

画像情報特論 (5)

- デジタル圧縮 (2) 音声・オーディオ圧縮

2001.05.15

電子情報通信学科 甲藤二郎

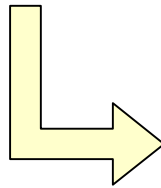
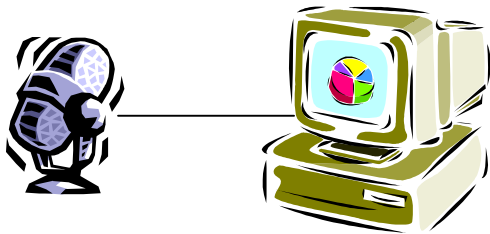
E-Mail: katto@katto.comm.waseda.ac.jp

音声・オーディオ圧縮の 原理

デジタルオーディオ

• キャプチャ&圧縮

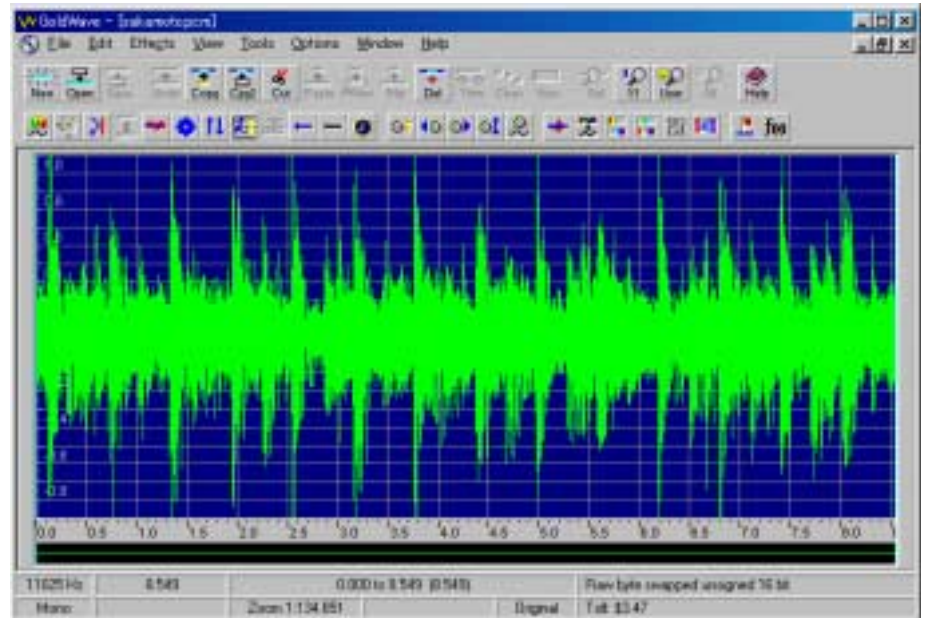
マイク サウンドキャプチャ



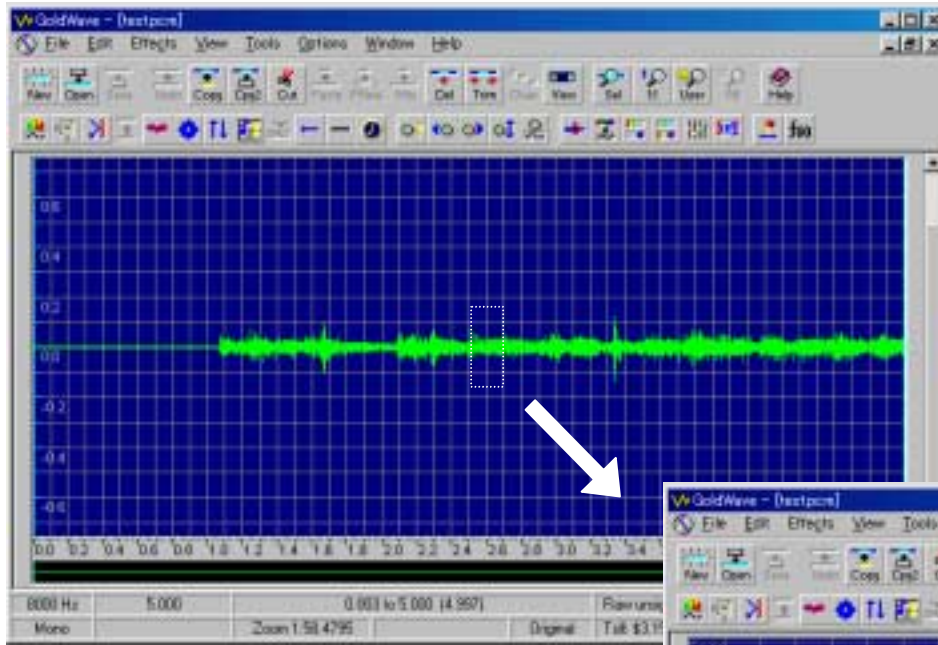
典型的なサンプリングレート

音声：
8 kHz、8 ビット

オーディオ：
22.5, 44.1, 48 kHz、16 ビット

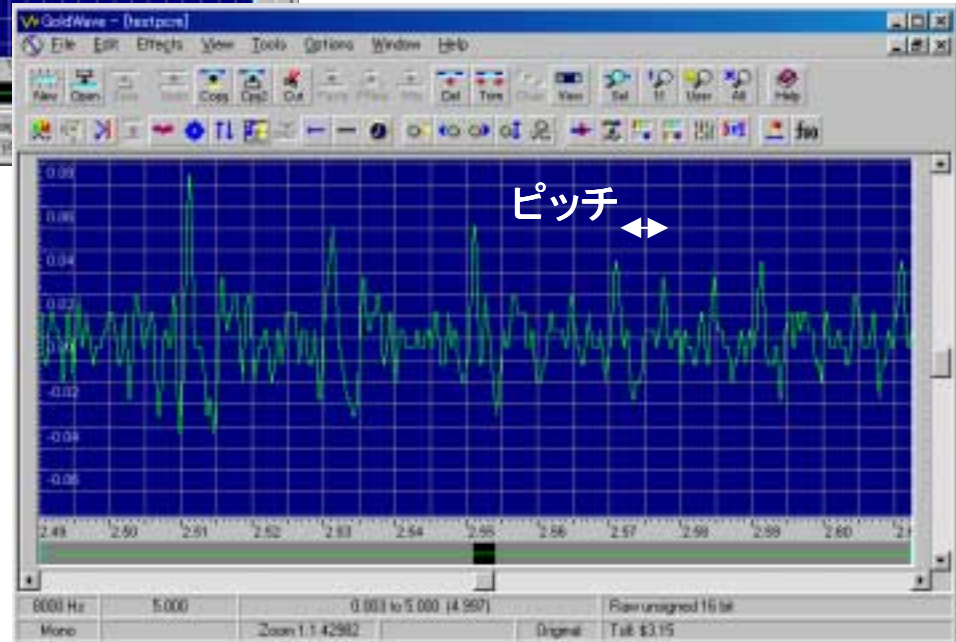


音声波形の例



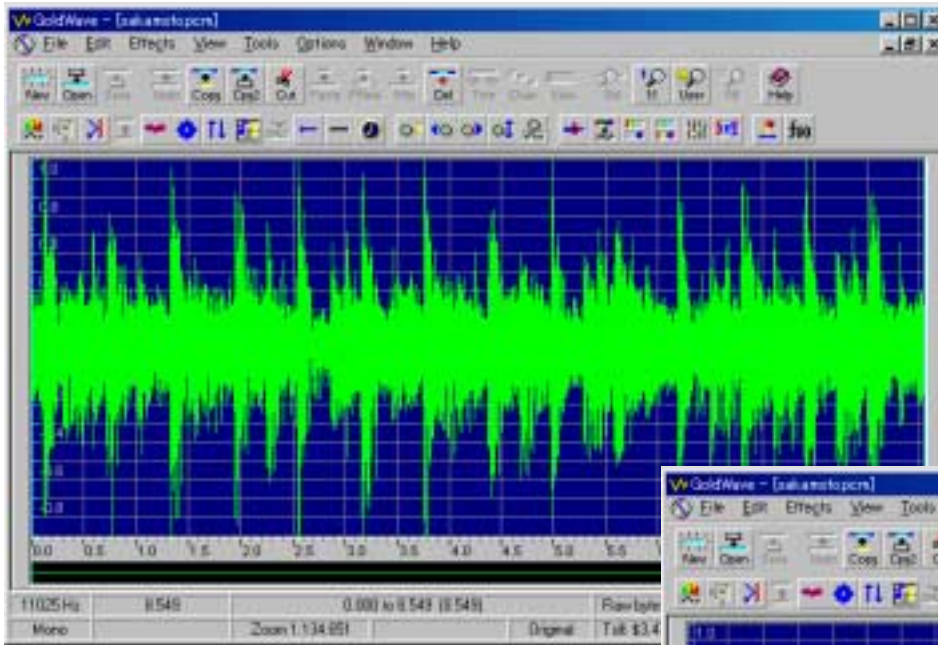
5秒間の波形

0.1 秒間の波形



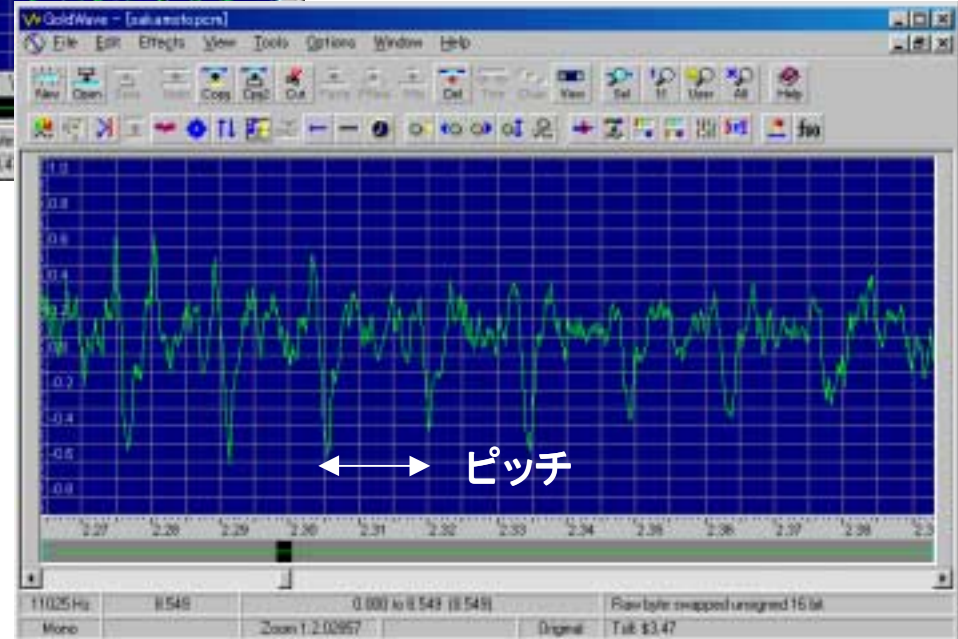
ピッチ (2~18ms): 右の例では約7ms

オーディオ波形の例

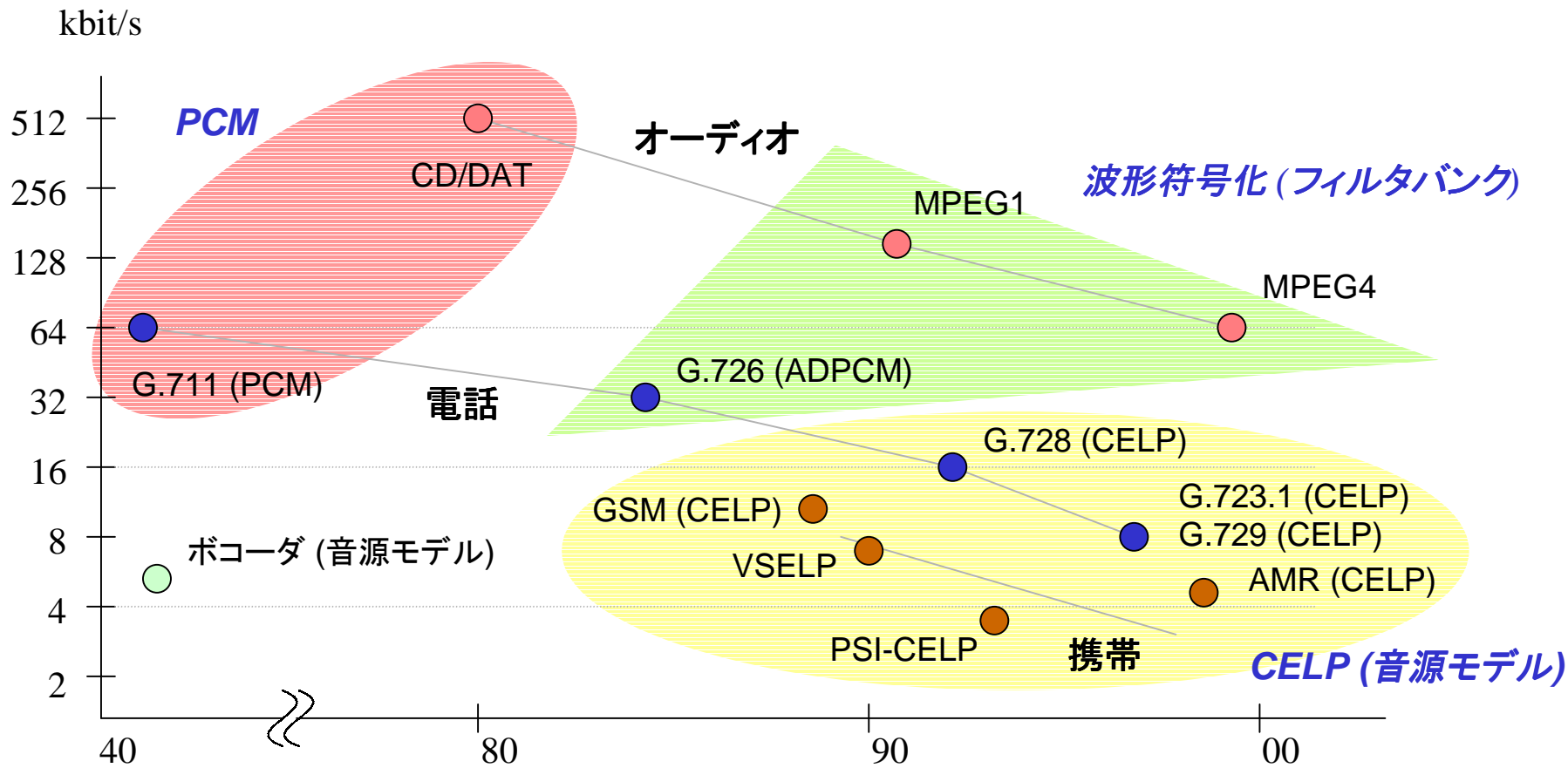


8秒間の波形

0.1 秒間の波形



音声・オーディオ符号化の歴史

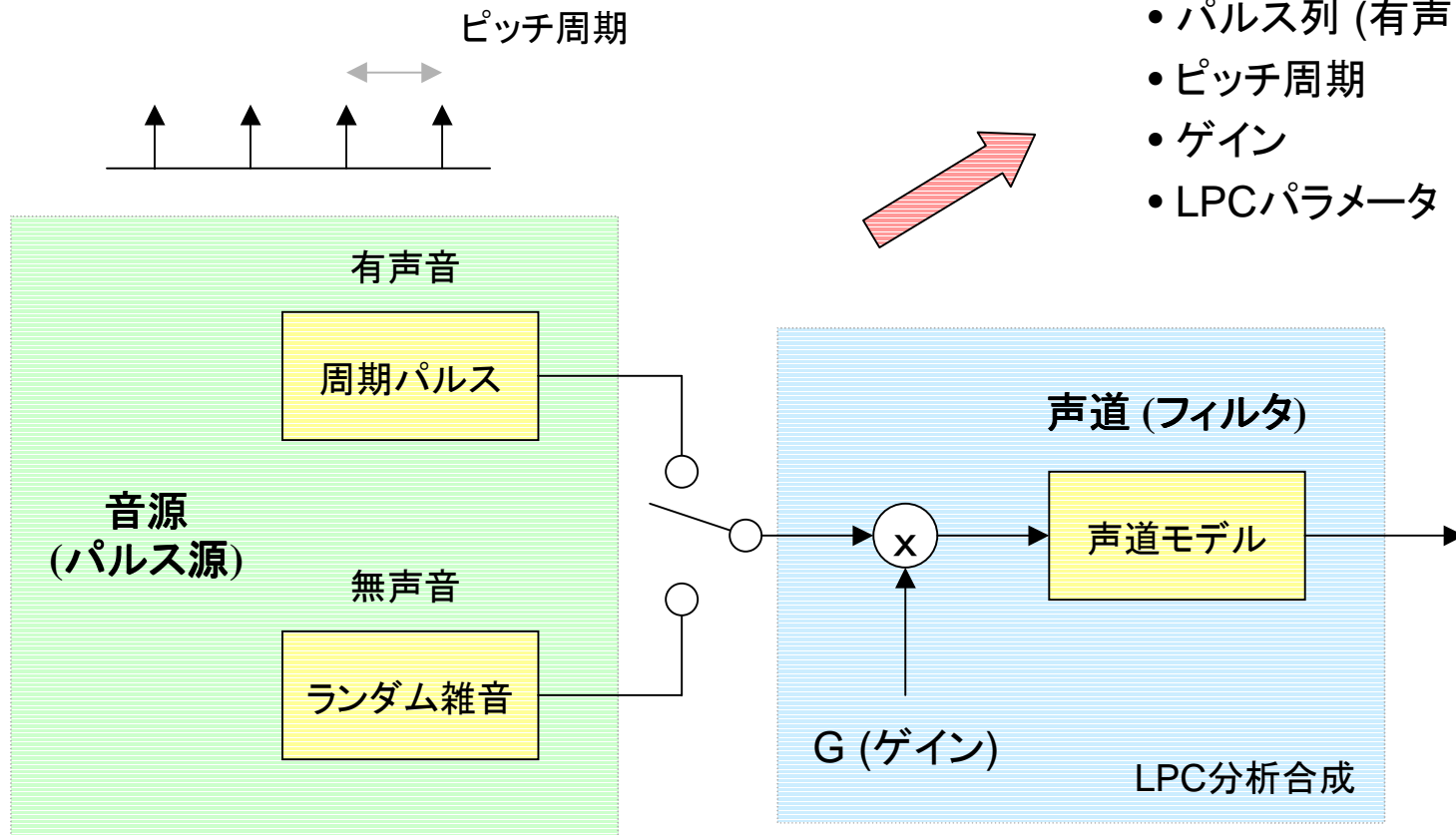


音声符号化 (1)

• 音声合成モデル

以下のパラメータを推定 (予測) して送信する

- パルス列 (有声・無声)
- ピッチ周期
- ゲイン
- LPCパラメータ

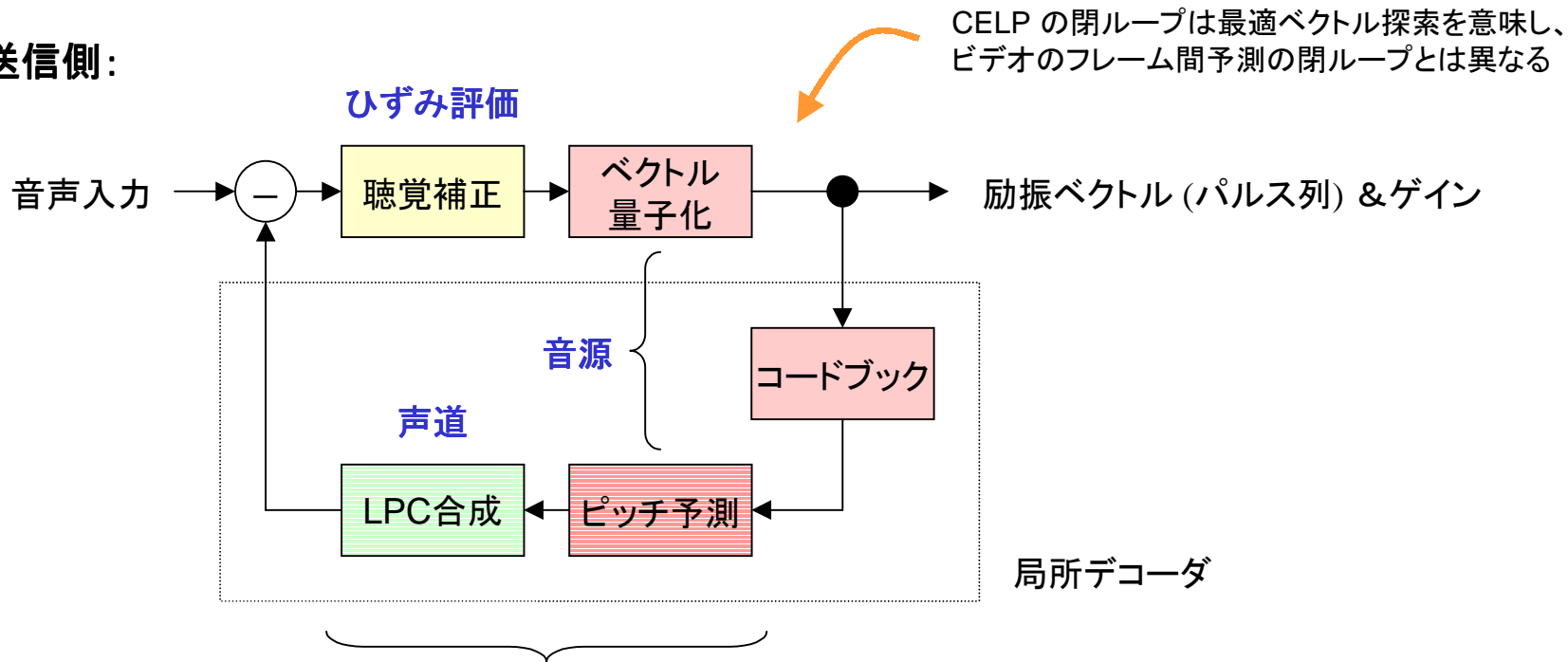


音声符号化 (2)

CELP: Code Excitation Linear Prediction

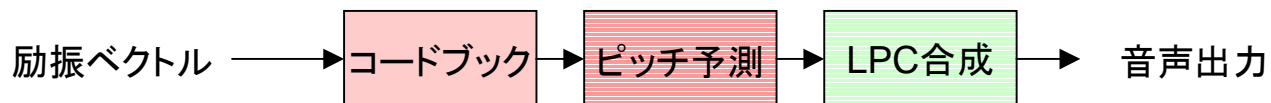
• CELP (LD-CELP 方式)

送信側:



LPCパラメータとピッチは送らずに復号信号から推定 (もともとのCELP方式では送信する)

受信側:



音声符号化 (3)

LPC: Linear Prediction Coding

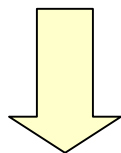
• LPC 分析 (線形予測分析) : 声道モデル

$$s(n) = \sum_{k=1}^p a_k s(n-k) + G \cdot u(n)$$

過去の k 個のサンプル値から線形予測

(注) 通常、画像のモデルでは雑音と扱う

$s(n)$: 音声サンプル
 a_k : LPC係数
 p : LPC分析次数
 G : 励振ゲイン
 $u(n)$: 正規化励振項



予測誤差二乗平均の最小化

$$\frac{\partial e(n)}{\partial a_k} = 0$$

$$\sum_{k=1}^p r_n(i-k) \hat{a}_k = r_n(i)$$

$r(k)$: 自己相関係数
 \hat{a}_k : 推定LPC係数

自己相関法 (Durbinのアルゴリズム)

音声符号化 (4)

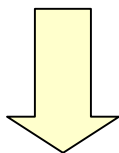
• ベクトル量子化 : 音源パルス列

励振ベクトルとゲインの探索:

$$d = \|x - gAc\| \rightarrow \min$$

となる励振ベクトルとゲインを探索

さまざまな探索手法 ...

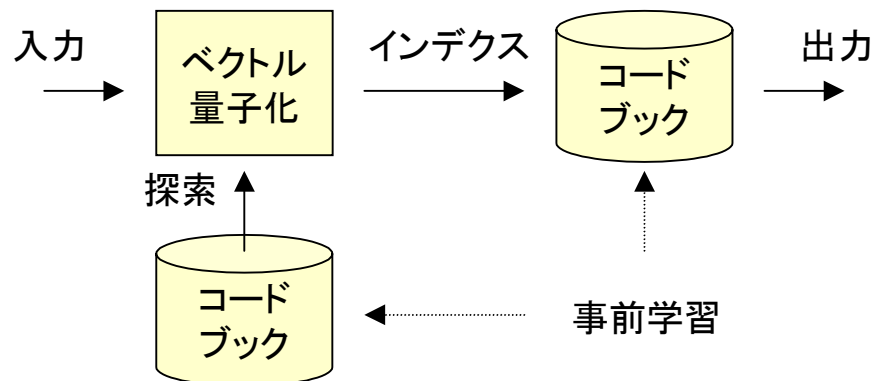


励振ベクトル → ベクトル量子化

ゲイン → スカラー量子化

(声道パラメータ → ベクトル量子化)

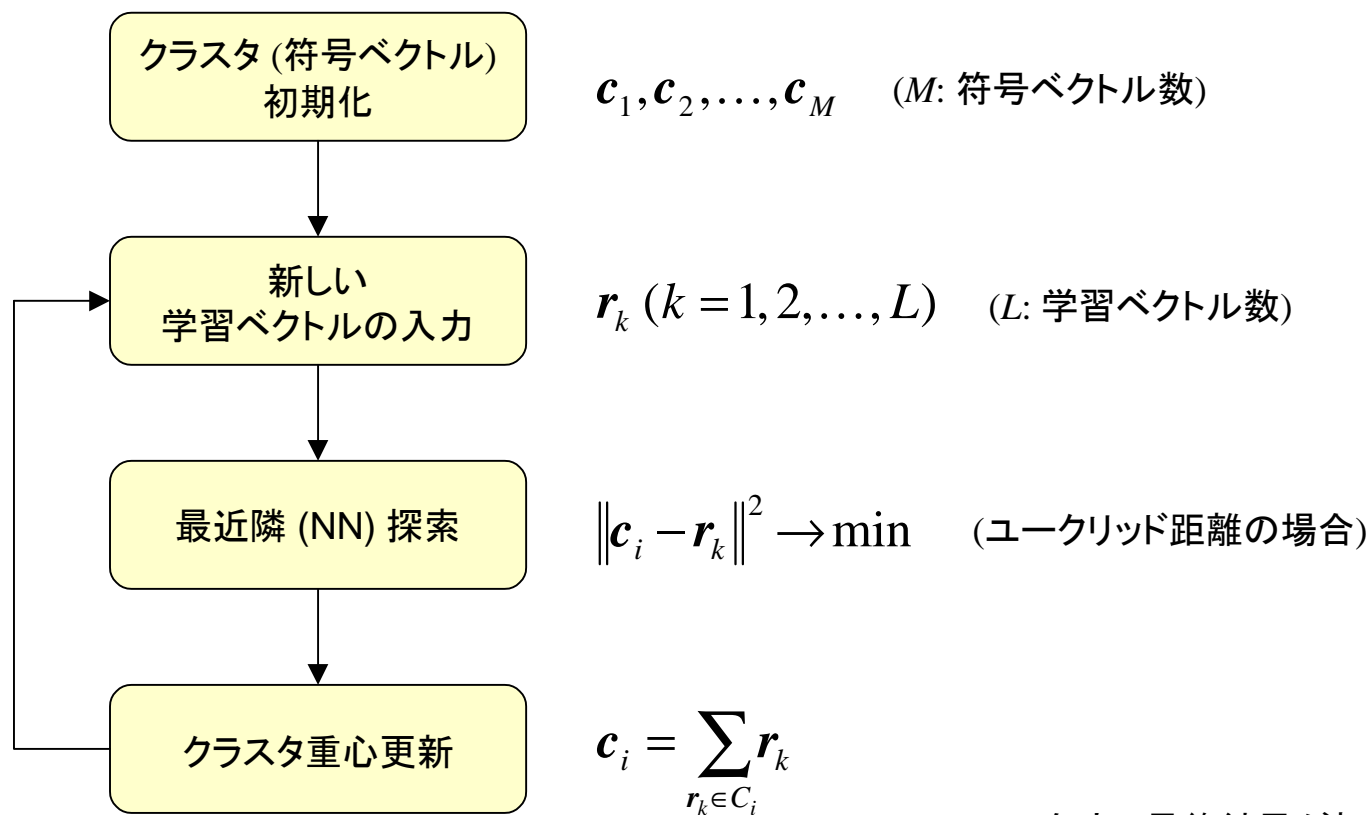
d : ひずみ
 x : 目標ベクトル (入力音声)
 A : LPC係数行列
 g : ゲイン
 c : 励振ベクトル (パルス列)



音声符号化 (5)

•ベクトル量子化 :コードブックの学習 (1)

K-平均アルゴリズム (一般化 Lloyd アルゴリズム)

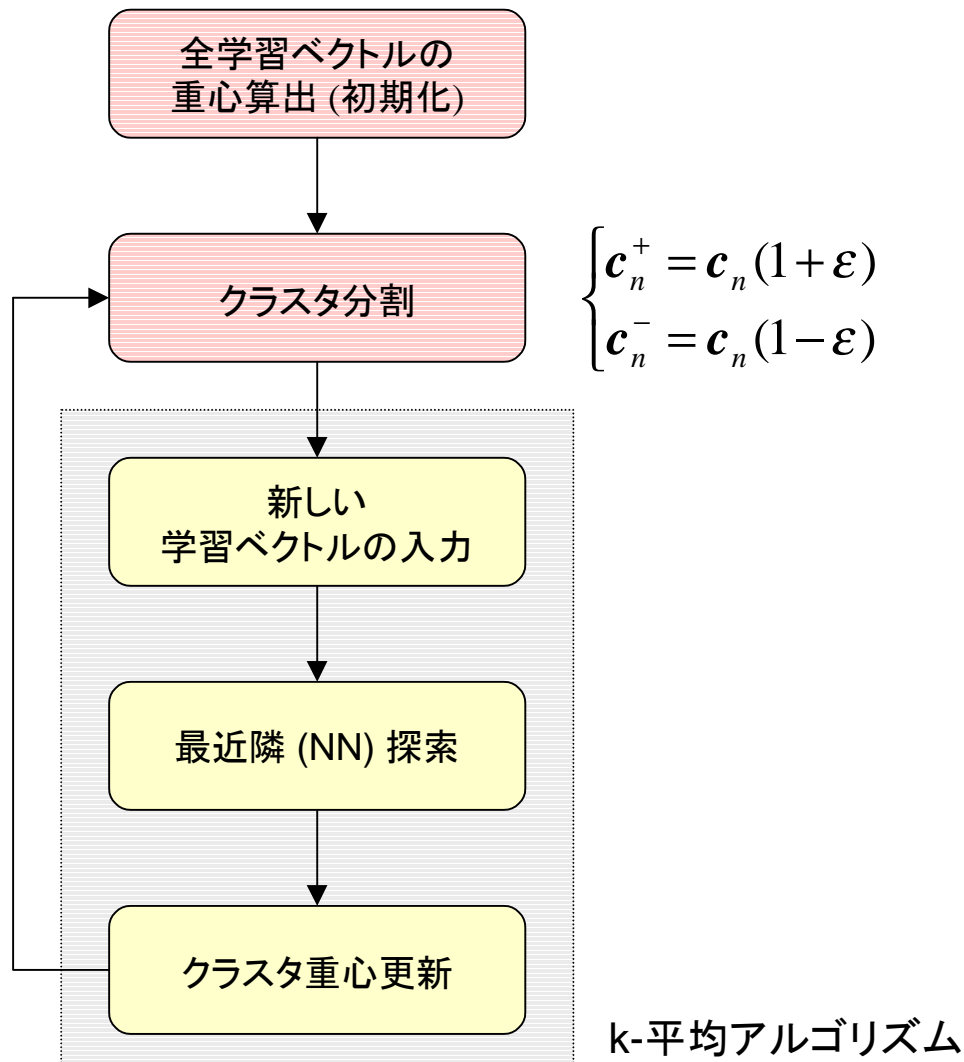


欠点: 最終結果が初期ベクトルに依存

音声符号化 (6)

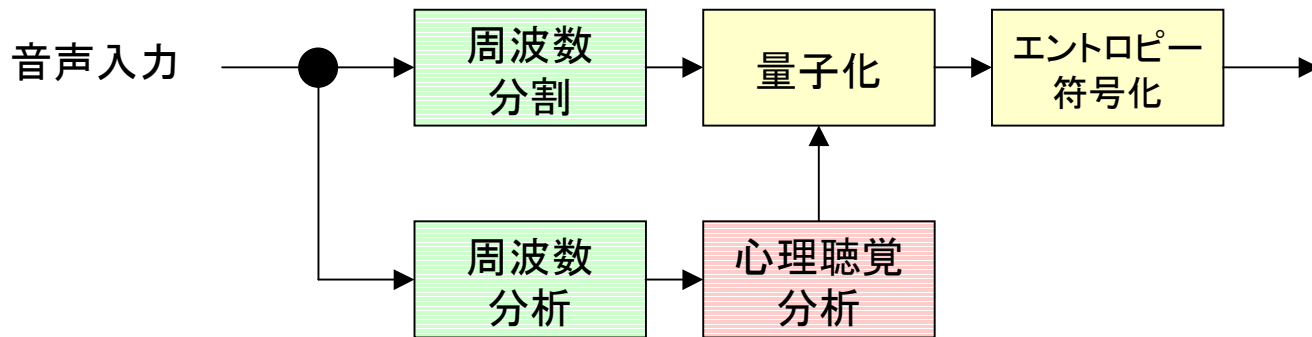
•ベクトル量子化 :コードブックの学習 (2)

LBG アルゴリズム



オーディオ符号化 (1)

• オーディオ符号化の基本

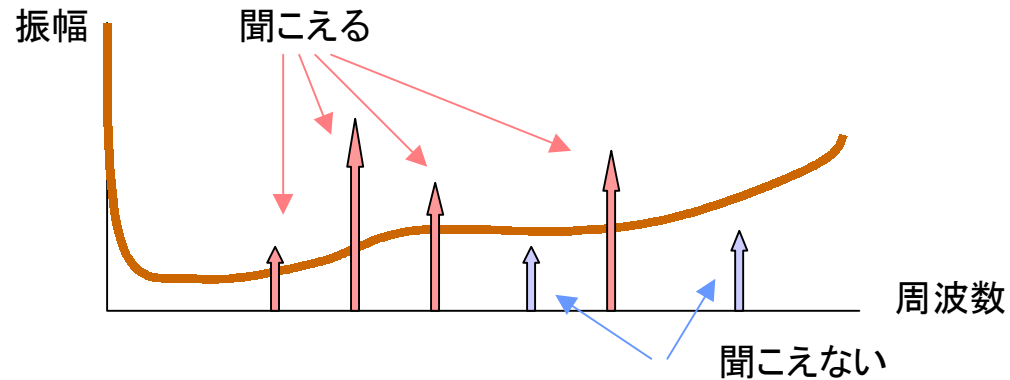


- 周波数分割、周波数分析: FFT、サブバンド分割 (QMF)、MDCT
- 心理聴覚分析: 絶対閾値とマスキング
- 量子化、エントロピー符号化: スカラー量子化とハフマン符号

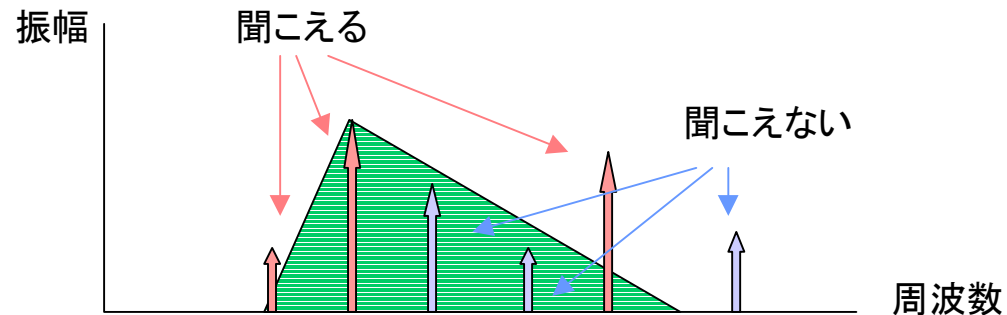
オーディオ符号化 (2)

• 心理聴覚分析

絶対閾値：人間は絶対可聴閾値よりも大きな音しか知覚できない

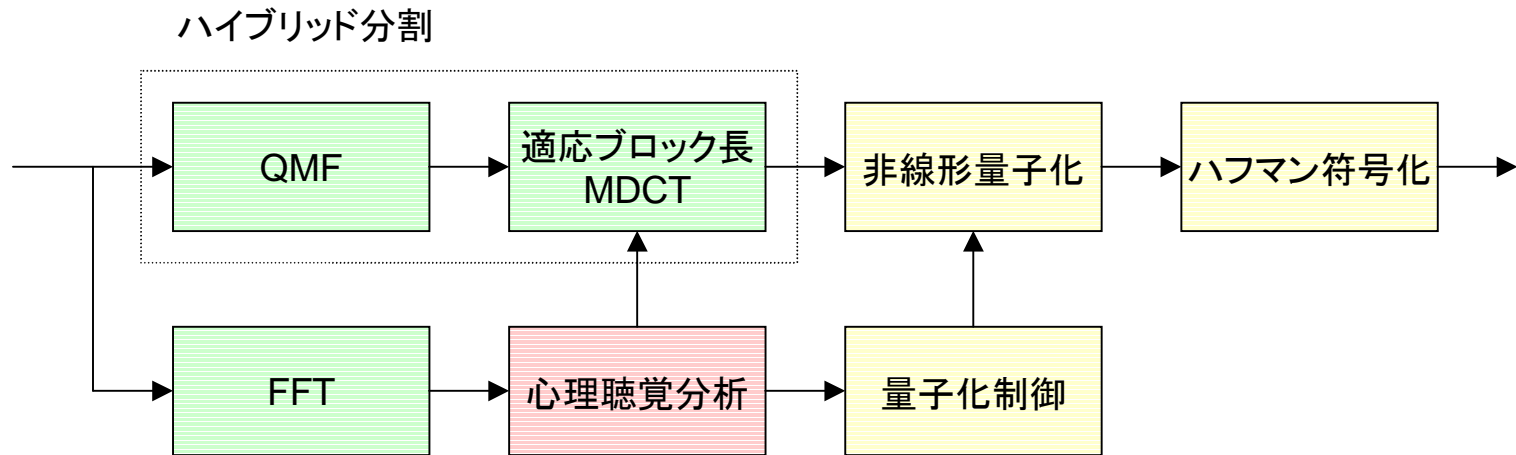


マスキング (相対閾値)：大きな音の周波数の近傍の小さな音の周波数は知覚できない

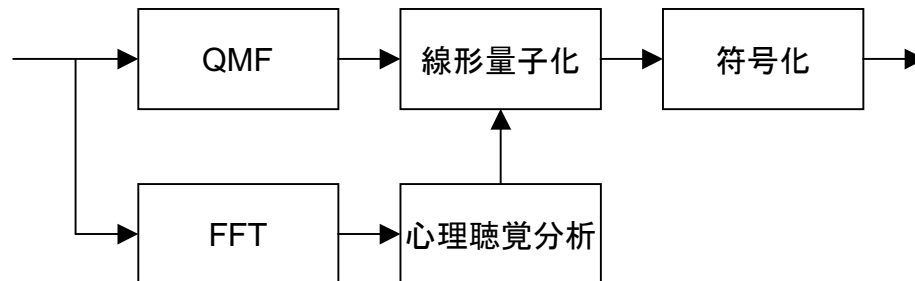


オーディオ符号化 (3)

- MP3 (MPEG-1 Layer III)

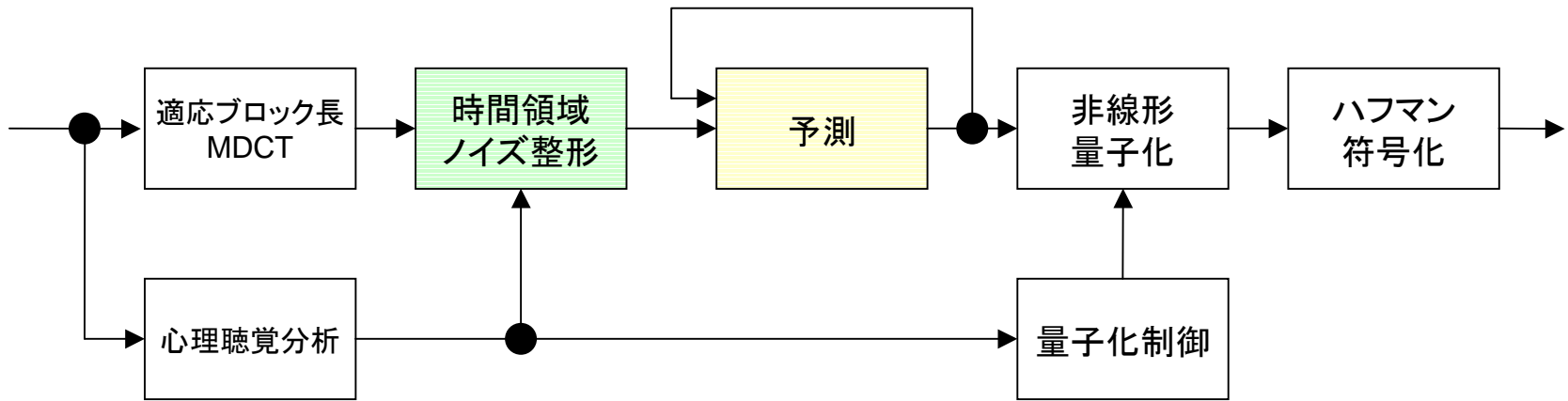


cf. Layer I, II



オーディオ符号化 (4)

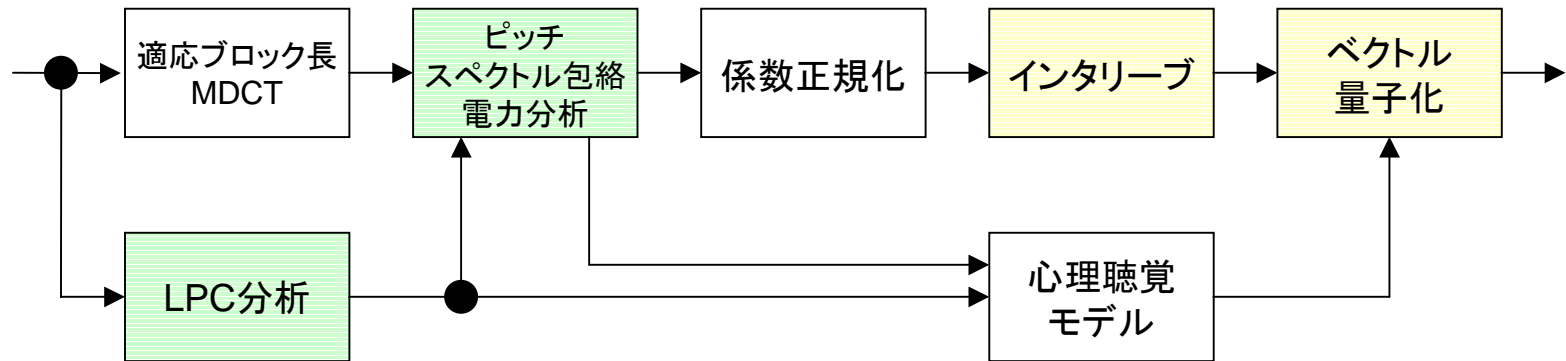
• MPEG-2 AAC



- 時間領域ノイズ整形 (for transient signals): 一部のMDCT係数を時系列とみなして線形予測 (LPC) 分析。振幅の大きい部分に量子化雑音が集中する (ノイズ整形)。
- 予測 (for stationary signals): MDCT係数毎に、過去2フレームのMDCT係数から予測。入力が定常的な場合に有効。

オーディオ符号化 (5)

• Twin VQ



- LPC分析、ピッチ・スペクトル包絡・電力分析：MDCT係数の平坦化。ベクトル量子化のコードブック削減。
- インターリーブベクトル量子化：適応量子化に替わるひずみの最小化手法。傾向の似た変換係数のグルーピング。

音声とオーディオ、ビデオの対比

- 音声符号化

PCM → 波形符号化 → 分析合成符号化 (音声合成モデル)

- オーディオ符号化、ビデオ符号化

PCM → 波形符号化 ...

オーディオ合成モデル: 楽器 (+ ボーカル)

ビデオ合成モデル: コンピュータグラフィックス?

分析合成手法の試み (ブレークスルーにはなっていない):

オーディオ符号化: 音源分離

ビデオ符号化: 知的符号化 (顔画像アニメーション)