音メディアと信号処理

- オーディオ符号化について
  - ー標本化、量子化、パルス符号変調(PCM)
  - ー圧縮符号化
    (相関符号化、エントロピー符号化)
  - ーマスキング
- 音声符号化について(音声の生成モデル)
- 文ー音声変換(テキスト音声合成)
- 音声認識(パターン認識の原理)

オーディオ符号化(1)

(1) オーディオ波形 $x(t)$ の標本化

可聴周波数(人の耳に知覚できる音の周波数)

20 Hz ～ 20 kHz

$x(t)$は $W=20$kHz に帯域制限されているとみなせるので、標本化定理より、$f_s \geqq 2W=40$kHz なる標本化周波数で標本化することができる。

標本値 $\{x(nT_s),\ n=0,\pm1,\pm2,\cdots\}$

($T_s=1/f_s$ [s](秒):標本化周期)

オーディオの標本化周波数 f_s の規格

DAT 48kHz CD、MD 44.1kHz

参考資料

可聴周波数（人の耳に知覚できる音の周波数）

20 Hz ～ 20 kHz

音波

http://www-antenna.ee.titech.ac.jp/~hira/hobby/edu/sonic_wave/index-j.html

いろいろな周波数（可聴周波数などの実験）

http://www-antenna.ee.titech.ac.jp/~hira/hobby/edu/sonic_wave/sine_wave/frequency/index-j.html

平野拓一氏（東京工業大学）作成

オーディオ符号化（2）

（2）標本値のディジタル記録（量子化）

標本値 x の範囲 $r_1 \leqq x \leqq r_{N+1}$ を有限個 $N=2^B$（B：量子化ビット数）に分割して、1つの値の範囲に入る標本値 x には同じ代表値を与え、この代表値を Bビットの2進符号で表す。

（例）線形量子化の場合

区間幅： $\Delta = (r_{N+1} - r_1)/N = (r_{N+1} - r_1)/2^B$

$r_1 \leq x < r_2 (= r_1 + \Delta) \Rightarrow y_1 \Rightarrow 00\cdots000$

$r_2 \leq x < r_3 (= r_2 + \Delta) \Rightarrow y_2 \Rightarrow 00\cdots001$

$\vdots$

$r_N \leq x \leq r_{N+1} (= r_N + \Delta) \Rightarrow y_N \Rightarrow 11\cdots111$

代表値＝中央値

標本値の分布＝一様分布

量子化誤差　$\sigma_B^2 = \int_{r_k}^{r_{k+1}} (x - y_k)^2 \, p(x) dx = \Delta^2 / 12$

$10\log_{10}\left(\frac{\sigma_B^2}{\sigma_{B+1}^2}\right) \fallingdotseq 6[\mathrm{dB}]$

*B*が1ビット増えるごとに量子化誤差が約6dB小さくなる

パルス符号変調 (Pulse Code Modulation, PCM) 符号化

線形PCM（線形量子化した符号）

ビットレート（1秒間に送られる情報量）

$$I = f_s \times B \ [\mathrm{bit/s}] \ (\mathrm{bps})$$

CD（16ビット線形PCM）のビットレート（2チャンネルの場合）

I＝44.1kHz×16bit×2≒1.41Mbit/s≒176.4kbyte/s

オーディオ符号化（3）

（3）圧縮符号化

オーディオ品質の低下を伴わない、あるいは低下を聴覚に感じさせないという条件での圧縮技術

(a) 標本値を量子化し、その代表値の発生確率を調べて2進符号に圧縮符号化する。（⇒ハフマン符号化、エントロピー符号化）

(b) 標本値を直交変換し、変換係数を人の聴覚特性に合わせてカットする。次に、変換係数を量子化し、その代表値の発生確率を調べて2進符号に圧縮符号化する。

オーディオ信号（標本値）の直交変換

変換係数は周波数領域の信号（スペクトル）とみなせるので、人の聴覚特性に合わせた処理が変換領域で容易にできる。

直交変換：離散コサイン変換
(Discrete Cosine Transform, DCT)

① 高い周波数成分をカットする。（標本値間の相関が大きければ、高い周波数成分は少ない。）

② 人の聴覚特性：
- 最小可聴限界（最小可聴曲線）
- マスキング（時間マスキング、周波数マスキング）

を利用して、周波数成分をカットする。

オーディオ符号化の標準化

- MPEG オーディオ
  （MPEG-1, MPEG-2, MPEG-4)
  ⇒インターネット、携帯電話、テレビ電話、放送（CS（通信衛星）放送、BS（放送衛星）放送）の規格
  (MPEG: Moving Picture Experts Group)
- AC-3 (Audio Coder 3)
  ⇒DVDのオーディオ符号化方式
- ATRAC (Adaptive Transform Audio Coder)
  ⇒民生用MD

音声符号化（1）

- 線形PCM（標本化 8 kHz、量子化 8 bit ⇒ ビットレート 64 kbps）
- 予測符号化（⇒ 音声信号の標本値間には大きな相関があることを利用）

線形予測分析

$$x(n) = \hat{x}(n) + e(n)$$

予測残差（予測誤差）

$x(n)$ の予測値： $\hat{x}(n) = \sum_{k=1}^{p} a_k x(n-k)$

音声符号化（2）

● 差分PCM (differential PCM, DPCM)

$$e(n) = x(n) - \hat{x}(n) = x(n) - x(n-1) \quad (p = 1, a_1 = 1)$$

音声の差分値（予測残差）を符号化する。

● 適応差分PCM (adaptive PCM, ADPCM)

$$e(n) = x(n) - \hat{x}(n) = x(n) - \sum_{k=1}^{p} a_k x(n-k)$$

音声を長さNのブロックに分割し、最小二乗法で予測残差$\{e(n)\}$が最小になるように予測係数$\{a_k\}$を求め、Nとpの情報と$\{e(n)\}$と$\{a_k\}$を符号化する。