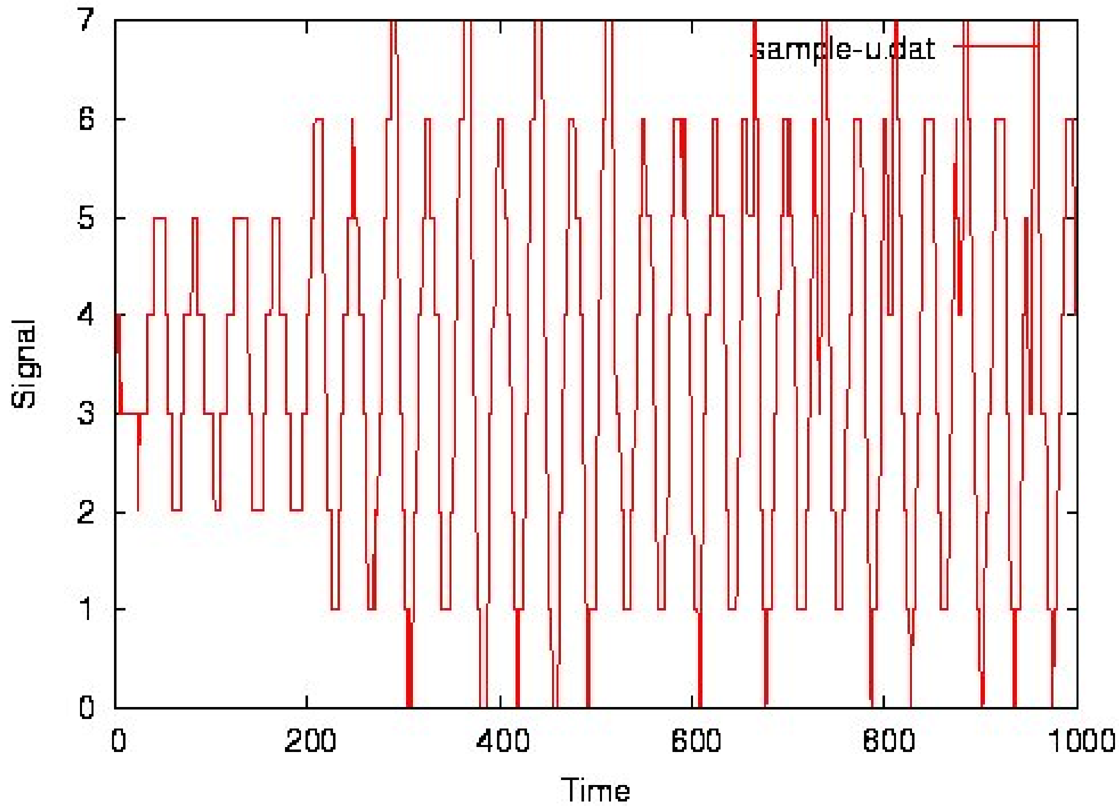
2009年　卒業研究

隠れマルコフモデルによる時間領域での音声信号の学習と生成

宮崎大学工学部情報システム工学科
黒田　優士

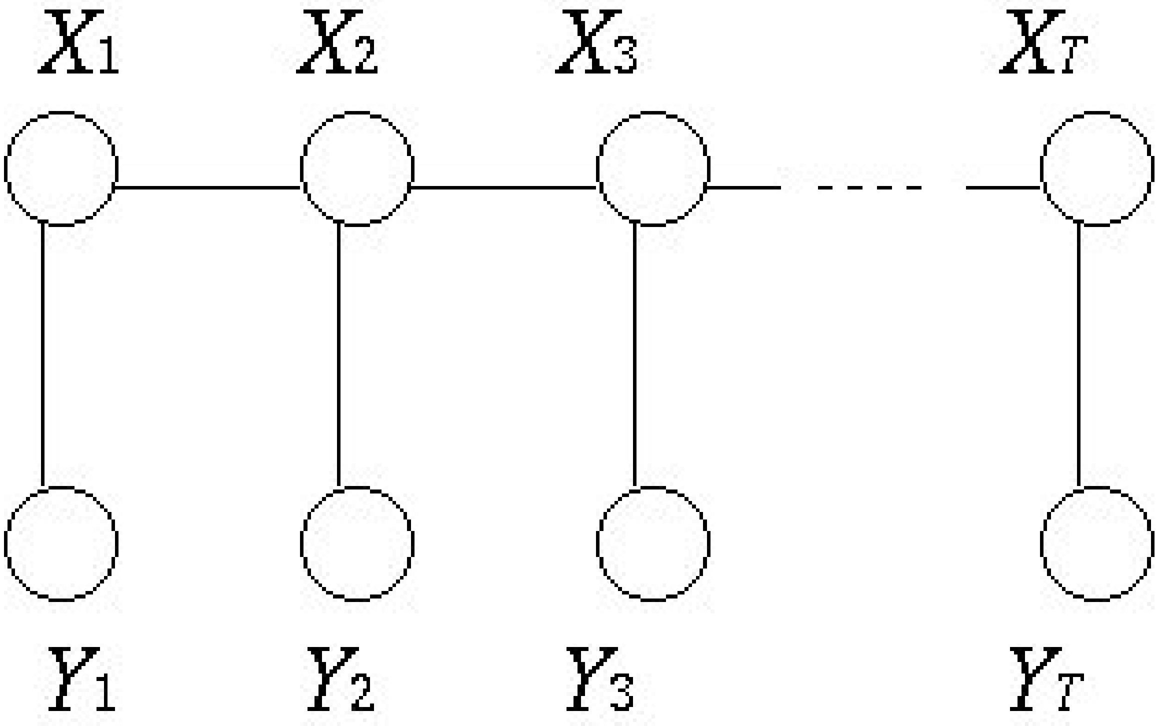
音声信号をモデル化したい

[u]



音声信号　→　複雑な確率分布をもっている

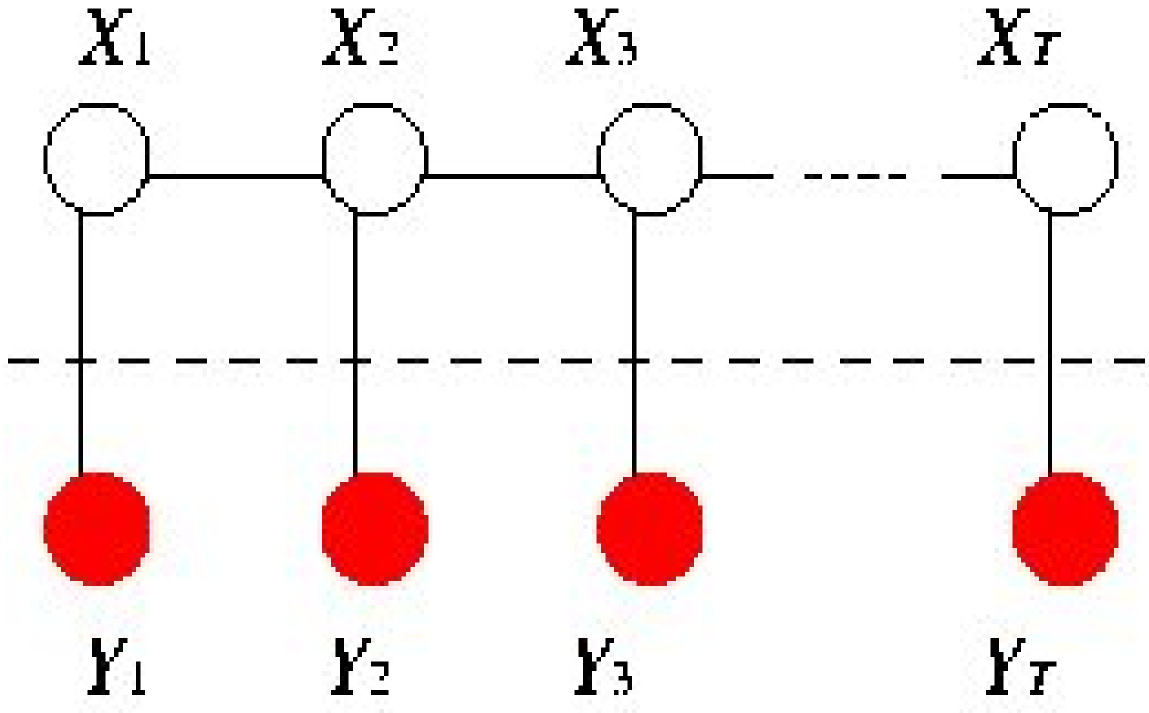
背景



単純な隠れマルコフモデル

背景

観測値：Y



単純な隠れマルコフモデル

→　複雑な同時確率分布 $P(Y_1, Y_2, \ldots, Y_T)$ をモデル化できる

目的

隠れマルコフモデル

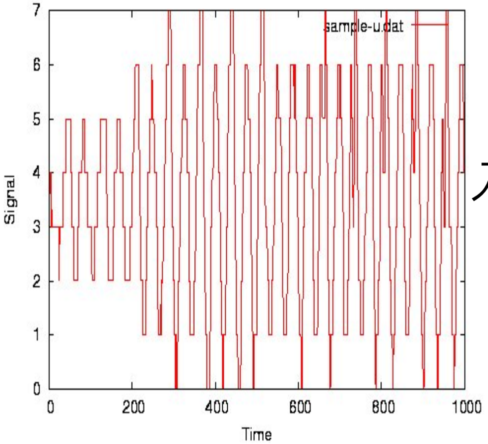
観測値 Y についての確率分布を精度良くモデル化できる

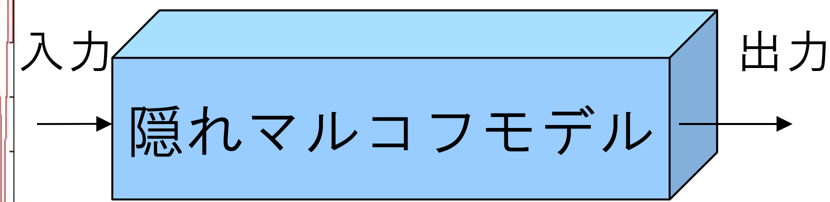
↓

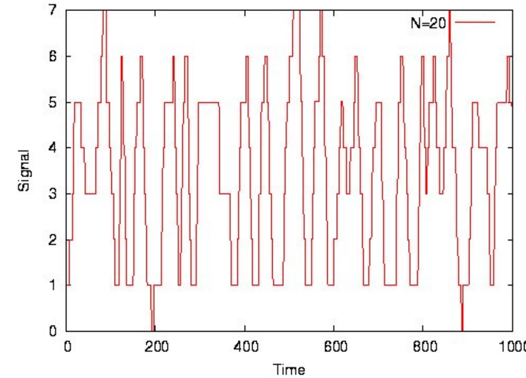
実際の音声信号を Y に与えて

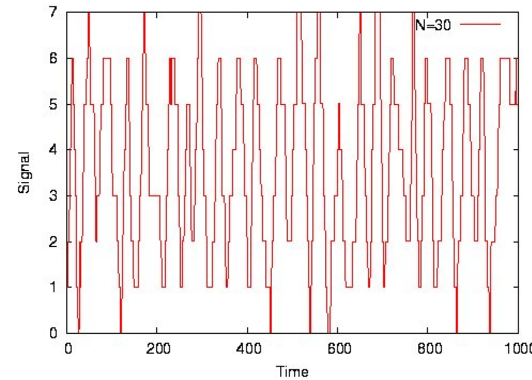
その音声信号をモデル化できるか調べる

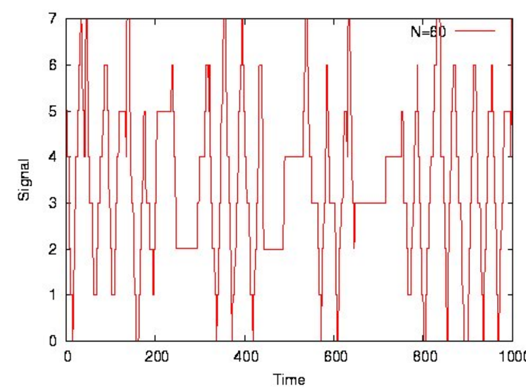
全体の流れ











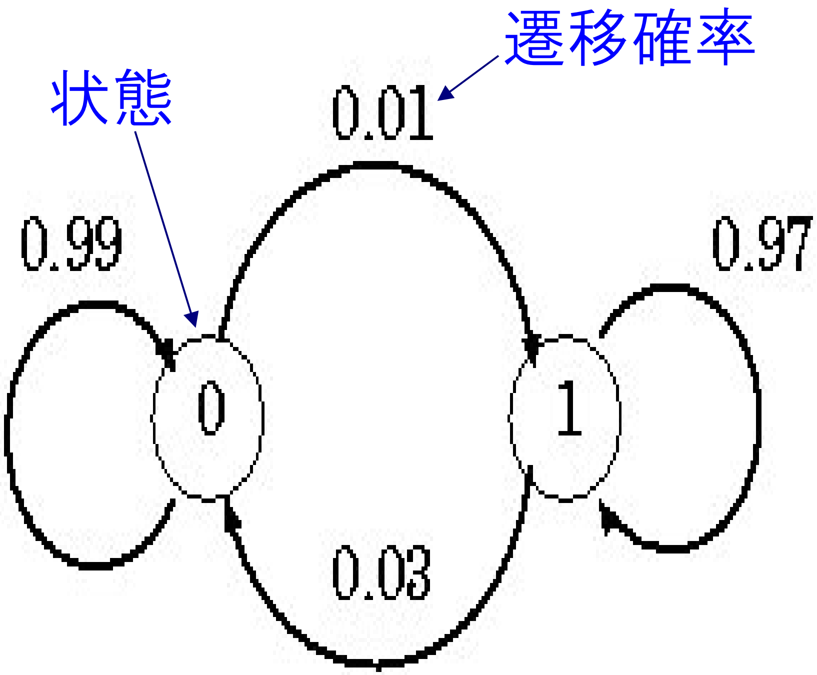
パラメータを学習

データを確率的に生成

目次

1. 隠れマルコフモデル
2. 学習アルゴリズム
3. 実験結果
4. まとめ

マルコフ過程



状態遷移図

確率変数 $X \in \{0,1\}$

遷移確率行列 P

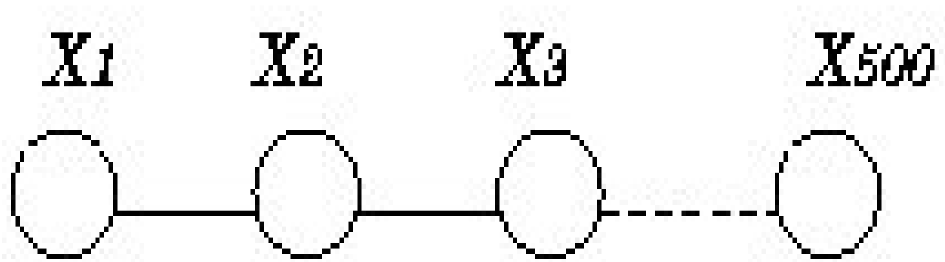
$$P = \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix} = \begin{bmatrix} 0.99 & 0.01 \\ 0.03 & 0.97 \end{bmatrix}$$

ランダムサンプル

500回遷移

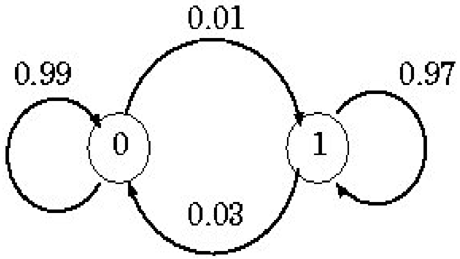
$$2^{500} \left\{ \begin{array}{l} 00000.......000000 \\ \vdots \\ 10011.......000000 \\ \vdots \\ 11111.......111111 \end{array} \right.$$

マルコフ過程

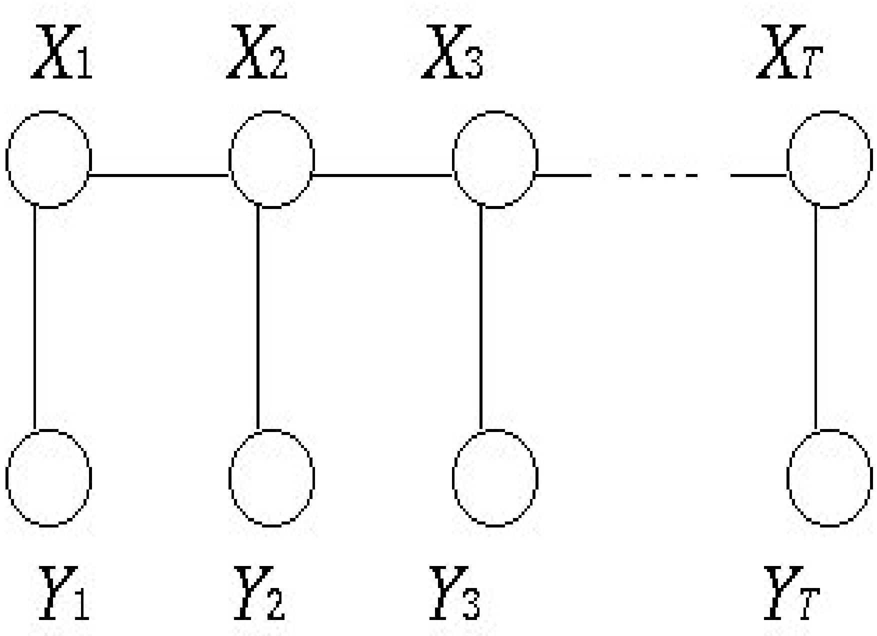


確率変数 X の依存性　$X = X_1, X_2, X_3, \ldots, X_{500}$

$$P(X_2 | X_1, X_3, \ldots, X_{500}) = P(X_2 | X_1, X_3)$$



隠れマルコフモデル



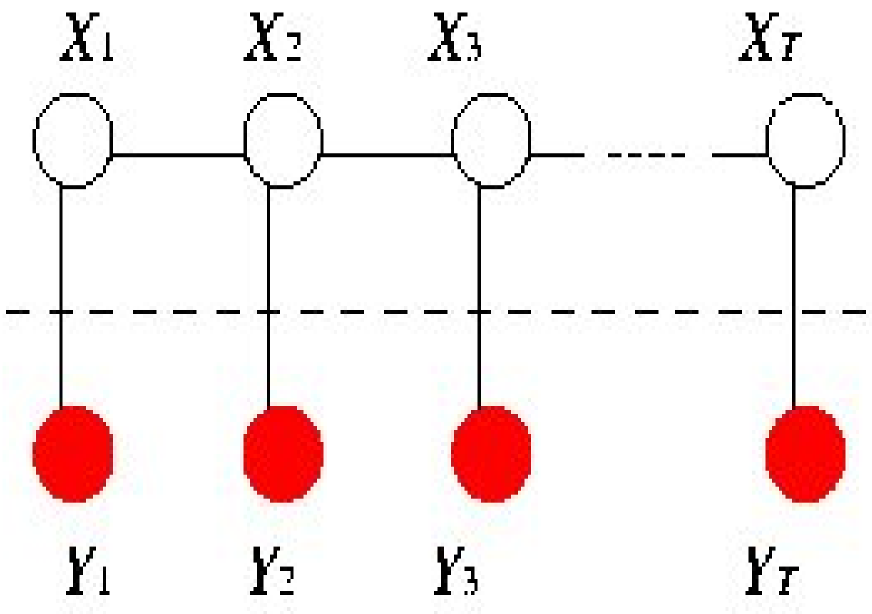
X ：マルコフ過程

Y ：マルコフ過程ではない

確率変数 $Y: Y_1, Y_2, Y_3, \ldots, Y_T$

$$Y_t = f(X_t)$$

隠れマルコフモデル



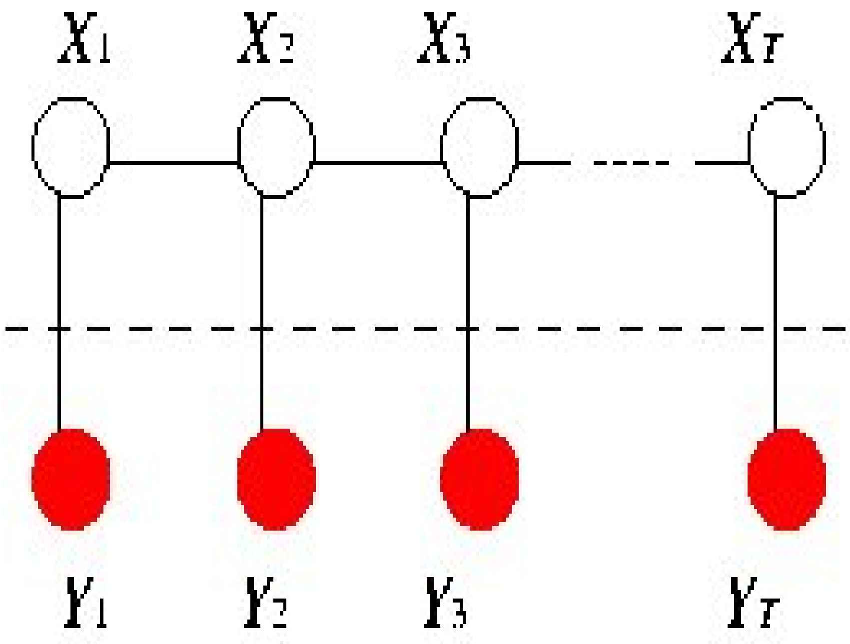
Xの状態遷移確率$\{p_{ij}\}$を学習

赤：観測値：Y

確率変数 $Y = Y_1, Y_2, Y_3, \ldots, Y_T$

$$Y_t = X_t \bmod 8$$

学習アルゴリズム



赤：観測値

$X_t \in \{1,\ldots,N\}$

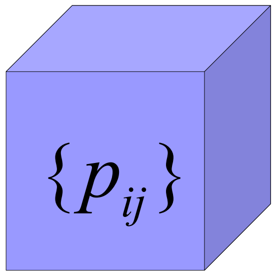
$Y_t = X_t \bmod 8$

$i, j \in \{1,\ldots,N\}$

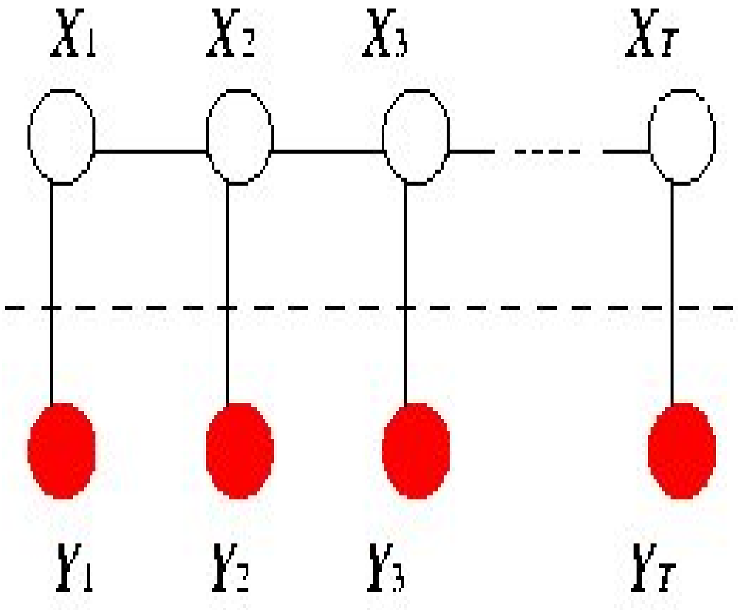
状態 i から j への遷移確率：p_{ij}

$N \times N$ 個の p_{ij} を求める

$X_1 = i$ である確率：p_i



学習アルゴリズム



$$p_{ij} = \frac{\sum_{t=1}^{T-1} p(X_t = i, X_{t+1} = j, \mathbf{y})}{\sum_{t=1}^{T-1} p(X_t = i, \mathbf{y})}$$

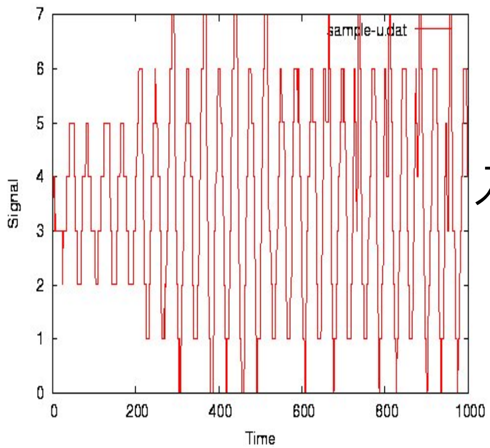
$$p_i = \frac{\sum_{t=1}^{T-1} p(X_t = i | \mathbf{y})}{\sum_{j=1}^{N} \sum_{t=1}^{T-1} p(X_t = j | \mathbf{y})}$$

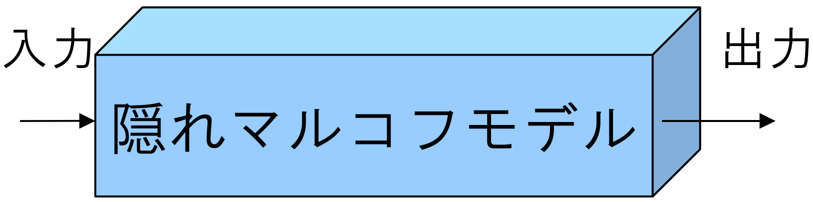
$$i, j \in \{1, \ldots, N\}$$

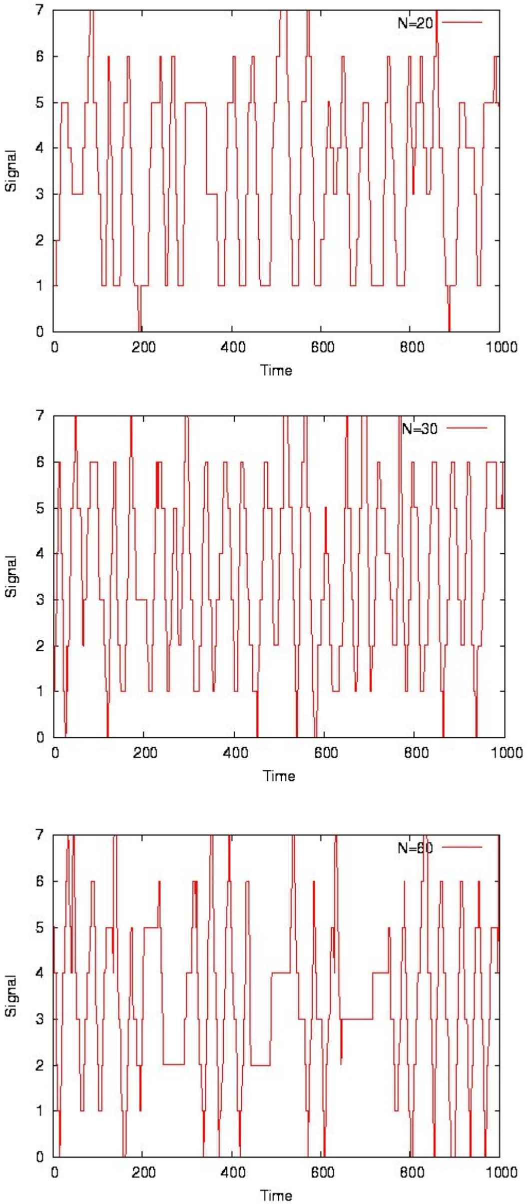
$$\mathbf{y} = (y_1, y_2, \ldots, y_T)$$

実験

実験の流れ



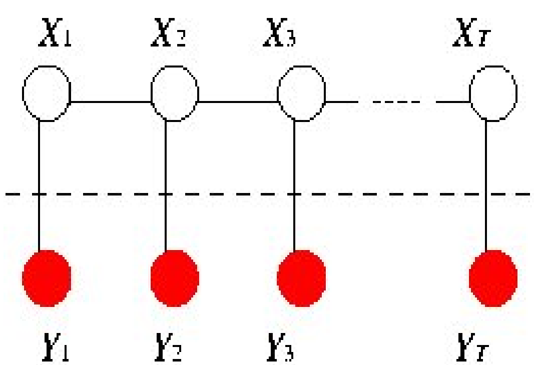




$\{p_{ij}\}$ を学習

↓

確率的に波形を生成



音声信号の学習と生成

- 学習データ
  1. 母音：[a], [i]
  2. 子音：[k], [m]

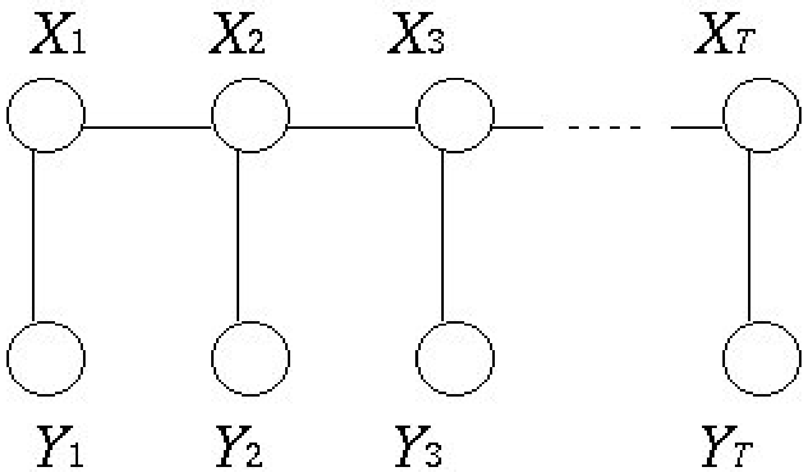
$$Y_t = X_t \bmod 8 \ (X_t \in 1,\ldots,N)$$

パラメータの数：$N \times N$

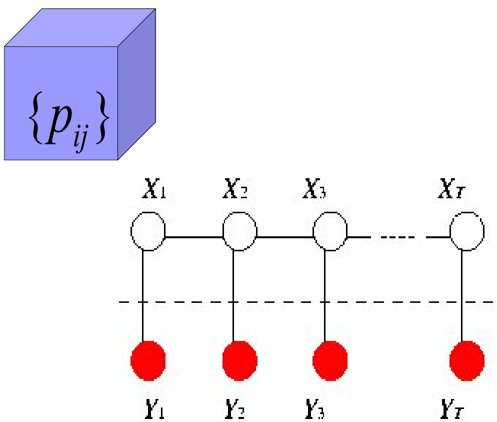
$$N = 10,30,60,100$$

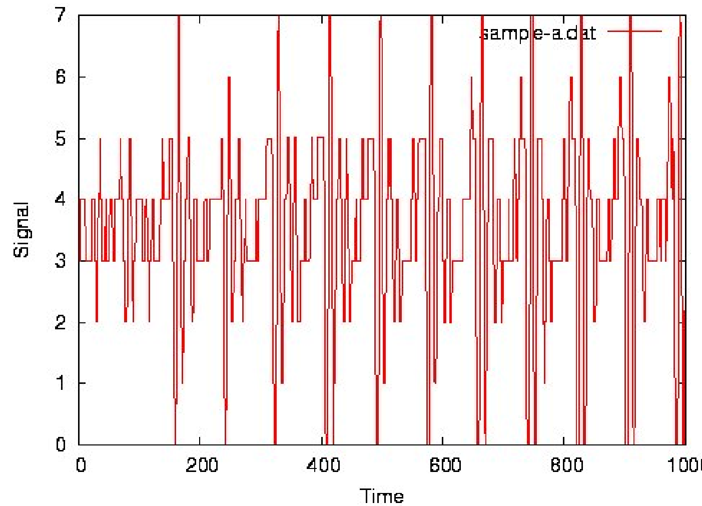
パラメータの初期値：ランダム

$$0 \le p_{ij} \le 1,\ \sum_j p_{ij} = 1$$



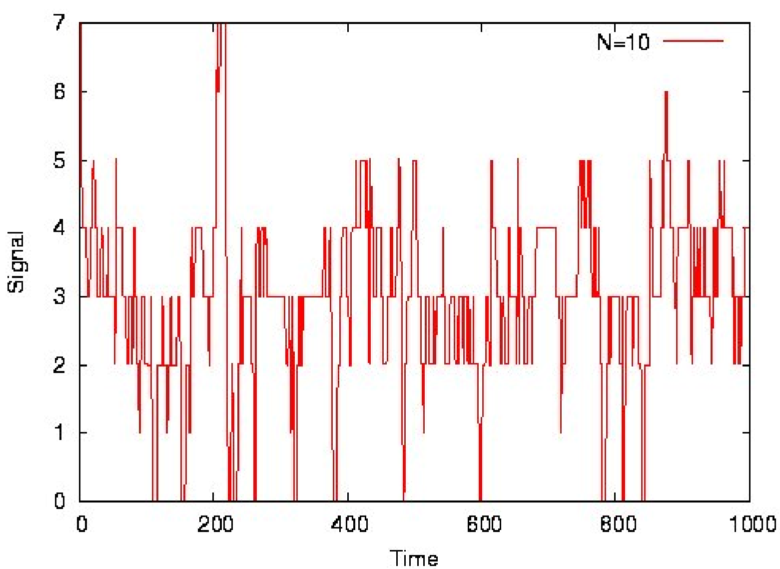
波形[a]

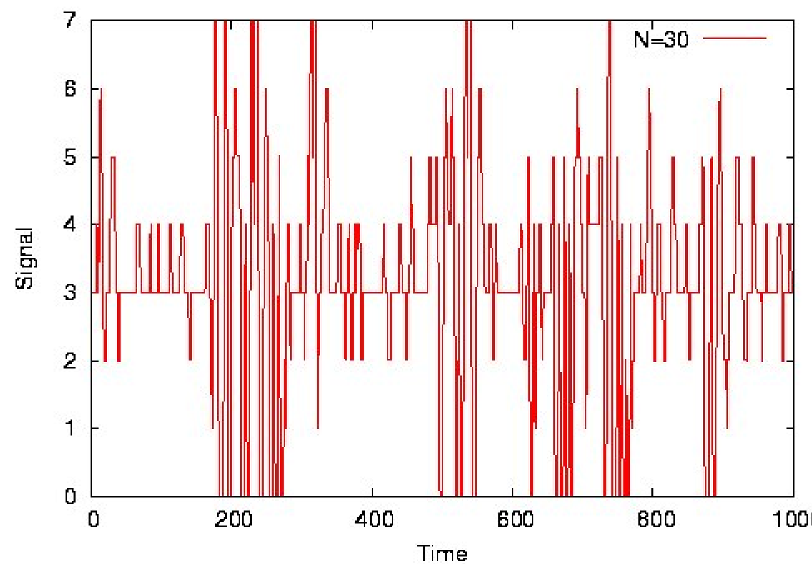




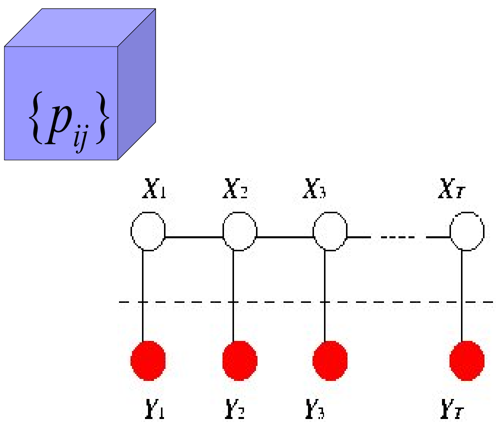
学習データ
10kHz,　8値
データ数T =2250

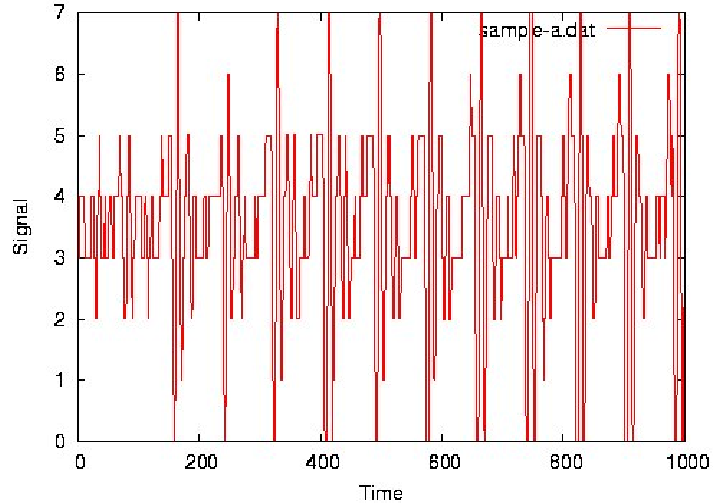
N = 10　　生成された波形の例　　N = 30





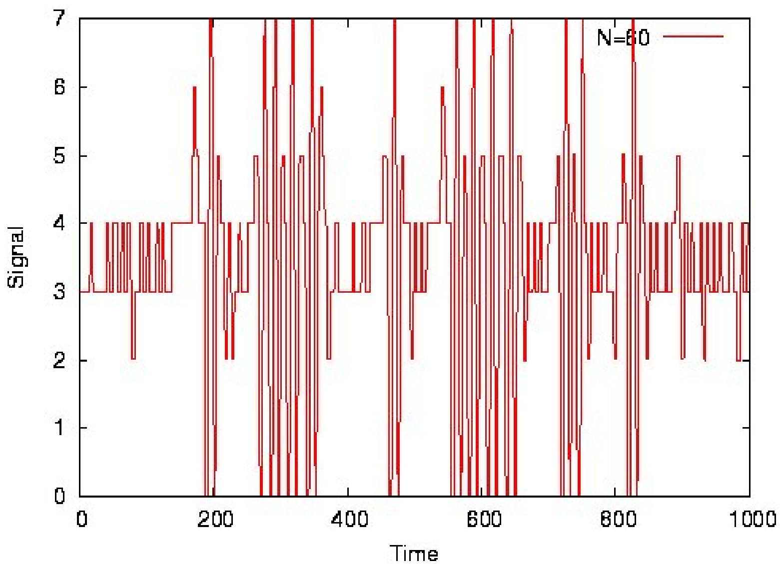
波形[a]

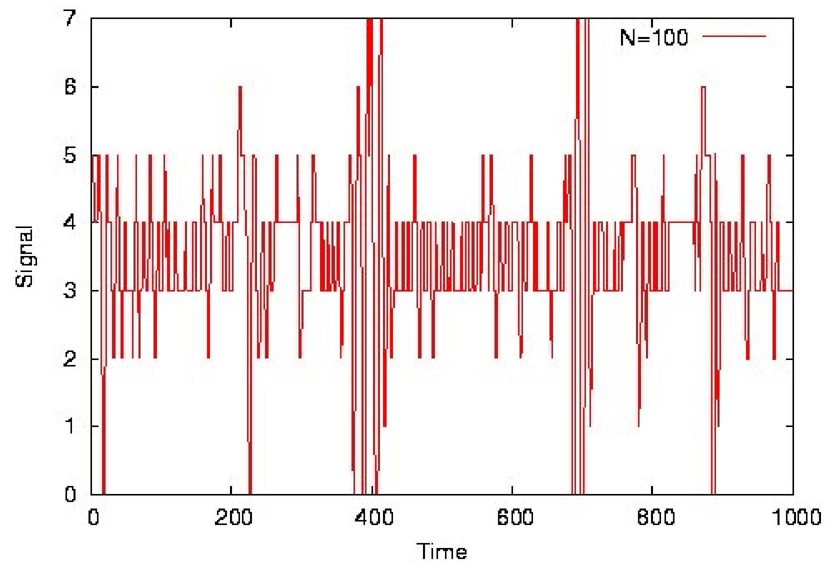




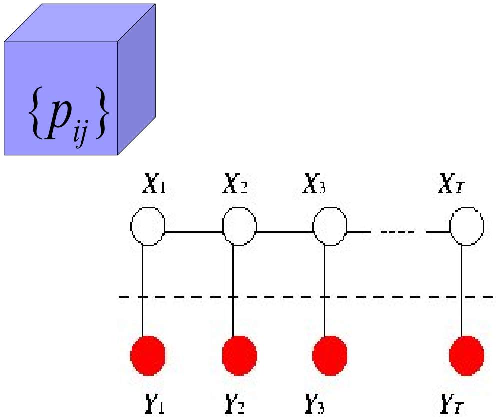
学習データ
10kHz,　8値
データ数T =2250

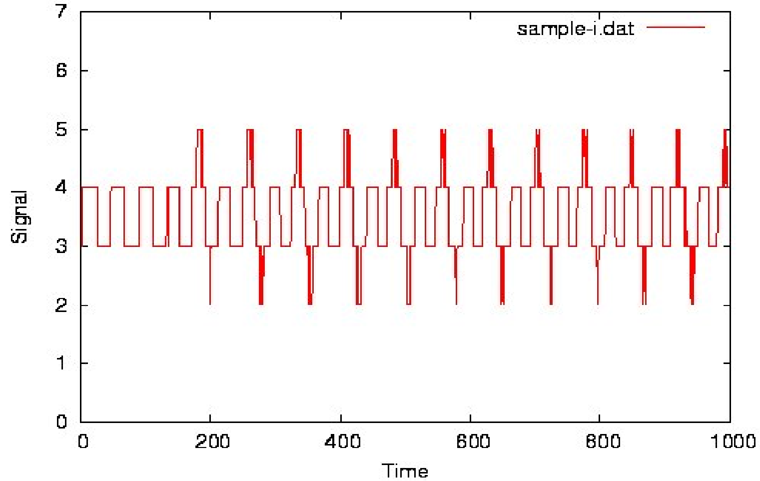
N = 60　　生成された波形の例　　N = 100





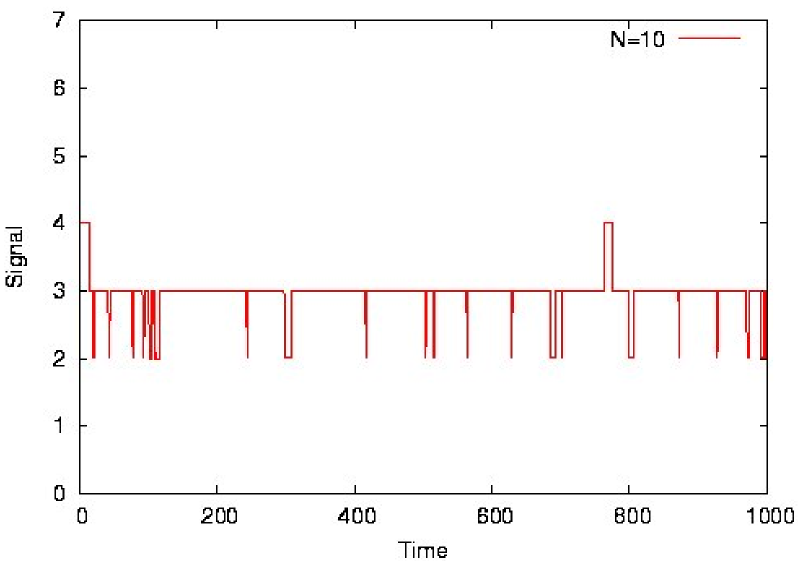
波形[i]

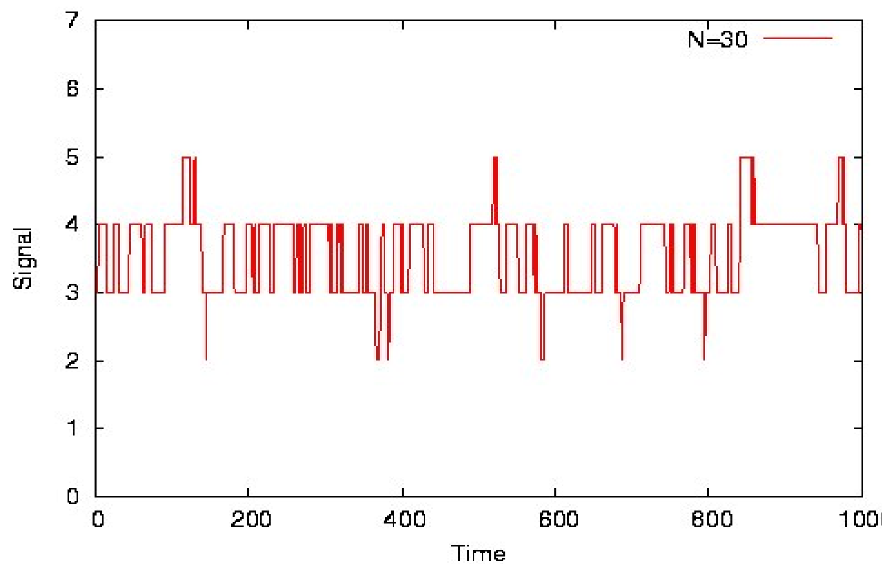




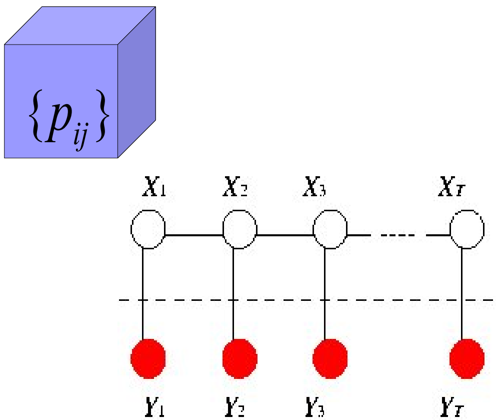
学習データ
10kHz,　8値
データ数T =2250

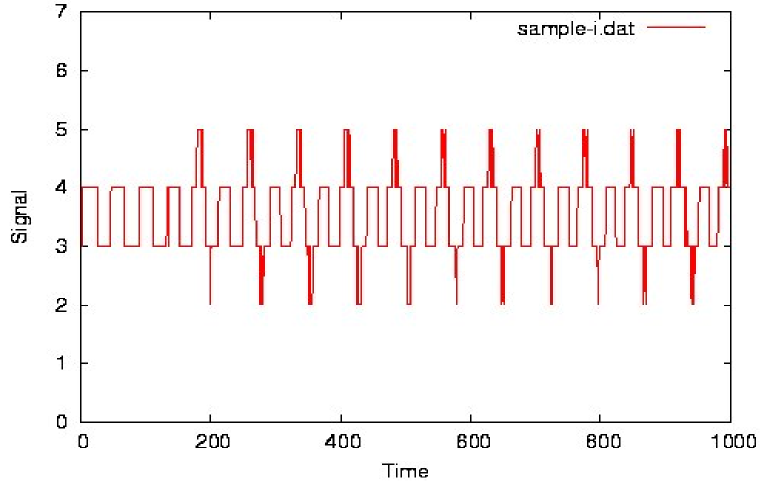
N = 10　　生成された波形の例　　N = 30





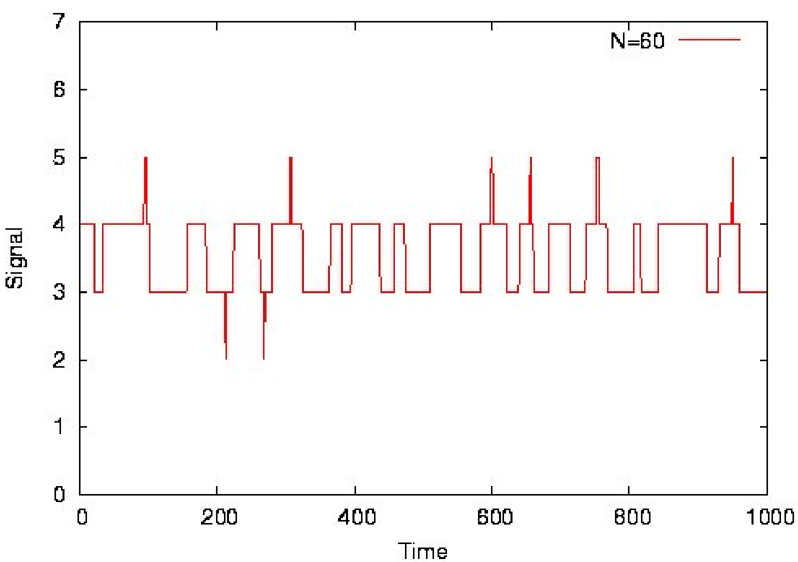
波形[i]

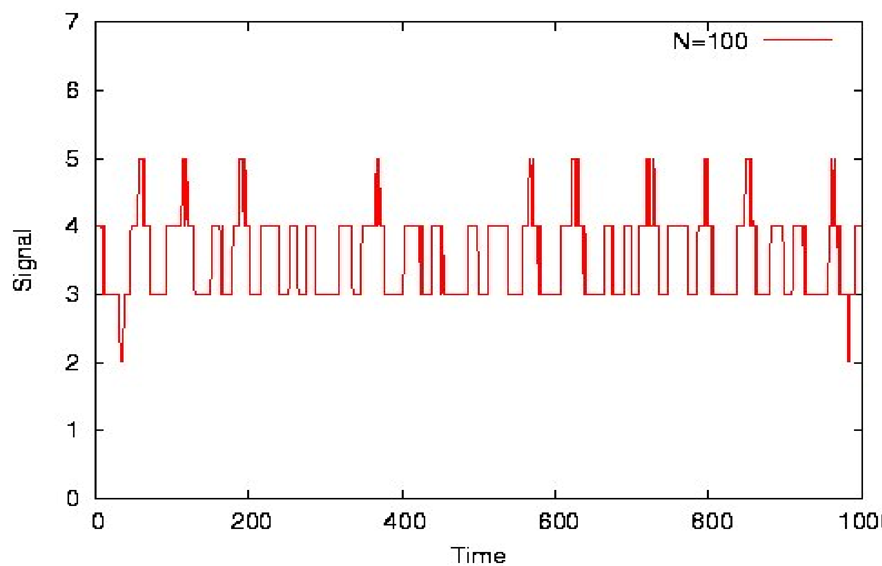




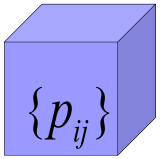
学習データ
10kHz,　8値
データ数T =2250

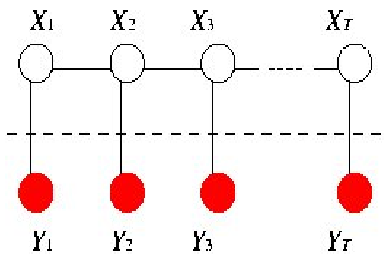
N = 60　　生成された波形の例　　N = 100

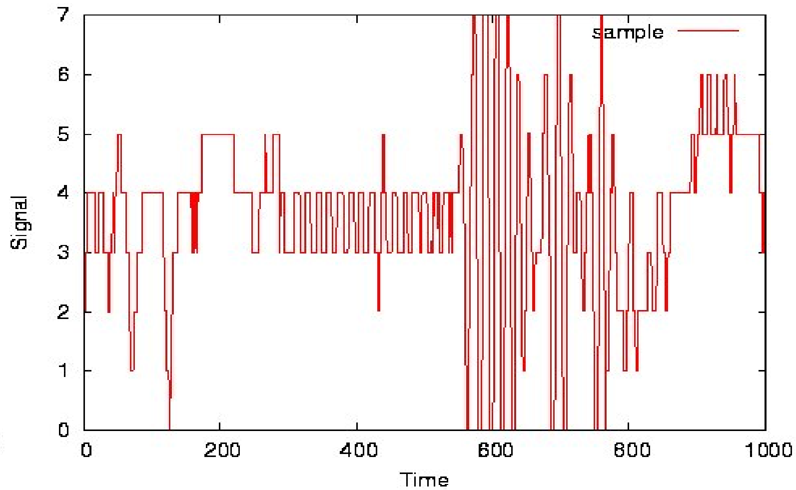




波形[k]





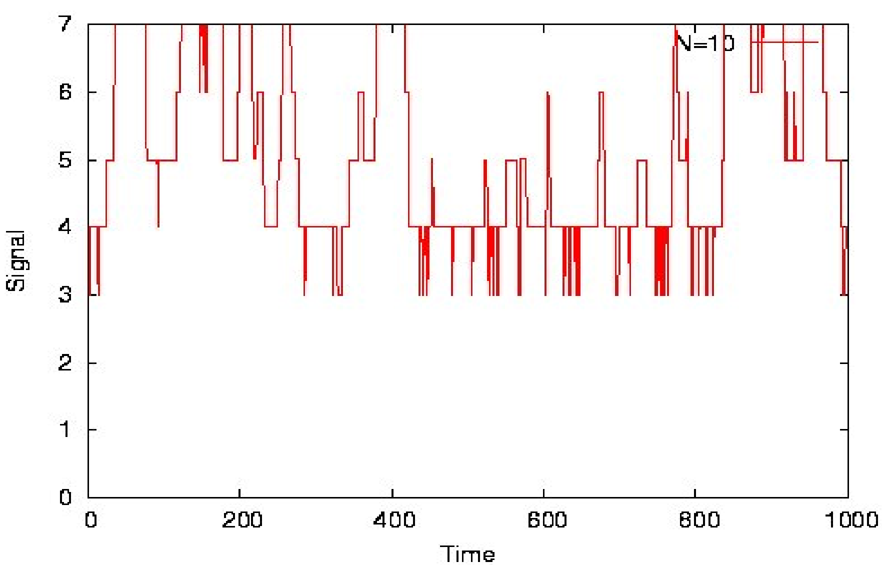


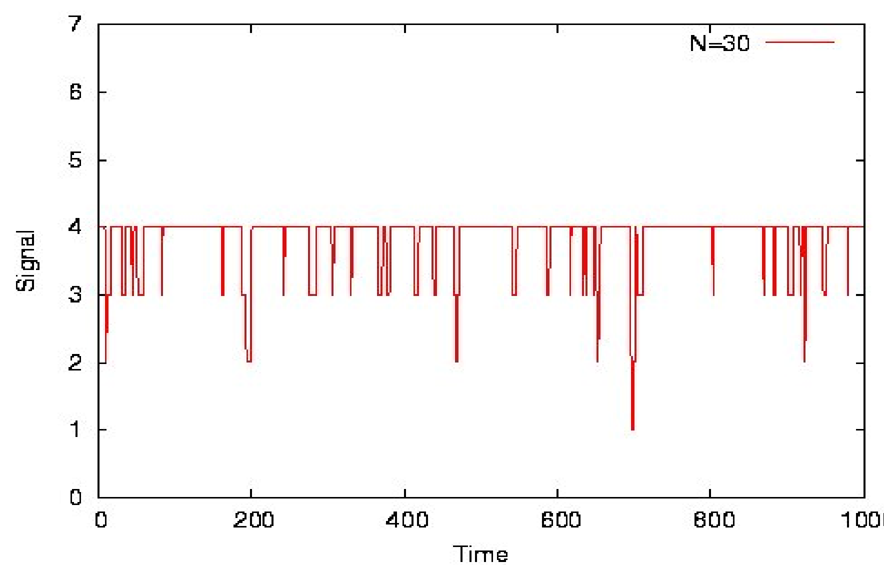
学習データ
10kHz,　8値
データ数T =3486

N = 10

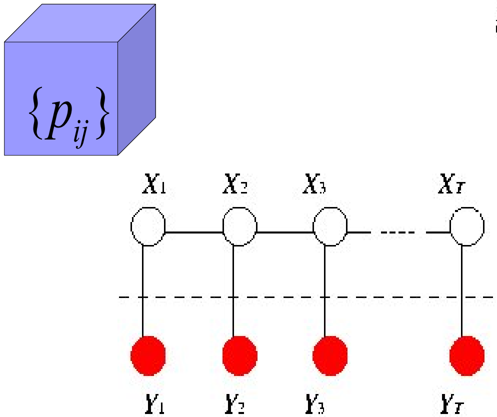
生成された波形の例

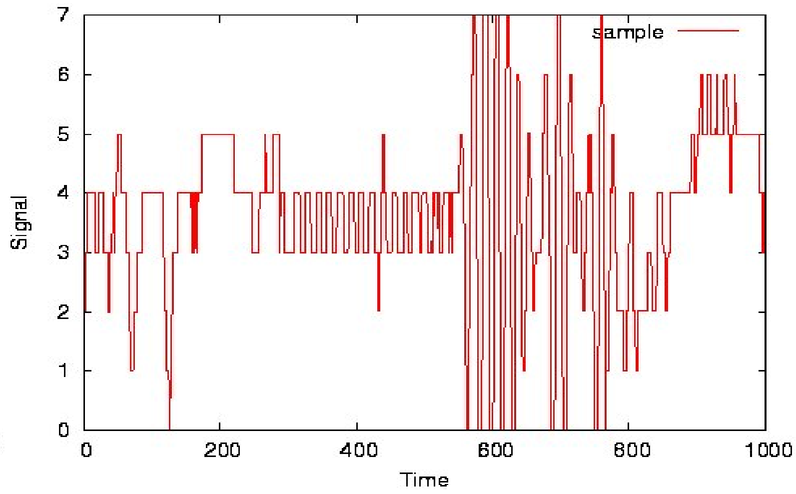
N = 30





波形[k]



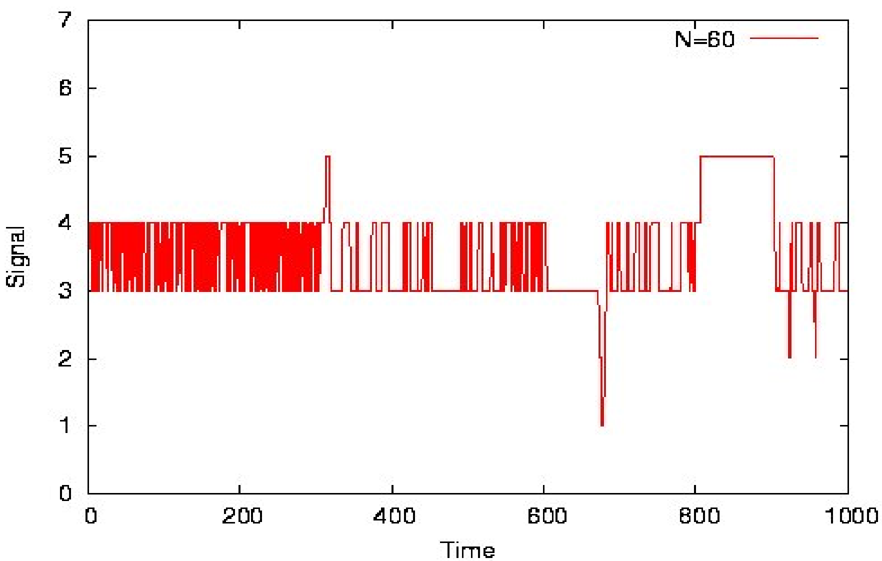


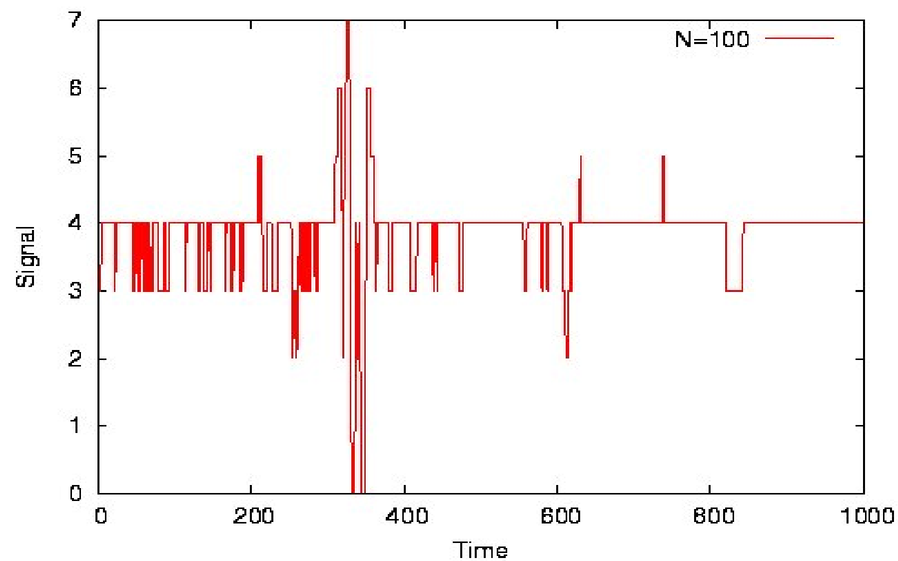
学習データ
10kHz,　8値
データ数 T =3486

N = 60

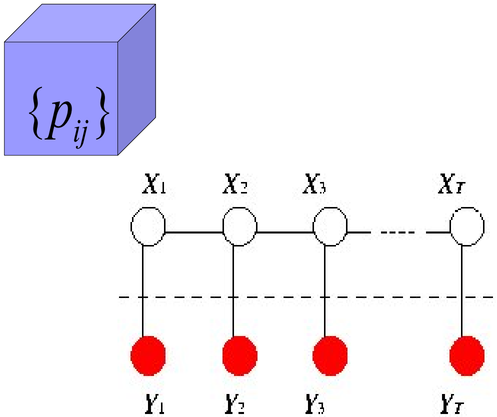
生成された波形の例

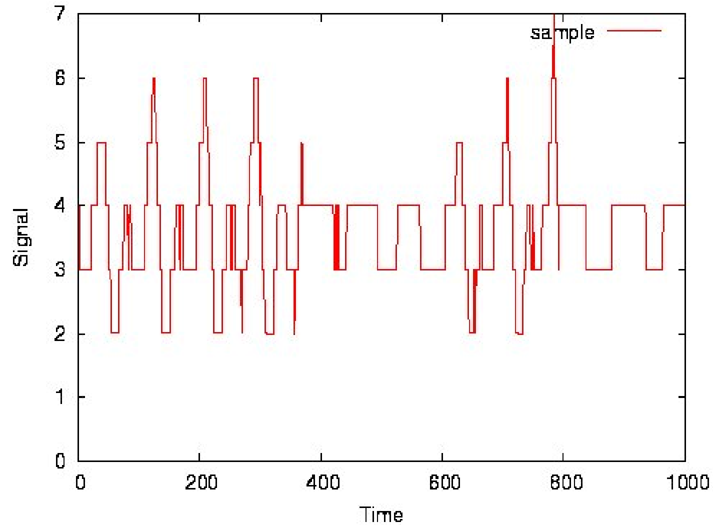
N = 100





波形[m]



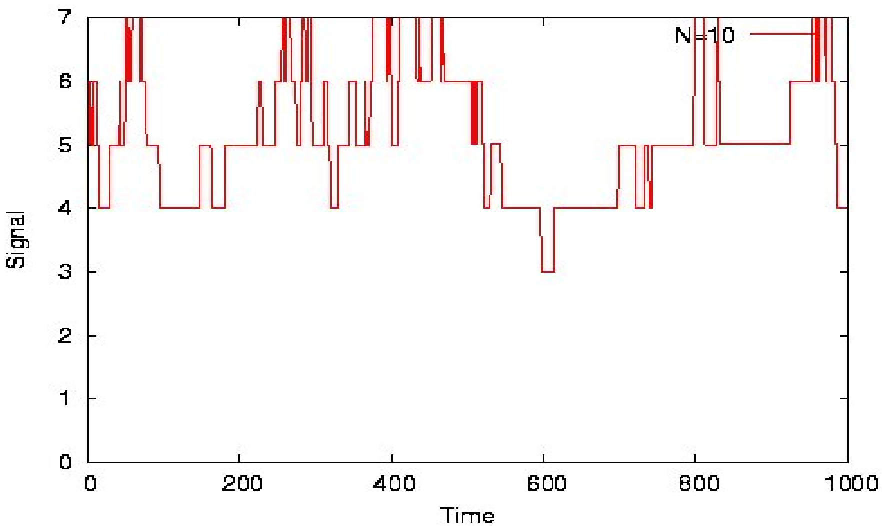


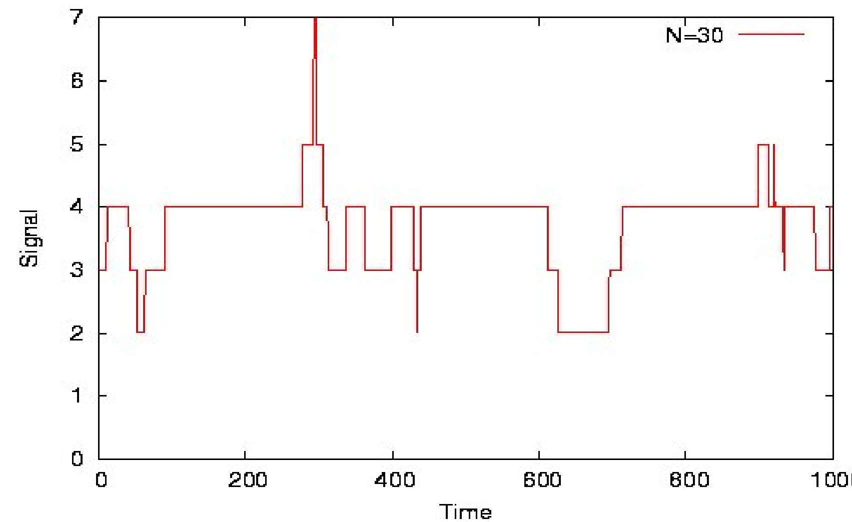
学習データ
10kHz,　8値
データ数T =2911

N = 10

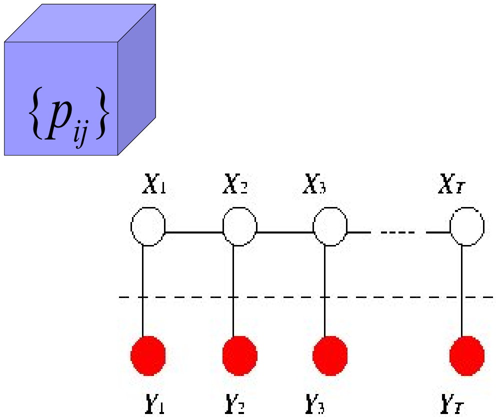
生成された波形の例

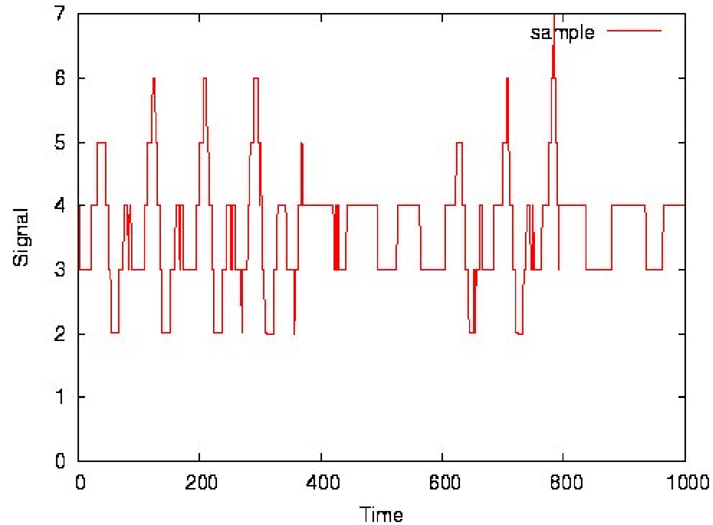
N = 30





波形[m]



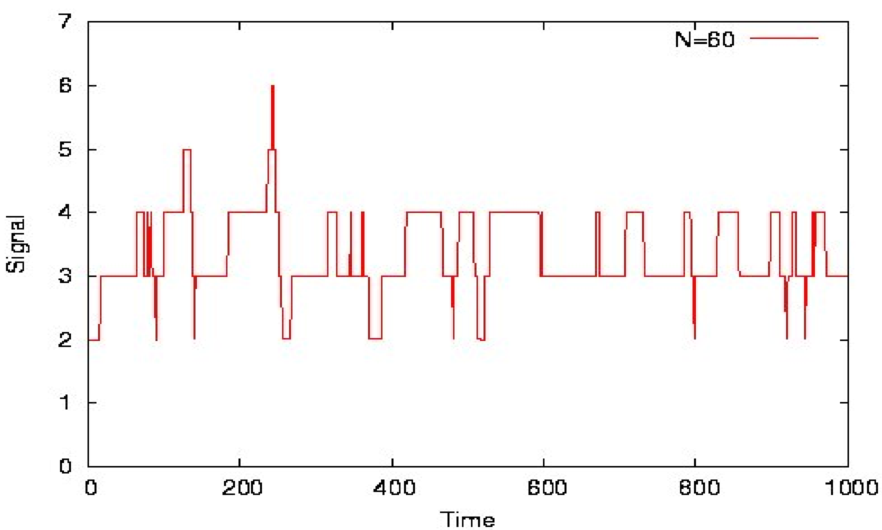


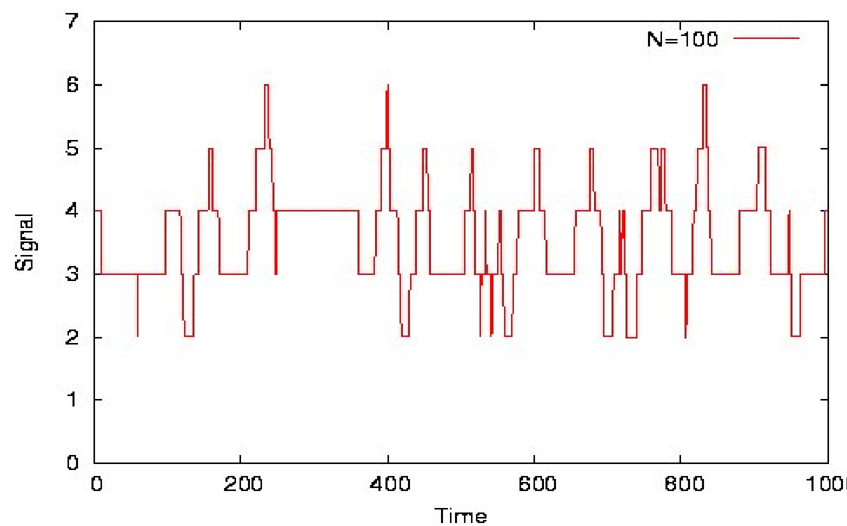
学習データ
10kHz,　8値
データ数T =2911

N = 60

生成された波形の例

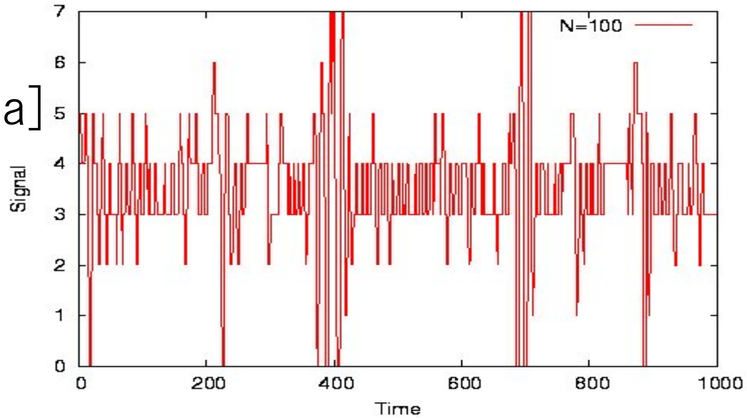
N = 100



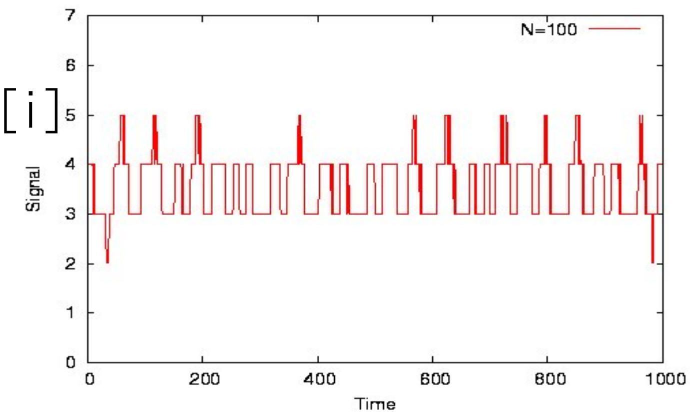


考察

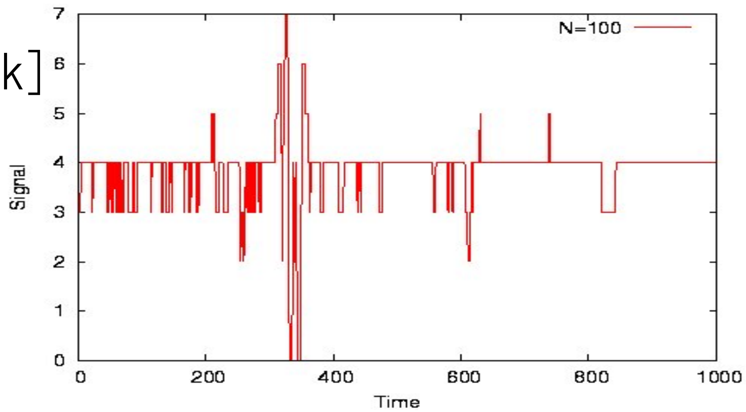
[a]



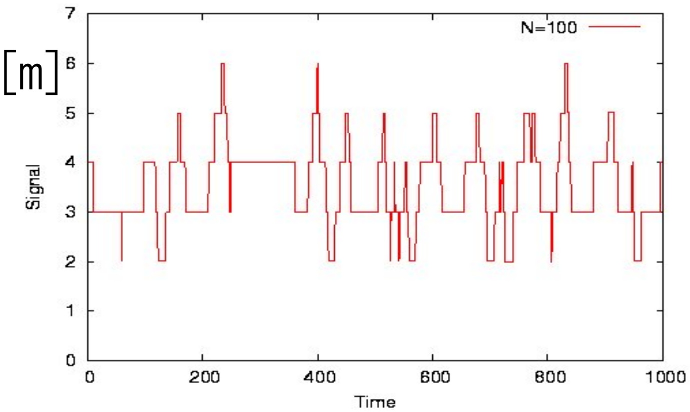
[i]



[k]



[m]



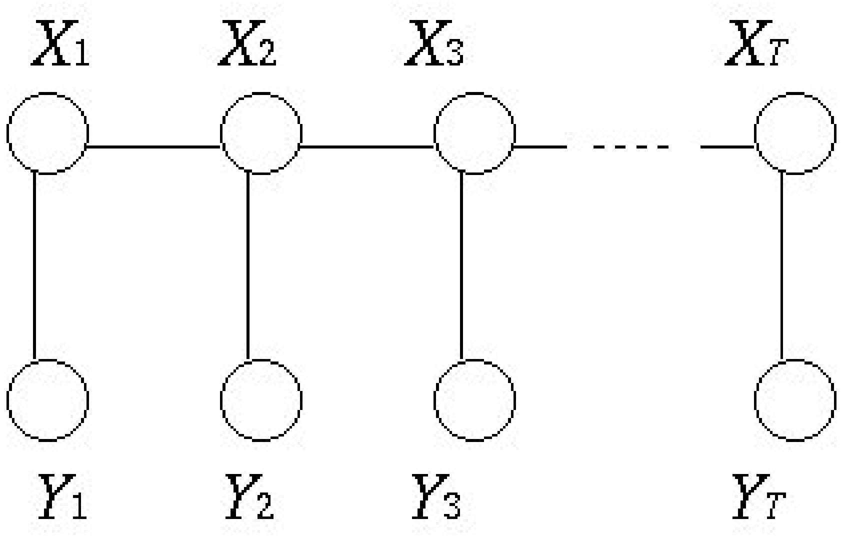
- Nが大きいと学習データに近づく
- 学習データにより生成される波形は違う

まとめ

- 隠れマルコフモデルで音声信号が表現できるのかを調べた
- 現実の音声信号を与え、モデル化した

- モデルから生成された波形から、学習データの違いが判別できた

→音声信号をモデル化できる

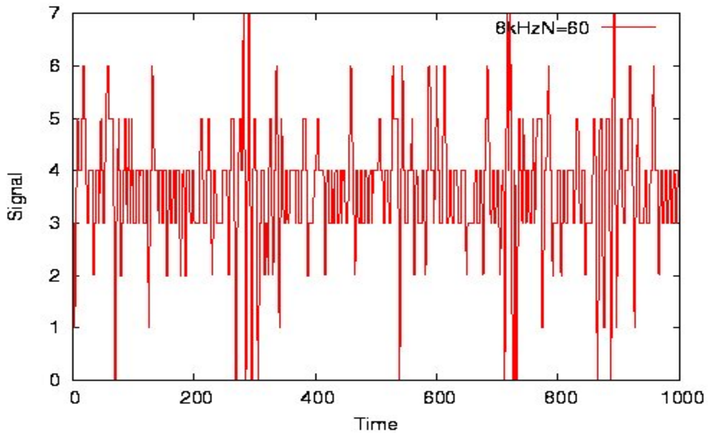


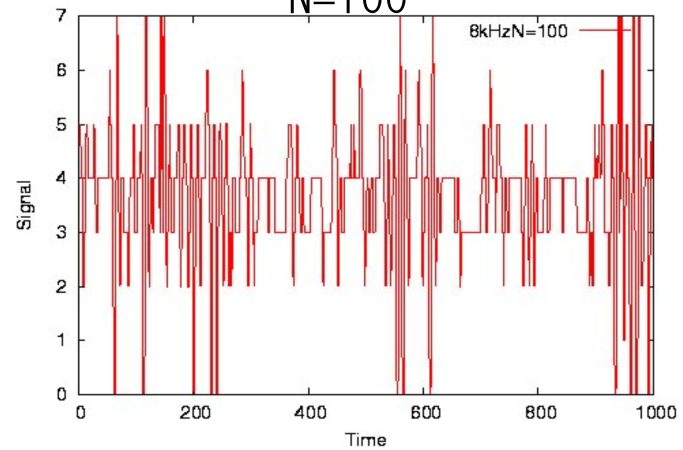
ありがとうございました

8kHz, 16kHz [a]

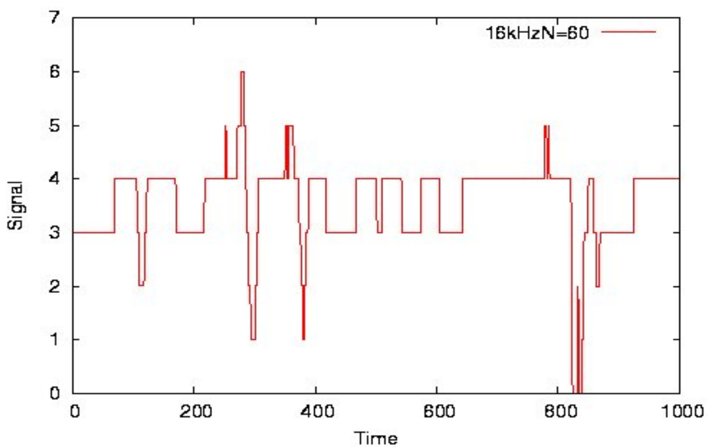
N=60 | N=100

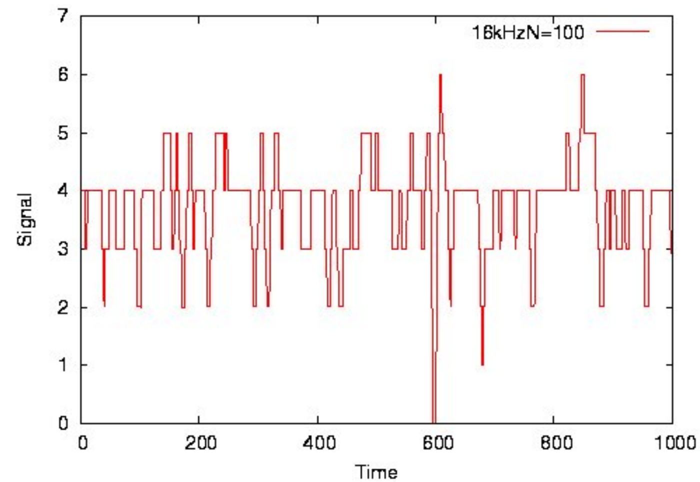
8kHz





16kHz





EMアルゴリズム

1. $\{p_{ij}^m\}$ の初期値を乱数に設定する

$$0 \leq p_{ij} \leq 1, \sum_j p_{ij} = 1$$

2. $X_t = i, X_{t+1} = j$ となる回数の期待値を計算する

$$n_{ij}^m = \mathrm{E}_{\{p_{ij}^m\}}[n_{ij} | y_1, \ldots, y_T]$$

3. $$\{p_{ij}^{m+1}\} = \left\{ \frac{n_{ij}^m}{\sum_k n_{ik}^m} \right\}$$

4. が収束するまで、2，3 を繰り返す

ＥＭアルゴリズム

$$p_{ij}^{m+1} = \frac{n_{ij}^{m}}{\sum_{k} n_{ik}^{m}} = \frac{\mathrm{E}[n_{ij} | \mathbf{y}]}{\sum_{k} \mathrm{E}[n_{ik} | \mathbf{y}]}$$

$$= \frac{\sum_{t=1}^{T-1} p(X_t = i, X_{t+1} = j, \mathbf{y})}{\sum_{t=1}^{T-1} p(X_t = i, \mathbf{y})}$$

計算式

$$p_{ij}^{m+1} = \frac{\sum_{t=1}^{T-1} L_t(i) p_{ij}^m q_{j,y_{t+1}} R_{t+1}(j)}{\sum_{t=1}^{T-1} L_t(i) R_t(i)}$$

$$L_{t+1}(i) = \sum_{j=1}^{N} L_t(j) p_{ji}^m q_{i,y_{t+1}}$$

$$R_t(i) = \sum_{j=1}^{N} R_{t+1}(j) p_{ij}^m q_{j,y_{t+1}}$$

$$L_1(i) = p_i^m q_{i,y_{t+1}},\ R_T(i) = 1$$

$$p_i^{m+1} = \frac{\sum_{t=1}^{T-1} L_t(i) R_t(i)}{\sum_{j=1}^{N} \sum_{t=1}^{T-1} L_t(j) R_t(j)}$$

$$p_{ij}^{m+1} = \frac{\sum_{t=1}^{T-1} L_t(i)\, p_{ij}^m\, q_{j,y_{t+1}}\, R_{t+1}(j)}{\sum_{t=1}^{T-1} L_t(i) R_t(i)}$$

$$L_{t+1}(i) = \sum_{j=1}^{N} L_t(j)\, p_{ji}^m\, q_{i,y_{t+1}}$$

$$R_t(i) = \sum_{j=1}^{N} R_{t+1}(j)\, p_{ij}^m\, q_{j,y_{t+1}}$$

$$L_1(i) = p_i^m q_{i,y_{t+1}},\ R_T(i) = 1$$

i, j

$$Yt = f(Xt)$$

$$(X_t \in 1,\ldots,N)$$

$$p_i^m = \frac{\sum_{t=1}^{T-1} L_t(i) R_t(i)}{\sum_{j=1}^{N} \sum_{t=1}^{T-1} L_t(j) R_t(j)}$$

p_{ij}^m

$$Y_t = X_t \bmod 8$$

$X_1 = i$

p_i^m

X Y X_t

$P(Y \mid X)$

$t = 1,\ldots,T$

$f(i) = y$

$N = 10$

$q_{i,y}$