画像情報特論 (4)

- ディジタル圧縮とメディア表現 (2)
  RD最適化
  音声、オーディオ、レイアウト

情報ネットワーク専攻　甲藤二郎
E-Mail: katto@waseda.jp

RD最適化と各種の応用

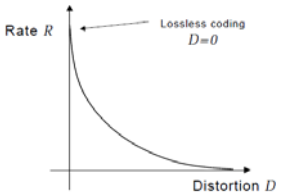
R-D最適化 (1)

目的: 与えられたレート制約下のひずみの最小化 (最適化問題)

minimize $D(\pi)$
subject to $R(\pi) \le R_{\max}$

← すべてのデジタル圧縮の基本

- D: Distortion (ひずみ)
- R: Rate (符号量)
- $R_{\max}$: 最大レート(制約条件)
- π: パラメータ



http://www.stanford.edu/class/ee368b/Handouts/04-RateDistortionTheory.pdf, ほか

R-D最適化 (2)

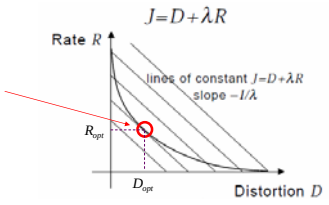
解法例: ラグランジュの未定乗数法

minimize $J = D(\pi) + \lambda \cdot R(\pi)$

λ: ラグランジュの未定乗数

R で微分すると、極小値では
以下が成立

$$\frac{dJ}{dR} = \frac{dD}{dR} + \lambda = 0$$



http://www.stanford.edu/class/ee368b/Handouts/04-RateDistortionTheory.pdf, ほか

R-D最適化 (3)

具体例: DCT変換係数への量子化ビット割り当ての最適化

minimize $\sigma_q^2 = \frac{1}{N}\sum_{k=0}^{N-1}\sigma_{qk}^2$
subject to $R = \frac{1}{N}\sum_{k=0}^{N-1} R_k \le R_{\max}$

$J = \sigma_q^2 + \lambda \cdot R$

$\frac{\partial J}{\partial R_k} = \frac{\partial \sigma_q^2}{\partial R_k} + \lambda \cdot \frac{1}{N} = 0$

← $\sigma_{qk}^2 = \varepsilon^2 2^{-2R_k}\sigma_k^2$
(情報理論のRDモデル)

$R_{k,opt} = R + \frac{1}{2}\log_2 \frac{\sigma_k^2}{\left[\prod_{j=0}^{N-1}\sigma_j^2\right]^{1/N}}$

$\sigma_{q,opt}^2 = $(自習)

- k: 変換係数のインデクス
- σ_q^2: 量子化誤差分散 (ひずみ)
- σ_k^2: 変換係数kの信号分散
- σ_{qk}^2: 変換係数kの量子化誤差分散
- R_k: 変換係数kのレート (符号量)

NS.Jayant & P.Noll: "Digital Coding of Waveforms", Prentice Hall

R-D最適化 (4)

RD-最適化の種々の拡張

minimize $J = D(\pi) + \lambda \cdot R(\pi)$

1. 動き補償における予測モード・動きベクトルの選択
2. マクロブロック量子化におけるモード選択
3. 複数参照フレームにおける参照フレーム選択
4. Rate-Distortion 最適化ストリーミング (RaDiO)
5. Congestion-Distortion 最適化ストリーミング (CoDiO)
6. ひずみ・経路ジョイント最適化
7. ひずみ・消費電力ジョイント最適化

"Rate-Distortion Optimization for Video Compression", IEEE Signal Processing Magazine, Nov.1998, ほか

R-D最適化 (5)

動き補償における予測モード・動きベクトルの選択

minimize $J = D(\pi) + \lambda \cdot R(\pi)$

D: 動き補償予測誤差

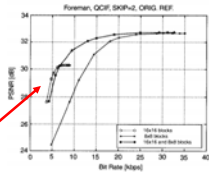
R: 予測モードと動きベクトルの送信に必要なオーバーヘッド

【予測モード】

Variable block-size: 16x16 ～ 4x4

・Pel-accuracy: integer, 1/2 pel, 1/4 pel

ブロックサイズが小さいほど予測効率は上がるが
オーバーヘッドが増える ⇒ RD-最適化



"Rate-Distortion Optimization for Video Compression", IEEE Signal Processing Magazine, Nov.1998.

R-D最適化 (6)

マクロブロック量子化におけるモード選択

minimize $J = D(\pi) + \lambda \cdot R(\pi)$

D: 量子化誤差分散

R: マクロブロックのモードと量子化結果の送信に必要な符号量

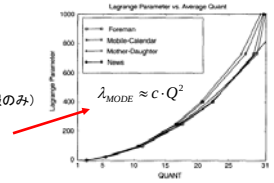
【マクロブロック・モード】

・Intra: 種々のイントラ予測

・Inter: 種々のインター予測

・Skip: マクロブロック・スキップ（モード情報のみ）

量子化パラメータ Q と λ の関係 (実測値)



"Rate-Distortion Optimization for Video Compression", IEEE Signal Processing Magazine, Nov.1998.

R-D最適化 (7)

複数参照フレームにおける参照フレーム選択

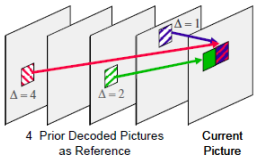
minimize $J = D(\pi) + \lambda \cdot R(\pi)$

D: 量子化誤差分散

R: 参照ピクチャの指定と動きベクトルの送信に必要な符号量

【実験報告例】

参照枚数を50枚にすることで動き補償予測誤差を 1～2dB 改善できたが、オーバーヘッドも30%増加 ⇒ RD-最適化の対象

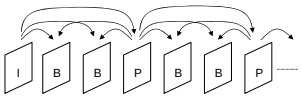


"Rate-Distortion Optimization for Video Compression", IEEE Signal Processing Magazine, Nov.1998.

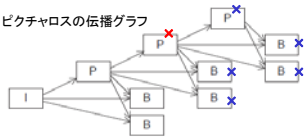
R-D最適化 (8)

RD-Optimized ストリーミング (1) RaDiO

IPB 予測



Dependency graph　ピクチャロスの伝播グラフ



Transmission policy　$\pi = [1,0,0,1,0,0,1,1,0]$　ピクチャ送信パターン
1: 送信する, 0: 送信しない

P.Chou & Z.Miao: "Rate-Distortion Optimized Streaming of Packetized Media", IEEE Trans on Multimedia.

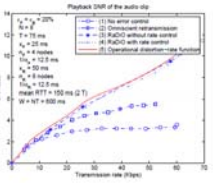
R-D最適化 (9)

RD-Optimized ストリーミング (2)

minimize $J = D(\pi) + \lambda \cdot R(\pi)$

$$D(\pi) = D_0 - \sum_l \Delta D_l \prod_{l' \le l} (1 - \varepsilon(\pi_{l'}))$$

$$R(\pi) = \sum_l \rho(\pi_l) B_l$$



ロスの発生を前提に、I ピクチャ、P ピクチャから優先的に送信する戦略の数学基盤

D_0 ：何も送らないときのひずみ（最悪値）
ΔD_l ：ピクチャ l の復号で改善されるひずみ
$\varepsilon(\pi_l)$ ：ピクチャ l の到着率（復号される確率）
$\rho(\pi_l)$ ：ピクチャ l の送信回数（再送を含む）
B_l ：ピクチャ l の符号量

P.Chou & Z.Miao: "Rate-Distortion Optimized Streaming of Packetized Media", IEEE Trans on Multimedia.

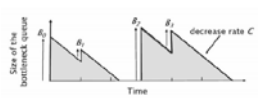
R-D最適化 (10)

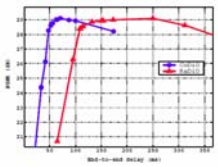
"Congestion-Distortion" Optimized ストリーミング: CoDiO

minimize $J = D(\pi) + \lambda \cdot X(\pi)$

$$D(\pi) = D_0 - \sum_l \Delta D_l \prod_{l' \le l} (1 - \varepsilon(\pi_{l'}))$$

$$X(\pi) = delay \cong \sum_l \frac{\rho(\pi_l) B_l}{C}$$





D_0 ：何も送らないときのひずみ（最悪値）
ΔD_l ：ピクチャ l の復号で改善されるひずみ
$\varepsilon(\pi_l)$ ：ピクチャ l の到着率（復号される確率）
$\rho(\pi_l)$ ：ピクチャ l の送信有無（0 or 1）
B_l ：ピクチャ l の符号量
C ：ボトルネックリンクのキャパシティ

E.Setton & B.Girod: "Congestion-distortion Optimized Scheduling of Video over a Bottleneck Link", MMSP 2004

R-D最適化 (11)

ひずみ・経路ジョイント最適化 (1)

リンク集合: $L = \{(i, j) \mid \text{node } j \text{ can hear node } i\}$

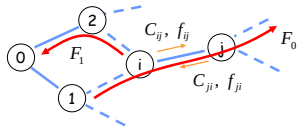
リンクキャパシティ: $C = \{C_{ij} \mid (i, j) \in L\}$

フロー集合: $F = \{f_{ij} \mid (i, j) \in L\}$

ストリーム集合: $S = \{s \mid \text{collection of streams}\}$

ストリーム s のフロー集合: $F_s = \{f_{ij}^s \mid (i, j) \in L\}$

ストリーム s のビットレート: R_s



"Error Resilient Concurrent Video Streaming ...", PV 2006 ほか

R-D最適化 (12)

ひずみ・経路ジョイント最適化 (2)

minimize $J = D(\pi) + \lambda \cdot X(\pi)$

D: ビデオ品質（量子化誤差分散など）

X: ネットワーク輻輳状況（遅延など） $e.g.\ X = \sum_{(i,j)\in L} \frac{f_{ij}}{C_{ij} - f_{ij}}$

その他の拘束条件

レート拘束: $0 \le f_{ij} = \sum_s f_{ij}^s < C_{ij}$

ループ回避: $\sum_{k:(i,k)\in L} f_{ik}^s - \sum_{k:(k,i)\in L} f_{ki}^s = \begin{cases} -R_s & (i = src) \\ R_s & (i = dst) \\ 0 & (otherwise) \end{cases}$

干渉緩和（無線の場合）: $\frac{f_{ij}}{C_{ij}} + \sum_{(m,n)\neq(i,j)} \frac{f_{mn}}{C_{mn}} \le 1$

"Error Resilient Concurrent Video Streaming ...", PV 2006 ほか

R-D最適化 (13)

ひずみ・消費電力ジョイント最適化

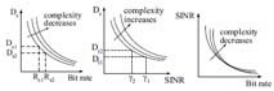
minimize $J = D(R, \pi) + \lambda \cdot P(\pi)$

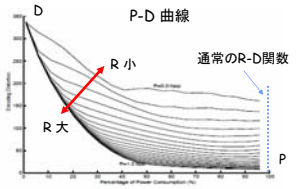
D: ビデオ品質（量子化誤差分散など）

R: 圧縮符号量

P: 消費電力（圧縮＋送信）

π: 処理の複雑度 (complexity)





"Power Efficient Wireless Video Communications ...", PCS 2006 ほか

音声・オーディオ圧縮

ディジタルオーディオ

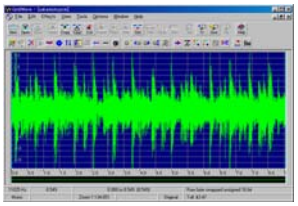
・キャプチャ＆圧縮

マイク　サウンドキャプチャ

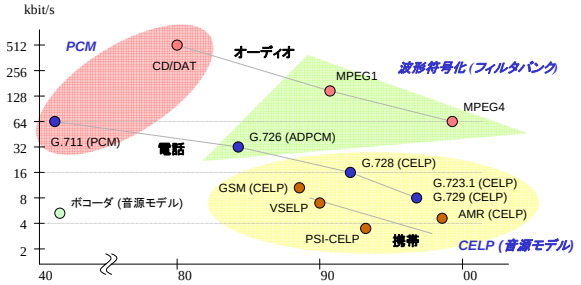
典型的なサンプリングレート

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・オーディオ符号化の歴史



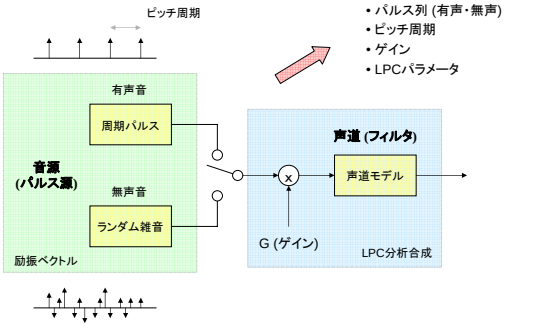
守谷："音声符号化"

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

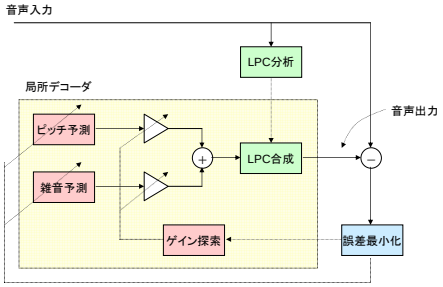


L.Rabiner et al: "Fundamentals of Speech Recognition"

音声符号化 (2)

CELP: Code Excitation Linear Prediction

• **CELP**



守谷: "音声符号化"

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

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

$r(k)$: 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**

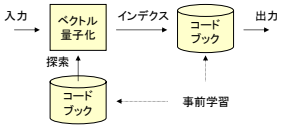
励振ベクトルとゲインの探索:

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

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 ...

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)



励振ベクトル → ベクトル量子化

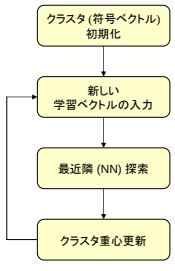
ゲイン → スカラー量子化

(声道パラメータ → ベクトル量子化)

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)



$\boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_M$ (M: 符号ベクトル数)

$\boldsymbol{r}_k \ (k=1,2,\ldots,L)$ (L: 学習ベクトル数)

$\|\boldsymbol{c}_i-\boldsymbol{r}_k\|^2 \rightarrow \min$ (ユークリッド距離の場合)

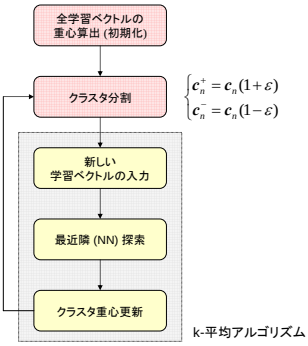
$\boldsymbol{c}_i=\sum_{\boldsymbol{r}_k\in C_i}\boldsymbol{r}_k$

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)**

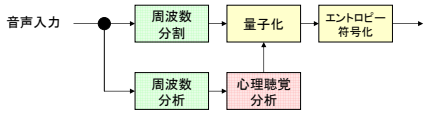
LBG アルゴリズム



k-平均アルゴリズム

オーディオ符号化 (1)

• **オーディオ符号化の基本**

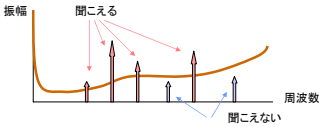


• 周波数分割、周波数分析：FFT、サブバンド分割 (QMF)、MDCT
• 心理聴覚分析：絶対閾値とマスキング
• 量子化、エントロピー符号化：スカラー量子化とハフマン符号

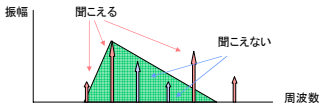
オーディオ符号化 (2)

• **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

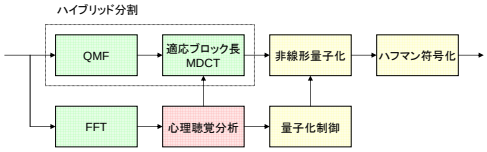


マスキング (相対閾値)：大きな音の周波数の近傍の小さな音の周波数は知覚できない

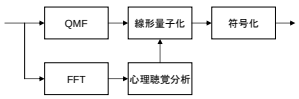


オーディオ符号化 (3)

• **MP3 (MPEG-1 Layer III)**

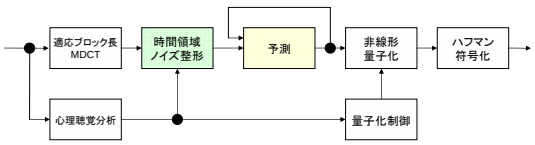


cf. Layer I, II



オーディオ符号化 (4)

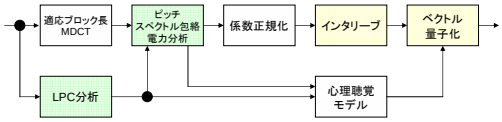
• **MPEG-2 AAC**



• 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)

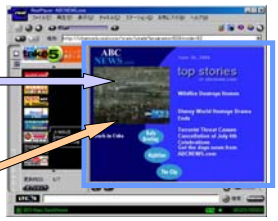
SMIL

SMIL

＊ Synchronized Multimedia Integration Language

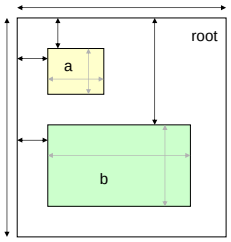
・ストリーミングのためのレイアウト記述言語

```
<smil>
<head>
   <layout>
      レイアウト記述
   </layout>
</head>
<body>
   <par>
      メディア記述
   </par>
</body>
</smil>
```



＊ XML ベース... HTML に慣れていれば習得は簡単

レイアウト記述




```
<root-layout width="500" height="400"/>
<region id="a" top="50" left="50"
                width="100" height="80" />
<region id="b" top="200" left="50"
                width="400" height="200" />
```

表示画面　　レイアウト記述

メディア記述

ストリーミング

```
<par>
   <video region="b" src="rtsp://www.foo.ac.jp/guide.sdp" />
   <seq>
      <img region="a" src="http://www.foo.ac.jp/point0.jpg" dur="10s" />
      <img region="a" src="http://www.foo.ac.jp/point1.jpg" dur="10s" />
      <img region="a" src="http://www.foo.ac.jp/point2.jpg" dur="10s" />
   </seq>
</par>
```

<par> メディア1, メディア2, … </par>　　複数メディアの「**並列**」再生

<seq> メディア1, メディア2, … </seq>　　複数メディアの「**逐次**」再生

<video>, <audio>, <img>, ...　　各種メディアタグ