$$
I_n := \sum_{k=0}^{N-1}\cos\left[\frac{\pi}{N}\left(k+\frac{1}{2}\right)n\right] \tag{55}
$$

を計算していく

MDCT・IMDCTによる再構成信号

$$
\begin{aligned}
I_n &= \sum_{k=0}^{N-1} \cos\left[\frac{\pi}{N}\left(k+\frac{1}{2}\right)n\right] = \sum_{k=0}^{N-1} \cos\left[\frac{\pi}{2N}(2k+1)n\right] \\
&= \frac{1}{2}\sum_{k=0}^{N-1}\left\{\exp\left[j\frac{\pi}{2N}(2k+1)n\right] + \exp\left[-j\frac{\pi}{2N}(2k+1)n\right]\right\} \\
&= \frac{1}{2}\sum_{k=0}^{N-1}\left(W_{2N}^{-\frac{2k+1}{2}n} + W_{2N}^{\frac{2k+1}{2}n}\right) = \frac{1}{2}\left(W_{2N}^{-\frac{n}{2}}\sum_{k=0}^{N-1} W_{2N}^{-nk} + W_{2N}^{\frac{n}{2}}\sum_{k=0}^{N-1} W_{2N}^{nk}\right)
\end{aligned}
$$

ここで，

$$
\sum_{k=0}^{N-1} W_{2N}^{nk} = \sum_{k=0}^{N-1} \left(W_{2N}^{n}\right)^k = \begin{cases} \displaystyle\sum_{k=0}^{N-1} 1^k = N & (n \text{ が } 2N \text{ の倍数}) \\ \dfrac{1\left\{\left(W_{2N}^{n}\right)^N - 1\right\}}{W_{2N}^{n} - 1} = \dfrac{(-1)^n - 1}{W_{2N}^{n} - 1} & (n \text{ が } 2N \text{ の倍数ではない}) \end{cases}
$$

から，場合分けして考える

MDCT・IMDCTによる再構成信号

n が $2N$ の倍数のとき，$n = 2Nm \ (m \in \mathbb{Z})$ と書けて，

$$I_n = \frac{N}{2}\left(W_{2N}^{-2Nm} + W_{2N}^{2Nm}\right) = \frac{N}{2}\left\{(-1)^{-m} + (-1)^{m}\right\}$$
$$= \begin{cases} N & (m\text{：偶数} \iff n = 0, \pm 4N, \pm 8N) \\ -N & (m\text{：奇数} \iff n = \pm 2N, \pm 6N, \pm 10N) \end{cases}$$

n が $2N$ の倍数ではないとき，

$$I_n = \frac{1}{2}\left\{W_{2N}^{-\frac{n}{2}}\frac{(-1)^{-n}-1}{W_{2N}^{-n}-1} + W_{2N}^{\frac{n}{2}}\frac{(-1)^{n}-1}{W_{2N}^{n}-1}\right\}$$
$$= \frac{W_{2N}^{-\frac{n}{2}}\left\{(-1)^{-n}-1\right\}(W_{2N}^{n}-1) + W_{2N}^{\frac{n}{2}}\left\{(-1)^{n}-1\right\}(W_{2N}^{-n}-1)}{2(W_{2N}^{-l}-1)(W_{2N}^{n}-1)}$$
$$= \frac{W_{2N}^{-\frac{n}{2}}\left\{(-1)^{-n}W_{2N}^{n} - (-1)^{-n} - W_{2N}^{n} + 1\right\} + W_{2N}^{\frac{n}{2}}\left\{(-1)^{n}W_{2N}^{-n} - (-1)^{n} - W_{2N}^{-n} + 1\right\}}{2(W_{2N}^{-l}-1)(W_{2N}^{n}-1)}$$
$$= \frac{(-1)^{-n}W_{2N}^{\frac{n}{2}} - (-1)^{-n}W_{2N}^{-\frac{n}{2}} - W_{2N}^{\frac{n}{2}} + W_{2N}^{-\frac{n}{2}} + (-1)^{n}W_{2N}^{-\frac{n}{2}} - (-1)^{n}W_{2N}^{\frac{n}{2}} - W_{2N}^{-\frac{n}{2}} + W_{2N}^{\frac{n}{2}}}{2(W_{2N}^{-l}-1)(W_{2N}^{n}-1)}$$
$$= 0$$

MDCT・IMDCTによる再構成信号

まとめると，

$$I_n = \begin{cases} N & (n = 0, \pm 4N, \pm 8N, \ldots) \\ -N & (n = \pm 2N, \pm 6N, \pm 10N, \ldots) \\ 0 & (\text{otherwise}) \end{cases} \tag{56}$$

となる．式 (56) を使えば，

$$\sum_{m=0}^{2N-1} x[m] I_{m-n} = \begin{cases} x[n] I_0 = N x[n] & (n = 0, \ldots, N-1) \\ x[n] I_0 = N x[n] & (n = N, \ldots, 2N-1) \end{cases}$$

$$\sum_{m=0}^{2N-1} x[m] I_{n+m+1+N} = \begin{cases} x[N-1-n] I_{2N} = -N x[N-1-n] & (n = 0, \ldots, N-1) \\ x[3N-1-n] I_{4N} = N x[3N-1-n] & (n = N, \ldots, 2N-1) \end{cases}$$

だから，

$$\begin{aligned} y[n] &= \frac{1}{N} \sum_{m=0}^{2N-1} x[m] (I_{n+m+1+N} + I_{m-n}) \\ &= \begin{cases} x[n] - x[N-1-n] & (n = 0, \ldots, N-1) \\ x[n] + x[3N-1-n] & (n = 0, \ldots, N-1) \end{cases} \end{aligned}$$

となる．

知覚エントロピー導出

聴覚しきい値 T_b に量子化分散（パワー）を合わせる

- 各周波数ビンを量子化ステップ幅 Δ_b で一様量子化 $\Rightarrow$ 量子化誤差が一様に生起するなら，量子化誤差分散はビンあたり $\frac{\Delta_b^2}{12}$
- パーティション b に含まれるビン数を n_b とすると，ビンあたりの聴覚しきい値は $\frac{T_b}{n_b}$

これらを等しいとすると，

$$\frac{T_b}{n_b} = \frac{\Delta_b^2}{12} \implies \Delta_b = \sqrt{\frac{12T_b}{n_b}} \tag{57}$$

知覚エントロピー導出

振幅スペクトル（$\approx$ DCT スペクトル）の符号化に必要なビット数を考える

- ビンあたりの平均振幅スペクトルは $\sqrt{I_b/n_b}$
- 平均振幅スペクトルを整数量子化．振幅スペクトルは $[-\lceil\sqrt{I_b/n_b}\rceil, \lceil\sqrt{I_b/n_b}\rceil]$ にある整数 $\Rightarrow$ 整数個数は高々 $2\lceil\sqrt{I_b/n_b}\rceil + 1$（$+1$ は 0）

整数個数をステップ幅 Δ_b で割ることで符号化シンボル個数になる．符号化に必要なビット幅 (bit) は，

$$\log_2 \frac{2\lceil\sqrt{I_b/n_b}\rceil + 1}{\Delta_b} \tag{58}$$

知覚エントロピー導出

(58) 式を変形

$$\log_2 \frac{2\lceil\sqrt{I_b/n_b}\rceil + 1}{\Delta_b} \approx \log_2 \frac{2\sqrt{I_b/n_b}}{\Delta_b} \quad (I_b\text{ は大と仮定})$$
$$= \frac{1}{2}\log_2 \frac{4I_b}{n_b\Delta_b^2} = \frac{1}{2}\log_2 \frac{I_b}{3T_b}$$
$$\propto \frac{1}{2}\log \frac{I_b}{3T_b} = \frac{1}{2}\log \frac{I_b}{T_b} + \text{const.}$$

全周波数ビンの符号化に必要なビット数は

$$\sum_b n_b \log \frac{I_b}{T_b} + \text{const.} \tag{59}$$

に比例

知覚エントロピー導出

規格では知覚エントロピー PE を

$$\mathrm{PE} = \sum_b n_n \log \frac{I_b + 1}{T_b} \tag{60}$$

で定義．dist10 では無音領域で $T_b = 0$ となるため，

$$\mathrm{PE} = \sum_b n_n \log \frac{I_b + 1}{T_b + 1} \tag{61}$$

で計算 [*7]

[*7] 有音区間で I_b, T_b は大きいので，近似としては問題ない想定