第3回：　今日の目標

- 平均情報量を説明し、計算できる
- シャノンの通信モデルを説明できる
- 情報源符号化の条件を示せる
- 通信路符号化の意味を示せる
- 標本化定理を説明できる
- AD変換における量子化を説明できる
- 人間の五感による情報処理能力を推測できる

平均情報量

完全事象系（全ての要素の確率の和が1）

$$\mathbf{E} = \begin{bmatrix} E_1, E_2, E_3, \cdots, E_n \\ p_1, p_2, p_3, \cdots, p_n \end{bmatrix} \qquad i(E_i) = -\log_2 p_i$$

事象の発生

$E_2\ E_1\ E_1\ E_3\ E_1\ E_2\ E_3\ \cdots\ E_n$ N個の事象

例: a big earthquake occurred ・・・

事象E_iが発生した数 : m_i ($N=\Sigma m_i$)

事象E_iの情報量 : $-m_i \log_2 p_i$

$$\text{平均情報量} = \frac{\text{情報量の総和}}{\text{発生した事象の数}} \quad : H(\mathbf{E})$$

$$H(\mathbf{E}) = \frac{-\Sigma m_i \log_2 p_i}{N}$$

$$\lim_{N\to\infty} \frac{m_i}{N} = p_i$$

$H(\mathbf{E}) = -\Sigma p_i \log_2 p_i$ [bit/事象] ：完全事象系Eのエントロピー

（平均情報量）

英文（アルファベット＋スペース：27文字）

（１）出現確率が全て等しいなら

$$H_0 = -27 \times \frac{1}{27} \log_2 \frac{1}{27} = \log_2 27 = 4.75 \text{ bit}$$

（２）現実の出現確率を使うと

$$H_1 = -\Sigma p_i \log_2 p_i = 4.08 \text{ bit}$$

例：コインを投げた時の事象

A_1：表が出る、　　　A_2：裏が出る

$p(A_1) = p(A_2) = 0.5$のとき　⇦ 裏表対称なコイン

$$\mathbf{A} = \begin{bmatrix} A_1 & A_2 \\ 0.5 & 0.5 \end{bmatrix}$$

$$H(\mathbf{A}) = -0.5\log_2 0.5 - 0.5\log_2 0.5 = -\log_2 2^{-1} = 1 \text{ bit}$$

$p(B_1) = 0.75$、$p(B_2) = 0.25$のとき　⇦ 裏表いびつなコイン

$$\mathbf{B} = \begin{bmatrix} B_1 & B_2 \\ 0.75 & 0.25 \end{bmatrix}$$

$$H(\mathbf{B}) = -0.75\log_2 0.75 - 0.25\log_2 0.25 = 0.811 \text{ bit}$$

従って、

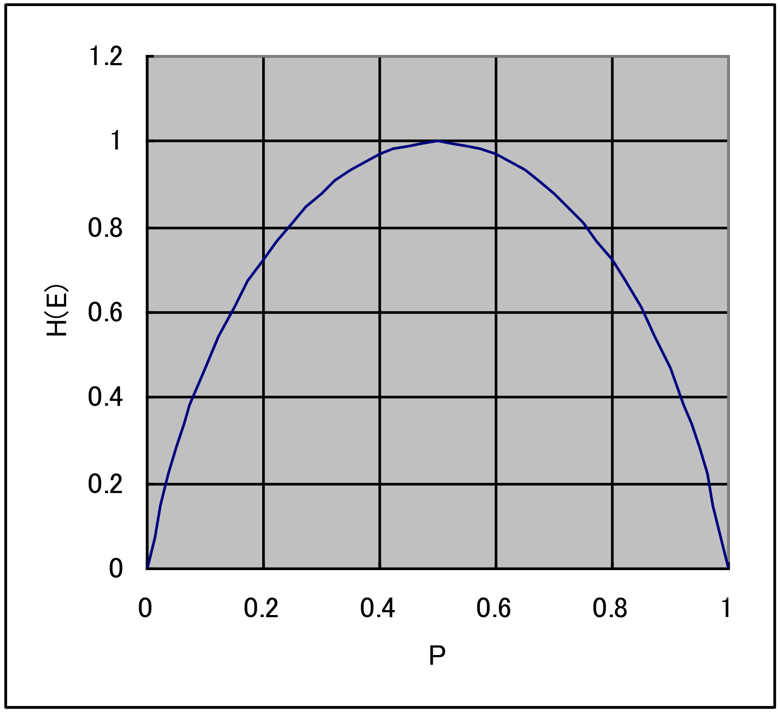
$$H(\mathbf{A}) \geqq H(\mathbf{B})$$

Aの方が予想がつき難い（不確かさが大きい）

無記憶二元情報源

$$\mathbf{E} = \begin{bmatrix} E_1 & E_2 \\ p & 1\text{-}p \end{bmatrix}$$

$$H(\mathbf{E}) = -p \log_2 p - (1-p) \log_2 (1-p)$$



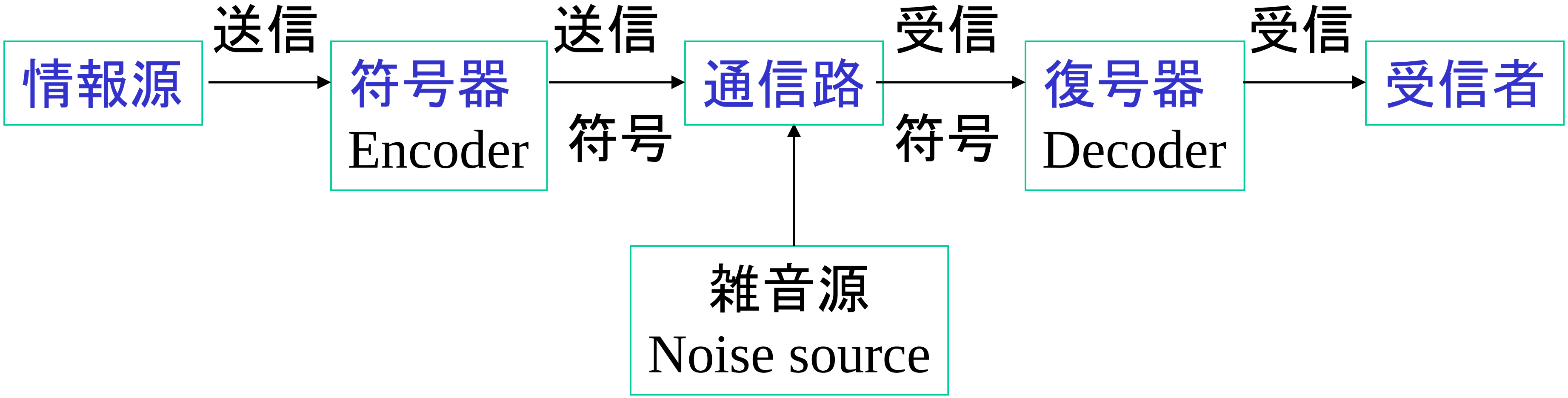
一般に

$$\mathbf{E} = \begin{bmatrix} E_1 & E_2 & \cdots\cdots\cdots & E_n \\ p_1 & p_2 & \cdots\cdots\cdots & p_n \end{bmatrix}$$

$p_1 = p_2 = \cdots = p_n$ のとき、$H(E) = \log_2 n$で最大

通信のモデル

;Shannon



情報伝達の基本定理

①通信路の容量

②雑音の混入

③通信路符号化定理

通信路の容量＞情報発生量

⇒雑音混入情報を100%復元できる

情報源符号化

記号　0　1　符号器

Vボルト

電気信号　0ボルト　通信路

情報	発生率	符号C_1	符号C_2
yes	0.3	0	1
no	0.2	1	10
Neutral	0.5	10	0

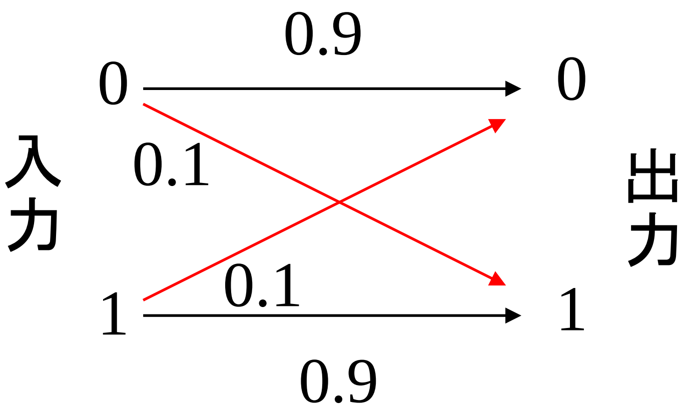
発生率の高い記号は短い符号にする

情報源を効率のよい符号に変換すること

＝1情報源記号当りの平均符号長を短くする（データ圧縮）

可逆符号：一意に復元可能

通信路符号化



雑音など

↓

誤った情報

情報源符号系列 ACAABC

情報系列 011001011

符号語系列 000111111000…

情報源 → 情報源符号器 → 通信路符号器 →

情報源符号器:
A⇒0
B⇒10
C⇒11

通信路符号器:
0⇒000
1⇒111

冗長化

情報源符号	通信路符号	受信
0	000	2つ又は3つ0ならば 0
1	111	2つ又は3つ1ならば 1

情報源　　情報系列　　受信系列　　復号情報

ACAABC⇒　011001011 —通信路→ 0|11|1|0|10|11 ⟶ ACBBC

雑音

通信路符号化

000111111000000111000111111　　000|011|111|000|100|111|001|111|101

0 | 1 1 | 0 | 0 | 1 0 | 1 1

A | C | A | A | B | C

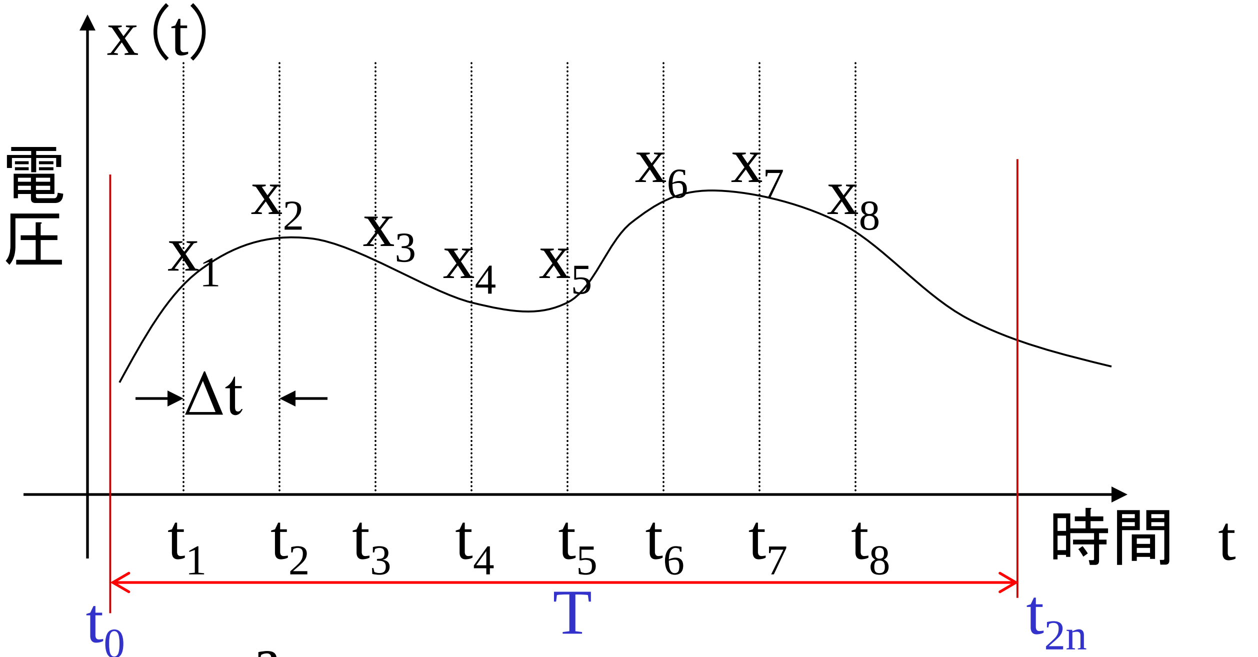
000,001,010,100=>0

$0.9^3+3\times0.9^2\times0.1=0.972$

誤り確率：0.1⇒0.028：信頼性の向上

標本化定理

アナログ情報:音、明るさ、温度、・・・⇒電圧



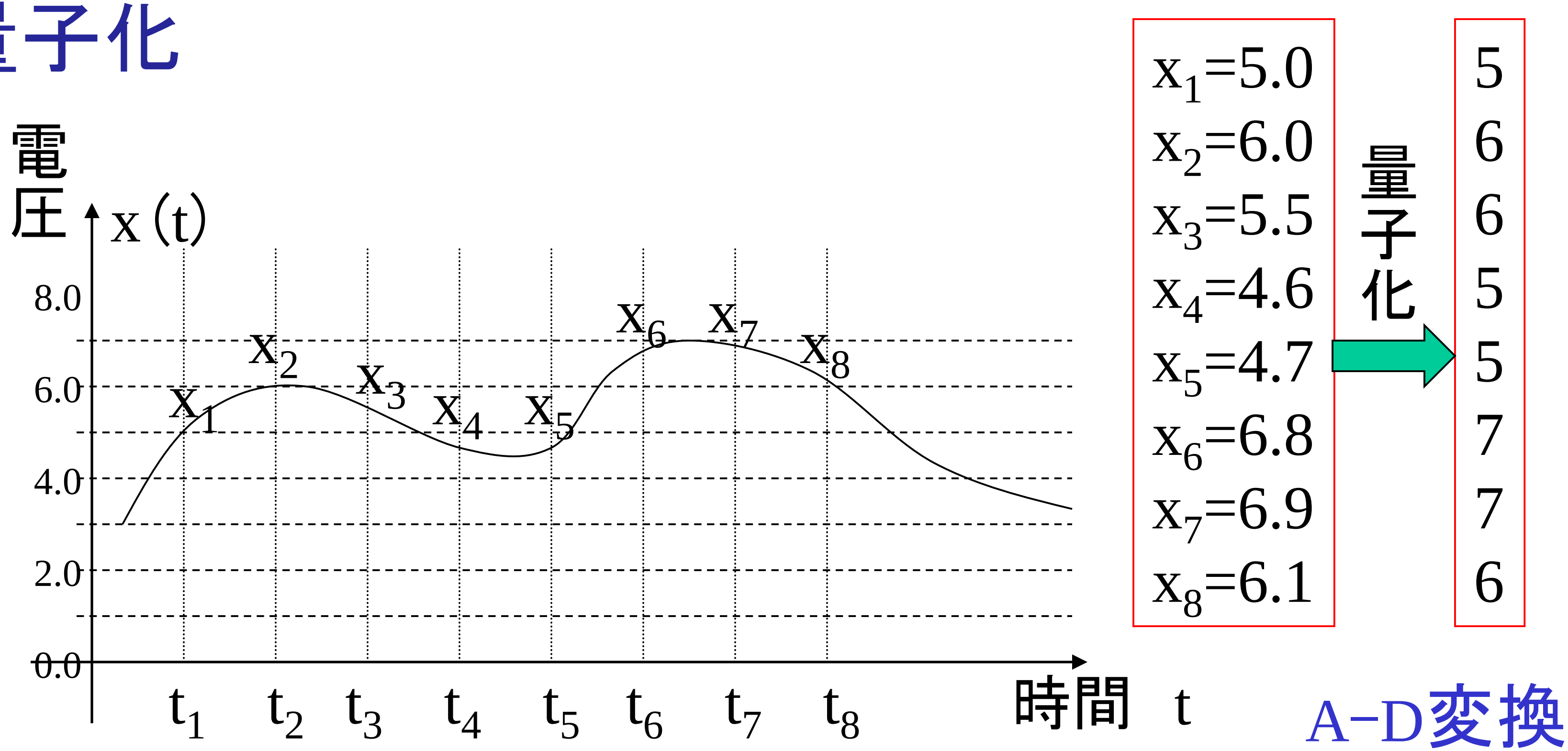
$$x(t) = \frac{a_0}{2} + \sum_{n=1}(a_n \cos 2\pi nft + b_n \sin 2\pi nft) \qquad f = 1/T$$

f: 基本周波数

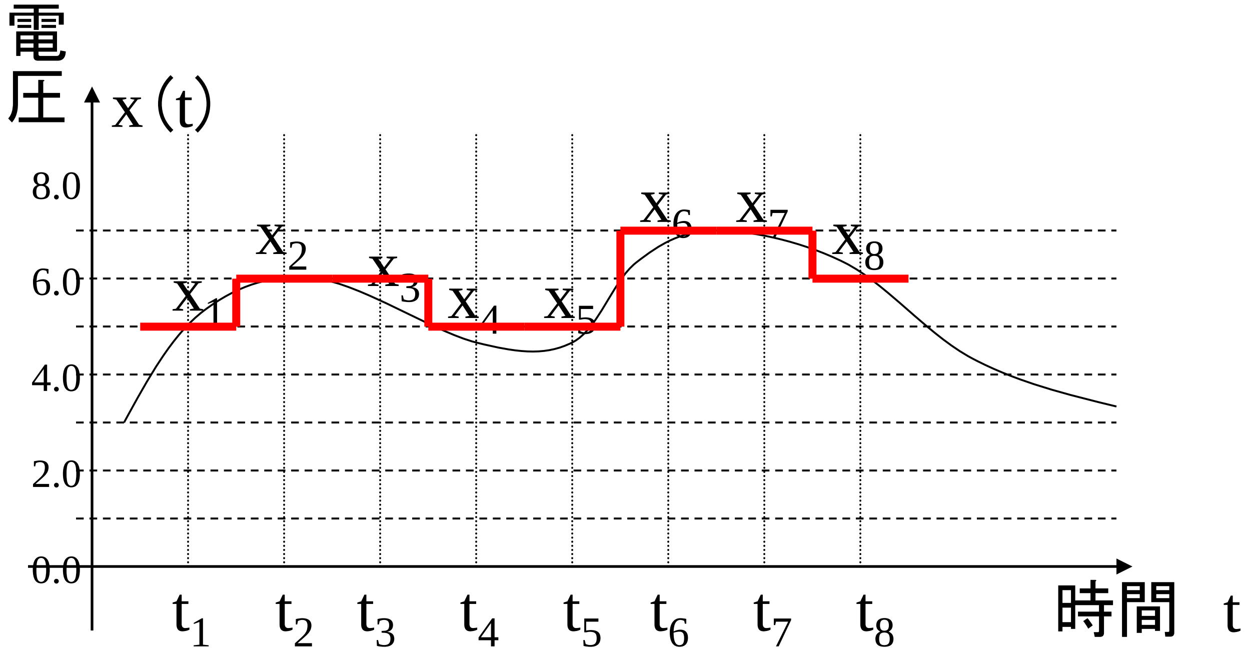
w≧nfならば、$\Delta t = 1/2w = T/2n$でサンプリングすれば、x (t) を完全に再現できる。

$\frac{1}{2w}$:ナイキスト間隔

量子化


電圧
x (t)
8.0
6.0
4.0
2.0
0.0
x_1 x_2 x_3 x_4 x_5 x_6 x_7 x_8
t_1 t_2 t_3 t_4 t_5 t_6 t_7 t_8
時間 t
x_1=5.0
x_2=6.0
x_3=5.5
x_4=4.6
x_5=4.7
x_6=6.8
x_7=6.9
x_8=6.1
量子化
5
6
6
5
5
7
7
6
A-D変換

D-A変換

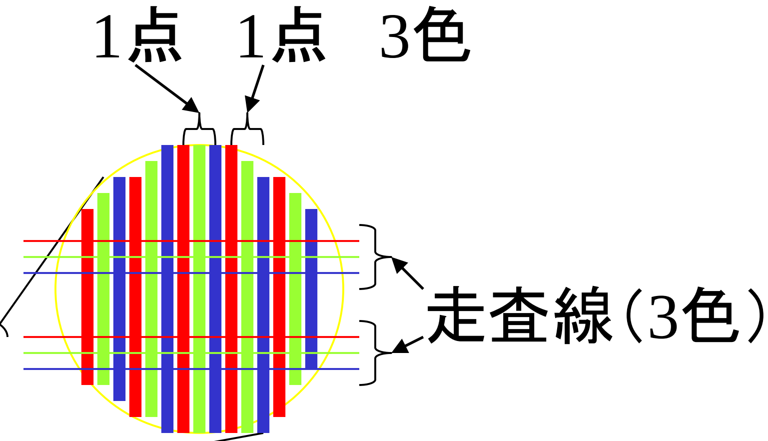


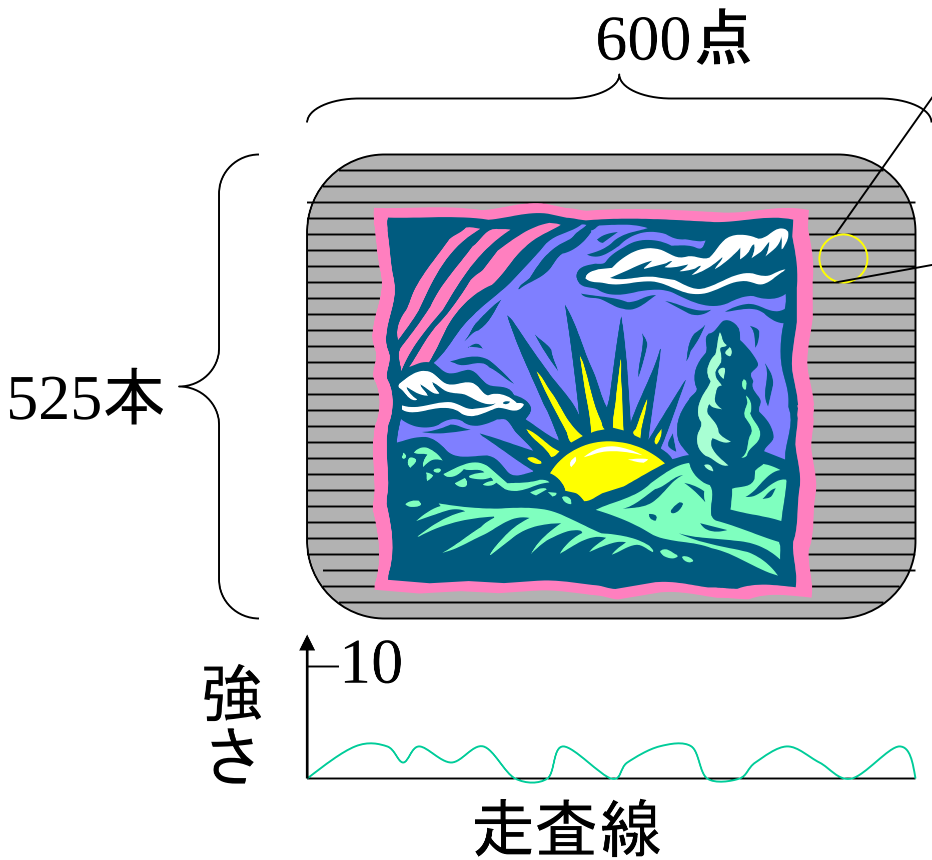
歪⇒量子化雑音

人間の情報処理能力

（1）目で見える能力

テレビ画面





1点の場合の数：10×3

画面全体の場合の数

$(10 \times 3)^{525 \times 600}$

1画面の情報量

$$I = \log_2 (10 \times 3)^{525 \times 600}$$
$$= 525 \times 600 \times \log_2 (30)$$
$$\fallingdotseq 1.5 \times 10^6 \text{ bit}$$

1秒間30コマ

$4 \sim 5 \times 10^7$ bit/s

Windows:96d/i, Mac:72d/i

（２）目で見る能力

日本語を意識して読む:　10文字／s

文章:　漢字＋ひらがな　＝　約3000個

1文字あたりの情報量

$I = \log_2 3000 \fallingdotseq 12$ bit/文字

10 [文字／s]× 12 [bit/文字] = 120 bit/s

（３）聞こえる能力

FFT

耳の感度:　　音の振幅30dB～90dB

周波数100Hz～10000Hz（会話）

10Hz～50000Hz（全身）:ピーク5000Hz

$a\,[\mathrm{dB}] = 20 \log_{10}(A/A_0)$、　$A_0 = 2 \times 10^{-4}\,\mathrm{dyn/cm^2}$

60dB幅　$\Rightarrow$　$A/A_0 = 10^{60/20} = 10^3$

1秒間の音のサンプル:1～10000 Hz

1秒間の情報量:　$\log_2 1000^{10000} = 10^4 \log_2 10^3 \fallingdotseq 10^5$ bit/s

演習

1. いろは48文字が独立で等確率で出現する場合の平均情報量はいくらか。

2. 1と6の目が出る確率がそれぞれ1/4、他の目の出る確率はすべて1/8のサイコロの平均情報量はいくらか。

3. 振幅を1024のレベルで区別し、0.1ms ごとにサンプリングした音声波形1分間のデータは何ビットか。

4. カラーディスプレイの画面を横600点、縦400点に分けて1点の３原色のレベルをそれぞれ8ビットで区別すると、1画面当り何バイトの情報になるか。また、1点では何色を区別できるか。

情報科学概論のトップへ
明治薬科大学のホームへ