

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

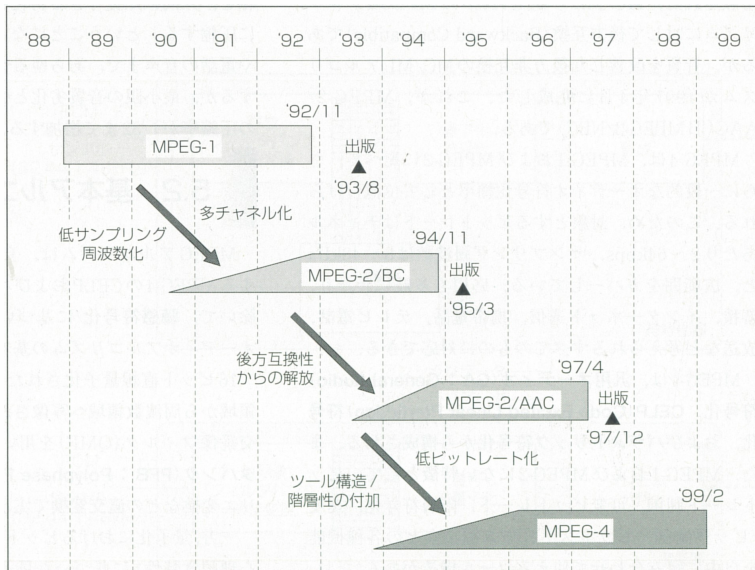
- フィルタバンク
- MDCT (Modified Discrete Cosine Transform)
- エイリアス削減バタフライ演算

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

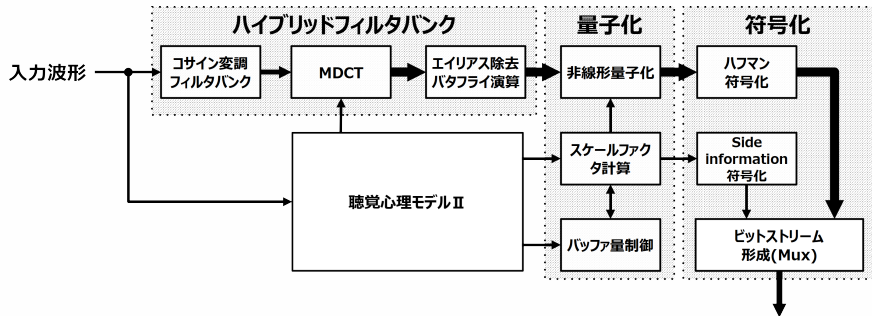
MPEG1の要素技術

- ▶ 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

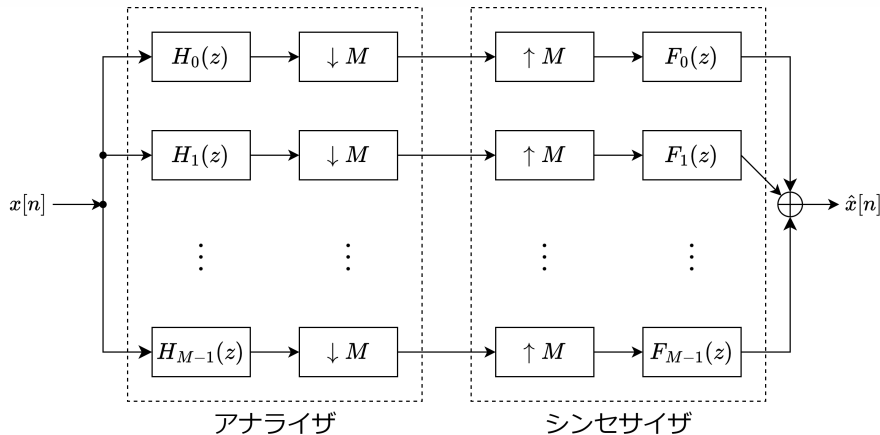
- フィルタバンク
- MDCT (Modified Discrete Cosine Transform)
- エイリアス削減バタフライ演算

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：

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

$$f_k[n] = h_k[L-1-n] \quad (2)$$

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！)

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

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

$$p_0[n] = \begin{cases} -C_n & [n/64] \text{ が偶数} \\ C_n & [n/64] \text{ が奇数} \end{cases} \quad (5)$$

C_n ($n = 0, \dots, 511$) 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 $k = 0, \dots, 15$ の周波数特性



フィルタバンクの周波数特性

バンク $k = 16, \dots, 31$ 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

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

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、

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

ここで、

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

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

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

フィルタバンクは完全再構成か？

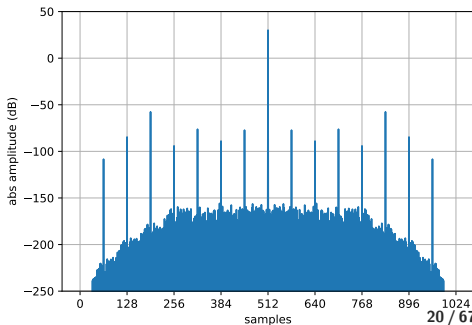
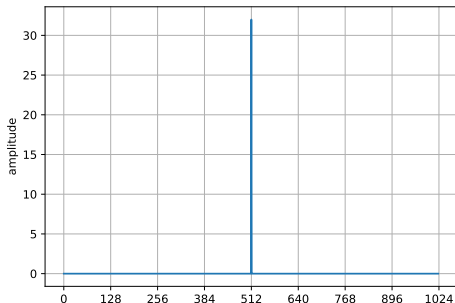
$$\begin{aligned}\hat{x}[n] &= \sum_{k=0}^{31} z_k[n] = \sum_{k=0}^{31} \left(\sum_{i=0}^{511} f_k[i] y_k[n-i] \right) \\&= \sum_{k=0}^{31} \left\{ \sum_{i=0}^{511} f_k[i] \left(\sum_{j=0}^{511} h_k[j] x[n-i-j] \right) \right\} \\&= \sum_{k=0}^{31} \sum_{i=0}^{511} \sum_{j=0}^{511} f_k[i] h_k[j] x[n-i-j]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

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)

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

IMDCT (Inverse MDCT)

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

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

MDCT と完全再構成条件

(10) 式に (9) 式を代入して整理すると

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

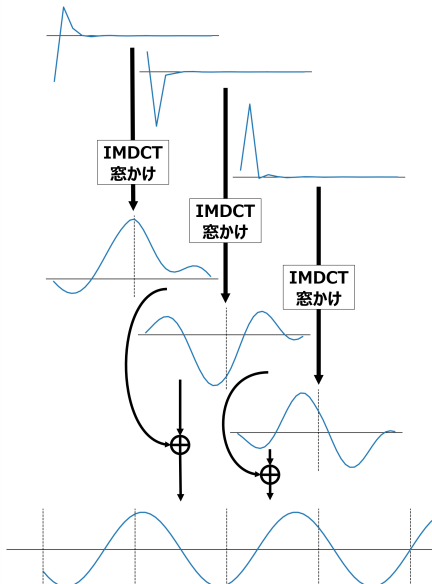
となる（証明は補足）．

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



エンコード



デコード

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，

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

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

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

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

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

が成り立つとする．窓かけした信号 $g_m[n]$ を

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

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、

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

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，

$$w[n]^2 + w[n + N]^2 = 1 \quad (17)$$

$$w[n + N]w[2N - 1 - n] = w[n]w[N - 1 - n] \quad (18)$$

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

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}

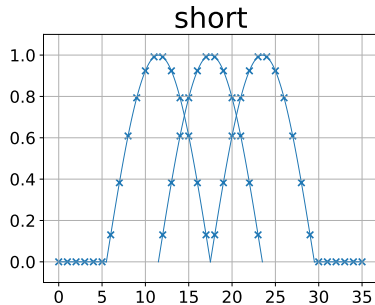
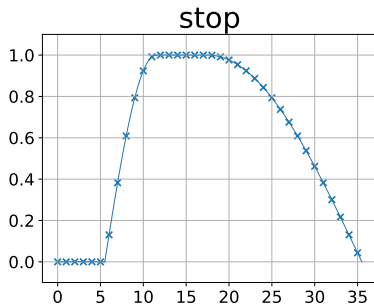
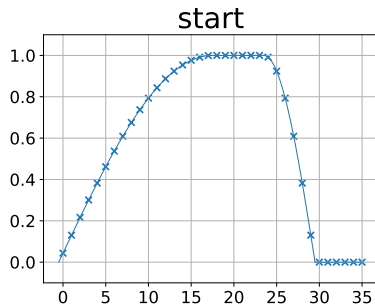
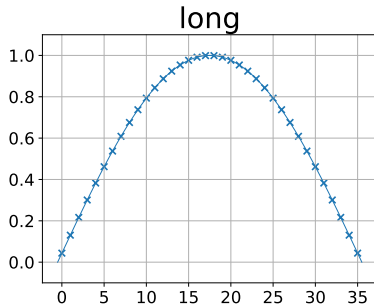
^{*1}逆に，式 (18) を満たしても対称とは限らない． $w[n] = \frac{w[n+N]w[2N-1-n]}{w[N-1-n]}$
だが，一般に $w[n + N] \neq w[N - 1 - n]$ だから $w[n] \neq w[2N - 1 - n]$ ．

MP3 と MDCT – 4 種類の窓関数

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

▶ 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



MP3 と Princen–Bradley 条件

サイン窓

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

は Princen–Bradley 条件を満たす。

(証明) 式 (17) は：

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

式 (18) は、サイン窓が対称であることより示される：

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

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 \rightarrow start
 2. stop \rightarrow long : 1. の対称ケース
 3. start \rightarrow short
 4. short \rightarrow 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

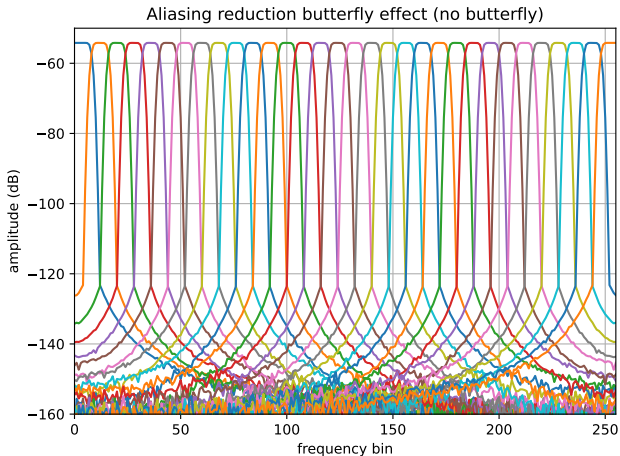
3. start \rightarrow short

- ▶ $n = 0, \dots, 5$: start 窓が1, short 窓が0なので完全再構成
- ▶ $n = 6, \dots, 11$: サイン窓になっているため完全再構成
- ▶ $n = 12, \dots, 17$: 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,

$$X_{18k-i} \leftarrow cs_i X_{18k-i} - ca_i X_{18k+i+1} \quad (20)$$

$$X_{18k+i+1} \leftarrow ca_i X_{18k-i} + cs_i X_{18k+i+1} \quad (21)$$

$$k = 1, \dots, 31, \quad i = 0, \dots, 7$$

- ▶ 係数 cs_i, ca_i の定義

$$cs_i := \frac{1}{\sqrt{1 + c_i^2}}, \quad ca_i := \frac{c_i}{\sqrt{1 + c_i^2}} \quad (22)$$

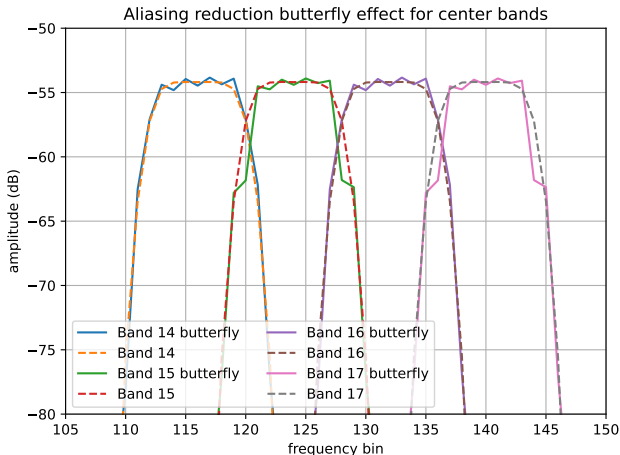
c_i ($i = 0, \dots, 7$) は規格で設定



[6] より引用

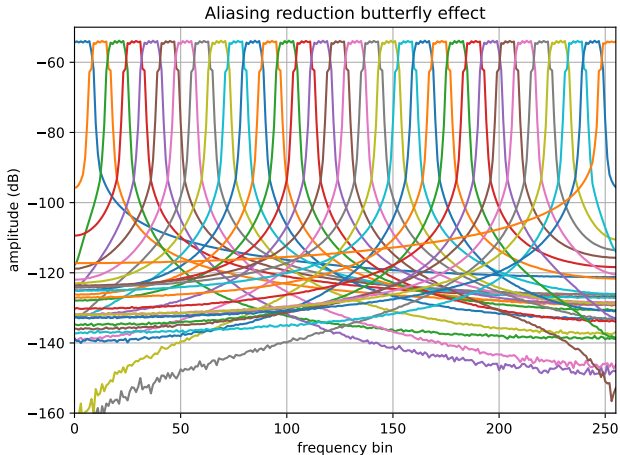
エイリアス削減バタフライ演算

- ▶ 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

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

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

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

- ▶ 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)

*2round を除けば，逆量子により元に戻る

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

非線形量子化

ブロック長が 12(short) のとき,

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

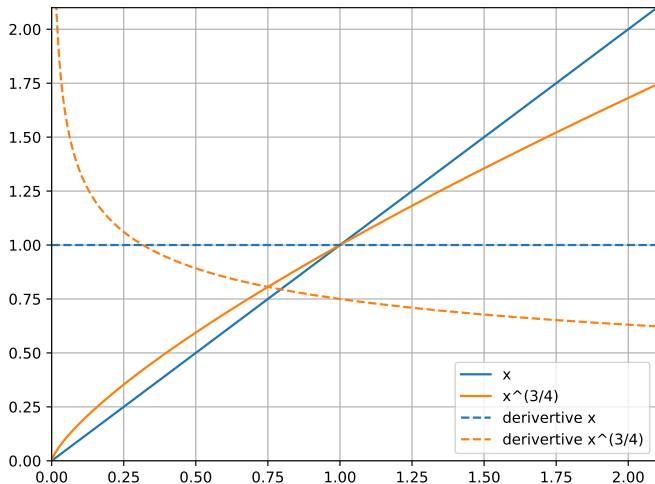
$$\hat{X}_i = \text{sign}(X_i^q) |X_i^q|^{\frac{4}{3}} G_i^s \quad (27)$$

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

- ▶ 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

$x, x^{\frac{3}{4}}$ とそれらの微分



- ▶ $x^{\frac{3}{4}}$ は0近傍で高感度 (logに近い)
- ▶ 0近傍は細かく, ≈ 0.3 以上は荒く量子化

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

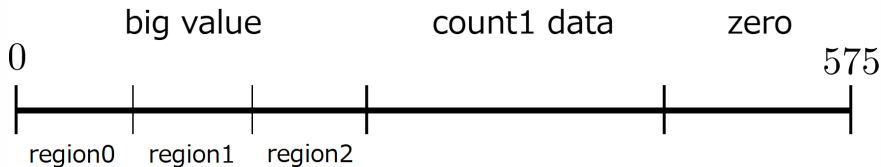
3. 量子化

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を兼用

- ▶ 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

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

として符号化．非0の場合に符号 bit を付加

- ▶ 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

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

として符号化 ($s(x) : x \neq 0$ なら 1, $x = 0$ なら 0)．
非0の場合に符号 bit を付加

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

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・ブロックタイプ判定
 8. 信号対マスク比 (SMR) 計算

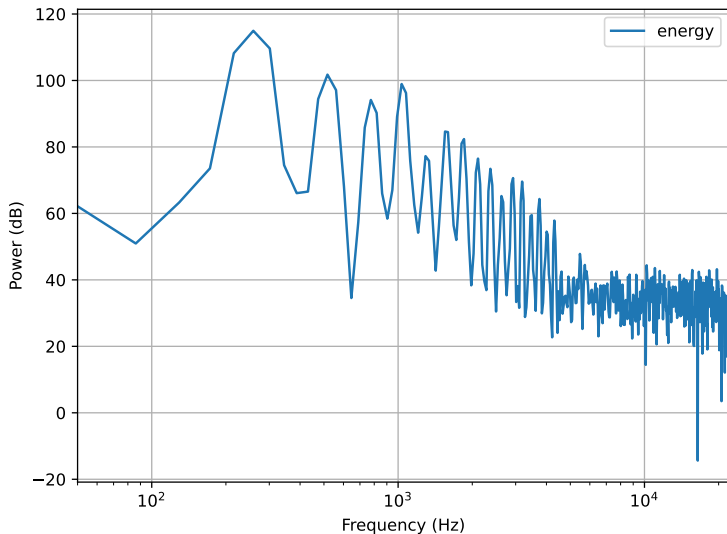
^{*3}Signal-to-Masking Ratio

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 $w^l(\text{long})$, $w^{sb}(\text{short}, b = 1, 2, 3)$ と表記

フレーム切り出し・窓かけ・FFT



ピアノ ($F_0=220\text{Hz}$) の long のエネルギー (パワースペクトル)

Unpredictability 計算

Unpredictability cw = 予測しずらさの尺度

$$cw[n] = \begin{cases} |w^l[n] - w^{l*}[n]| & 0 \leq n \leq 5 \\ \frac{|w^l[n]| + |w^{l*}[n]|}{|w^{s2}[j] - w^{s*}[j]|} & j = \lfloor (n+2)/4 \rfloor, 6 \leq n \leq 205 \\ 0.4 & 206 \leq n \leq 512 \end{cases} \quad (31)$$

- ▶ w^{l*}, w^{s*} : 振幅と位相を直線予測したスペクトル
- ▶ 直線予測の式 ($w^{l'}$ は前フレームの w^l) :

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

$$w^{s*}[n] = (2|w^{s1}[n]| - |w^{s3}[n]|) \exp[j(2 \arg(w^{s1}[n]) - \arg(w^{s3}[n]))] \quad (33)$$

- ▶ 予測が当たれば $cw[n] = 0$, 外れると $cw[n] = 1$

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

- ▶ eb^l, eb^s : パーティション内のエネルギーを合算
- ▶ cb : Unpredictability で重みづけしたエネルギーを合算

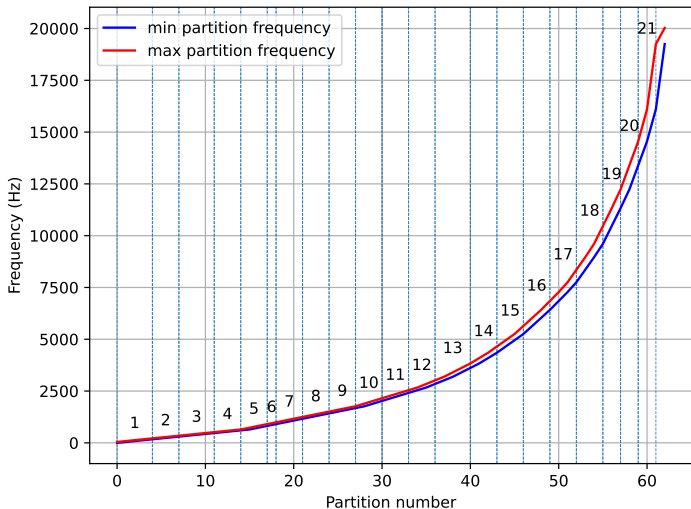
long, short のパーティション b を P_b^l, P_b^s と書くと,

$$eb^l[b] = \sum_{n \in P_b^l} |w^l[n]|^2 \quad (34)$$

$$cb[b] = \sum_{n \in \min P_b^l} cw[n] |w^l[n]|^2 \quad (35)$$

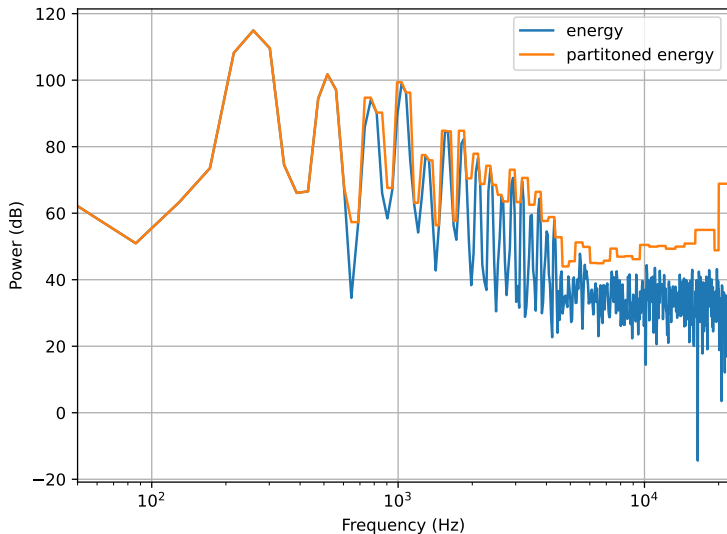
$$eb^s[b] = \sum_{n \in \min P_b^s} |w^s[n]|^2 \quad (36)$$

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 パーティション. Bark スケール細分化とみなせる. 数値ラベルはスケールファクタ番号.

パーティションごとのエネルギー計算



ピアノ ($F_0=220\text{Hz}$) 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

エネルギーに広がり関数 (Spreading function) SF_b^l, SF_b^s を畳み込み，調波成分のマスキングを考慮

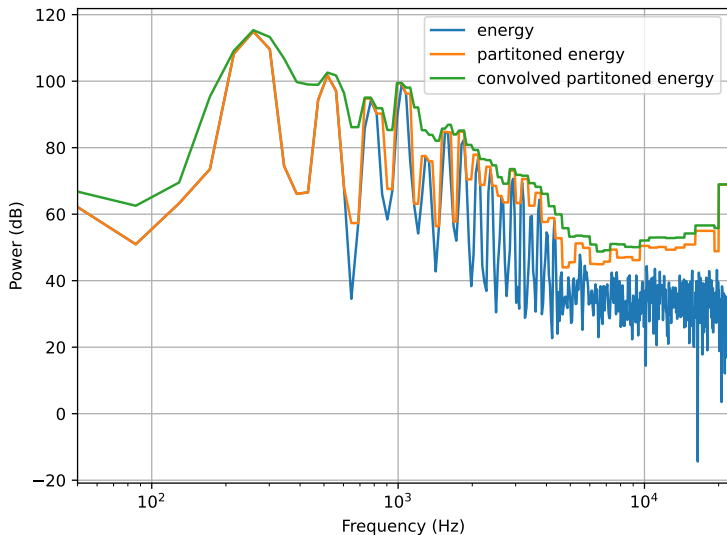
- ▶ $ecb^l, ecb^s : SF_b^l, SF_b^s$ を畳みこんだ eb^l, eb^s
- ▶ $ctb : SF_b^l$ を畳みこんだ cb

$$ecb^l[b] = \sum_k SF_b^l[k] eb^l[k] \quad (37)$$

$$ctb[b] = \sum_k SF_b^l[k] cb[k] \quad (38)$$

$$ecb^s[b] = \sum_k SF_b^s[k] eb^s[k] \quad (39)$$

広がり関数の畳み込み



ピアノ ($F_0=220\text{Hz}$) の広がり関数を畳みこんだエネルギー ecb^l

Schroeder の広がり関数 $SF_{\text{schroeder}}$ [7]

バークスケールに変換した周波数の差

$\Delta z = z(f_{\text{マスクー}}) - z(f_{\text{マスカー}})$ を用いて,

$$\begin{aligned} 10 \log_{10} SF_{\text{schroeder}}(\Delta z) \\ = 15.81 + 7.5(\Delta z + 0.474) - 17.5 \sqrt{1 + (\Delta z + 0.474)^2} \end{aligned} \quad (40)$$

(40) 式で求まる値は dB スケール. $10^{\frac{10 \log_{10} SF_{\text{schroeder}}(\Delta z)}{10}}$
で振幅スケールに変換

MP3の広がり関数 SF

Schroeder の広がり関数を修正

$$t_x = \begin{cases} 3\Delta z & \Delta z \leq 0 \\ 1.5\Delta z & \Delta z > 0 \end{cases} \quad (41)$$

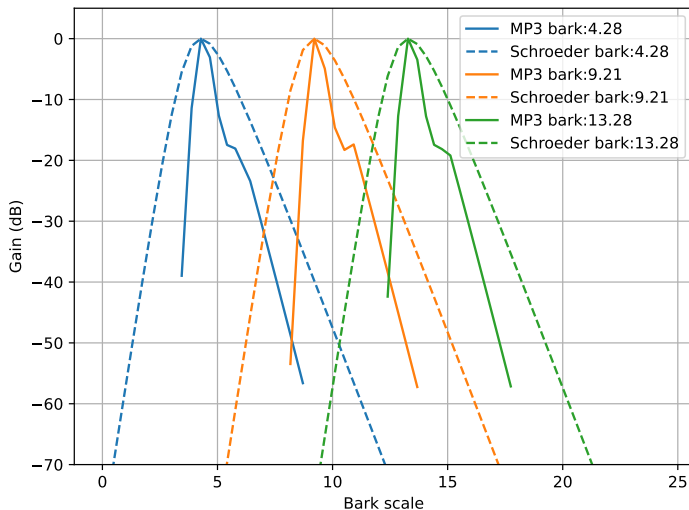
$$x = 8 \min\{(t_x - 0.5)^2 - 2(t_x - 0.5), 0\} \quad (42)$$

$$y = 15.811389 + 7.5(t_x + 0.474) - 17.5\sqrt{1 + (t_x + 0.474)^2} \quad (43)$$

として,

$$10 \log_{10} SF(\Delta z) = \begin{cases} x + y & y \geq -60 \\ 0 & \text{otherwise} \end{cases} \quad (44)$$

広がり関数の概形



いくつかのバーク値での広がり関数。破線は Schroeder[7]。MP3
では-60dB 以下のスケールは 0 に丸め込むため線が途切れている。

ノイズ許容レベル計算

パーティション b の量子化ノイズ許容レベルを計算^{*4}

tbb 計算 $tbb = 0$ はノイズ, $tbb = 1$ は調波成分を示す尺度

$$cbb = \log \left(\max \left\{ \frac{ctb[b]}{ecb^l[b]}, 0.01 \right\} \right) \quad (45)$$

$$tbb = \min \{ 1.0, \max \{ 0.0, -0.299 - 0.43cbb \} \} \quad (46)$$

SNR 計算 マスクの加重和で SNR(dB) を計算. 29.0 は調波によるノイズのマスク, 6.0 はノイズによる純音のマスク

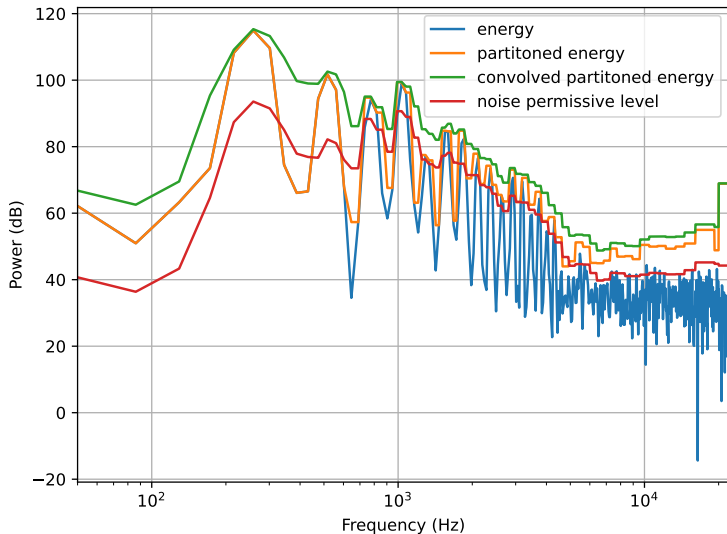
$$snr = \max \{ \minval_b, 29.0tbb + 6.0(1 - tbb) \} \quad (47)$$

許容レベル計算 SNR にエネルギーを乗じて許容レベル nb^l を得る. 広がり関数で増えたエネルギーを戻す

$$nb^l[b] = \frac{ecb^l[b]}{\sum_k SF_b[k]} \times 10^{-\frac{snr}{10}} \quad (48)$$

^{*4}short は省略. SNR をテーブル引きして計算する以外同様のため

ノイズ許容レベル計算



ピアノ ($F_0=220\text{Hz}$) のノイズ許容レベル nb^l

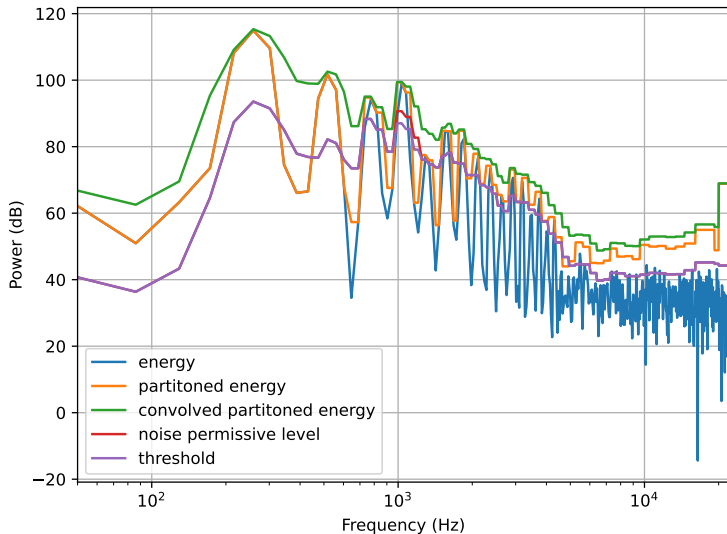
聴覚しきい値計算

前フレームのノイズ許容レベルと最小可聴パワーを考慮し，これを聴覚しきい値 thr^l とする

$$thr^l[b] = \max \{ qthr_b, \min \{ 2nb^{l'}[b], 16nb^{l''}[b], nb^l[b] \} \} \quad (49)$$

- ▶ $qthr_b$: 最小可聴パワー
- ▶ $nb^{l'}, nb^{l''}$: 前とさらにその前のノイズ許容レベル
 - ▶ 前フレームのノイズを残す．プリエコー対策

聴覚しきい値計算



ピアノ ($F_0=220\text{Hz}$) の聴覚しきい値 thr^l

知覚エントロピー計算

知覚エントロピー PE を以下で計算

$$PE = \sum_b |P_b^l| \log \frac{eb^l[b] + 1}{thr^l[b]} \quad (50)$$

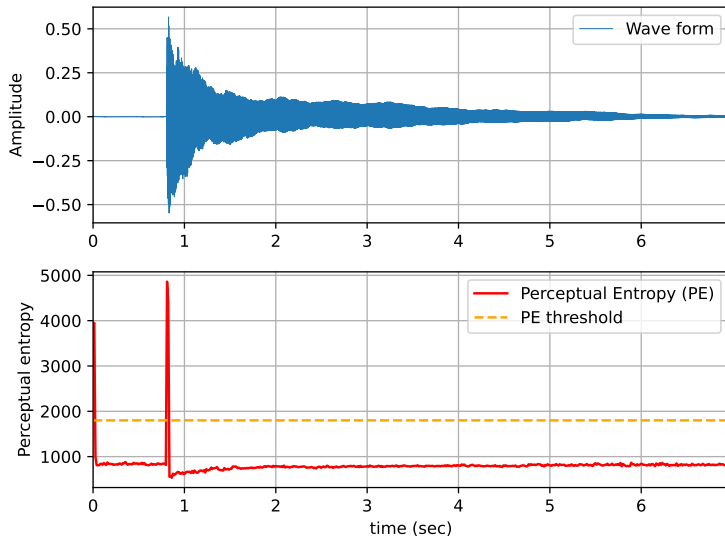
フレームの符号化に必要なビット数の目安^{*5}

- ▶ $PE \geq 1800$ ならば short ブロックと判定．さもなければ long ブロックと判定^{*6}

^{*5}導出は補足 3 参照

^{*6}dist10 では以前のブロック判定結果を保持．判定とブロックの状態遷移に従い結果を start,stop に変更

知覚エントロピー計算



ピアノ ($F_0=220\text{Hz}$) の波形と聴覚エントロピー PE の変化

信号対マスク比 (SMR) 計算

(聴覚しきい値)/(信号パワー) を計算．同時に周波数の分割をクリティカルバンドに揃える．クリティカルバンド k に属すパーティションの集合を P_k と書くと，

$$en[k] = w_1[k]eb[\min P_k] + w_2[k]eb[\max P_k] + \sum_{P \in P_k} \sum_{b \in P} eb[b] \quad (51)$$

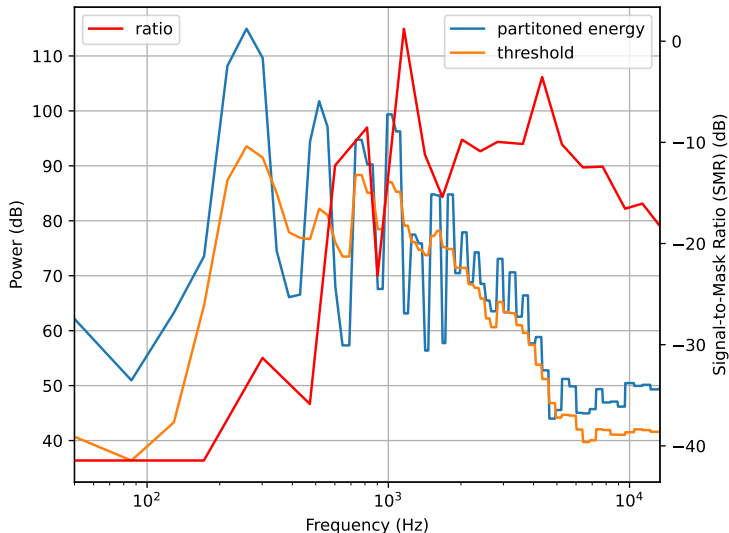
$$thm[k] = w_1[k]thr[\min P_k] + w_2[k]thr[\max P_k] + \sum_{P \in P_k} \sum_{b \in P} thr[b] \quad (52)$$

として (w_1, w_2 は隣接バンド間の重み定数^{*7})，

$$\text{ratio}[k] = \begin{cases} \frac{thm[k]}{en[k]} & en[k] > 0 \\ 0 & \text{otherwise} \end{cases} \quad (53)$$

^{*7} $w_1[k] + w_2[k-1] = 1$ が成立．導出方法は不明...

信号対マスク比 (SMR) 計算



ピアノ ($F_0=220\text{Hz}$) の信号対マスク比 ratio

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 [17]

参考文献の紹介

- ▶ [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, 理解しづらい.
- ▶ [5, 16] : エイリアス削減の論文
- ▶ [17] : ノイズマスキングを用いた知覚エントロピーの考え方についての原論文
- ▶ [7] : マスキング曲線のモデル化とそれを用いた音声符号化の原論文

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] Manfred R Schroeder, Bishnu S Atal, and JL Hall. "Optimizing digital speech coders by exploiting masking properties of the human ear". In: *The Journal of the Acoustical Society of America* 66.6 (1979), pp. 1647–1652.
- [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7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.
- [16] 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.

参考文献 III

- [17] James D Johnston. “Estimation of perceptual entropy using noise masking criteria”. In: *Icassp-88., international conference on acoustics, speech, and signal processing*. IEEE. 1988, pp. 2524–2527.

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 [17]

完全再構成

遅延・定数倍を除き入出力が一致すること：

$$\hat{x}[n] = cx[n - n_0], \quad c \neq 0 \quad (54)$$

これは z 領域で，

$$\hat{X}(z) = cz^{-n_0}X(z) \quad (55)$$

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

$$E_{k,l}(z) = \sum_{n=-\infty}^{\infty} h_k[nM + l] z^{-n} \quad (56)$$

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

$$R_{k,l}(z) = \sum_{n=-\infty}^{\infty} f_k[nM + M - 1 - l] z^{-n} \quad (57)$$

ポリフェーズ行列表現

(56) 式を l について並べ，行列表現すると

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

(57) 式も同様にして，以下のように書ける

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

ポリフェーズ行列表現

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

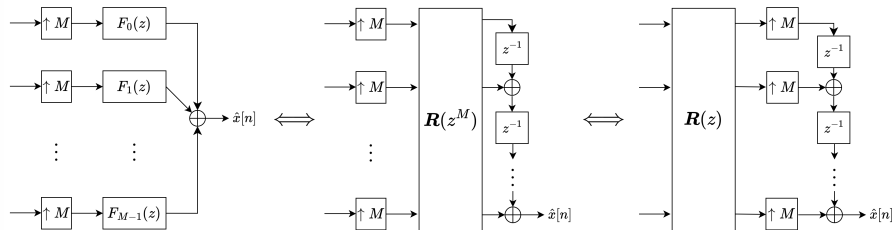
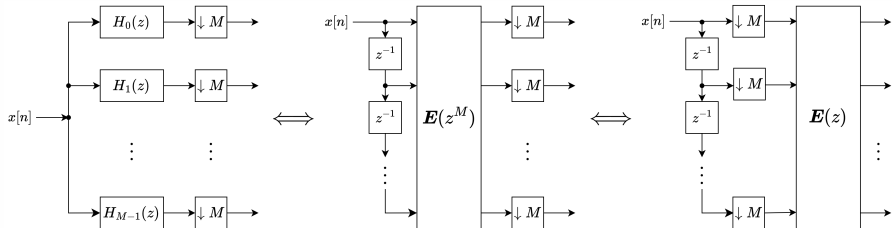
$$h(z) = E(z)e(z) \quad (58)$$

$$f(z)^\top = z^{-(M-1)}\tilde{e}(z)R(z) \quad (59)$$

とまとめられる． $E(z), R(z)$ をポリフェーズ行列という

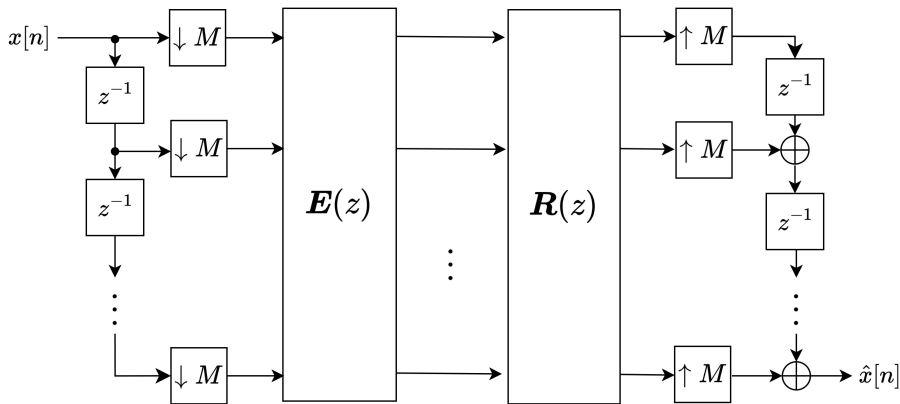
ポリフェーズ行列表現

$E(z), R(z)$ 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

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

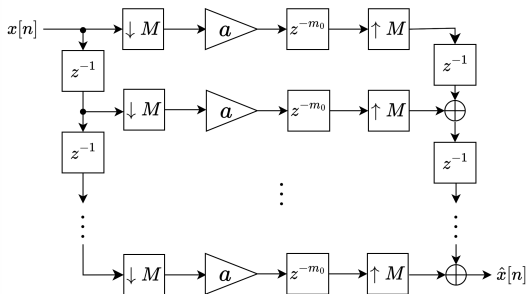


完全再構成 M 分割フィルタバンク

$$\mathbf{R}(z)\mathbf{E}(z) = az^{-m_0}\mathbf{I} \quad (a \neq 0, m_0 \in \mathbb{N}) \quad (60)$$

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, $\hat{X}(z) = aMz^{-(M-1+K)}X(z)$



$\mathbf{R}(z)\mathbf{E}(z) = az^{-m_0}\mathbf{I}$ 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

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

▶ しかし, $E(z)^{-1}$ の計算に問題を孕む.

代わりに, $E(z)$ がパラユニタリ^{*8}

$$\tilde{E}(z)E(z) = dI, \quad d \neq 0 \quad (61)$$

ならば,

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

とするとフィルタバンクは完全再構成.

^{*8} $\tilde{E}(z) = E_*(z^{-1})$ 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}^{*9} より

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

これを (1) 式に代入すると，

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

$$^{*9} W_{2M} = \exp \left(-j \frac{2\pi}{2M} \right) = \exp \left(-j \frac{\pi}{M} \right)$$

コサイン変調フィルタバンク

さらに (69) 式を代入すると *10

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

$${}^{*10} \left(W_{2M}^{\pm(k+\frac{1}{2})} \right)^{2M} = \exp \left[\mp j \frac{\pi}{M} \left(k + \frac{1}{2} \right) 2M \right] = \exp[\mp j\pi(2k+1)] = -1 \text{ を}$$

使用

コサイン変調フィルタバンク

さらに変形すると

$$H_k(z) = \sum_{l=0}^{M-1} z^{-l} \{t_{k,l}G_l(-z^{2M}) + z^{-M}t_{k,M+l}G_{M+l}(-z^{2M})\} \quad (62)$$

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

(56) 式と (62) 式を見比べると,

$$E_{k,l}(z) = t_{k,l}G_l(-z^2) + z^{-1}t_{k,M+l}G_{M+l}(-z^2) \quad (63)$$

コサイン変調フィルタバンク

(63) 式より，アナライザのポリフェーズ行列は，

$$\begin{aligned} \mathbf{E}(z) &= \underbrace{\begin{bmatrix} t_{0,0} & \cdots & t_{0,2M-1} \\ t_{1,0} & \cdots & t_{1,2M-1} \\ \vdots & \cdots & \vdots \\ t_{M-1,0} & \cdots & t_{M-1,2M-1} \end{bmatrix}}_{\mathbf{T}} \begin{bmatrix} G_0(-z^2) & \cdots & \cdots \\ & \ddots & \ddots & G_{M-1}(-z^2) \\ z^{-1}G_M(-z^2) & \cdots & \cdots \\ & & \ddots & \vdots \\ & & & z^{-1}G_{2M-1}(-z^2) \end{bmatrix} \\ &= \mathbf{T} \begin{bmatrix} \mathbf{G}_0(z^2) \\ z^{-1}\mathbf{G}_1(z^2) \end{bmatrix} \end{aligned} \quad (64)$$

$$\mathbf{G}_i(z) := \text{diag} \left[G_{Mi}(-z) \quad G_{Mi+1}(-z) \cdots G_{Mi+M-1}(-z) \right] \quad (65)$$

と書ける

コサイン変調フィルタバンク

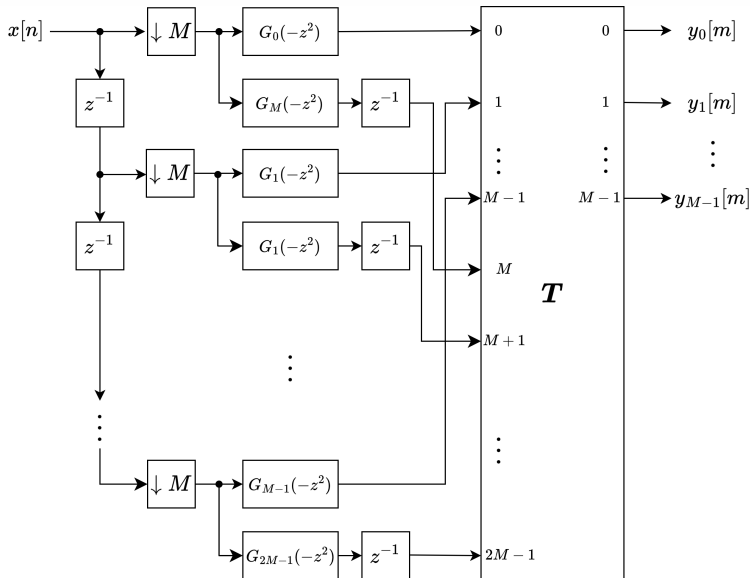
シンセサイザを構成する．フィルタ係数は実だから，

$$\tilde{\mathbf{E}}(z) = \mathbf{E}_*(z^{-1})^T = \begin{bmatrix} \mathbf{G}_0(z^{-1}) & z\mathbf{G}_1(z^{-1}) \end{bmatrix} \mathbf{T}^T \quad (66)$$

$G_l(z)$ 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

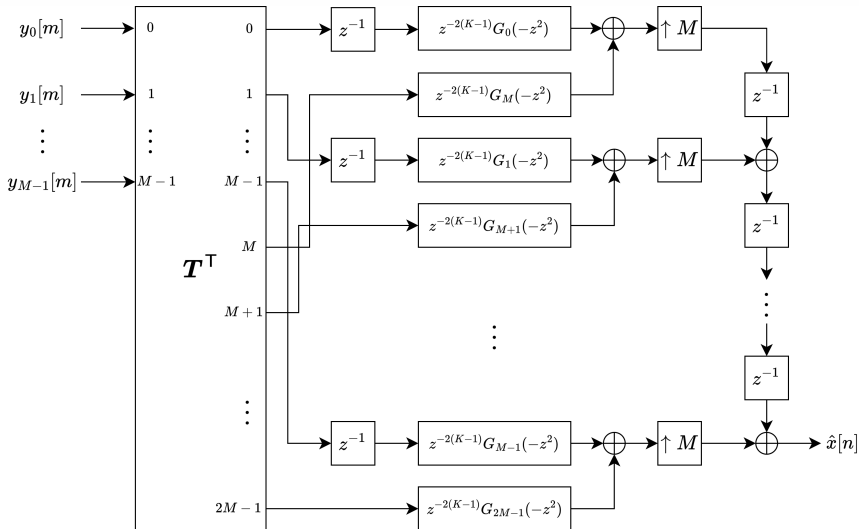
$$\begin{aligned} \mathbf{R}(z) &= z^{-(2K-1)} \tilde{\mathbf{E}}(z) \\ &= z^{-(2K-1)} \begin{bmatrix} \mathbf{G}_0(z^{-1}) & z\mathbf{G}_1(z^{-1}) \end{bmatrix} \mathbf{T}^T \end{aligned} \quad (67)$$

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．(64)，(67)式より，

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

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

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

$$G_k(z^{-1})G_k(z) + G_{M+k}(z^{-1})G_{M+k}(z) = \alpha \quad (68)$$

となること. ここで,

$$G_k(z) = \sum_{n=-\infty}^{\infty} p_0[2Mn + k]z^{-n} \quad (69)$$

$G_k(z)$ 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

$T^\top T = 2MI$ の証明

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

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

$$i + j - (L-1) = A \text{ とおくと,}$$

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

$T^{\top}T = 2MI$ の証明

ここで,

$$\sum_{k=0}^{M-1} \cos \left[\frac{\pi}{M} \left(k + \frac{1}{2} \right) (i - j) \right] = M \delta_{ij} \quad (70)$$

$$\sum_{k=0}^{M-1} (-1)^k \sin \left[\frac{\pi}{M} \left(k + \frac{1}{2} \right) A \right] = 0 \quad (71)$$

(δ_{ij} : クロネッカーのデルタ) を示せば,

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

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

(70) 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，

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

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

$T^T T = 2MI$ の証明

さらに式変形すると,

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

よって (70) 式が示された

$T^\top T = 2MI$ の証明

(71) 式を示す.

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

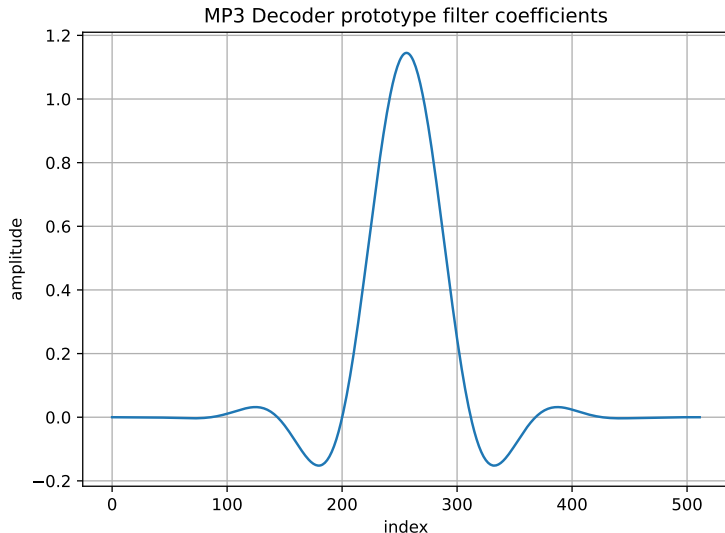
$T^\top T = 2MI$ の証明

さらに計算を進めると、

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

よって、(71) 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

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

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

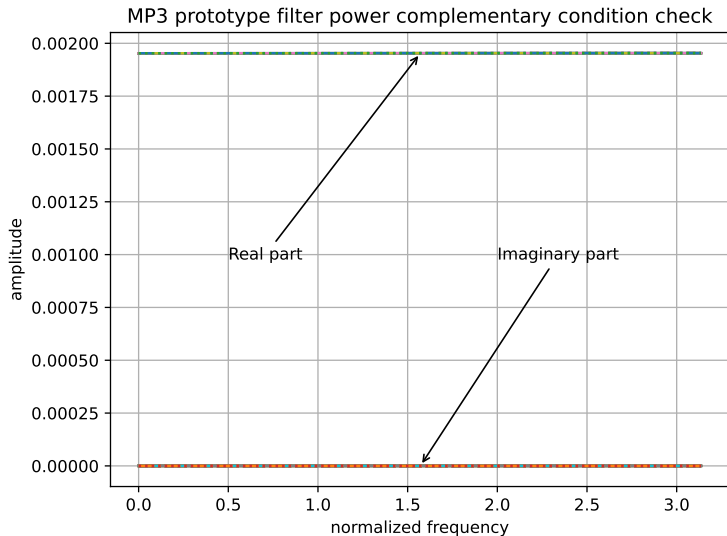
$$G_k(z^{-1}) = \sum_{n=-\infty}^{\infty} p_0[64n + k](z^{-1})^{-n} = \sum_{n=-\infty}^{\infty} p_0[-64n + k]z^{-n} = \sum_{n=-7}^0 (-1)^n C_{-64n+k} z^{-n}$$

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

$k = 0, \dots, 31$ 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,

$$I_n := \sum_{k=0}^{N-1} \cos \left[\frac{\pi}{N} \left(k + \frac{1}{2} \right) n \right] \quad (72)$$

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,

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

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,

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

n が $2N$ の倍数ではないとき,

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

MDCT・IMDCTによる再構成信号

まとめると,

$$I_n = \begin{cases} N & (n = 0, \pm 4N, \pm 8N, \dots) \\ -N & (n = \pm 2N, \pm 6N, \pm 10N, \dots) \\ 0 & (\text{otherwise}) \end{cases} \quad (73)$$

となる. 式 (73) を使えば,

$$\begin{aligned} \sum_{m=0}^{2N-1} x[m] I_{m-n} &= \begin{cases} x[n] I_0 = Nx[n] & (n = 0, \dots, N-1) \\ x[n] I_0 = Nx[n] & (n = N, \dots, 2N-1) \end{cases} \\ \sum_{m=0}^{2N-1} x[m] I_{n+m+1+N} &= \begin{cases} x[N-1-n] I_{2N} = -Nx[N-1-n] & (n = 0, \dots, N-1) \\ x[3N-1-n] I_{4N} = Nx[3N-1-n] & (n = N, \dots, 2N-1) \end{cases} \end{aligned}$$

だから,

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

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，

$$\frac{T_b}{n_b} = \frac{\Delta_b^2}{12} \implies \Delta_b = \sqrt{\frac{12T_b}{n_b}} \quad (74)$$

知覚エントロピー導出

スペクトルの符号化に必要なビット数を導出

- ▶ ビンあたりの平均振幅スペクトルは $\sqrt{I_b/n_b}$
- ▶ スペクトルは $[-\lceil \sqrt{I_b/n_b} \rceil, \lceil \sqrt{I_b/n_b} \rceil]$ の範囲 \Rightarrow
符号化範囲の幅は $2\sqrt{I_b/n_b}$

範囲をステップ幅 Δ_b で割ると符号化シンボル個数になる．実虚両軸で符号化に必要なビット数は

$$\log_2 \frac{2\sqrt{I_b/n_b}}{\Delta_b} + \log_2 \frac{2\sqrt{I_b/n_b}}{\Delta_b} = 2 \log_2 \frac{2\sqrt{I_b/n_b}}{\Delta_b} \quad (75)$$

知覚エントロピー導出

(75) 式を変形

$$\begin{aligned} 2 \log_2 \frac{2\sqrt{I_b/n_b}}{\Delta_b} &= \log_2 \frac{4I_b}{n_b \Delta_b^2} = \log_2 \frac{I_b}{3T_b} \\ &\propto \log \frac{I_b}{3T_b} = \log \frac{I_b}{T_b} + \text{const.} \end{aligned}$$

全周波数ビンの符号化に必要なビット数は

$$\sum_b n_b \log \frac{I_b}{T_b} + \text{const.} \quad (76)$$

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

$$\text{PE} = \sum_b n_n \log \frac{I_b + 1}{T_b} \quad (77)$$

で定義．dist10 では無音領域で $T_b = 0$ となるため，

$$\text{PE} = \sum_b n_n \log \frac{I_b + 1}{T_b + 1} \quad (78)$$

で計算^{*11}

^{*11}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