

情報ネットワーク (第6回)

音声のデジタル化

2011年 11月17日

音声のデジタル化(標本化)

- 標本化: アナログ信号に対して標本化パルスのタイミングで振幅を抜き出す操作

- 標本化周波数: f

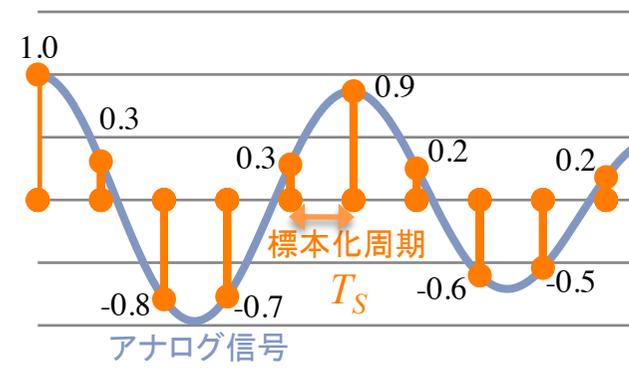
$$f = 1/T_s$$

- 元のアナログ信号を完全に復元するためには、アナログ信号の持つ最大周波数 (f_{\max}) の2倍以上の周波数で標本化しなければならない

$$f \geq 2f_{\max}$$

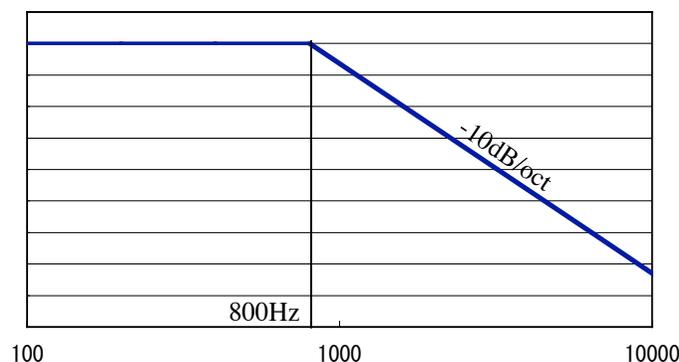


標本化定理



音声信号のスペクトル分布

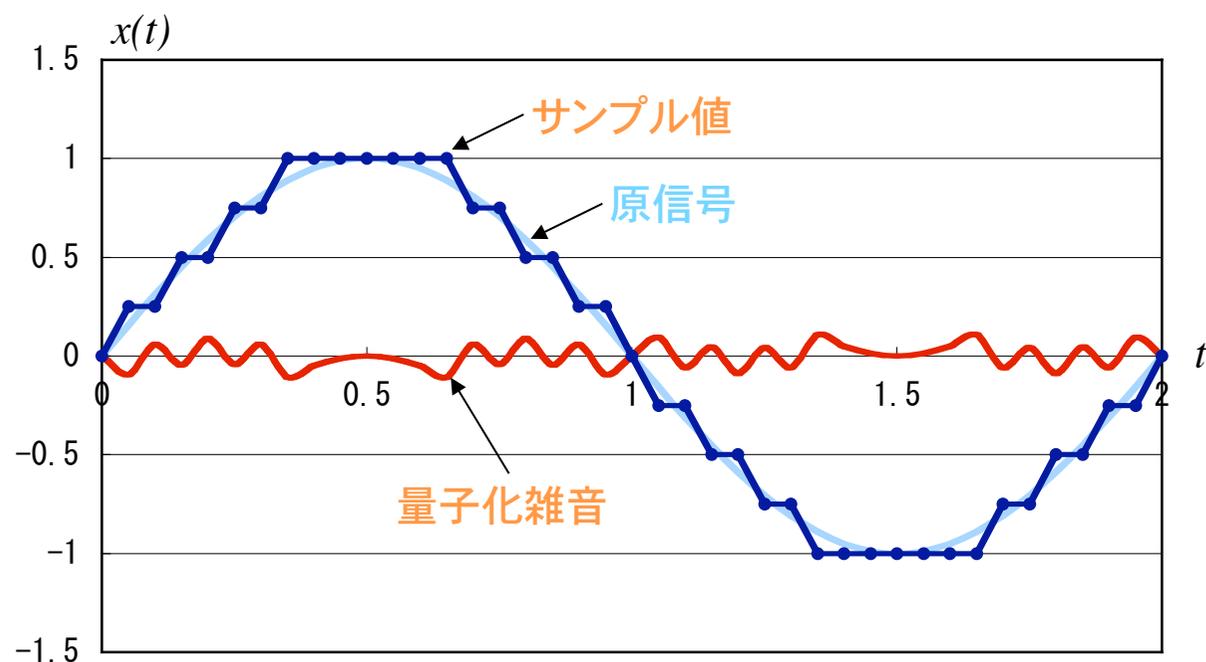
- 音声のスペクトルは800Hzまでほぼ平坦
- 800Hz以上は、オクターブ当たり10dBの傾斜で減衰
- 800Hz以下の成分が全エネルギーの80%を占める。



- 電話音声のサンプリング周波数は最高周波数の倍 (8kHz) :エイリアシングを避けるため。
- エネルギーの大半を占める800Hz以下の成分に対しては無駄に細かいサンプリングが行われている。

音声のデジタル化(量子化)

- 量子化による符号化波形の歪: アナログ信号波形とサンプル値との差の2乗平均値



量子化雑音の電力

- 量子化雑音の電力: N_Q
 - 信号の存在が量子化範囲内で一様
 - 線形符号化を仮定 (量子化刻み ΔL)

$$N_Q = \frac{1}{\Delta L} \int_{-\Delta L/2}^{\Delta L/2} x^2 dx = \frac{\Delta L^2}{12}$$

- 信号の存在範囲: L
- 2進 B ビット符号化を仮定

$$\Delta L = \frac{L}{2^B}$$



量子化後の信号雑音比

- 信号の実効振幅： σ
- 信号の波高値(ピーク値／実効値)： p

$$L = 2p\sigma$$

$$\begin{aligned} S/N|_{dB} &= 10\log_{10}\left[\frac{\sigma^2}{\Delta L^2/12}\right] = 10\log_{10}\left[\frac{\sigma^2}{(2p\sigma/2^B)/12}\right] \\ &= 10\log_{10}\left[\frac{12}{4p^2}2^{2B}\right] = 2B \cdot 10\log_{10} 2 + 10\log_{10}\left(\frac{3}{p^2}\right) \\ &= 6.02B + 4.77 - 20\log_{10} p \end{aligned}$$



音声の信号雑音比

- 音声信号の波高値: $p=12\text{dB}$

$$S/N|_{dB} = 6.02B - 7.23$$

- 情報伝送率: R
- サンプル周波数: $2\omega_M$

$$R = 2\omega_M B$$

サンプル当たりのビット数 B を増やして S/N を改善すると情報伝送率が増え、 B を少なくして情報を圧縮すると S/N が劣化する。



音声の振幅分布

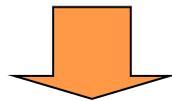
- 音声信号の振幅分布

$$b(x) = \frac{a_1}{\sqrt{2\pi\sigma_1}} e^{-\frac{x^2}{2\sigma_1^2}} : x \text{が小さいとき} \quad b(x) = a_2 e^{-\frac{|x|}{\sigma_2}} : x \text{が大きいとき}$$

- x が小さいとき: 子音
- x が大きいとき: 母音

- 時間的に母音区間の出現頻度が高いため、振幅分布を指数分布で近似する。

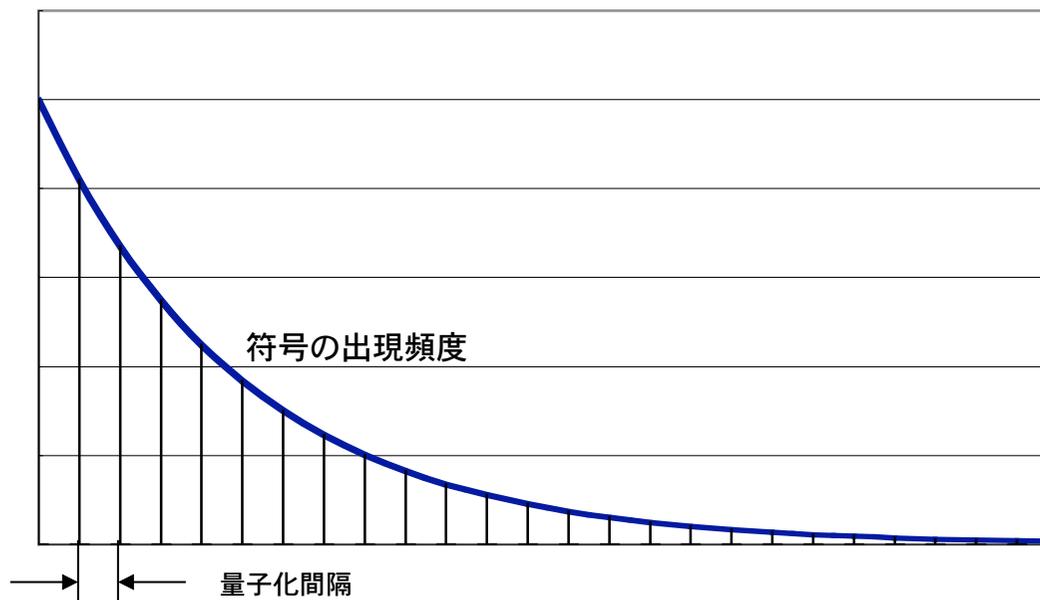
$$b(x) = a e^{-\frac{|x|}{\sigma}} \quad \int_{-\infty}^{\infty} b(x) dx = 1, \quad \int_{-\infty}^{\infty} x^2 b(x) dx = P$$



$$b(x) = \frac{1}{\sqrt{2P}} e^{-\sqrt{\frac{2}{P}}|x|}$$



音声の線形符号化



- 線形量子化を行うと各符号の出現頻度に大きな隔たりができ、情報理論的には、能率の悪い符号化になる。
- 信号の振幅を変換して分布が一定になるようにしてから線形量子化を行うと能率的に符号化できる。



圧伸PCM

- 信号の振幅を変換して分布が一定になるようにしてから線形量子化を行うと能率的に符号化できる。

振幅分布 = 指数分布 \longrightarrow $y = \log|x|$

$$y = \operatorname{sgn}(x) \cdot V \cdot \frac{\log(1 + \mu \frac{|x|}{V})}{\log(1 + \mu)}$$

$$\lim_{\mu \rightarrow 0} y = \operatorname{sgn}(x) \cdot |x| = x$$

$$\lim_{\mu \rightarrow \infty} y = \operatorname{sgn}(x) \cdot V$$

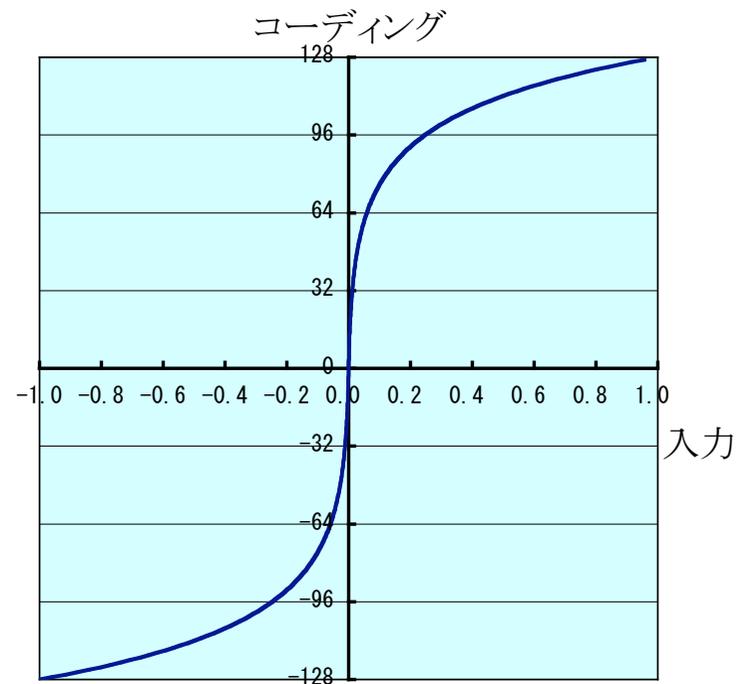
変換後の分布が一定となる μ

$$\log(1 + \mu) = \left(\frac{V}{\sqrt{P}} \right) \sqrt{2}$$

音声符号化

○ PCM μ 法則

- 線形PCMと比べて3-4ビットの改善効果がある。
- 8ビットの圧伸PCMは12ビットの線形PCMに匹敵する。



宿題

- CDでは、音声信号を44.1kHzで標本化し、これを16ビットで符号化する。ステレオ信号の場合、符号化後の伝送速度を求めよ。
- 授業の感想、要望等も記入してください。
- アンケートのエクセルファイルをダウンロードして、上記、感想、要望に加えアンケートにも回答してください。文章で回答する設問の場合、文字数制限は有りません。自由に記述してください。アンケートのファイル名は、今日の日付＋学籍番号としてください。

例:2011117+0929xx.xls

- メールで宿題の解答とアンケートを提出してください。
- 提出期限:12月1日午前10時30分。
- メールのタイトル:今日の日付＋学籍番号。

例:20111117+0929xx

- メールの宛先: information.net.2011@gmail.com

